Please check our wiki for help on navigating the form.

Horizon 2020

Call: H2020-INFRAEOSC-2018-2020

(Implementing the European Open Science Cloud)

Topic: INFRAEOSC-02-2019

Type of action: RIA

Proposal number: 863421

Proposal acronym: KOINE

Deadline Id: H2020-INFRAEOSC-2019-1

Table of contents

How to fill in the forms

The administrative forms must be filled in for each proposal using the templates available in the submission system. Some data fields in the administrative forms are pre-filled based on the steps in the submission wizard.

1 - General information

Topic	INFRAEOSC-02-2019	Type of Action	RIA
Call Identifier	H2020-INFRAEOSC-2018-2020	Deadline Id	H2020-INFRAEOSC-2019-1

Acronym: KOINE

Proposal title: Interdisciplinary data interoperability, discovery and exploitation in the EOSC

Note that for technical reasons, the following characters are not accepted in the Proposal Title and will be removed: < > " &

Duration in months: *36*

Free keywords: *Data Interoperability; Terminology Services; Terminology Interoperability; Machine Actionability; Automation; Data Discoverability; Data Exploitation; B2FIND; B2TERM*

Abstract

Today, knowledge- and information-based economies rely on readily exploitable data. However, most available data is siloed due to heterogeneous standards and practices, making such exploitation inefficient or impossible. In response, multiple communities are adopting the FAIR Data Principles. Of these, Findability and Accessibility have received much attention, leading to notable increases in Reusability. Interoperability, however, has proven to be the hardest principle to implement across initiatives and infrastructures. FAIR Interoperability relies on the use of formal and machine-readable languages for knowledge representation to 1) overcome ambiguities in natural language and 2) promote the use of semantic technologies in tasks including data discovery, quality control, integration, and analysis.

In this project, we aim to develop B2TERM, an innovative EOSC service to address terminology-based and semantic interoperability. The B2TERM service will build on existing, well-used FAIR terminologies and services to enhance the interoperability of EOSC data holdings with research data e-Infrastructures and research communities. KOINE will also impact the terminologies themselves, increasing their FAIRness where needed and aligning them through a novel Terminology Interoperability Framework to support global reporting mechanisms including the Essential Variable frameworks for Climate, Biodiversity, Oceans, and Ecosystems. We thus envisage B2TERM as a cornerstone of interoperability in the EOSC ecosystem, promoting automated data discovery and exploitation across disciplines, while supporting user-centric applications interlinking data from science, industry, and society. Looking forward, we see B2TERM as a foundation on which existing and future EOSC services can build interoperable solutions. We will demonstrate this vision through integration of B2TERM with disciplinary pan-European research infrastructures and the interdisciplinary data discovery service B2FIND.

Remaining characters 2

Has this proposal (or a very similar one) been submitted in the past 2 years in response to a call for proposals under Horizon 2020 or any other EU programme(s)? ○ Yes ◉ No

Please give the proposal reference or contract number.	
xxxxxx-x	

Declarations

1) The coordinator declares to have the explicit consent of all applicants on their participation and on the content of this proposal.	☒
2) The information contained in this proposal is correct and complete.	☒
3) This proposal complies with ethical principles (including the highest standards of research integrity — as set out, for instance, in the European Code of Conduct for Research Integrity — and including, in particular, avoiding fabrication, falsification, plagiarism or other research misconduct).	☒

4) The coordinator confirms:

- to have carried out the self-check of the financial capacity of the organisation on http://ec.europa.eu/research/participants/portal/desktop/en/organisations/lfv.html or to be covered by a financial viability check in an EU project for the last closed financial year. Where the result was "weak" or "insufficient", the coordinator confirms being aware of the measures that may be imposed in accordance with the H2020 Grants Manual (Chapter on Financial capacity check); or	○
- is exempt from the financial capacity check being a public body including international organisations, higher or secondary education establishment or a legal entity, whose viability is guaranteed by a Member State or associated country, as defined in the H2020 Grants Manual (Chapter on Financial capacity check); or	◉
- as sole participant in the proposal is exempt from the financial capacity check.	○

5) The coordinator hereby declares that each applicant has confirmed:

- they are fully eligible in accordance with the criteria set out in the specific call for proposals; and	☒
- they have the financial and operational capacity to carry out the proposed action.	☒

The coordinator is only responsible for the correctness of the information relating to his/her own organisation. Each applicant remains responsible for the correctness of the information related to him and declared above. Where the proposal to be retained for EU funding, the coordinator and each beneficiary applicant will be required to present a formal declaration in this respect.

According to Article 131 of the Financial Regulation of 25 October 2012 on the financial rules applicable to the general budget of the Union (Official Journal L 298 of 26.10.2012, p. 1) and Article 145 of its Rules of Application (Official Journal L 362, 31.12.2012, p.1) applicants found guilty of misrepresentation may be subject to administrative and financial penalties under certain conditions.

Personal data protection

The assessment of your grant application will involve the collection and processing of personal data (such as your name, address and CV), which will be performed pursuant to Regulation (EC) No 45/2001 on the protection of individuals with regard to the processing of personal data by the Community institutions and bodies and on the free movement of such data. Unless indicated otherwise, your replies to the questions in this form and any personal data requested are required to assess your grant application in accordance with the specifications of the call for proposals and will be processed solely for that purpose. Details concerning the purposes and means of the processing of your personal data as well as information on how to exercise your rights are available in the privacy statement. Applicants may lodge a complaint about the processing of their personal data with the European Data Protection Supervisor at any time.

Your personal data may be registered in the Early Detection and Exclusion system of the European Commission (EDES), the new system established by the Commission to reinforce the protection of the Union's financial interests and to ensure sound financial management, in accordance with the provisions of articles 105a and 108 of the revised EU Financial Regulation (FR) (Regulation (EU, EURATOM) 2015/1929 of the European Parliament and of the Council of 28 October 2015 amending Regulation (EU, EURATOM) No 966/2012) and articles 143 - 144 of the corresponding Rules of Application (RAP) (COMMISSION DELEGATED REGULATION (EU) 2015/2462 of 30 October 2015 amending Delegated Regulation (EU) No 1268/2012) for more information see the Privacy statement for the EDES Database.

2 - Participants & contacts

#	Participant Legal Name	Country	Action
1	TECHNISCHE INFORMATIONSBIBLIOTHEK (TIB)	DE	
2	CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE CNRS	FR	
3	UMWELTBUNDESAMT GESELLSCHAFT MIT BESCHRANKTER HAFTUNG (UBA GMBH)	AT	
4	UNIVERSITAET BREMEN	DE	
5	ALFRED-WEGENER-INSTITUT HELMHOLTZ-ZENTRUM FUR POLAR- UND MEERESFORSCHUNG	DE	
6	NORSK INSTITUTT FOR LUFTFORSKNING STIFTELSE	Norway	
7	UNIVERSITEIT VAN AMSTERDAM	NL	
8	UNIVERSITEIT MAASTRICHT	NL	
9	METEOROLOGISK INSTITUTT	Norway	
10	E-SCIENCE EUROPEAN INFRASTRUCTURE FOR BIODIVERSITY AND ECOSYSTEM RESEARCH	ES	
11	UNITED KINGDOM RESEARCH AND INNOVATION	UK	
12	ILMATIETEEN LAITOS	Finland	
13	CONSIGLIO NAZIONALE DELLE RICERCHE	IT	
14	BARCELONA SUPERCOMPUTING CENTER - CENTRO NACIONAL DE SUPERCOMPUTACION	ES	
15	DEUTSCHES KLIMARECHENZENTRUM GMBH	DE	
16	EUDAT OY	FI	
17	CSC-TIETEEN TIETOTEKNIIKAN KESKUS OY	FI	
18	INSTITUT FRANCAIS DE RECHERCHE POUR L'EXPLOITATION DE LA MER	FR	
19	INSTITUT NATIONAL DE LA RECHERCHE AGRONOMIQUE	FR	
20	TECHNISCHE UNIVERSITEIT DELFT	NL	
21	OESTERREICHISCHE AKADEMIE DER WISSENSCHAFTEN	AT	

22	Australian National Data Service	Australia	
23	SISTEMA GMBH	Austria	

2 - Administrative data of participating organisations

PIC ***Legal name***

925088419 *TECHNISCHE INFORMATIONSBIBLIOTHEK (TIB)*

Short name: TIB

Address of the organisation

Street WELFENGARTEN 1B

Town HANNOVER

Postcode 30167

Country Germany

Webpage www.tib.eu

Legal Status of your organisation

Research and Innovation legal statuses

Public body .....yes

Non-profit .....yes

International organisation .....no

International organisation of European interest .....no

Secondary or Higher education establishment .....no

Research organisation .....no

Legal person .....yes

Industry (private for profit).....no

Enterprise Data

SME self-declared status..... unknown

SME self-assessment ..... unknown

SME validation sme..... unknown

Based on the above details of the Beneficiary Registry the organisation is not an SME (small- and medium-sized enterprise) for the call.

Department(s) carrying out the proposed work

No department involved

Department name: *Name of the department/institute carrying out the work.* ☒ not applicable

☐ Same as proposing organisation's address

Street: *Please enter street name and number.*

Town: *Please enter the name of the town.*

Postcode: *Area code.*

Country: *Please select a country*

Dependencies with other proposal participants

Character of dependence	***Participant***	

Person in charge of the proposal

The name and e-mail of contact persons are read-only in the administrative form, only additional details can be edited here. To give access rights and basic contact details of contact persons, please go back to Step 4 of the submission wizard and save the changes.

Title: Dr. Sex: (•) Male ( ) Female

First name: **Markus** Last name: **Stocker**

E-Mail: **markus.stocker@tib.eu**

Position in org.: Research Group Leader

Department: TECHNISCHE INFORMATIONSBIBLIOTHEK (TIB) ☒ Same as organisation name

☒ Same as proposing organisation's address

Street: WELFENGARTEN 1B

Town: HANNOVER Post code: 30167

Country: Germany

Website: www.tib.eu

Phone: +49 511 762 14204 Phone 2: +xxx xxxxxxxxx Fax: +49 511 762 14728

Other contact persons

First Name	***Last Name***	***E-mail***	***Phone***
Gino	ERKELING	gino.erkeling@tib.eu	*+49 511 762 19652*
Angelina	Kraft	angelina.kraft@tib.eu	*+xxx xxxxxxxxx*
Sören	Auer	soeren.auer@tib.eu	*+xxx xxxxxxxxx*
Alexandra	Garatzogianni	alexandra.garatzogianni@tib.eu	*+xxx xxxxxxxxx*

PIC	***Legal name***
999997930	*CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE CNRS*

Short name: CNRS

Address of the organisation

Street RUE MICHEL ANGE 3

Town PARIS

Postcode 75794

Country France

Webpage www.cnrs.fr

Legal Status of your organisation

Research and Innovation legal statuses

Public body .....yes

Non-profit .....yes

International organisation .....no

International organisation of European interest .....no

Secondary or Higher education establishment .....no

Research organisation .....yes

Legal person .....yes

Industry (private for profit).....no

Enterprise Data

SME self-declared status.....18/11/2008 - no

SME self-assessment ..... unknown

SME validation sme.....18/11/2008 - no

Based on the above details of the Beneficiary Registry the organisation is not an SME (small- and medium-sized enterprise) for the call.

Department(s) carrying out the proposed work

Department 1

Department name: Laboratoire d'Informatique, de Robotique et de Microélectronique ☐ not applicable

☐ Same as proposing organisation's address

Street: 161 rue Ada CC477

Town: MONTPELLIER CEDEX 5

Postcode: 34095/5

Country: France

Department 2

Department name: Centre d'Ecologie Fonctionnelle et Evolutive ☐ not applicable

☐ Same as proposing organisation's address

Street: 1919 route de Mende

Town: MONTPELLIER CEDEX 5

Postcode: 34293

Country: France

Department 3

Department name: Institut Pierre-Simon-Laplace ☐ not applicable

☐ Same as proposing organisation's address

Street: SORBONNE UNIVERSITE - 4 Place Jussieu

Town: PARIS CEDEX 05

Postcode: 75252

Country: France

Dependencies with other proposal participants

Character of dependence	***Participant***	

Person in charge of the proposal

The name and e-mail of contact persons are read-only in the administrative form, only additional details can be edited here. To give access rights and basic contact details of contact persons, please go back to Step 4 of the submission wizard and save the changes.

Title: Dr. Sex: (•) Male ( ) Female

First name: **Clement** Last name: **Jonquet**

E-Mail: **jonquet@lirmm.fr**

Position in org.: Assistant Professor

Department: Informatics Department of LIRMM ☐ Same as organisation name

☐ Same as proposing organisation's address

Street: 161 Rue Ada CC477

Town: MONTPELLIER Post code: 34095/5

Country: France

Website: http://www.lirmm.fr/lirmm_eng

Phone: +334 67418509 Phone 2: +334 67418585 Fax: +334 67418500

Other contact persons

First Name	***Last Name***	***E-mail***	***Phone***
Marie-Christine	Pothier	marie-christine.pothier@lirmm.fr	*+334 67149723*
Elodie	Vatonne	elodie.vatonne@dr13.cnrs.fr	*+334 67613502*
Guillaume	ROCHET	rspv@dr13.cnrs.fr	*+xxx xxxxxxxxx*
Marie Hélène	PAPILLON	dr04spv-europe@cnrs.fr	*+xxx xxxxxxxxx*

PIC	***Legal name***
999452014	*UMWELTBUNDESAMT GESELLSCHAFT MIT BESCHRANKTER HAFTUNG (UBA GMBH)*

Short name: EAA

Address of the organisation

Street SPITTELAUER LANDE 5

Town WIEN

Postcode 1090

Country Austria

Webpage www.umweltbundesamt.at

Legal Status of your organisation

Research and Innovation legal statuses

Public body .....no

Non-profit .....yes

International organisation .....no

International organisation of European interest .....no

Secondary or Higher education establishment .....no

Research organisation .....no

Legal person .....yes

Industry (private for profit).....no

Enterprise Data

SME self-declared status.....04/01/1999 - no

SME self-assessment ..... unknown

SME validation sme..... unknown

Based on the above details of the Beneficiary Registry the organisation is not an SME (small- and medium-sized enterprise) for the call.

Department(s) carrying out the proposed work

Department 1

Department name: Ecosystem Research & Environmental Information Management ☐ not applicable

☒ Same as proposing organisation's address

Street: SPITTELAUER LANDE 5

Town: WIEN

Postcode: 1090

Country: Austria

Dependencies with other proposal participants

Character of dependence	***Participant***	

Person in charge of the proposal

The name and e-mail of contact persons are read-only in the administrative form, only additional details can be edited here. To give access rights and basic contact details of contact persons, please go back to Step 4 of the submission wizard and save the changes.

Title: Mrs

Sex: ○ Male ◉ Female

First name: **Barbara** Last name: **Magagna**

E-Mail: **barbara.magagna@umweltbundesamt.at**

Position in org.: Senior Expert

Department: Ecosystem Research & Environmental Information Management ☐ Same as organisation name

☒ Same as proposing organisation's address

Street: SPITTELAUER LANDE 5

Town: WIEN Post code: 1090

Country: Austria

Website: www.umweltbundesamt.at

Phone: +431313043447 Phone 2: *+xxx xxxxxxxxx* Fax: *+xxx xxxxxxxxx*

Other contact persons

First Name	***Last Name***	***E-mail***	***Phone***
Herbert	Haubold	herbert.haubold@umweltbundesamt.at	*+431313045910*

PIC	***Legal name***
999987454	*UNIVERSITAET BREMEN*

Short name: UBREMEN

Address of the organisation

Street Bibliothekstrasse 1

Town BREMEN

Postcode 28359

Country Germany

Webpage www.uni-bremen.de

Legal Status of your organisation

Research and Innovation legal statuses

Public body .....yes

Non-profit .....yes

International organisation .....no

International organisation of European interest .....no

Secondary or Higher education establishment .....yes

Research organisation .....yes

Legal person .....yes

Industry (private for profit).....no

Enterprise Data

SME self-declared status.....20/05/2016 - no

SME self-assessment .....20/05/2016 - no

SME validation sme.....28/01/2009 - no

Based on the above details of the Beneficiary Registry the organisation is not an SME (small- and medium-sized enterprise) for the call.

Department(s) carrying out the proposed work

Department 1

Department name: MARUM – Cen-ter for Mar-ine En-vir-on-mental Sci-ences ☐ not applicable

☐ Same as proposing organisation's address

Street: Leo-bener Str. 8

Town: Bremen

Postcode: 28359

Country: Germany

Dependencies with other proposal participants

Character of dependence	***Participant***	

Person in charge of the proposal

The name and e-mail of contact persons are read-only in the administrative form, only additional details can be edited here. To give access rights and basic contact details of contact persons, please go back to Step 4 of the submission wizard and save the changes.

Title: Dr.

Sex: (•) Male ( ) Female

First name: **Michael**

Last name: **Diepenbroek**

E-Mail: **mdiepenbroek@pangaea.de**

Position in org.: Managing Director of PANGAEA

Department: MARUM

[ ] Same as organisation name

[ ] Same as proposing organisation's address

Street: FVG Ost building, Leobener Str. 2

Town: Bremen

Post code: 28359

Country: Germany

Website: https://www.marum.de/en/Michael-Diepenbroek.html

Phone: +49 421 218 65590

Phone 2: *+xxx xxxxxxxxx*

Fax: *+xxx xxxxxxxxx*

Other contact persons

First Name	***Last Name***	***E-mail***	***Phone***
Robert	Huber	rhuber@uni-bremen.de	*+49 421 218 65593*
Anusuriya	Devaraju	adevaraju@marum.de	*+49 421 218 65676*
Martin	Mehrtens	eu@vw.uni-bremen.de	*+xxx xxxxxxxxx*

PIC	***Legal name***
999497507	*ALFRED-WEGENER-INSTITUT HELMHOLTZ-ZENTRUM FUR POLAR- UND MEERESFORSCHUNG*

Short name: AWI

Address of the organisation

Street AM HANDELSHAFEN 12

Town BREMERHAVEN

Postcode 27570

Country Germany

Webpage www.awi.de

Legal Status of your organisation

Research and Innovation legal statuses

Public body .....yes

Non-profit .....yes

International organisation .....no

International organisation of European interest .....no

Secondary or Higher education establishment .....no

Research organisation .....yes

Legal person .....yes

Industry (private for profit).....no

Enterprise Data

SME self-declared status.....31/12/2015 - no

SME self-assessment .....17/03/1986 - no

SME validation sme..... unknown

Based on the above details of the Beneficiary Registry the organisation is not an SME (small- and medium-sized enterprise) for the call.

Department(s) carrying out the proposed work

Department 1

Department name: Biosciences, Deep Sea Ecology and Technology ☐ not applicable

☒ Same as proposing organisation's address

Street: AM HANDELSHAFEN 12

Town: BREMERHAVEN

Postcode: 27570

Country: Germany

Dependencies with other proposal participants

Character of dependence	***Participant***	

Person in charge of the proposal

The name and e-mail of contact persons are read-only in the administrative form, only additional details can be edited here. To give access rights and basic contact details of contact persons, please go back to Step 4 of the submission wizard and save the changes.

Title: Dr. Sex: (•) Male ( ) Female

First name: **Pier Luigi** Last name: **Buttigieg**

E-Mail: **pbuttigi@mpi-bremen.de**

Position in org.: Research Scientist

Department: Biosciences, Deep Sea Ecology and Technology ☐ Same as organisation name

☒ Same as proposing organisation's address

Street: AM HANDELSHAFEN 12

Town: BREMERHAVEN Post code: 27570

Country: Germany

Website: www.awi.de

Phone: +49(471)4831-2268 Phone 2: +49(0)4212028984 Fax: +49(471)4831-177

Other contact persons

First Name	***Last Name***	***E-mail***	***Phone***
Tordis	Hellmann	tordis.hellmann@awi.de	*+49(471)4831-2356*
EU	Grants	eu-grants@awi.de	*+xxx xxxxxxxxx*

PIC	***Legal name***
999654162	*NORSK INSTITUTT FOR LUFTFORSKNING STIFTELSE*

Short name: NORWEGIAN INSTITUTE FOR AIR RESEARCH NILU

Address of the organisation

Street INSTITUTTVEIEN 18

Town KJELLER

Postcode 2027

Country Norway

Webpage www.nilu.no

Legal Status of your organisation

Research and Innovation legal statuses

Public body .....no

Non-profit .....yes

International organisation .....no

International organisation of European interest .....no

Secondary or Higher education establishment .....no

Research organisation .....yes

Legal person .....yes

Industry (private for profit).....no

Enterprise Data

SME self-declared status..... unknown

SME self-assessment ..... unknown

SME validation sme..... unknown

Based on the above details of the Beneficiary Registry the organisation is not an SME (small- and medium-sized enterprise) for the call.

Department(s) carrying out the proposed work

Department 1

Department name: Atmospheric and Climate Research Department (ATMOS) ☐ not applicable

☐ Same as proposing organisation's address

Street: Instituttveien 18

Town: Kjeller

Postcode: 2007

Country: Norway

Dependencies with other proposal participants

Character of dependence	***Participant***	

Person in charge of the proposal

The name and e-mail of contact persons are read-only in the administrative form, only additional details can be edited here. To give access rights and basic contact details of contact persons, please go back to Step 4 of the submission wizard and save the changes.

Title: Dr. Sex: ○ Male ◉ Female

First name: **Cathrine Lund** Last name: **Myhre**

E-Mail: **clm@nilu.no**

Position in org.: Senior Scientist

Department: Atmospheric and Climate Research Department (ATMOS) ☐ Same as organisation name

☐ Same as proposing organisation's address

Street: Instituttveien 18

Town: Kjeller Post code: 2007

Country: Norway

Website: www.nilu.no

Phone: +4763898000 Phone 2: +4763898042 Fax: +xxx xxxxxxxxx

Other contact persons

First Name	*Last Name*	*E-mail*	*Phone*
Berit	Modalen	bmo@nilu.no	*+4763898071*
Markus	Fiebig	markus.fiebig@nilu.no	*+4763898235*
Heidi	Fjeldstad	hfj@nilu.no	*+4763898101*
Eva Beate	Andresen	eba@nilu.no	*+4763898020*

PIC	***Legal name***
999985708	*UNIVERSITEIT VAN AMSTERDAM*

Short name: UNIVERSITEIT VAN AMSTERDAM

Address of the organisation

Street SPUI 21

Town AMSTERDAM

Postcode 1012WX

Country Netherlands

Webpage www.uva.nl

Legal Status of your organisation

Research and Innovation legal statuses

Public body .....yes

Non-profit .....yes

International organisation .....no

International organisation of European interest .....no

Secondary or Higher education establishment .....yes

Research organisation .....yes

Legal person .....yes

Industry (private for profit).....no

Enterprise Data

SME self-declared status.....04/07/2008 - no

SME self-assessment ..... unknown

SME validation sme.....04/07/2008 - no

Based on the above details of the Beneficiary Registry the organisation is not an SME (small- and medium-sized enterprise) for the call.

Department(s) carrying out the proposed work

Department 1

Department name: Institute of information technology ☐ not applicable

☐ Same as proposing organisation's address

Street: science park 904

Town: Amsterdam

Postcode: 1098XH

Country: Netherlands

Dependencies with other proposal participants

Character of dependence	***Participant***	

Person in charge of the proposal

The name and e-mail of contact persons are read-only in the administrative form, only additional details can be edited here. To give access rights and basic contact details of contact persons, please go back to Step 4 of the submission wizard and save the changes.

Title: Dr. Sex: (•) Male ( ) Female

First name: **Zhiming** Last name: **Zhao**

E-Mail: **z.zhao@uva.nl**

Position in org.: University of Amsterdam

Department: University of Amsterdam ☐ Same as organisation name

☐ Same as proposing organisation's address

Street: science park 904

Town: Amsterdam Post code: 1098XH

Country: Netherlands

Website:

Phone: +31641265121 Phone 2: +xxx xxxxxxxxx Fax: +xxx xxxxxxxxx

Other contact persons

First Name	***Last Name***	***E-mail***	***Phone***
Paul	Martin	p.w.martin@uva.nl	+xxx xxxxxxxxx
Vanessa	Wolters	v.m.wolters@uva.nl	+xxx xxxxxxxxx

PIC	***Legal name***
999975911	*UNIVERSITEIT MAASTRICHT*

Short name: UNIVERSITEIT MAASTRICHT

Address of the organisation

Street Minderbroedersberg 4-6

Town MAASTRICHT

Postcode 6200 MD

Country Netherlands

Webpage http://www.maastrichtuniversity.nl

Legal Status of your organisation

Research and Innovation legal statuses

Public body .....yes

Non-profit .....yes

International organisation .....no

International organisation of European interest .....no

Secondary or Higher education establishment .....yes

Research organisation .....no

Legal person .....yes

Industry (private for profit).....no

Enterprise Data

SME self-declared status.....01/01/1976 - no

SME self-assessment .....01/01/1976 - no

SME validation sme..... unknown

Based on the above details of the Beneficiary Registry the organisation is not an SME (small- and medium-sized enterprise) for the call.

Department(s) carrying out the proposed work

Department 1

Department name: Institute of Data Science ☐ not applicable

☐ Same as proposing organisation's address

Street: Universiteitssingel 60

Town: Maastricht

Postcode: 6229 ER

Country: Netherlands

Dependencies with other proposal participants

Character of dependence	***Participant***	

Person in charge of the proposal

The name and e-mail of contact persons are read-only in the administrative form, only additional details can be edited here. To give access rights and basic contact details of contact persons, please go back to Step 4 of the submission wizard and save the changes.

Title: Prof.

Sex: (•) Male ( ) Female

First name: **Michel**

Last name: **Dumontier**

E-Mail: **michel.dumontier@maastrichtuniversity.nl**

Position in org.: Professor

Department: Institute of Data Science

☐ Same as organisation name

☐ Same as proposing organisation's address

Street: Universiteitssingel 60

Town: Maastricht

Post code: 6229 ER

Country: Netherlands

Website:

Phone: +xxx xxxxxxxxx

Phone 2: +xxx xxxxxxxxx

Fax: +xxx xxxxxxxxx

PIC	***Legal name***
999510893	*METEOROLOGISK INSTITUTT*

Short name: METEOROLOGISK INSTITUTT

Address of the organisation

Street HENRIK MOHNS PLASS 1

Town OSLO

Postcode 0313

Country Norway

Webpage www.met.no

Legal Status of your organisation

Research and Innovation legal statuses

Public body .....yes

Non-profit .....yes

International organisation .....no

International organisation of European interest .....no

Secondary or Higher education establishment .....no

Research organisation .....yes

Legal person .....yes

Industry (private for profit).....no

Enterprise Data

SME self-declared status.....12/03/1995 - no

SME self-assessment ..... unknown

SME validation sme.....12/03/1995 - no

Based on the above details of the Beneficiary Registry the organisation is not an SME (small- and medium-sized enterprise) for the call.

Department(s) carrying out the proposed work

Department 1

Department name: Research and Development ☐ not applicable

☐ Same as proposing organisation's address

Street: P.O.BOX 43

Town: Blindern

Postcode: N-0313

Country: Norway

Dependencies with other proposal participants

Character of dependence	***Participant***	

Person in charge of the proposal

The name and e-mail of contact persons are read-only in the administrative form, only additional details can be edited here. To give access rights and basic contact details of contact persons, please go back to Step 4 of the submission wizard and save the changes.

Title: Dr.

Sex: (•) Male ( ) Female

First name: **Øystein**

Last name: **Godøy**

E-Mail: **o.godoy@met.no**

Position in org.: Senior Scientist

Department: Research and Development ☐ Same as organisation name

☐ Same as proposing organisation's address

Street: P.O.BOX 43 Blindern

Town: Oslo Post code: N-0313

Country: Norway

Website:

Phone: +47 9802 4433 Phone 2: +xxx xxxxxxxxx Fax: +xxx xxxxxxxxx

Other contact persons

First Name	***Last Name***	***E-mail***	***Phone***
Michael	Schulz	michael.schulz@met.no	+xxx xxxxxxxxx

PIC	***Legal name***
909022018	*E-SCIENCE EUROPEAN INFRASTRUCTURE FOR BIODIVERSITY AND ECOSYSTEM RESEARCH*

Short name: E-SCIENCE EUROPEAN INFRASTRUCTURE FOR BIODIVERSITY AND ECOSYSTEM RESEARCH

Street PLAZA DE ESPANA S/N, SECTOR II-III

Town SEVILLA

Postcode 41071

Country Spain

Webpage https://www.lifewatch.eu/

Legal Status of your organisation

Research and Innovation legal statuses

Public body .....yes

Non-profit .....yes

International organisation .....no

International organisation of European interest .....no

Secondary or Higher education establishment .....no

Research organisation .....yes

Legal person .....yes

Industry (private for profit).....no

Enterprise Data

SME self-declared status..... unknown

SME self-assessment ..... unknown

SME validation sme..... unknown

Based on the above details of the Beneficiary Registry the organisation is not an SME (small- and medium-sized enterprise) for the call.

Department(s) carrying out the proposed work

No department involved

Department name: *Name of the department/institute carrying out the work.* ☒ not applicable

☐ Same as proposing organisation's address

Street: *Please enter street name and number.*

Town: *Please enter the name of the town.*

Postcode: *Area code.*

Country: *Please select a country*

Dependencies with other proposal participants

Character of dependence	***Participant***	

Person in charge of the proposal

The name and e-mail of contact persons are read-only in the administrative form, only additional details can be edited here. To give access rights and basic contact details of contact persons, please go back to Step 4 of the submission wizard and save the changes.

Title: Dr. Sex: (•) Male ( ) Female

First name: **Nicola** Last name: **Fiore**

E-Mail: **nicola.fiore@lifewatch.eu**

Position in org.: SERVICE CENTRE ICT COORDINATOR

Department: E-SCIENCE EUROPEAN INFRASTRUCTURE FOR BIODIVERSITY AND EC ☒ Same as organisation name

☐ Same as proposing organisation's address

Street: Via per Monteroni sn

Town: LECCE Post code: 73100

Country: Italy

Website: http://www.lifewatch.eu

Phone: +390832294819 Phone 2: +393939109170 Fax: *+xxx xxxxxxxxx*

PIC	***Legal name***
906446474	*UNITED KINGDOM RESEARCH AND INNOVATION*

Short name: UKRI

Address of the organisation

Street POLARIS HOUSE NORTH STAR AVENUE

Town SWINDON

Postcode SN2 1FL

Country United Kingdom

Webpage https://www.ukri.org/

Legal Status of your organisation

Research and Innovation legal statuses

Public body .....yes

Non-profit .....yes

International organisation .....no

International organisation of European interest .....no

Secondary or Higher education establishment .....no

Research organisation .....yes

Legal person .....yes

Industry (private for profit).....no

Enterprise Data

SME self-declared status..... unknown

SME self-assessment ..... unknown

SME validation sme..... unknown

Based on the above details of the Beneficiary Registry the organisation is not an SME (small- and medium-sized enterprise) for the call.

Department(s) carrying out the proposed work

Department 1

Department name: National Oceanography Centre - British Oceanographic Data Centre ☐ not applicable

☐ Same as proposing organisation's address

Street: 6 Brownlow Street

Town: Liverpool

Postcode: L3 5DA

Country: United Kingdom

Dependencies with other proposal participants

Character of dependence	***Participant***	

Person in charge of the proposal

The name and e-mail of contact persons are read-only in the administrative form, only additional details can be edited here. To give access rights and basic contact details of contact persons, please go back to Step 4 of the submission wizard and save the changes.

Title: Dr. Sex: ○ Male ◉ Female

First name: **Louise J.** Last name: **Darroch**

E-Mail: **louise.darroch@bodc.ac.uk**

Position in org.: Senior Data Scientist

Department: National Oceanography Centre - British Oceanographic Data Centre ☐ Same as organisation name

☐ Same as proposing organisation's address

Street: 6 Brownlow Street

Town: Liverpool Post code: L3 5DA

Country: United Kingdom

Website:

Phone: +44 151 795 4883 Phone 2: *+xxx xxxxxxxxx* Fax: *+xxx xxxxxxxxx*

Other contact persons

First Name	***Last Name***	***E-mail***	***Phone***
Gwenaelle	Moncoiffe	gmon@bodc.ac.uk	*+44 151 795 4880*
Stephanie	Walsh	stls@noc.ac.uk	*+44 151 795 4846*
Philip	Worrall	pgwo@noc.ac.uk	*+44 151 795 4842*
Alexandra	Kokkinaki	alexk@bodc.ac.uk	*+44 151 795 4874*

PIC	***Legal name***
999591306	*ILMATIETEEN LAITOS*

Short name: FMI

Address of the organisation

Street Erik Palmenin aukio 1

Town HELSINKI

Postcode 00560

Country Finland

Webpage www.fmi.fi

Legal Status of your organisation

Research and Innovation legal statuses

Public body .....yes

Non-profit .....yes

International organisation .....no

International organisation of European interest .....no

Secondary or Higher education establishment .....no

Research organisation .....yes

Legal person .....yes

Industry (private for profit).....no

Enterprise Data

SME self-declared status..... unknown

SME self-assessment ..... unknown

SME validation sme..... unknown

Based on the above details of the Beneficiary Registry the organisation is not an SME (small- and medium-sized enterprise) for the call.

Department(s) carrying out the proposed work

No department involved

Department name: *Name of the department/institute carrying out the work.* ☒ not applicable

☐ Same as proposing organisation's address

Street: *Please enter street name and number.*

Town: *Please enter the name of the town.*

Postcode: *Area code.*

Country: *Please select a country*

Dependencies with other proposal participants

Character of dependence	***Participant***	

Person in charge of the proposal

The name and e-mail of contact persons are read-only in the administrative form, only additional details can be edited here. To give access rights and basic contact details of contact persons, please go back to Step 4 of the submission wizard and save the changes.

Title: Dr.

Sex: (•) Male ( ) Female

First name: **Ewan**

Last name: **O'Connor**

E-Mail: **ewan.oconnor@fmi.fi**

Position in org.: Tenure-track Professor

Department: ILMATIETEEN LAITOS ☒ Same as organisation name

☒ Same as proposing organisation's address

Street: Erik Palmenin aukio 1

Town: HELSINKI Post code: 00560

Country: Finland

Website:

Phone: +358 505907973 Phone 2: *+xxx xxxxxxxxx* Fax: *+xxx xxxxxxxxx*

PIC	***Legal name***
999979500	*CONSIGLIO NAZIONALE DELLE RICERCHE*

Short name: CNR

Address of the organisation

Street PIAZZALE ALDO MORO 7

Town ROMA

Postcode 00185

Country Italy

Webpage www.cnr.it

Legal Status of your organisation

Research and Innovation legal statuses

Public body .....yes

Non-profit .....yes

International organisation .....no

International organisation of European interest .....no

Secondary or Higher education establishment .....no

Research organisation .....yes

Legal person .....yes

Industry (private for profit).....no

Enterprise Data

SME self-declared status.....18/05/2016 - no

SME self-assessment ..... unknown

SME validation sme.....05/12/2008 - no

Based on the above details of the Beneficiary Registry the organisation is not an SME (small- and medium-sized enterprise) for the call.

Department(s) carrying out the proposed work

Department 1

Department name: Istitute of Methodologies for Environmental Analysis ☐ not applicable

☐ Same as proposing organisation's address

Street: C.da S. Loja

Town: Tito (Potenza)

Postcode: I-85050

Country: Italy

Dependencies with other proposal participants

Character of dependence	***Participant***	

Person in charge of the proposal

The name and e-mail of contact persons are read-only in the administrative form, only additional details can be edited here. To give access rights and basic contact details of contact persons, please go back to Step 4 of the submission wizard and save the changes.

Title: Dr. Sex: ○ Male ◉ Female

First name: **Lucia** Last name: **Mona**

E-Mail: **lucia.mona@cnr.it**

Position in org.: Senior Researcher

Department: Istitute of Methodologies for Environmental Analysis ☐ Same as organisation name

☐ Same as proposing organisation's address

Street: C.da S. Loja

Town: Tito (Potenza) Post code: I-85050

Country: Italy

Website:

Phone: +39 09714272657 Phone 2: *+xxx xxxxxxxxx* Fax: *+xxx xxxxxxxxx*

PIC	***Legal name***
999655520	*BARCELONA SUPERCOMPUTING CENTER - CENTRO NACIONAL DE SUPERCOMPUTACION*

Short name: BSC

Address of the organisation

Street Calle Jordi Girona 31

Town BARCELONA

Postcode 08034

Country Spain

Webpage www.bsc.es

Legal Status of your organisation

Research and Innovation legal statuses

Public body .....yes

Non-profit .....yes

International organisation .....no

International organisation of European interest .....no

Secondary or Higher education establishment .....no

Research organisation .....yes

Legal person .....yes

Industry (private for profit).....no

Enterprise Data

SME self-declared status.....01/03/2005 - no

SME self-assessment ..... unknown

SME validation sme..... unknown

Based on the above details of the Beneficiary Registry the organisation is not an SME (small- and medium-sized enterprise) for the call.

Department(s) carrying out the proposed work

Department 1

Department name: Earth Scince Department ☐ not applicable

☐ Same as proposing organisation's address

Street: NEXUS II building, Jordi Girona 29

Town: Barcelona

Postcode: 08034

Country: Spain

Dependencies with other proposal participants

Character of dependence	***Participant***	

Person in charge of the proposal

The name and e-mail of contact persons are read-only in the administrative form, only additional details can be edited here. To give access rights and basic contact details of contact persons, please go back to Step 4 of the submission wizard and save the changes.

Title: Dr. Sex: (•) Male ( ) Female

First name: **Carlos Pérez** Last name: **García-Pando**

E-Mail: **carlos.perez@bsc.es**

Position in org.: Atmospheric Compostion Group Manager

Department: Earth Science Department ☐ Same as organisation name

☐ Same as proposing organisation's address

Street: NEXUS II building, Jordi Girona 29

Town: Barcelona Post code: 08034

Country: Spain

Website: www.bsc.es

Phone: +34934137722 Phone 2: *+xxx xxxxxxxxx* Fax: *+xxx xxxxxxxxx*

Other contact persons

First Name	***Last Name***	***E-mail***	***Phone***
Sara	Basart	sara.basart@bsc.es	*+34934134038*
Mar	Rodríguez	mar.rodriguez@bsc.es	*+34934137566*

PIC	***Legal name***
998692310	*DEUTSCHES KLIMARECHENZENTRUM GMBH*

Short name: DEUTSCHES KLIMARECHENZENTRUM GMBH

Address of the organisation

Street BUNDESSTRASSE 45A

Town HAMBURG

Postcode 20146

Country Germany

Webpage http:/www.dkrz.de

Legal Status of your organisation

Research and Innovation legal statuses

Public body .....yes

Non-profit .....yes

International organisation .....no

International organisation of European interest .....no

Secondary or Higher education establishment .....no

Research organisation .....yes

Legal person .....yes

Industry (private for profit).....no

Enterprise Data

SME self-declared status.....03/11/2008 - no

SME self-assessment ..... unknown

SME validation sme.....03/11/2008 - no

Based on the above details of the Beneficiary Registry the organisation is not an SME (small- and medium-sized enterprise) for the call.

Department(s) carrying out the proposed work

Department 1

Department name: Data Management Department ☐ not applicable

☒ Same as proposing organisation's address

Street: BUNDESSTRASSE 45A

Town: HAMBURG

Postcode: 20146

Country: Germany

Dependencies with other proposal participants

Character of dependence	***Participant***	

Person in charge of the proposal

The name and e-mail of contact persons are read-only in the administrative form, only additional details can be edited here. To give access rights and basic contact details of contact persons, please go back to Step 4 of the submission wizard and save the changes.

Title: Mr.

Sex: (•) Male ( ) Female

First name: **Hannes**

Last name: **Thiemann**

E-Mail: **thiemann@dkrz.de**

Position in org.: Deputy Department Head

Department: DEUTSCHES KLIMARECHENZENTRUM GMBH ☒ Same as organisation name

☒ Same as proposing organisation's address

Street: BUNDESSTRASSE 45A

Town: HAMBURG

Post code: 20146

Country: Germany

Website:

Phone: +49 40 460094113

Phone 2: +xxx xxxxxxxxx

Fax: +xxx xxxxxxxxx

Other contact persons

First Name	***Last Name***	***E-mail***	***Phone***
Heinrich	Widmann	widmann@dkrz.de	+xxx xxxxxxxxx

PIC	***Legal name***
907014021	*EUDAT OY*

Short name: EUDAT OY

Address of the organisation

Street KEILARANTA 14

Town ESPOO

Postcode 02100

Country Finland

Webpage

Legal Status of your organisation

Research and Innovation legal statuses

Public body .....no

Non-profit .....yes

International organisation .....no

International organisation of European interest .....no

Secondary or Higher education establishment .....no

Research organisation .....no

Legal person .....yes

Industry (private for profit).....no

Enterprise Data

SME self-declared status.....28/02/2018 - no

SME self-assessment .....28/02/2018 - no

SME validation sme..... unknown

Based on the above details of the Beneficiary Registry the organisation is not an SME (small- and medium-sized enterprise) for the call.

Department(s) carrying out the proposed work

No department involved

Department name: *Name of the department/institute carrying out the work.* ☒ not applicable

☐ Same as proposing organisation's address

Street: *Please enter street name and number.*

Town: *Please enter the name of the town.*

Postcode: *Area code.*

Country: *Please select a country*

Dependencies with other proposal participants

Character of dependence	***Participant***	

Person in charge of the proposal

The name and e-mail of contact persons are read-only in the administrative form, only additional details can be edited here. To give access rights and basic contact details of contact persons, please go back to Step 4 of the submission wizard and save the changes.

Title: Dr. Sex: (•) Male ( ) Female

First name: **Per** Last name: **Öster**

E-Mail: **per.oster@csc.fi**

Position in org.: Chair of the Board

Department: EUDAT OY ☒ Same as organisation name

☒ Same as proposing organisation's address

Street: KEILARANTA 14

Town: ESPOO Post code: 02100

Country: Finland

Website: eudat.eu

Phone: +358503819030 Phone 2: *+xxx xxxxxxxxx* Fax: *+xxx xxxxxxxxx*

Other contact persons

First Name	***Last Name***	***E-mail***	***Phone***
Damien	Lecarpentier	damien.lecarpentier@csc.fi	*+358503819515*
Heli	Autere	heli.autere@csc.fi	*+358503819065*

PIC	***Legal name***
999645820	*CSC-TIETEEN TIETOTEKNIIKAN KESKUS OY*

Short name: CSC-TIETEEN TIETOTEKNIIKAN KESKUS OY

Address of the organisation

Street Keilaranta 14

Town ESPOO

Postcode 02101

Country Finland

Webpage www.csc.fi

Legal Status of your organisation

Research and Innovation legal statuses

Public body .....no

Non-profit .....yes

International organisation .....unknown

International organisation of European interest .....unknown

Secondary or Higher education establishment .....unknown

Research organisation .....unknown

Legal person .....yes

Industry (private for profit).....no

Enterprise Data

SME self-declared status..... unknown

SME self-assessment ..... unknown

SME validation sme..... unknown

Based on the above details of the Beneficiary Registry the organisation is not an SME (small- and medium-sized enterprise) for the call.

Department(s) carrying out the proposed work

No department involved

Department name: *Name of the department/institute carrying out the work.* ☒ not applicable

☐ Same as proposing organisation's address

Street: *Please enter street name and number.*

Town: *Please enter the name of the town.*

Postcode: *Area code.*

Country: *Please select a country*

Dependencies with other proposal participants

Character of dependence	***Participant***	

Person in charge of the proposal

The name and e-mail of contact persons are read-only in the administrative form, only additional details can be edited here. To give access rights and basic contact details of contact persons, please go back to Step 4 of the submission wizard and save the changes.

Title: Mr. Sex: (•) Male ( ) Female

First name: **Antti** Last name: **Pursula**

E-Mail: **antti.pursula@csc.fi**

Position in org.: Program Director

Department: CSC-TIETEEN TIETOTEKNIIKAN KESKUS OY ☒ Same as organisation name

☒ Same as proposing organisation's address

Street: Keilaranta 14

Town: ESPOO Post code: 02101

Country: Finland

Website:

Phone: +xxx xxxxxxxxx Phone 2: +xxx xxxxxxxxx Fax: +xxx xxxxxxxxx

Other contact persons

First Name	***Last Name***	***E-mail***	***Phone***
Tiina	Timonen	tiina.timonen@csc.fi	+xxx xxxxxxxxx
Minna	Hahl	minna.hahl@csc.fi	+xxx xxxxxxxxx
Inkeri	Utunen	inkeri.utunen@csc.fi	+xxx xxxxxxxxx

PIC	***Legal name***
999630300	*INSTITUT FRANCAIS DE RECHERCHE POUR L'EXPLOITATION DE LA MER*

Short name: IFREMER

Address of the organisation

Street 155 rue Jean Jacques Rousseau

Town ISSY-LES-MOULINEAUX

Postcode 92138

Country France

Webpage http://www.ifremer.fr

Legal Status of your organisation

Research and Innovation legal statuses

Public body .....yes

Non-profit .....yes

International organisation .....unknown

International organisation of European interest .....unknown

Secondary or Higher education establishment .....unknown

Research organisation .....yes

Legal person .....yes

Industry (private for profit).....no

Enterprise Data

SME self-declared status..... unknown

SME self-assessment ..... unknown

SME validation sme..... unknown

Based on the above details of the Beneficiary Registry the organisation is not an SME (small- and medium-sized enterprise) for the call.

Department(s) carrying out the proposed work

Department 1

Department name: Research Infrastructures and Information Systems ☐ not applicable

☐ Same as proposing organisation's address

Street: 1625 Route de Sainte-Anne - CS 10070

Town: PLOUZANE

Postcode: 29280

Country: France

Dependencies with other proposal participants

Character of dependence	***Participant***	

Person in charge of the proposal

The name and e-mail of contact persons are read-only in the administrative form, only additional details can be edited here. To give access rights and basic contact details of contact persons, please go back to Step 4 of the submission wizard and save the changes.

Title: Mr.

Sex: (•) Male ( ) Female

First name: **Thierry**

Last name: **Carval**

E-Mail: **thierry.carval@ifremer.fr**

Position in org.: Head of Information Systems Engineering Services

Department: Research Infrastructures and Information Systems

☐ Same as organisation name

☐ Same as proposing organisation's address

Street: 1625 Route de Sainte-Anne - CS 10070

Town: PLOUZANE

Post code: 29280

Country: France

Website: http://wwz.ifremer.fr/

Phone: +33 298224597

Phone 2: *+xxx xxxxxxxxx*

Fax: *+xxx xxxxxxxxx*

Other contact persons

First Name	***Last Name***	***E-mail***	***Phone***
Sebastien	Poulain	projet.europe@ifremer.fr	*+33 298224365*

PIC	***Legal name***
999993274	*INSTITUT NATIONAL DE LA RECHERCHE AGRONOMIQUE*

Short name: INRA

Address of the organisation

Street Rue De L'Universite 147

Town PARIS CEDEX 07

Postcode 75338

Country France

Webpage www.inra.fr

Legal Status of your organisation

Research and Innovation legal statuses

Public body .....yes

Non-profit .....yes

International organisation .....no

International organisation of European interest .....no

Secondary or Higher education establishment .....no

Research organisation .....yes

Legal person .....yes

Industry (private for profit).....no

Enterprise Data

SME self-declared status.....10/07/2018 - no

SME self-assessment ..... unknown

SME validation sme.....25/09/2008 - no

Based on the above details of the Beneficiary Registry the organisation is not an SME (small- and medium-sized enterprise) for the call.

Department(s) carrying out the proposed work

Department 1

Department name: Délégation à l'Information Scientifique et Technique ☐ not applicable

☐ Same as proposing organisation's address

Street: INRA- Route de Saint Cyr

Town: Versailles

Postcode: 78026

Country: France

Dependencies with other proposal participants

Character of dependence	***Participant***	

Person in charge of the proposal

The name and e-mail of contact persons are read-only in the administrative form, only additional details can be edited here. To give access rights and basic contact details of contact persons, please go back to Step 4 of the submission wizard and save the changes.

Title: Ms

Sex: ○ Male ◉ Female

First name: **Sophie**

Last name: **Aubin**

E-Mail: **sophie.aubin@inra.fr**

Position in org.: Chargée de ressources

Department: DIST INRA ☐ Same as organisation name

☐ Same as proposing organisation's address

Street: 42, Rue Georges Morel

Town: Beaucouzé Post code: 49070

Country: France

Website:

Phone: +33241225657 Phone 2: *+xxx xxxxxxxxx* Fax: *+xxx xxxxxxxxx*

Other contact persons

First Name	***Last Name***	***E-mail***	***Phone***
Odile	Hologne	odile.hologne@inra.fr	*+xxx xxxxxxxxx*

PIC	***Legal name***
999977366	*TECHNISCHE UNIVERSITEIT DELFT*

Short name: TU Delft

Address of the organisation

Street STEVINWEG 1

Town DELFT

Postcode 2628 CN

Country Netherlands

Webpage www.tudelft.nl

Legal Status of your organisation

Research and Innovation legal statuses

Public body .....yes

Non-profit .....yes

International organisation .....no

International organisation of European interest .....no

Secondary or Higher education establishment .....yes

Research organisation .....yes

Legal person .....yes

Industry (private for profit).....no

Enterprise Data

SME self-declared status.....20/05/2016 - no

SME self-assessment .....20/05/2016 - no

SME validation sme..... unknown

Based on the above details of the Beneficiary Registry the organisation is not an SME (small- and medium-sized enterprise) for the call.

Department(s) carrying out the proposed work

Department 1

Department name: TU Delft Library ☐ not applicable

☐ Same as proposing organisation's address

Street: Prometheusplein 1

Town: Delft

Postcode: 2628 ZC

Country: Netherlands

Dependencies with other proposal participants

Character of dependence	***Participant***	

Person in charge of the proposal

The name and e-mail of contact persons are read-only in the administrative form, only additional details can be edited here. To give access rights and basic contact details of contact persons, please go back to Step 4 of the submission wizard and save the changes.

Title: Dr.

Sex: (•) Male ( ) Female

First name: **Alastair** Last name: **Dunning**

E-Mail: **a.c.dunning@tudelft.nl**

Position in org.: Head of Research Data

Department: TU Delft Library ☐ Same as organisation name

☐ Same as proposing organisation's address

Street: Prometheusplein 1

Town: Delft Post code: 2628

Country: Netherlands

Website:

Phone: +31 634858734 Phone 2: *+xxx xxxxxxxxx* Fax: *+xxx xxxxxxxxx*

Other contact persons

First Name	***Last Name***	***E-mail***	***Phone***
Marta	Teperek	m.teperek@tudelft.nl	*+xxx xxxxxxxxx*

PIC	***Legal name***
999823912	*OESTERREICHISCHE AKADEMIE DER WISSENSCHAFTEN*

Short name: OEAW

Address of the organisation

Street DR. IGNAZ SEIPEL-PLATZ 2

Town WIEN

Postcode 1010

Country Austria

Webpage www.oeaw.ac.at

Legal Status of your organisation

Research and Innovation legal statuses

Public body .....yes

Non-profit .....yes

International organisation .....no

International organisation of European interest .....no

Secondary or Higher education establishment .....no

Research organisation .....yes

Legal person .....yes

Industry (private for profit).....no

Enterprise Data

SME self-declared status.....14/10/1921 - no

SME self-assessment ..... unknown

SME validation sme..... unknown

Based on the above details of the Beneficiary Registry the organisation is not an SME (small- and medium-sized enterprise) for the call.

Department(s) carrying out the proposed work

Department 1

Department name: Austrian Centre for Digital Humanities ☐ not applicable

☐ Same as proposing organisation's address

Street: Sonnenfelsgasse 19

Town: Vienna

Postcode: 1010

Country: Austria

Dependencies with other proposal participants

Character of dependence	***Participant***	

Person in charge of the proposal

The name and e-mail of contact persons are read-only in the administrative form, only additional details can be edited here. To give access rights and basic contact details of contact persons, please go back to Step 4 of the submission wizard and save the changes.

Title: Mr.

Sex: (•) Male ( ) Female

First name: **Matej**

Last name: **Durco**

E-Mail: **matej.durco@oeaw.ac.at**

Position in org.: Head of technical working group

Department: Austrian Centre for Digital Humanities ☐ Same as organisation name

☐ Same as proposing organisation's address

Street: Sonnenfelsgasse 19

Town: Vienna Post code: 1010

Country: Austria

Website: www.oeaw.ac.at/acdh/

Phone: +43 1515812212 Phone 2: +xxx xxxxxxxxx Fax: +xxx xxxxxxxxx

Other contact persons

First Name	***Last Name***	***E-mail***	***Phone***
Veronika	Gruendhammer	veronika.gruendhammer@oeaw.ac.at	*+43 1515812203*

PIC	***Legal name***
957358476	*Australian National Data Service*

Short name: Australian National Data Service

Address of the organisation

Street Blackburn Rd 680

Town Clayton

Postcode 3068

Country Australia

Webpage www.ands.org.au

Legal Status of your organisation

Research and Innovation legal statuses

Public body .....no

Non-profit .....no

International organisation .....no

International organisation of European interest .....no

Secondary or Higher education establishment .....no

Research organisation .....no

Legal person .....no

Industry (private for profit).....yes

Enterprise Data

SME self-declared status..... unknown

SME self-assessment ..... unknown

SME validation sme..... unknown

Based on the above details of the Beneficiary Registry the organisation is not an SME (small- and medium-sized enterprise) for the call.

Department(s) carrying out the proposed work

Department 1

Department name: Australian Research Data Commons ☐ not applicable

☒ Same as proposing organisation's address

Street: Blackburn Rd 680

Town: Clayton

Postcode: 3068

Country: Australia

Dependencies with other proposal participants

Character of dependence	***Participant***	

Person in charge of the proposal

The name and e-mail of contact persons are read-only in the administrative form, only additional details can be edited here. To give access rights and basic contact details of contact persons, please go back to Step 4 of the submission wizard and save the changes.

Title: Dr.

Sex: (•) Male ( ) Female

First name: **Adrian** Last name: **Burton**

E-Mail: **adrian.burton@ands.org.au**

Position in org.: Director, Data, Policy and Services

Department: Australian Research Data Commons ☐ Same as organisation name

☒ Same as proposing organisation's address

Street: Blackburn Rd 680

Town: Clayton Post code: 3068

Country: Australia

Website: ardc.edu.au

Phone: +61404016407 Phone 2: *+xxx xxxxxxxxx* Fax: *+xxx xxxxxxxxx*

PIC	***Legal name***
972772164	*SISTEMA GMBH*

Short name: SISTEMA GMBH

Address of the organisation

Street TIEFER GRABEN 19 TOP 2

Town WIEN

Postcode 1010

Country Austria

Webpage www.sistema.at

Legal Status of your organisation

Research and Innovation legal statuses

Public body .....no

Non-profit .....no

International organisation .....no

International organisation of European interest .....no

Secondary or Higher education establishment .....no

Research organisation .....no

Legal person .....yes

Industry (private for profit).....yes

Enterprise Data

SME self-declared status.....31/12/2016 - yes

SME self-assessment .....31/12/2016 - yes

SME validation sme.....28/08/2009 - yes

Based on the above details of the Beneficiary Registry the organisation is an SME (small- and medium-sized enterprise) for the call.

Department(s) carrying out the proposed work

No department involved

Department name: *Name of the department/institute carrying out the work.* ☒ not applicable

☐ Same as proposing organisation's address

Street: *Please enter street name and number.*

Town: *Please enter the name of the town.*

Postcode: *Area code.*

Country: *Please select a country*

Dependencies with other proposal participants

Character of dependence	***Participant***	

Person in charge of the proposal

The name and e-mail of contact persons are read-only in the administrative form, only additional details can be edited here. To give access rights and basic contact details of contact persons, please go back to Step 4 of the submission wizard and save the changes.

Title: Mr.

Sex: (•) Male ( ) Female

First name: **Stefano**

Last name: **Natali**

E-Mail: **natali@sistema.at**

Position in org.: Director

Department: SISTEMA GMBH ☒ Same as organisation name

☒ Same as proposing organisation's address

Street: TIEFER GRABEN 19 TOP 2

Town: WIEN

Post code: 1010

Country: Austria

Website: www.sistema.at

Phone: +43 1 8908788

Phone 2: *+xxx xxxxxxxxx*

Fax: *+xxx xxxxxxxxx*

3 - Budget

No	Participant	Country	(A) Direct personnel costs/€	(B) Other direct costs/€	(C) Direct costs of sub-contracting/€	(D) Direct costs of providing financial support to third parties/€	(E) Costs of inkind contributions not used on the beneficiary's premises/€	(F) Indirect Costs / € (=0.25(A+B-E))	(G) Special unit costs covering direct & indirect costs / €	(H) Total estimated eligible costs / € (=A+B+C+D+F+G)	(I) Reimburse-ment rate (%)	(J) Max.EU Contribution / € (=H*I)	(K) Requested EU Contribution/ €
1	Technische Informationsbibliothek (Tib)	DE	532000	11000	20000	0	0	135750,00	0	698750,00	100	698750,00	698750,00
2	Centre National De La Recherche	FR	412148	33400	20000	0	0	111387,00	0	576935,00	100	576935,00	576935,00
3	Umweltbundesamt Gesellschaft	AT	302400	53000	0	0	0	88850,00	0	444250,00	100	444250,00	444250,00
4	Universitaet Bremen	DE	274700	16000	0	0	0	72675,00	0	363375,00	100	363375,00	363375,00
5	Alfred-wegener-institut	DE	243200	19000	0	0	0	65550,00	0	327750,00	100	327750,00	327750,00
6	Norsk Institutt For Luftforskning	NO	41850	8000	0	0	0	12462,50	0	62312,50	100	62312,50	62312,50
7	Universiteit Van Amsterdam	NL	307500	11000	0	0	0	79625,00	0	398125,00	100	398125,00	398125,00
8	Universiteit Maastricht	NL	112500	4000	0	0	0	29125,00	0	145625,00	100	145625,00	145625,00
9	Meteorologisk Institutt	NO	224000	6000	0	0	0	57500,00	0	287500,00	100	287500,00	287500,00
10	E-science European Infrastructure	ES	507000	47000	0	0	0	138500,00	0	692500,00	100	692500,00	692500,00

11	United Kingdom Research And	UK	282748	27000	0	0	0	77437,00	0	387185,00	100	387185,00	387185,00
12	Ilmatieteen Laitos	FI	27500	2000	0	0	0	7375,00	0	36875,00	100	36875,00	36875,00
13	Consiglio Nazionale Delle Ricerche	IT	30000	2000	0	0	0	8000,00	0	40000,00	100	40000,00	40000,00
14	Barcelona Supercomputing Center -	ES	13500	2000	0	0	0	3875,00	0	19375,00	100	19375,00	19375,00
15	Deutsches Klimarechenzentrum Gmbh	DE	415000	8000	0	0	0	105750,00	0	528750,00	100	528750,00	528750,00
16	Eudat Oy	FI	0	0	0	0	0	0,00	0	0,00	100	0,00	0,00
17	Csc-tieteen Tietotekniikan Keskus Oy	FI	100000	4000	0	0	0	26000,00	0	130000,00	100	130000,00	130000,00
18	Institut Francais De Recherche	FR	54594	2000	0	0	0	14148,50	0	70742,50	100	70742,50	70742,50
19	Institut National De La Recherche	FR	156000	2000	0	0	0	39500,00	0	197500,00	100	197500,00	197500,00
20	Technische Universiteit Delft	NL	331926	11000	0	0	0	85731,50	0	428657,50	100	428657,50	428657,50
21	Oesterreichische Akademie Der	AT	88400	3000	0	0	0	22850,00	0	114250,00	100	114250,00	114250,00
22	Australian National Data Service	AU	0	0	0	0	0	0,00	0	0,00	100	0,00	0,00
23	Sistema Gmbh	AT	27500	2000	0	0	0	7375,00	0	36875,00	100	36875,00	36875,00
	Total		4484466	273400	40000	0	0	1189466,50	0	5987332,50		5987332,50	5987332,50

4 - Ethics

1. HUMAN EMBRYOS/FOETUSES		Page
Does your research involve Human Embryonic Stem Cells (hESCs)?	○ Yes ◉ No	
Does your research involve the use of human embryos?	○ Yes ◉ No	
Does your research involve the use of human foetal tissues / cells?	○ Yes ◉ No	
2. HUMANS		**Page**
Does your research involve human participants?	○ Yes ◉ No	
Does your research involve physical interventions on the study participants?	○ Yes ◉ No	
3. HUMAN CELLS / TISSUES		**Page**
Does your research involve human cells or tissues (other than from Human Embryos/ Foetuses, i.e. section 1)?	○ Yes ◉ No	
4. PERSONAL DATA		**Page**
Does your research involve personal data collection and/or processing?	◉ Yes ○ No	61
Does it involve the collection and/or processing of sensitive personal data (e.g: health, sexual lifestyle, ethnicity, political opinion, religious or philosophical conviction)?	○ Yes ◉ No	
Does it involve processing of genetic information?	○ Yes ◉ No	
Does it involve tracking or observation of participants?	○ Yes ◉ No	
Does your research involve further processing of previously collected personal data (secondary use)?	○ Yes ◉ No	
5. ANIMALS		**Page**
Does your research involve animals?	○ Yes ◉ No	
6. THIRD COUNTRIES		**Page**
In case non-EU countries are involved, do the research related activities undertaken in these countries raise potential ethics issues?	○ Yes ◉ No	
Do you plan to use local resources (e.g. animal and/or human tissue samples, genetic material, live animals, human remains, materials of historical value, endangered fauna or flora samples, etc.)?	○ Yes ◉ No	
Do you plan to import any material - including personal data - from non-EU countries into the EU?	○ Yes ◉ No	

Do you plan to export any material - including personal data - from the EU to non-EU countries?	◯ Yes ◉ No	
In case your research involves low and/or lower middle income countries, are any benefits-sharing actions planned?	◯ Yes ◉ No	
Could the situation in the country put the individuals taking part in the research at risk?	◯ Yes ◉ No	
7. ENVIRONMENT & HEALTH and SAFETY		Page
Does your research involve the use of elements that may cause harm to the environment, to animals or plants?	◯ Yes ◉ No	
Does your research deal with endangered fauna and/or flora and/or protected areas?	◯ Yes ◉ No	
Does your research involve the use of elements that may cause harm to humans, including research staff?	◯ Yes ◉ No	
8. DUAL USE		Page
Does your research involve dual-use items in the sense of Regulation 428/2009, or other items for which an authorisation is required?	◯ Yes ◉ No	
9. EXCLUSIVE FOCUS ON CIVIL APPLICATIONS		Page
Could your research raise concerns regarding the exclusive focus on civil applications?	◯ Yes ◉ No	
10. MISUSE		Page
Does your research have the potential for misuse of research results?	◯ Yes ◉ No	
11. OTHER ETHICS ISSUES		Page
Are there any other ethics issues that should be taken into consideration? Please specify	◯ Yes ◉ No	

I confirm that I have taken into account all ethics issues described above and that, if any ethics issues apply, I will complete the ethics self-assessment and attach the required documents. ☒

How to Complete your Ethics Self-Assessment

5 - Call-specific questions

Extended Open Research Data Pilot in Horizon 2020

If selected, applicants will by default participate in the Pilot on Open Research Data in Horizon 2020[1] , which aims to improve and maximise access to and re-use of research data generated by actions.

However, participation in the Pilot is flexible in the sense that it does not mean that all research data needs to be open. After the action has started, participants will formulate a Data Management Plan (DMP), which should address the relevant aspects of making data FAIR – findable, accessible, interoperable and re-usable, including what data the project will generate, whether and how it will be made accessible for verification and re-use, and how it will be curated and preserved. Through this DMP projects can define certain datasets to remain closed according to the principle "as open as possible, as closed as necessary". A Data Management Plan does not have to be submitted at the proposal stage.

Furthermore, applicants also have the possibility to opt out of this Pilot completely at any stage (before or after the grant signature). In this case, applicants must indicate a reason for this choice (see options below).

Please note that participation in this Pilot does not constitute part of the evaluation process. Proposals will not be penalised for opting out.

We wish to opt out of the Pilot on Open Research Data in Horizon 2020.	○ Yes	◉ No

Further guidance on open access and research data management is available on the participant portal: http://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-dissemination_en.htm and in general annex L of the Work Programme.

[1] *According to article 43.2 of Regulation (EU) No 1290/2013 of the European Parliament and of the Council, of 11 December 2013, laying down the rules for participation and dissemination in "Horizon 2020 - the Framework Programme for Research and Innovation (2014-2020)" and repealing Regulation (EC) No 1906/2006.*

KOINE: Interdisciplinary data interoperability, discovery and exploitation in the EOSC

Proposal addressing H2020 call INFRAEOSC-02-2019
Hannover, 28 January 2019

List of participants

No.	Participant organisation name (Acronym)	Country
1	TECHNISCHE INFORMATIONSBIBLIOTHEK (TIB) – (Coordinator)	Germany
2	CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE (CNRS)	France
3	UMWELTBUNDESAMT GESELLSCHAFT MIT BESCHRANKTER HAFTUNG (EAA)	Austria
4	UNIVERSITAET BREMEN (UHB)	Germany
5	ALFRED-WEGENER-INSTITUT HELMHOLTZ-ZENTRUM FUR POLAR- UND MEERESFORSCHUNG (AWI)	Germany
6	NORSK INSTITUTT FOR LUFTFORSKNING STIFTELSE (NILU)	Norway
7	UNIVERSITEIT VAN AMSTERDAM (UvA)	Netherlands
8	UNIVERSITEIT MAASTRICHT (UMAAS)	Netherlands
9	METEOROLOGISK INSTITUTT (MET)	Norway
10	E-SCIENCE EUROPEAN INFRASTRUCTURE FOR BIODIVERSITY AND ECOSYSTEM RESEARCH (LW ERIC)	Spain
11	UNITED KINGDOM RESEARCH AND INNOVATION (UKRI)	United Kingdom
12	ILMATIETEEN LAITOS (FMI)	Finland
13	CONSIGLIO NAZIONALE DELLE RICERCHE (CNR)	Italy
14	BARCELONA SUPERCOMPUTING CENTER - CENTRO NACIONAL DE SUPERCOMPUTACION (BSC)	Spain
15	DEUTSCHES KLIMARECHENZENTRUM GMBH (DKRZ)	Germany
16	EUDAT OY (EUDAT)	Finland
17	CSC-TIETEEN TIETOTEKNIIKAN KESKUS OY (CSC)	Finland
18	INSTITUT FRANCAIS DE RECHERCHE POUR L'EXPLOITATION DE LA MER (IFREMER)	France
19	INSTITUT NATIONAL DE LA RECHERCHE AGRONOMIQUE (INRA)	France
20	TECHNISCHE UNIVERSITEIT DELFT (TUD)	Netherlands
21	OESTERREICHISCHE AKADEMIE DER WISSENSCHAFTEN (OEAW)	Austria
22	AUSTRALIAN NATIONAL DATA SERVICE (ANDS)	Australia
23	SISTEMA GMBH (SISTEMA)	Austria

Contents

1. Excellence

Data interoperability is the hardest of the FAIR Data Principles to translate into practice and must be addressed with formal, machine-readable terminologies. In KOINE, we take on this challenge, which has yet to be addressed at scale.

1.1 Objectives

Global knowledge economies are driven by data openness, accessibility and exchange. As yet, academia and industry do not fully exploit the vast quantities of data (and thus information) available. Significant obstacles include: Natural language ambiguity, the consequent lack of semantic data interoperability and under-performing machine support in tasks ranging from discovering to cleaning and integrating data. Indeed, it is common practice to annotate data with their meanings using free text annotations. Free text annotations are not machine readable and are often ambiguous, even to human experts. Consequently, we have not reached our full potential when it comes to transforming the massive volumes of data produced in observation, experimentation or simulation to information and knowledge.

The research community is addressing these challenges by adopting, among other, the FAIR Data Principles[1] (Findable, Accessible, Interoperable, Re-Usable research data). Some of the principles have seen significant progress in their implementation. A good example is the resolvable persistent identification of data to ensure their findability and accessibility. Interoperability, however, is hindered by ambiguous words, phrases or

[1] Wilkinson, M. D. et al. (2016). The FAIR Guiding Principles for scientific data management and stewardship. Sci. Data 3:160018. https://doi.org/10.1038/sdata.2016.18

incomprehensible abbreviations used to encode semantic annotations for datasets. To overcome this obstacle and pave the way towards interoperability, the adoption of FAIR terminologies[2] (controlled vocabularies, thesauri, ontologies) is a requirement.

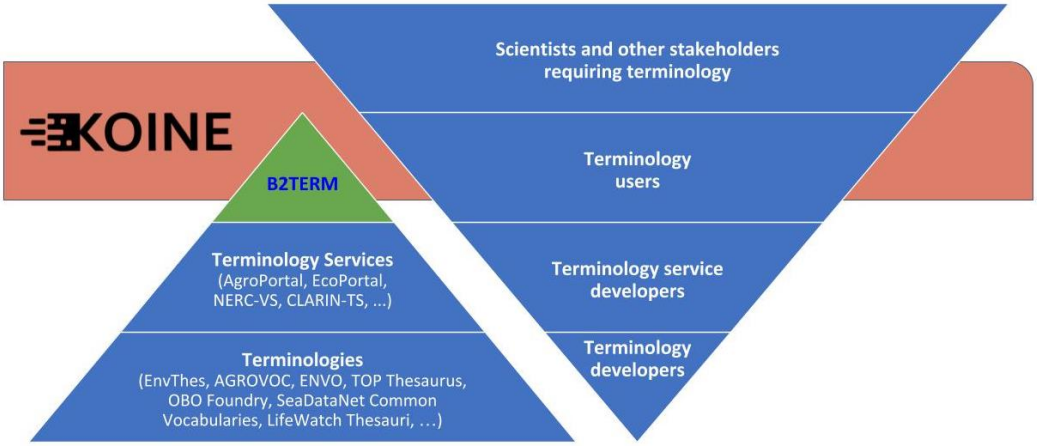


Figure 1: As its key service, KOINE develops B2TERM, the EOSC service for terminology that will harmonize access to, and exchange of, terminology served by established terminology services. By integrating B2TERM into (inter)disciplinary operational infrastructures and diverse communities, KOINE ensures that the many stakeholders (e.g., scientists) requiring terminology are connected with terminology developers (i.e., terminology supply meets terminology demand) and catalyses terminology use. With this sociotechnical integration of terminology and data, KOINE will contribute substantially to improved data interoperability, semantic findability and re-usability in the EOSC.

The overall objective of the KOINE[3] project is to develop **B2TERM[4], a novel and innovative EOSC service for terminology, and to integrate B2TERM into (inter)disciplinary operational infrastructures and diverse communities in research, industry or society.** Figure 1 illustrates that KOINE implements B2TERM on established terminology services and ensures that terminology demand by stakeholders is met by terminology supply in order to catalyse terminology use and significantly contribute to data interoperability in the EOSC. B2TERM will be the EOSC cornerstone that enables **semantic data interoperability**, **discovery and exploitation** across disciplines and thus supports user-centric applications that maximize the scientific and societal impact of research data. Such a service is instrumental in establishing a comprehensive, integrated Internet of FAIR Data and Services[5] through semantic representation, terminology-based interoperability, accessibility and interlinking of research data. Additionally, B2TERM will serve as a foundation for existing and future services of the EOSC ecosystem that rely on terminology exploitation. KOINE exemplifies this, for example, by extending B2FIND, EOSC's

[2] Throughout the document, terminology is used as the overarching name for any set of fixed denotations that are used to describe something with the goal to reduce ambiguity and facilitate interoperability. A terminology can range from a simple controlled vocabulary (a simple list of terms) to a complex ontology (formal definitions of terms and their relations semantically expressed in a machine readable way). This term may also include taxonomies, thesauri or any other kind of knowledge organisation sources. In the document, we will also use the expression "terminology services" to identify portals or platforms serving/hosting terminologies such as ontology repositories (e.g., NCBO BioPortal, AgroPortal, EBI OLS, etc.), terminology services (e.g., EnvThes, CLARIN-TS) or vocabulary services (e.g., NERC-VS, ANDS Research Vocabularies Australia).

[3] In linguistics, a koiné language, koiné dialect, or simply koiné (Ancient Greek κοινή, "common [language]") is a standard language "that has arisen as a result of contact between two or more mutually intelligible varieties (dialects) of the same language" (https://en.wikipedia.org/wiki/Koin%C3%A9_language). We think this is fitting for the project goal to catalyse the use of a common language for interdisciplinary data interoperability, discovery and exploitation in the EOSC.

[4] Named B2TERM, the new EOSC service is envisioned as an element of the service portfolio originally developed during the EUDAT and EUDAT 2020 projects and now included and maintained in the EOSC Service Catalogue.

[5] Mons, B. et al. (2017). Cloudy, increasingly FAIR; revisiting the FAIR Data guiding principles for the European Open Science Cloud. Information Services & Use, 37(1):49-56. https://doi.org/10.3233/isu-170824

existing generic metadata and discovery service for research datasets. B2FIND will be extended with semantics by interfaces to B2TERM and by enriching indexed datasets with annotations and subscriptions[6].

The project's specific objectives are to:

1. **Develop a Terminology Interoperability Framework.** KOINE specifies and implements a Terminology Interoperability Framework (TIF) that (1) harmonizes programmatic persistent access to terminologies served by established terminology services (e.g., NERC-VS[7], EcoPortal[8], AgroPortal[9], CLARIN-TS[10]); (2) develops a Common Semantic Model that allows harmonizing terminologies; and (3) facilitates the exchange of terminology required to describe research data semantics. The TIF determines the Compliance Standards for connecting current and future FAIR terminology to the EOSC.
2. **Create B2TERM as a novel EOSC service for terminology.** KOINE designs, develops and devises a model to sustainably operate B2TERM as a novel and innovative service for terminology in the EOSC that builds on the TIF and provides comprehensive support for terminology use by communities working with research data.
3. **Make B2TERM available to researchers and other stakeholders by establishing the service in the EOSC Service Catalogue.** B2TERM will be an EOSC cornerstone service that (1) is accessible by users outside its original community; (2) is operated in a production environment; (3) is well described and documented; (4) has well-defined access rules; and (5) has helpdesk channels. KOINE establishes B2TERM as an element of the EOSC Service Catalogue.
4. **Integrate B2TERM with disciplinary operational data research infrastructures or e-Infrastructures researchers already use.** KOINE integrates B2TERM with data providers (e.g., PANGAEA, 4TU.Centre for Research Data) and research infrastructures (e.g., ACTRIS, SIOS, eLTER, CLARIN). This integration catalyses the transition from the current practice of using ambiguous labels for describing research data towards using unambiguous, persistently identified and machine-readable terminology.
5. **Integrate B2TERM with interdisciplinary operational data discovery services.** B2FIND's current metadata schema is extended to enable semantic data discovery and exploitation across disciplines for stakeholders including individual researchers, research communities, public sector, industry and educators. Such integration significantly improves the precision and recall of search results, enables annotating indexed data records with terminologies and reduces the resources required in data exploitation activities, e.g., data cleaning, integration, and analysis.
6. **Integration of the Data Subscription Service.** KOINE enhances, customizes and devises a model to sustainably operate the novel and innovative Data Subscription Service (DSS), integrated with B2TERM and B2FIND, to enable semantic and interdisciplinary data subscription and notification. DSS notifies users when their datasets-of-interest are available or have been changed.
7. **Enable exploitation of disciplinary metadata in generic data discovery services.** KOINE implements the syndication of metadata enriched with terminology between disciplinary infrastructures (e.g., PANGAEA) and generic data discovery services (e.g., B2FIND) to ensure that disciplinary-specific data curation is fully exploited in interdisciplinary data discovery and usage activities, allowing interoperable and cross-domain re-use of data.
8. **Embed B2TERM in data producer/consumer communities.** KOINE integrates B2TERM into research communities and other social (human) infrastructure stakeholders (e.g., educators, policy makers, journalists). This supports individual scientists and communities as well as non-academic

[6] B2FIND, http://b2find.eudat.eu, is developed within the EUDAT and EUDAT 2020 projects and is the generic service for 'Metadata based Discovery' in the EOSC-hub service catalogue, https://eosc-hub.eu/catalogue/B2FIND .

[7] NERC Vocabulary Server (http://vocab.nerc.ac.uk/)

[8] EcoPortal is currently in development (as priority for 2019) at LW ERIC. Its goal is to design and develop a repository for the semantic resources in the ecological domain supporting the community in the management and alignment of their semantics. The first release will be published for the end of February 2019 to the address http://ecoportal.lifewatchitaly.eu and will be based on the NCBO technology.

[9] LIRMM AgroPortal (http://agroportal.lirmm.fr/)

[10] CLARIN Terminology Services, specifically CLARIN Concept Registry (https://concepts.clarin.eu/ccr/browser/), CLARIN Vocabulary Service (https://vocabularies.clarin.eu/clavas/), and ACDH Vocabs (https://vocabs.acdh.oeaw.ac.at/en/)

actors in creating high-quality (meta)data, and thus strengthens the role of these stakeholders in the FAIRification of research data. The project thus facilitates collaborative (meta)data annotation by social infrastructures.

9. **Pilot the integration of B2TERM in the research lifecycle.** Finally, KOINE prototypes the integration of B2TERM in relevant phases of the research lifecycle, e.g., design of experiments, data analysis, or scholarly communication as well as in relevant information objects other than (meta)data, e.g., Data Management Plans (DMPs), provenance, software, articles or Electronic Lab Notebooks (ELNs). The project thus allows for open interfaces enabling the use of terminologies by other services supporting the research data lifecycle, e.g., (discipline-specific) DMP tools and ELNs.

Figure 2 provides an overview of the key systems and services KOINE uses and advances, and their roles in supporting the project's specific objectives. The project will also engage with other terminology services (e.g., NCBO BioPortal, EBI OLS, ANDS Research Vocabularies Australia [ANDS-RVA], and SKOSMOS) and additional pan-European or global research infrastructures. The diagram illustrates how interoperability among terminology services in the EOSC builds a foundation for B2TERM and how this novel and innovative EOSC service is integrated with socio-technical infrastructures to catalyse research data interoperability and machine support in semantic data discoverability and exploitation.

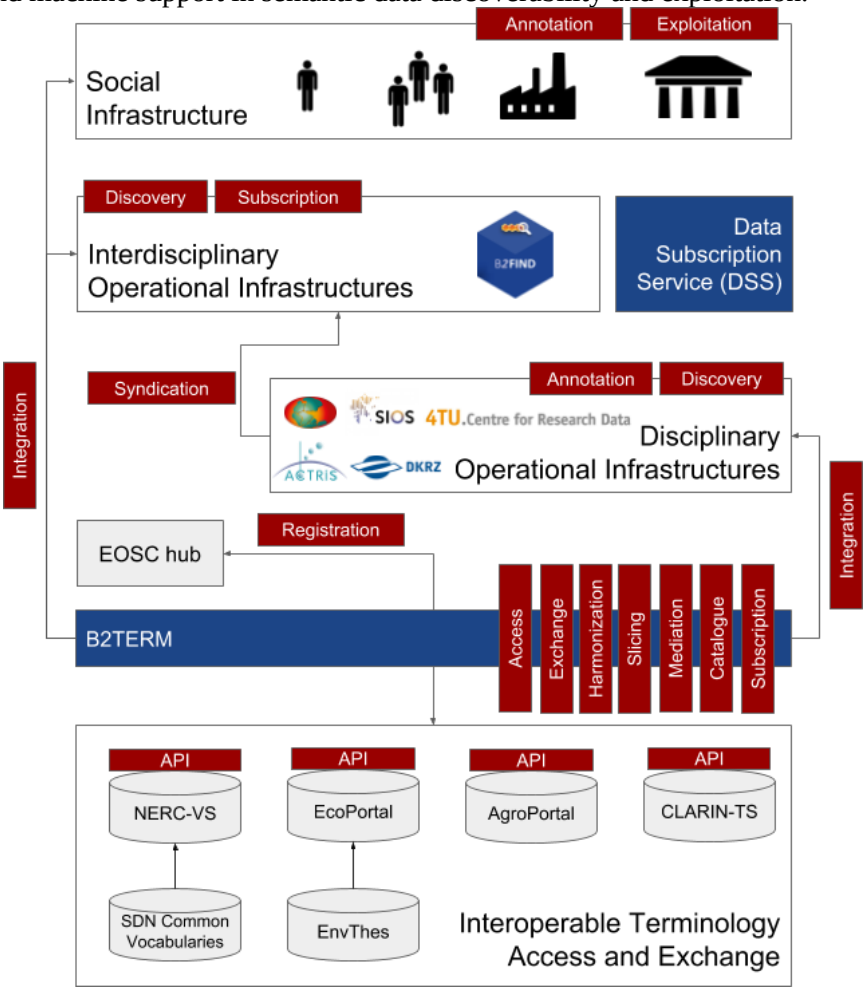


Figure 2: Illustration of selected systems and services and their role in supporting KOINE's objectives. The EOSC-catalogued service B2TERM provides harmonized access to and exchange of terminology served by established community-driven terminology services. These services implement a common API as determined by the KOINE Terminology Interoperability Framework (TIF). Additional B2TERM features include mediation of terminology curation requests, the community-driven creation of terminology slices (views), terminology cataloguing and subscription and notification to changes. As a key aspect, KOINE ensures to integrate B2TERM in (inter)disciplinary operational infrastructures researchers already use. As a second key aspect, KOINE also ensures that B2TERM is integrated in social infrastructure by actively engaging stakeholders (e.g., researchers, industry, educators) and thus drive the adoption of terminology in the EOSC to ensure, in particular, data interoperability, for humans and machines.

1.2 Relation to the work programme

KOINE addresses the INFRAEOSC-02-2019 Call *Prototyping new innovative services* of the H2020 Work Programme European Research Infrastructures (including e-Infrastructures).

By developing a novel service for interdisciplinary data interoperability and integrating the service in (inter)disciplinary operational infrastructures (WP4) and research communities (WP5), KOINE provides an essential component for the realization of the EOSC vision.[11] **Data interoperability is a key challenge in the EOSC and up to now this aspect of the FAIR Data Principles has been addressed poorly**. Indeed, according to the EOSC Declaration,[12] "Research data must be both syntactically and semantically understandable, allowing meaningful data exchange and reuse among scientific disciplines and countries." KOINE builds the foundations and develops the EOSC cornerstone service that is urgently needed to make semantic data interoperability in the EOSC a reality and catalyse the effective use of global scientific data. By tackling interoperability, KOINE directly impacts on semantic data findability and re-usability aspects as well. In alignment with the EOSC vision, KOINE puts great emphasis on enabling actual usage of the service. This will ensure that the project not only advances technical infrastructure but also research community literacy in using innovative technologies in research lifecycles and in conducting state-of-the-art research data management.

KOINE aligns with the INFRAEOSC-02-2019 call objectives and more generally with the EOSC Implementation Roadmap[13]. Table 1 details how KOINE addresses specific challenges and objectives of the INFRAEOSC-02-2019 call. Table 2 details how KOINE aligns with the EOSC Declaration[14].

KOINE sees many ongoing and upcoming projects as major enablers of the EOSC and will pursue an active engagement strategy with them. A list of relevant ongoing and upcoming projects, presented with the focus and relevant action line they address, is presented in Table 3.

A number of research communities from different scientific domains have joined the consortium as full partners and beneficiaries, through institutions and infrastructures that officially represent the community in the consortium. They bring specific requirements and knowledge as well as commitment to pilot and take up B2TERM within their own research infrastructure or e-Infrastructure. KOINE will focus on facilitating the uptake of B2TERM by these communities by actively involving community experts in the development, implementation and deployment of KOINE technologies and approaches as well as by dedicated training and outreach activities. By exposing KOINE solutions to others, we expect to enrich developed solutions so that they match the needs of a broad range of stakeholders and demonstrate the potential of the services for EOSC. The selected communities and the foreseen joint activities are described in Table 4.

Table 1: How KOINE addresses INFRAEOSC-02-2019 specific challenges and objectives.

Specific challenges of INFRAEOSC-02-2019	**How KOINE addresses the challenge**
Develop an agile, fit-for-purpose and sustainable service offering	We will develop and sustain the innovative service B2TERM, which will be offered as an EOSC service. B2TERM enables essential interoperability features for research data (and other scientific artefacts) by supporting data producers and data consumers in the agile use, adaptation, harmonization, and curation of existing terminologies. B2TERM will be co-developed with communities in an agile manner taking community requirements and feedback into account. The project coordinator, TIB, is committed to sustain the service beyond the project.
Service offering is accessible through the EOSC hub	B2TERM will be made available as a generic service through EOSC hub and will be registered in the **EOSC Service Catalogue** at the EOSC Portal (https://eosc-portal.eu/).

[11] https://ec.europa.eu/research/openscience/pdf/realising_the_european_open_science_cloud_2016.pdf
[12] https://ec.europa.eu/research/openscience/pdf/eosc_declaration.pdf
[13] https://ec.europa.eu/research/openscience/pdf/swd_2018_83_f1_staff_working_paper_en.pdf
[14] https://ec.europa.eu/research/openscience/pdf/eosc_declaration.pdf

Service offering satisfies the evolving needs of the scientific community	As the **need for multidisciplinary research** increases, the urgency of data interoperability increases, too. KOINE directly addresses this need. The development, implementation and long-term sustainability of the **B2TERM service together with its integration in (inter)disciplinary data infrastructures and research communities** will represent the major outcome of the KOINE project.
Design and prototype novel innovative digital services	KOINE develops a unique and novel service, namely **B2TERM**, for terminology exploitation within EOSC. The project validates the service with community-driven use cases and demonstrates the impact of leveraging B2TERM within research infrastructures and B2FIND, an important and existing EOSC service.
Innovative collaboration models with incentive mechanisms for user oriented open science	Through B2TERM, users will be able to identify and contribute **community best practices for the curation and use of terminology**, raising the general quality and interoperability of research data and services. In particular, B2TERM enables the collaborative curation of terminology in the EOSC and significantly eases stakeholder contributions. In addition, semantic data discovery, e.g., by the integration of B2TERM in B2FIND will **facilitate effective and more frequent re-use of available research outputs, thus stimulating the transition to open science across all disciplines**.
Target gap in service offering of the EOSC hub	A fundamental requirement for efficient interdisciplinary data discovery and exploitation is semantically-rich data. There is currently no service for semantic interoperability at the interdisciplinary, EOSC-wide level. **B2TERM aspires to catalyse research data interoperability in the EOSC by adopting persistently identified and unambiguous terminology in describing research data semantics.**
Develop innovative services that address relevant aspects of the research data cycle	It is of great importance to generate high quality (meta)data early in the research data lifecycle. To meet this requirement, the need for semantics has been clearly established by research infrastructures and research communities in virtually all scientific domains. The **B2TERM service is relevant for all stages of the research data lifecycle to semantically improve research data** with persistently identified and harmonised terminologies.
Allow for implementation of new scientific data-related developments	The use of precise terminology is fundamental to high-quality data science; the service provided by KOINE will underpin an **entire spectrum of innovative research activities** in the future.
Intelligent linking and discovering of all research artefacts	KOINE will significantly **enhance interdisciplinary linkage, precision and recall** of existing data discovery services such as B2FIND by enriching metadata with semantic annotations. This will enable semantic and cross-border data discovery and exploitation within and across disciplines for stakeholders including individual researchers, research communities, public sector, industry and educators.
Initially the service offering needs to respond to specific needs of particular scientific communities	KOINE will focus the development of its **initial service offering in partner communities from within the project (ecosystem and biodiversity, marine and cryosphere, atmospheric and climate sciences)**, as well as services of relevance to participating data centres/publishers, research infrastructures and e-Infrastructures.
At the end of the project the service offering is leveraged to foster	KOINE will leverage networks of the consortium partners to ensure uptake of the service beyond the initial KOINE partner communities. In particular, coastal engineering, agronomy, humanities and social sciences will be engaged in the

interdisciplinary research	second stage of the project. Furthermore, the consortium is connected to numerous other domains (e.g., health, biomedicine). Interdisciplinary data discovery services as well as interdisciplinary data consumers (urban studies) are an additional set of B2TERM stakeholders. The use of unambiguous terminology is particularly relevant in an interdisciplinary setting, since descriptive labels used in a discipline are often unintelligible by researchers from a different discipline. **By including partners from different science domains in the co-development and uptake of B2TERM it is ensured that the service will address the needs from various science domains to facilitate interdisciplinary research.**
Service offering serves new users like industry and the public sector	Similar to other EOSC services, B2TERM will be accessible and provide value to a broad range of stakeholders, including industry and the public sector. In particular, **semantic data discovery will support users from outside of academia in accessing and re-using data**. More generally, better interoperability of open scientific data will have an important impact on industry and the public sector, inside EU and internationally. This is one of the main motivations behind the European Open Science roadmap.
Scalability tested by user communities from different disciplines	KOINE involves **23 partners** representing different scientific communities, including earth and environmental sciences (multiple domains), agronomy, engineering, humanities and social sciences. Each of these communities will test, evaluate and validate aspects of KOINE's outcomes.
Service offering based on systems and technologies that have reached TRL 6 before the start of the project and will be brought to at least TRL 8 by the end of the project	**KOINE builds B2TERM on existing technologies and providers that are already at TRL 6-7 (e.g., AgroPortal, NERC-VS, and CLARIN-TS) and integrates B2TERM in existing services at TRL 6-9 (e.g., PANGAEA, B2FIND).** The TRL of B2TERM will progress during the project with its integration in disciplinary operational data research infrastructures (e.g., ACTRIS, SIOS, eLTER, CLARIN) and interdisciplinary operational data discovery services (e.g., B2FIND). These integrations will contribute to lifting the TRL of some of the existing technologies and providers from TRL 6-8.
Demonstrate how the service offering complements, enriches and could potentially be integrated into the EOSC hub	B2TERM will be part of the well-established EUDAT portfolio, which provides a core set of services for the EOSC Catalogue. **KOINE will benefit from EUDAT's expertise as the B2TERM service will be offered as a generic service through the EOSC hub.** As such, the service will be easily integrated within other infrastructure services. To demonstrate this, the B2TERM service will be integrated within the EOSC's B2FIND discovery service and other community data services.
Consider accessibility requirements set under the projects funded under EINFRA-12-2017 topic that may be available at the time the call will be open, in view of their integration into the mainstream services of the EOSC hub	The relation of KOINE to the projects EOSC-hub, OpenAIRE Advance and FREYA is shown in Table 3.

Involve SMEs	EUDAT Ltd. and SISTEMA[15] GmbH are KOINE partners. SISTEMA is an SME in Environmental Information Mining, and will evaluate if semantic research data supports information mining. KOINE will also engage with Dutch SME companies (e.g., Evides[16]), which are involved in water management (as part of its consultations with coastal engineering communities) in gathering the requirements and validating KOINE outputs. In addition, the consortium will engage with SMEs through Deltares[17] (closely connected with TU Delft), which is an independent applied research institute in coastal engineering, and which works very closely with various businesses.

Table 2: How KOINE addresses applicable aspects of the EOSC Declaration.

EOSC Declaration Aspect	**How it is addressed by KOINE**
Data culture and FAIR data	
Data culture	KOINE significantly contributes to state-of-the-art research data curation throughout the research lifecycle and supports a common culture of data stewardship.
Open access by-default	KOINE's service and functionalities will be offered openly, accessibly and sustainably through EOSC Service Catalogue so that they can be used and adopted by any research community, SMEs and public organisations.
Skills	Through research community engagement and capacity building (WP5), KOINE supports the development of necessary skills in research data management.
Data stewardship	By ensuring that research data are widely comprehensible and correctly interpreted by both humans and machines, KOINE facilitates a better and higher quality data stewardship.
Rewards and incentives	Researchers should be rewarded for creating FAIR datasets. However, before this can happen, tools and services need to be provided so that research datasets can be easily made FAIR. KOINE services are therefore part of the necessary ecosystem needed to start rewarding researchers for creating reusable research data.
FAIR principles	KOINE tackles a neglected aspect of the FAIR principles, namely data interoperability and directly impacts on data findability, accessibility and reusability.
Standards	KOINE services and approaches are underpinned by global and open standards, such as languages for knowledge representation recommended by W3C and OGC.
FAIR Data governance	The design and implementation of KOINE's service and functionalities are built upon inclusive stakeholder participation (researchers from different scientific disciplines). Furthermore, KOINE supports making FAIR data a reality by extensive education and training as well as by establishing a related RDA WG *Harmonizing FAIR Descriptions of Observational Data* that will act as a global platform for knowledge exchange.
Implementation & transition to FAIR	By catalysing terminology-based data annotation in (inter)disciplinary operational infrastructures as well as in user communities, thus enabling data interoperability and

[15] http://sistema.at
[16] https://www.evides.nl/english
[17] https://www.deltares.nl/en/

	semantic data discovery and exploitation, KOINE contributes substantially to implementing and transitioning to FAIR.
Research data repositories	KOINE advances the annotation of research data in operational data centres and publishers with unambiguous terminology thereby contributing significantly to the implementation of the FAIR data principles in trusted research data repositories that ensure long-term sustainability of research data across disciplines.
Technical implementation: Semantic layer	This aspect of the EOSC Declaration is the key challenge addressed by KOINE.
Data expert organisations	KOINE uses the Research Data Alliance and other organisations active in the research communities (e.g., ESIP) for knowledge exchange, to reach consensus and maximise impact at the global scale.
Research data services and architecture	
EOSC architecture	KOINE develops a service and service integrations that support a data infrastructure commons that serves the needs of scientists. The project will federate existing terminology services and makes terminology discoverable and available through B2TERM at pan-European level.
Implementation	Stakeholders will participate iteratively to drive the development of KOINE services. With B2TERM and its integrations, KOINE contributes substantially to advancing EOSC functionality, especially regarding the provision of core common services and the definition of minimum quality standards of service (e.g., through the Terminology Interoperability Framework). Furthermore, it advances the capabilities of data cataloguing as well as those in computing/analytic services and tools.
Legacy	KOINE builds on existing building blocks (e.g., research infrastructures and data centres), state-of-the-art services and solutions (e.g., terminology services, B2FIND, and DSS) delivered by past and ongoing projects at a national, European and international levels.
User needs	B2TERM is the one-stop-shop to find, discover and access terminology in the EOSC, built on existing terminology services maintained and provided by scientific communities. To meet user needs, KOINE ensures that service implementation and its integrations in data repositories and discovery services are user-driven.
Service provision	KOINE's service and functionalities will be offered at highest Technology Readiness Levels (TRLs) and kept future-proof through a sustainability plan (WP6). B2TERM will be operated sustainably by TIB.
Thematic areas	KOINE covers numerous thematic areas including environment (marine, atmosphere, climate), food, humanities, and engineering. Via the TIF a common method will be established which enables the discovery, access and sharing of terminology across all scientific disciplines.
Research infrastructures	KOINE deeply integrates several research infrastructures including ESFRIs (LW ERIC, ACTRIS, eLTER, CLARIN ERIC, SIOS) both for technical integrations in data centres as well as for their role as 'the stewards of the community of standards'.

Table 3: Provisional list of EOSC-relevant projects.

H2020 Call	Title (Project/Call)	Start	Related activities	KOINE partners involved
Ongoing projects				
EINFRA-12-2017	EOSC-hub	Q1/2018	B2TERM integration with the marketplace and EOSC core functions provided by the hub and with B2FIND, which is positioned as central data discovery service of EOSC-hub. Furthermore, coordination of training and events to extend the training portfolio of EOSC-hub on data interoperability.	CSC, BSC, DRKZ, EAA, CNRS-AERIS, IFREMER
EINFRA-12-2017	OpenAIRE Advance	Q1/2018	Collaboration on FAIR & interoperability practices and training activities related to open access and data management.	UHB
EINFRA-21-2017	FREYA	Q1/2018	Via the FREYA PID Forum, collaboration on the role of persistent identifiers for unambiguous reference to terminology and connecting terminology with the FREYA PID Graph.	UHB
INFRASUPP-03-2016	eInfraCentral	Q1/2017	Re-use and contribution to the development of eInfraCentral Service Description Template.	EUDAT
INFRASUPP - 2017-1	RDA-Europe	Q1/2018	Participation of project partners in relevant RDA working groups, in particular on semantic interoperability. RDA WG/IG and their outcomes that will be considered include (among others) Vocabulary and Semantics Services IG[18], Agrisemantics WG[19], Data Type Registry WG[20], and Data Foundations and Terminology IG[21]. Moreover, selected partners will hold a P13 BoF[22] for a prospective WG on harmonisation of research observation semantics.	TIB, UHB, EAA, UKRI, AWI, LW ERIC, CNRS, SIOS
INFRAEOSC-04-2018	SSHOC	Q1/2019	Social Sciences & Humanities Open Cloud, tasks on multilingual terminologies, ontologies and vocabularies used in SSH domain, making SSH data FAIR, semantic annotation of Heritage Science Data	OEAW
Upcoming projects (INFRAEOSC & other relevant calls)				
INFRAEOSC-	Connecting	Q1/2019	Collaboration on semantics (coordination of	EAA,

[18] https://www.rd-alliance.org/groups/vocabulary-services-interest-group.html
[19] https://www.rd-alliance.org/groups/agrisemantics-wg.html
[20] https://www.rd-alliance.org/groups/data-type-registries-wg.html
[21] https://www.rd-alliance.org/groups/data-foundations-and-terminology-ig.html
[22] https://rd-alliance.org/groups/harmonizing-fair-descriptions-observational-data

04-2018	ESFRI Infrastructures through cluster projects		domain specific input), dissemination of project results and solutions tested by research communities, e.g. with selected KOINE partners in the ENVRI-FAIR cluster to provide customised consultation and support to RIs for FAIR data and service co-design as well as in the SSHOC cluster via OEAW. We will also engage with the other clusters – ESCAPE, EOSClife, PANOSC – to cover the widest range of disciplines.	TIB, LW ERIC, NILU, CNRS-AERIS, IFREMER, CNR, FMI, SIOS, UvA, INRA
INFRAEOSC-05- 2018-2019	Support to the EOSC governance	Q3/2019	Collaboration at policy level, dissemination of project output relevant to EOSC	
	FAIR data uptake and compliance in all scientific communities	Q3/2019	Collaboration on FAIR data. The proposal includes dedicated activities related to FAIR and has an ambitious engagement programme with repositories to improve their awareness of FAIR and their data management practices. These activities nicely complement the activities of the FAIRsFAIR project.	
INFRAEOSC-06-2020:	Enhancing the EOSC portal and connecting thematic clouds (a) Support to the EOSC portal (b) Connecting thematic clouds into the EOSC	Q1/2020	Integration of B2TERM with EOSC portal marketplace.	
Ongoing submissions to important EOSC Implementation Roadmap calls - *our consortium is involved in proposal currently being evaluated for both the Food Cloud and Blue Cloud.*				
DT-SFS-26-2019	Food & Nutrition Security Data Commons (FNS-DC)	Q1/2020	FNS-DC will contribute to the FOOD 2030 vision by engaging European and international FNS research communities into EOSC, through the conception and development of a distributed e-Infrastructure for open science and FAIR data. KOINE will engage with this initiative on semantic interoperability aspect relevant to this community and seek to integrate their input into our terminology service	INRA, CNRS-LIRMM
BG-07-2019-2020 -	The Future of Seas and Oceans Flagship	Q3/2019	The Blue Cloud initiative aims to address the need to build and demonstrate the Pilot Blue Cloud as a thematic EOSC cloud to support research for understanding & better manage the many aspects of	UKRI

	Initiative		ocean sustainability. We will engage with this initiative on semantic interoperability aspect relevant to this community and seek to integrate their input into our terminology service.	

Table 4: KOINE's relation to research infrastructures and initiatives.

Initiative	**Discipline**	**Connections with research infrastructures (ESFRIs and others) and other national or international research initiatives** KOINE will receive from (→) or provide to (←)	**WPs**	**Partners involved**
ACTRIS	Earth and Environmental Sciences, Atmospheric Physics	→ ACTRIS will contribute with its expertise on data curation for observation of atmospheric parameters → ACTRIS will provide domain specific expertise on instruments needed in creating terminology ← Via B2TERM, KOINE will provide terminology to be used in the next generation data portal of ACTRIS ← KOINE will transfer know-how on terminology construction to be used in ACTRIS	4,5	NILU, MET, FMI, CNR, BSC, CNRS-AERIS
eLTER	Earth and Environmental Sciences, Ecosystem Research	→ eLTER will provide requirements and practical experience on the harmonisation of parameters as well as providing a terminology (EnvThes) → eLTER will provide practical experience in organising community processes for the establishment and alignment of terminologies → eLTER will test semantic services and integration of semantic resources within the eLTER Information Infrastructure → eLTER will provide DEIMS-SDR as test bed for site and dataset documentation ← KOINE will provide B2TERM which will be used in DEIMS-SDR to enable metadata annotation and discovery ← KOINE will provide interaction with a wider user community from different domains	2,3,4,5	EAA, UKRI
IS-ENES	Earth and Environmental Sciences, Climate Science	→ IS-ENES will provide requirements from the Earth System Modelling Community for B2TERM → IS-ENES will exemplary test B2TERM within an encapsulated ESGF-environment → IS-ENES will show in how far B2TERM can be adopted within the global ESGF ← KOINE will improve the interoperability of IS-ENES data within an interdisciplinary context by adopting B2TERM ← KOINE contributes to an enhanced acceptance of FAIR data principles	4,5	DKRZ
PANGAEA	Earth and Environmental Sciences, Data Publisher	→ PANGAEA, the data infrastructure managed by UHB and AWI, will provide practical insights based on experiences made within national infrastructures (GFBio & de.NBI) to support the development of B2TERM → PANGAEA will provide actual settings in which	4	UHB

		KOINE's B2TERM will be applied and tested in a 'real-life scenario' ← KOINE will provide B2TERM, which will be used by PANGAEA to produce high quality of metadata annotations and crosslinking of terms to users and to improve the data discovery on the its portal ← KOINE will provide resources to PANGAEA to test new applications, e.g., data analytics ← Interaction with data providers from various domains in particular marine and environmental sciences		
LW ERIC	Biodiversity and Ecosystem Sciences e-infrastructure	→ LifeWatch will provide requirements and practical experience on the harmonisation of parameters as well as providing a terminology resource (LifeWatch Thesauri and ontologies) → LifeWatch will provide a framework to turn biodiversity and ecosystem data into semantic, interoperable, actionable, open knowledge → LifeWatch will provide methods for the management, alignment and service of semantic resources in biodiversity, ecosystems, and in general in ecology with the EcoPortal ontology repository → LifeWatch will provide workflows for the production of FAIR and semantically annotated –with ontologies – linked open ecology data → LifeWatch will test semantic services and integration of semantic resources → Lifewatch will coordinate all the communication and dissemination activities ← KOINE will provide B2TERM which will be used in LifeWatch Data Portal to enable metadata annotation and discovery ← KOINE will contribute to an enhanced acceptance of FAIR data principles ← KOINE will enhance the interaction with related EU research infrastructures	2,3,4,5,6,7	LW ERIC
SIOS	Earth and Environmental Science, Arctic Research	→ SIOS will contribute with expertise in data curation of meteorological, oceanographical and cryospheric parameters → SIOS will contribute with expertise on harmonisation of parameters through engagement in WMO efforts as well as Polar Semantics activities ← KOINE will provide B2TERM which will be evaluated and exploited in SIOS search and evaluated for integration in data submission services ← KOINE will provide interaction with a wider community for semantic integration across disciplines	4,5	MET, NILU
EUDAT Ltd.	e-Infrastructure	→ EUDAT Ltd. provides support with the integration of KOINE services in the EOSC portal → EUDAT Ltd. supports the KOINE B2TERM service as a sustainable service as part of the EUDAT Service Portfolio	4,6	EUDAT

		← KOINE service will extend the discoverability and exploitation of research data on basis of terminology ← KOINE contributes to the interoperability and further FAIRification of research data within maintained within the EUDAT Collaborative Data Infrastructure		
EOSC-hub	e-Infrastructure	→ EOSC hub provides a catalogue of services to register B2TERM → EOSC hub ensures the visibility of B2TERM and supports its adoption in the EOSC ← KOINE will extend the EOSC hub catalogue of services	6	EUDAT
ENVRI-FAIR	Earth and Environmental Sciences, Cluster Project	→ ENVRI-FAIR will provide a forum for interacting with a cluster of environmental RIs on terminology needs. → ENVRI-FAIR will develop metrics for evaluating FAIRness of data, which is also applicable to terminologies. ← KOINE will provide a standard terminology service for environmental RIs to use. ← KOINE will contribute to ENVRI pilot studies with overlapping concerns, such as data subscription.	2, 3, 4, 5, 6, 7	TIB, EAA, NILU, UvA, LW ERIC, CNRS-AERIS, UKRI
SSHOC	Social Sciences and Humanities	→ SSHOC will share consolidated landscaping over data repositories in SSH domain, with focus on terminologies used to describe research data → SSHOC will provide guidelines on making SSH data FAIR	4,5	OEAW
CLARIN	Social Sciences and Humanities	→ CLARIN will contribute terminologies and terminology services → CLARIN will extend existing terminology services with B2TERM-compliant API enabling their integration → CLARIN will contribute expertise on terminology creation & curation ← Via B2TERM, KOINE will provide terminology to be used in ARCHE one of the data portals of CLARIN ← KOINE will transfer know-how on terminology construction to be used in CLARIN	3,4,5	OEAW
DARIAH	Humanities	→ DARIAH will contribute terminologies and terminology services (via OEAW's ACDH Vocabs service) → DARIAH will extend existing terminology services (ACDH Vocabs service) with B2TERM-compliant API enabling their integration → DARIAH will contribute expertise/experience on terminology curation (Working Group Thesaurus Maintenance)	3,4,5	OEAW
FAIRsFAIR	Interdisciplinary	→ Will develop and support the implementation of a common scheme to ensure data development, wide uptake of and compliance with FAIR data principles and practices through the EOSC		UHB

		→ Provide a sustainable certification mechanism for trustworthy repositories of FAIR research data ← KOINE will supply use cases for the application & alignment of the interoperability of FAIR research resources & repositories		
ELIXIR	Life Sciences	→ ELIXIR will link to life science communities → KOINE will benefit from ELIXIR expertise and experience ← KOINE will engage and seek input from relevant ELIXIR core resources (e.g., EBI Ontology Lookup Service) ← Members of KOINE will participate in ELIXIR events, such as the BioHackathon, to further develop and promote KOINE services	5	UMAAS, INRA
French ANR D2KAB (d2kab.org)	Informatics (Semantic Web), Agronomy, Biodiversity	→ A framework to turn agronomy and biodiversity data into –semantically described, interoperable, actionable, open– knowledge → An important national network (30 pers.) of agronomy and biodiversity semantic experts (INRA, IRSTEA, CNRS-CEFE) → Methods for the management, alignment and service of semantic resources in agronomy, agriculture, plant sciences, food and biodiversity with the AgroPortal ontology repository → Workflows for the production of FAIR and semantically annotated –with ontologies– linked open agri-food data → Expertise on ontology repositories and semantic resources management. ← Interaction with multiple related EU research infrastructures ← EOSC service to leverage impact of national project at EU level	2,3,4	CNRS-LIRMM, INRA
GFBio (gfbio.org)	Biodiversity	→ Comprising 20 partners in Germany and data archives in the environmental, collection, and genome data domain, GFBio will link its data services to B2TERM → Will supply additional terminologies through the GFBio terminology service (https://terminologies.gfbio.org/) ← KOINE will provide B2TERM, which will be used by GFBio to produce high quality of metadata annotations and crosslinking of terms to users to improve data integration and discovery on the its portal	3	UHB
SeaData Cloud	Earth and Environmental Sciences, Marine Sciences	→ The SeaDataNet pan-European infrastructure for marine and ocean data comprises over 100 marine data centres and research institutes from 34 countries and will be connected to EUDAT services by 2020/Q1 (H2020-INFRAIA-2016-2017) ← KOINE will gain access to SDN Common Vocabularies through the NERC-VS	3,4	UKRI

		← Access to SDN common Vocabularies in KOINE B2TERM will enhance SDN data discoverability in the EOSC via B2FIND		
ANDS	e-Infrastructure	→ ANDS will provide KOINE access to Australian research communities and infrastructures → ANDS will trial interoperability of Australian content, metadata, and/or tools with B2TERM from the perspective of global standards ← KOINE will benefit from ANDS terminology service integrations in B2TERM ← KOINE and EOSC will benefit from global integrations	2,5	ANDS
SISTEMA	Earth and Environmental Sciences	→ SISTEMA will provide requirements and use cases for stakeholders other than research communities → SISTEMA will test and validate B2TERM and KOINE outputs ← KOINE will provide SISTEMA with highly interoperable data that will make information mining more efficient.	5	SISTEMA

1.3 Concept and methodology

Given the importance of unambiguous and persistently identified terminology to data interoperability, we argue that the EOSC will not attain FAIR data without a generic terminology service which eases the use of terminologies by and across EOSC stakeholders. This service shall provide unified, harmonized and simplified access to terminology across scientific disciplines.

(a) Concept

1.3.1 Overall concept

Research data are rapidly being produced in such big volumes that it challenges our ability to transform data into open, semantically rich and automatically exploitable information and knowledge. To tackle this challenge, the scientific community is embracing the **FAIR** data principles.

Some aspects of the **FAIR** data principles, in particular **metadata findability and accessibility**, have received considerably more attention than others (in particular, data interoperability and reusability). As a result, their implementation is today more advanced. Indeed, metadata describing digital objects often follow (community) standards (e.g., DataCite Schema[23], DCAT[24], Schema.org[25], etc.) and are uniquely and persistently identified and retrieved using standardized communication protocols (e.g., HTTP Content Negotiation and OAI-PMH). Important examples are the services provided by Crossref and DataCite for the governed persistent identification of articles and (primarily) datasets, respectively, and the systematic registration of metadata following open schemas. Such infrastructures have significantly contributed to the findability, accessibility and to some extent interoperability of metadata. By supporting the resolution of persistent identifiers, the infrastructures have also contributed to the findability and accessibility of the objects themselves, notably digital objects such as articles and datasets, or digital representations of (i.e., metadata about) physical entities, such as the IGSN[26] (International Geo Sample Number) for samples or ORCID[27] (Open Researcher Contributor Identification Initiative) for people. Certain principles of **reusability** have also advanced significantly in their implementation. Specifically, the importance of releasing metadata

[23] https://schema.datacite.org/
[24] https://www.w3.org/TR/vocab-dcat/
[25] https://schema.org/
[26] http://www.igsn.org/
[27] https://orcid.org/

and data with clear and accessible usage licenses is commonly accepted and increasingly broadly implemented among research infrastructures. The implementation of provenance is also underway.

The elusive principle is data interoperability. Arguably the hardest principle to translate into practice, the key to data interoperability is the use of **FAIR terminology (vocabulary)**. Defying the principle, data semantics continue to be represented using natural language words or phrases. Indeed, data and their descriptive "column headers" routinely use textual labels (such as "sea surface temperature", "prunus persica", "tropical storm", or "CTDCond [S/m]") instead of references to unambiguous and persistently identified terms.

There exist numerous general and domain-specific terminologies. The Climate and Forecast (CF) Standard Names[28], SeaDataNet Common Vocabularies[29] served by the NERC-VS, EnvThes Environmental Thesaurus[30] or AGROVOC thesaurus[31] are few among many concrete examples of discipline-specific terminologies that can be used to describe – with less ambiguity and in machine readable form – research data semantics (e.g., observed properties, methods, units, etc.). Terminologies can be placed at different levels along the semantic gradient/spectrum[32,33]. Thus, controlled vocabularies, taxonomies, thesauri, and ontologies have different degrees of expressiveness. It is generally easier to create and use a taxonomy than an ontology. Harmonizing, aligning or otherwise integrating terminology of heterogeneous expressiveness are complex processes. Furthermore, terminologies reside on distributed and heterogeneous systems and are aggregated differently by domain-specific terminology services (such as AgroPortal[34] and NCBO BioPortal[35]). By hosting and serving terminologies, such portals facilitate discovery, but using such resources remains a challenge as these platforms target knowledge engineers with skills in semantics technologies. Indeed, while understood by experts (e.g., data managers, data infrastructures), the use of terminologies is not easily understood and applied by non-experts (e.g., scientists). The agile access and exploitation of terminologies as well as tracking the evolution of individual terms (e.g., to record changes needed or made by users, for instance to meet application requirements) are further challenging activities. Furthermore, there is considerable variation in formats, size, and structure of these resources, even among those that address the same domain. This heterogeneity and the high number of published terminologies have led to new problems, e.g., their description, selection, evaluation, trust, and interconnection. **The complexity, multitude, diversity of these resources and the challenges of applying them, however, overwhelm most users** which, in practice, often employ a single resource that, however, does seldom satisfactorily meet their needs.

We respond to this challenge by proposing the project KOINE: Interdisciplinary data interoperability, discovery and exploitation in the EOSC.

KOINE specifies and implements a **Terminology Interoperability Framework** (TIF) that aims to harmonize access to and exchange of information about terminology in the EOSC. The TIF determines the Compliance Standards for terminology services in the EOSC to lower the barriers for discovery, access, exchange and harvesting of terminologies. KOINE aims to establish the TIF as a requirement to register such services in the EOSC Service Catalogue. The TIF is expected to enable persistent access to terminology over HTTP and HTTP content negotiation (i.e., web pages for humans, JSON-LD content for machines) for terminologies and terminology services in the EOSC and will introduce a Common Semantic Model that allows harmonizing terminologies, thus ensuring that – in the EOSC – human and machine readable descriptions of terminology follow a shared model. Furthermore, the TIF will be set-up along the guidelines

[28] http://cfconventions.org/

[29] https://www.seadatanet.org/Standards/Common-Vocabularies

[30] http://vocabs.ceh.ac.uk/evn/tbl/envthes.evn

[31] http://aims.fao.org/vest-registry/vocabularies/agrovoc

[32] McGuinness, D. (2003). Ontologies Come of Age. In Dieter Fensel, Jim Hendler, Henry Lieberman, and Wolfgang Wahlster, editors. Spinning the Semantic Web: Bringing the World Wide Web to Its Full Potential. MIT Press http://www.ksl.stanford.edu/people/dlm/papers/ontologies-come-of-age-mit-press-(with-citation).htm)

[33] Zeng, M.L. (2008). Knowledge Organization Systems (KOS). Knowl. Organ. 35, 160–182. http://doi.org/10.5771/0943-7444-2008-2-3-160

[34] Jonquet, C. et al. (2018). AgroPortal: A vocabulary and ontology repository for agronomy. Computers and Electronics in Agriculture, 144, 126-143. https://doi.org/10.1016/j.compag.2017.10.012

[35] Noy, N. F. et al. (2009). BioPortal: ontologies and integrated data resources at the click of a mouse. Nucleic acids research, 37(suppl_2), W170-W173. https://doi.org/10.1093/nar/gkp440

recommended by the New European Interoperability Framework which is currently adopted by the European member states and the EC, as described in Section 1.3.2.

As the project's central service, KOINE develops B2TERM. Terminologies are heterogeneous and data centres have developed bespoke solutions to access them to support interdisciplinary applications. Maintaining these solutions is tedious. To simplify terminology use, B2TERM will abstract this heterogeneity and provide unified access to terminology in the EOSC. The KOINE project will leverage user participation, e.g., by integrating B2TERM into the relevant phases of the research (meta)data lifecycle within research communities and other social infrastructure stakeholders. It is important to note that the project will not develop terminologies. Rather, KOINE enables communities in terminology curation and use. For example, B2TERM offers community-tailored views (slicing) of terminology served by terminology services in the EOSC. Such views are necessary in order to support efficient terminology search and retrieval. B2TERM thus facilitates the use of terminology by stakeholders (e.g., data curators, researchers) in annotating research data with unambiguous and persistently identified terminology. As a service within the EOSC Service Catalogue, B2TERM mediates requests for terminology creation and updates between terminology providers and consumers in the EOSC and offers technical and administrative support (e.g., helpdesk and issue tracker for terminology maintainers). It is the B2TERM service through which consumers, both human (e.g., data curators, researchers, businesses, policy makers, citizens) and machines, access terminologies needed to unambiguously describe research data semantics. In detail, the TIF and B2TERM implement and enable the features displayed in Table 5.

Table 5: Overview of B2TERM features.

Feature	Description	Result
Harmonization	The Terminology Interoperability Framework (TIF) harmonizes access and exchange of information about terminology in the EOSC. In addition to technical harmonization (API), KOINE prototypes the harmonization of terminology needed to describe research data semantics (in particular data variable descriptions, e.g. "essential variables" in biodiversity, climate, etc.)	Common Semantic Model that enables interoperable terminology in the EOSC, easily harvested and used by systems and services, including B2TERM.
Access	B2TERM makes terminologies discoverable and accessible across science domains and community based terminology services. Harmonized access to terminology is given both through B2TERM (centralized) as well as through terminology services (decentralized, discipline-specific).	Terminologies are made accessible in EOSC via a common HTTP REST API. The HTTP REST API will be adopted by B2TERM and community-driven terminology services.
Exchange	As for access to terminology, harmonized exchange of information about terminology is enabled through B2TERM by harmonizing the information exchange interfaces of existing partnering services (e.g., NERC-VS, CLARIN-TS, AgroPortal), partnering services to be developed (e.g., EcoPortal), and third-party services through co-operations (e.g., EBI OLS, SKOSMOS, NCBO BioPortal, ANDS Research Vocabularies Australia).	B2TERM will enable exchange of information about terminology and terminology services in the EOSC over HTTP and HTTP content negotiation, providing an endpoint ready to be used by data discovery services and other research data management tools and services within the research (data) lifecycle.
Slicing	B2TERM supports the community-driven definition and operation of tailored slices (views) over terminology in the EOSC. This feature is necessary to enable efficient terminology search and discovery.	Data curators, researchers and other stakeholders can specify their community driven views and use terminology of specific interest.

Mediation	B2TERM mediates requests for terminology updates between terminology consumers and providers, supported by B2TERM Helpdesk. Requests will be issued by stakeholders (data providers, curators, researchers, and other consumers). Established terminology workflows among terminology services are respected. KOINE will prototype engaging library subject specialists (at TIB, TUD) in B2TERM Helpdesk operation.	Effective and efficient interaction between terminology consumers and providers leading to an evaluated, high quality terminology service, e.g. featuring a helpdesk and issue tracker.
Catalogue	B2TERM fosters the provision of a common interface and catalogue. Given the massive volume of existing terminologies, we focus on high visibility terminology such as "essential variables" and terminology of specific relevance to participating partners. B2TERM assesses the FAIRness score for each of the terminology it serves.	Maximized findability and reusability of terminologies. Focus will be on the use cases within KOINE, to be extended by the end of the project.
Subscription	B2TERM can be connected to subscription and notification service for terminology updates, which can be used by all stakeholders (data providers, curators, researchers, and other consumers).	Reliably up-to-date terminologies and their use for a broad user community within the EOSC.

Parallel to the provision of B2TERM, KOINE integrates this novel and innovative service for terminology with (inter)disciplinary operational infrastructures and diverse communities in research, industry and society. This will be done using network partners and multipliers from three broad research disciplines (Environmental Sciences, Engineering, and Humanities and Social Sciences) as well as the established EUDAT service B2FIND and the Data Subscription Service already prototyped in Marine Environmental Science. As such, data providers (e.g., PANGAEA, DKRZ, 4TU.Centre for Research Data) and research infrastructures (e.g., ACTRIS, eLTER, SIOS, LW ERIC, CLARIN) apply the B2TERM service and thus catalyse the transition in the current practice of using ambiguous labels to describe research data semantics toward using unambiguous, persistently identified and machine-readable terminology. Within data discovery services such as B2FIND, the adoption of terminology improves precision and recall of search results, and enables semantic data discovery and exploitation across disciplines for stakeholders, including individual researchers, research communities, public sector, industry, and educators. Table 6 provides an overview for the B2TERM integrations in (inter)disciplinary operational activities and services executed and operated by existing sociotechnical infrastructures. Table 7 lists the network and multiplier organisations directly involved in KOINE as partners. **With the help of these organisations, which are also scientific networks and digital service providers, KOINE engages European and global research infrastructures serving thousands of scientists.** The networks and specific types of repositories and vocabularies used by the project partners are detailed in the following profiles.

Table 6: Overview of B2TERM integrations in (inter)disciplinary operational activities and services.

Integration	Description	Result
Annotation	B2TERM is integrated in research data annotation activities of data curation performed by data centres, researchers or other stakeholders of the KOINE Stakeholder Forum. Specifically, ambiguous labels used in research data to convey semantics, e.g. textual parameter descriptions, are annotated with (mapped to or entirely substituted with) corresponding unambiguous, persistently identified and machine readable information objects.	The sharing and reuse of unambiguous, persistently identified and machine readable information objects that describe research data semantics results in greater semantic data interoperability.

Discovery	Terminology services provided by B2TERM are integrated in disciplinary and interdisciplinary data discovery services.	Novel search functionalities with improved precision and recall. For instance, a user can search for data in an unambiguously identified unit.
Data Subscription	Given the increasing rate of data production, data consumers (e.g., researchers, industry, citizens, and educators) may want to be notified when relevant new data are published or existing data sets are extended or updated. As for discovery, terminology is integrated in subscription to enhance the service's semantic capabilities.	Novel subscription functionalities for personalized data discovery
Exploitation	Preparing data for analysis typically involves expensive data cleaning and integration activities. Machine support in data exploitation activities continues to be limited, largely because data semantics are unavailable to machines. The integration of B2TERM terminology enables novel machine support in data exploitation.	Technical infrastructure proactively supports users in data exploitation activities. For instance, systems may automatically translate data to shared units and thus support data integration.

Table 7: Network and multiplier organisations directly involved in KOINE.

Scientific domain	**Multiplier / Research Infrastructure**	**Area of expertise**	**Resources availability (repositories, vocabularies, data, knowledge)**
Atmosphere	ACTRIS, http://actris.nilu.no	Data Portal covering all atmospheric composition disciplines including aerosol, greenhouse gas, stratospheric composition, and reactive gas data.	Expertise on semantic description and repositories of atmospheric composition data, involved in WMO GAW, WIS, and WIGOS.
Ecosystem	LTER/eLTER, http://www.lter-europe.net/	Knowledge of the structure and functions of ecosystems and their long-term response to environmental, societal and economic drivers. eLTER is a network of sites in Europe, which can be searched and explored by DEIMS-SDR.	Expertise in the semantic description of scientific observations and measurement parameters in ecosystem research and biodiversity.
Biodiversity	LW ERIC, https://www.lifewatch.eu/catalogue-of-data	Global provider of e-Science research facilities, content and services for the Biodiversity research community.	Expertise in semantic data integration, semantic enrichment of data and data discovery. Development of the EcoPortal, a semantic repository focused on ecology and biodiversity as well as on ecosystem observation in the European context.
Arctic	SIOS, https://sios-svalbard.org	International interdisciplinary observing system for long-term	Using GCMD Science Keywords and Climate and

		measurements in and around the Norwegian archipelago of Svalbard focusing on Earth System Science and covering meteorology, oceanography, glaciology, biodiversity etc.	Forecast Standard names actively. Building services around GBIF and TDWG. Linked with WMO efforts of WMO Integrated Global Observing System (WIGOS) and WMO Information System (WIS).
Earth System	DRKZ, https://www.dkrz.de/	Globally established Climate Computing Center, operating a broad set of data related services.	Expertise in data, modelling and user driven service development. Operates B2FIND service within EUDAT which will adapt B2TERM services.
Earth & Environmental Science	PANGAEA, https://www.pangaea.de/	Globally established Data Publisher, operated as an Open Access library aimed at archiving, publishing and distributing georeferenced data from earth system research.	Expertise in all fields of earth and environmental science, data provision and review, modelling and user driven service development.
Marine	UKRI, https://www.bodc.ac.uk/resources/	Member of SeaDataNet pan-European infrastructure for ocean and marine management; UK National Oceanographic Data Centre and host of the NERC Vocabulary Server. Member of major European and global networks and projects including Argo, DIMES, GEOTRACES, GEBCO, GLOSS, EMODnet, IOC-IODE, SeaDataCloud, OBIS	Expertise in the semantic description of scientific observations and measurement parameters within marine sciences. Provider of vocabulary services to the marine community and host of the NERC-VS.
Marine and Cryosphere	AWI, https://www.awi.de/en.html	Centre for polar and marine research providing scientific knowledge and services, e.g., data portals from expeditions and infrastructures, and background information such as cruise reports.	Expertise in ontology development, e.g. the Environment Ontology (EnvO), which provides a controlled, structured vocabulary that is designed to support the annotation of any organism or biological sample with environment descriptors.
Agronomy	INRA, http://institut.inra.fr/en	French National Institute for Agricultural Research, providing scientific products and services, e.g. in modelling natural phenomena, managing regional land use and understanding how the environment influences gene expressions.	DataVerse, Food Cloud, Annotation of Terminology, Focus on Soil Data

Engineering	TU Delft and 4TU.Centre for Research Data, https://researchdata.4tu.nl/en/home/	Provide a Research Data Repository for engineering disciplines and expertise in Geosciences and Coastal Engineering.	Know-how in Geoscience Engineering, geodata and community outreach, particularly within CESAER[36]. Access to disciplinary experts such as data stewards and data champions[37].
Humanities and Social Sciences	OEAW, https://www.oeaw.ac.at/en member of CLARIN ERIC, https://www.clarin.eu/content/services	Common Language Resources and Technology Infrastructure, providing access to digital language data and advanced tools to discover, explore, exploit, annotate, analyse or combine data sets.	OEAW carries out innovative basic research in the arts and humanities, the social and the natural sciences and represents the dha - digital humanities Austria.
Informatics, Data Science	UvA, CNRS-LIRMM, UMAAS	Expertise in semantics and semantic technologies, and service development or deployment.	UvA curates the semantic resources developed in the 'Data for Science' theme of the ENVRI community cluster. CNRS-LIRMM develops and maintains two ontology repositories.
Information Technologies in the Natural Sciences	TIB, https://www.tib.eu/en	Information Centre for the digitisation of science and technology, conducting applied research and development in the areas of data science, Open Knowledge, Open Science, and Visual Analytics. Provides scientific content and digital services.	TIB contributes resources on the methods and techniques for the semantic networking, interdisciplinary community outreach, will integrate B2TERM in the TIB and EUDAT service portfolio and acts as coordinator.
Interdisciplinary Services	EUDAT B2FIND, http://b2find.eudat.eu	B2FIND is a discovery service based on metadata steadily harvested from research data collections from EUDAT data centres and other repositories. It is part of the EOSC Service Catalogue.	As an interdisciplinary operational infrastructure, B2FIND will adapt B2TERM services and thus facilitate the semantic discovery of data across disciplines.
Interdisciplinary Services	Data Subscription Service (DSS)	DSS is developed by CSC, in order to enable data users to be notified when relevant data are published or updated.	DSS will be integrated with B2TERM and B2FIND, and demonstrated in Argo Data and Metadata Discovery Service.

[36] https://www.cesaer.org/

[37] https://www.tudelft.nl/en/library/current-topics/research-data-management/research-data-management/data-stewardship/data-champions/our-data-champions/

Interdisciplinary Services	ANDS	Research data management.	Know-how in making research data assets more valuable for researchers and research institutions.
Earth & Environmental Science	SISTEMA	Mining of actionable environmental information from (Earth Observation) data for numerous stakeholder groups also outside academia.	ADAM (https://adamplatform.eu) is a cross-domain platform that provides access, processing and visualization services to a wide range of environmental data.

1.3.2 Related work

Discovery services are fundamental to finding and accessing research data. They have been developed since the first Digital Object Identifiers (DOIs) for datasets were indexed back in 2005, starting with local search engines at institutes (e.g., DKRZ, which registered the first DOIs for climate data in 2005). Nowadays, academic and commercial service providers operate discipline specific (e.g., Climate Data Store, ELIXIR Core Data Resources, PANGAEA) and interdisciplinary research data discovery services (e.g., B2FIND, DataONE, DataCite Metadata Search, and Elsevier Data search). Google, too, is now prominently on-board and driving these developments offering Google Dataset Search by leveraging Schema.org metadata.

In the life sciences, the Bioschemas.org initiative is a leading example for how to encourage and drive data interoperability across a diverse scientific community. By taking on the Schema.org approach, Bioschemas.org provides successful showcases and discipline-related, community-agreed specifications as additions to the Schema.org model. In addition to the directly via partners involved disciplines environmental science, engineering and social sciences, KOINE will aim to develop a close cooperation with Bioschemas.org and the ELIXIR initiative.

The EOSCpilot Interoperability Theme[38] **is a further related initiative**. In their "first draft of the strategy and recommendations to help users and services to find and access datasets across several scientific disciplines"[39] they stress the vision to enable a *'research and data interoperability across the diversity of existing (and potential future) research communities, RIs, and other research assets*'. In order to put this into practice and apply the FAIR principles, they strongly highlight the responsibility of research infrastructures, e-Infrastructures and research communities. Regarding the metadata and data interoperability that is desired according to the Interoperability Principle in FAIR, KOINE answers to these demands by:

- Forming a strong network across major scientific disciplines (Table 7)
- Streamlining the use of terminology and actively integrating terminology in operational infrastructures
- Creating B2TERM as a service that allows for adaptation and integration with existing data discovery architectures for producers and consumers (Table 5).

In 2017, the EU published the **New European Interoperability Framework (EIF)**[40], which aims at setting the basis for a harmonized approach to interoperable, digital public services that may form the backbone of the new digital single market in Europe. Being currently adopted by the European member states and the EC, the EIF provides 47 recommendations to improve the quality and interoperability of such services. The EIF describes interoperability areas with a particular focus on:

- Governmental mechanisms to interoperability (administrative, business and user view)
- Collaboration between organisations, institutions and stakeholders
- Raising awareness of the interoperability benefits and
- An interoperability model consisting of four layers (legal, organisational, semantic and technical).

[38] https://eoscpilot.eu/themes/wp6-interoperability

[39] https://eoscpilot.eu/themes/wp6-interoperability/1st-report-on-data-report-findability-interoperability

[40] https://ec.europa.eu/isa2/eif_en

The **EIF semantic interoperability** layer describes recommendations to ensure "*that the precise format and meaning of exchanged data and information is preserved and understood throughout exchanges between parties, in other words 'what is sent is what is understood.'*"[41] In this section of the EIF, there are three recommendations to enhance the semantic and syntactic aspects of interoperability:

- First, it strongly highlights the importance of the need for an enhanced perception that data and information are valuable public assets (*Recommendation 30*)
- Secondly, it is recommended to set-up an "*information management strategy*" which mandates community agreements on e.g., reference data, taxonomies, controlled vocabularies, thesauri, code, reusable data structures and their models, and others (*Recommendation 31*)
- Thirdly, it calls for the support and encouragement of discipline-specific and interdisciplinary communities which have the goal to create, harmonize and distribute "*open information specifications*" (*Recommendation 32*).

Within the KOINE project, we will answer these three recommendations by:

- Developing a Terminology Interoperability Framework (TIF) that harmonizes access to terminologies, develops a Common Semantic Model and facilitates the exchange of terminology required to describe research data semantics. As such, the TIF will also determine a set of Compliance Standards mandating the connection of FAIR terminology within the EOSC.
- Developing B2TERM as a novel EOSC service that significantly increases the 'user experience' when searching for and working with terminologies.
- Making B2TERM and TIF available to European researchers by establishing the service in the EOSC Service Catalogue and by integrating B2TERM with disciplinary operational data infrastructures and interdisciplinary operational data discovery services.

Table 8 overviews important initiatives that are related to KOINE and indicates how the project complements and further improves the state of the art.

Table 8: Initiatives related to KOINE.

Initiative	**How KOINE relates, complements, improves**
The ELIXIR's Bioschemas.org service and network in Life Science is developing a collection of specifications that provide guidelines to facilitate a more consistent adoption of schema.org markup within the life sciences.	A central task of the KOINE project is the active integration of terminologies in operational infrastructures to drive data interoperability. KOINE therefore complements the schema development driven by initiatives such as Bioschemas.org.
FAIRsharing (https://fairsharing.org) is a leading resource on (meta)data standards, databases, data collections and data policies, collecting also from major European and global data initiatives and projects.	FAIRsharing can inform KOINE on existing terminologies. Terminology services that are compliant with the Terminology Interoperability Framework developed by KOINE could be discoverable through FAIRsharing. In contrast to FAIRsharing, which functions as a registry, KOINE plays an active role in driving interoperability among terminologies.
FAIR metrics (http://fairmetrics.org) is an initiative with the goal to define metrics enabling both qualitative and quantitative assessment of the degree to which online resources comply with the 15 Principles of FAIR Data.	To support FAIR metrics, KOINE develops a FAIRness scorer, which specifies a set of indicators, specific to terminologies, to assess their FAIRness level.

[41] https://eur-lex.europa.eu/resource.html?uri=cellar:2c2f2554-0faf-11e7-8a35-01aa75ed71a1.0017.02/DOC_3&format=PDF

Odex4all tools and services in Life Sciences and eHealth: FAIRifier, FAIR Data Point, FAIR Search Engine, ORKA, Data FAIRport	Odex4all software will be able to reliably use KOINE's machine friendly terminology services to enable their users to convert data to existing standards and to use the latest terminologies.
The Open Research Knowledge Graph (http://orkg.org) is a project that aims to increase machine actionability of scientific information communicated in scholarly literature. The project is led and sustained by TIB.	Through B2TERM, KOINE will streamline access to terminology needed for machine readable representation of scientific information. KOINE will thus support the use of terminology also in research artefacts other than datasets, such as research papers.

(b) Methodology

Terminology Interoperability Framework (TIF)

The objective of TIF is to harmonize across participating services:

- How terms and information about them are accessed
- The protocol and the representational model of exchanged information about terms
- The metadata about terminologies
- The metadata about terminology services
- The notification interface for terminology curation requests.

KOINE proposes to achieve these objectives by ensuring that existing services expose their terminologies using a common interface. KOINE adopts a Schema.org approach to terminology interoperability implemented within existing partnering terminology services (e.g., NERC-VS, CLARIN-TS, AgroPortal), new services to be developed by partners (e.g., EcoPortal), and in collaborations with third-party partners and their services (e.g., NCBO BioPortal, EBI OLS SKOSMOS, ANDS Research Vocabularies Australia). KOINE operates on the principle not to mint new identifiers (Uniform Resource Identifiers) for existing terminology terms and to instead leave such activity at the discretion of resource maintainers. Thus, a key accessibility requirement of the TIF is term identifier persistence and resolution. KOINE builds on HTTP based best practices and will handle direct access (e.g., NERC-VS) and indirect access (through URL query parameter, e.g., AgroPortal) according to the service. Building on persistent identifier resolution, the TIF introduces a Common Semantic Model for harmonization of exchanged information about terms, for both humans and machines. As for the protocol, the TIF adopts an HTTP based approach with content negotiation, i.e., serve different information (JSON-LD, HTML) for the same URI. The framework specifies the required and suggested content types. In addition, the framework also specifies the Common Semantic Model of the exchanged information for each content type. As a result, both humans and machines will be able to access identified in a harmonized manner. Moreover, consumers can rely on Common Semantic Model of the exchanged information about the terms accessed. KOINE harmonizes both natural language labels served to human agents and formal (i.e., machine readable) descriptions (e.g., in JSON-LD) served to computer agents.

B2TERM as a new and innovative EOSC service

B2TERM is the key new and innovative EOSC service developed by KOINE. It is envisioned as a future service of the EUDAT service portfolio (hence the use of "B2") and will be sustainably operated by TIB. This approach allows for optimal integration in the EOSC (in particular the EOSC Service Catalogue) and dedicated development and operational support from EUDAT as a key European e-Infrastructure that can immediately leverage a strong network of service providers, experts, and user communities to facilitate the uptake of the service.

B2TERM is a key value-added service that builds on the interoperability enacted by the TIF. As a terminology service itself, B2TERM also implements the TIF and thus adheres to the same interoperability requirements. B2TERM has the following primary objectives:

- **Enable terminology search, discovery and use through the definition of community-driven tailored slices (views) over terminology in the EOSC.** Operations such as search are executed on slices. In contrast to massive term indexes where search generally results in hard to overview result

sets, slicing focuses results to those of specific interest to individual researchers and research communities, data centres/publishers, e-Infrastructures and other stakeholders. B2TERM thus aims at effective and efficient terminology search, discovery and use.

- **Mediate requests for terminology curation.** Terminology services such as NERC-VS serve atomic terminology used in complex (compound) variables (e.g., "temperature of water in degrees Celsius") that are used to describe the correct interpretation of data values in datasets. B2TERM enables users such as data curators to search and discover such terminology and, in case none applicable is found, in formulating corresponding curation requests. Thus, B2TERM standardizes the mediation of requests while leaving the implementation, tooling and workflows for curation at the discretion of terminology maintainers.

B2TERM integration with disciplinary operational infrastructures

One of KOINE's central aspects is the integration of B2TERM with (inter)disciplinary operational infrastructures that researchers use and thus the activity of putting terminology to use. KOINE disciplinary operational infrastructures, i.e., operational data centres of research infrastructures in several scientific disciplines used by thousands of researchers worldwide, are B2TERM stakeholders and early adopters that integrate the service into their data curation workflows to ensure that published data and their semantics will be more intelligible and less ambiguous to both human experts and, in particular, machines.

Infrastructures expose annotations to users, primarily on landing pages by turning textual labels into hyperlinks that refer the user to a description of the meaning of the selected label. This practice is already in production in such infrastructures for various entities, including people where names are increasingly hyperlinked with ORCID iDs. The KOINE TIF ensures that such hyperlinks are resolvable. Crucially, infrastructures expose the annotations also to machines, by including them in machine readable metadata. To this aim, KOINE develops extensions to metadata standards in use by partnering disciplinary operational infrastructures. Of particular focus is the adoption of unambiguous persistently identified terminology in Schema.org/Dataset[42] metadata, specifically as the value of the *variableMeasured*[43] property. Finally, infrastructures expose annotations, in human and machine readable form, also in published data files. This ensures that downloaded data, manually by users or programmatically by machines, retain the annotations with unambiguous terminology that are available on landing pages and in published metadata. Data semantics are thus unambiguous also in subsequent use, e.g., in data analysis.

This B2TERM integration will significantly improve the quality of published (meta)data and contribute substantially to the FAIRification of research data that infrastructures carry out on behalf and for the research community, especially along the data interoperability principles. This integration is one of KOINE's most important strengths.

B2TERM integration with interdisciplinary operational data discovery services

Interdisciplinary operational data discovery services are a second set of B2TERM stakeholders and early adopters of the service. With the B2FIND data discovery service, KOINE develops integrations also in *interdisciplinary* operational infrastructure and thus facilitates the semantic discovery of data across disciplines. The use of unambiguous terminology is particularly relevant in an interdisciplinary setting, since descriptive labels used in a discipline are often unintelligible by researchers of a different discipline. KOINE integrates B2TERM in B2FIND along two pathways. First, users are enabled in contributing direct annotations of data discovered in B2FIND. Second, B2FIND benefits from data curation by disciplinary operational infrastructures in that the data discovery service harvests metadata about published datasets that include terminology annotations, in particular for the measured variables.

Data Subscription Service

EUDAT data infrastructure partner CSC has developed a Data subscription service (DSS) that allows data users to make subscriptions to data of their interest. The DSS functions as a component that is integrated with thematic or interdisciplinary data discovery service. The subscription can be defined, for example, based on variable names or on the actual content of the data. DSS has been demonstrated in an operational

[42] https://schema.org/Dataset
[43] https://pending.schema.org/variableMeasured

environment in connection with the Euro-Argo portal on Argo marine observation data. The DSS is currently at TRL 6.

KOINE will provide innovative data exploration capabilities in extending the DSS with semantic information through integration to the B2TERM service. The work on DSS will start with targeted development of the tool to make it a generic service component easily integrated to different data portals. Next, integration with B2FIND is done to allow registering subscriptions from this metadata catalogue service. In parallel, DSS usage in the Euro-Argo case will be extended with semantic capabilities enabled by KOINE. These developments together present a demonstrator for the value of B2TERM in intelligent data discovery, and establish DSS as a TRL 8 component.

The design of DSS relies on well-defined APIs and data models. It can handle subscriptions on any digital objects that have a persistent identifier (PID) attached to it. In addition, the DSS can also be used as a tool for connecting data discovery to analytics, as the subscription can include instructions for automatic processing of the data. In connection with the climate research use case, we will utilize this functionality in triggering automatic calculation of standard parameters, and include Jupyter notebooks as an additional analytics component.

Community integration

To meet research community needs and develop a fit-for-purpose service, KOINE deeply integrates the community (WP5) early on in the project by establishing the KOINE Stakeholder Forum to gather the requirements of the Terminology Interoperability Framework (in particular the Common Semantic Model), the annotation strategies, and B2TERM.

The sociotechnical integration of B2TERM in research communities is a fundamental and central aspect of the KOINE project. Thus, the project ensures that the novel service is integrated in the scientific workflows of research communities. Essential in this regard is the annotation of research data with terminologies (WP4) published by operational (inter)disciplinary infrastructures, e.g. data centres and data discovery services. KOINE thus ensures that the research data produced and consumed in scientific workflows are of unprecedented semantic clarity, for both humans and machines.

Research life cycle adoption

Research data (observational, experimental, simulation, etc.) is not the only digital object that can benefit from B2TERM and its integration by annotation of labels with terms. Along the research lifecycle, experiments are designed and executed, data are generated, processed and analysed, and results are communicated. In all these phases, numerous digital objects are produced and consumed. Examples include data management plans, provenance, software, and articles. Each of these digital objects can benefit from annotating (human readable) labels with terminology.

Operations and Sustainability

To meet the quality standards and sustainability of the B2TERM service, KOINE designs and implements a robust sustainability model driven by the experience in operating sustained infrastructures by many project partners (Table 7). The model includes, in particular, mechanisms to interoperability (from an administrative, business and user view), and secures a strong and long-term partnership with the platforms that implement the B2TERM service.

After a review process for the TIF and associated Compliance Standards for terminologies included in the B2TERM service, a workflow model is specified. This model ensures a seamless interaction between components of the TIF and the overall B2TERM service, and is set-up according to the guidelines of the European Interoperability Framework (EIF), which also addresses governmental mechanisms to interoperability. As such, the workflow model supports efficient harmonization and user-friendliness across infrastructures and ensures the sustainable development of the TIF and B2TERM within and beyond the project.

Another task of the governance model is to ensure an effective collaboration between organisations, institutions and stakeholders. This process is composed of two focus areas:

- A method whereby priorities for development can be set so as to address emerging EOSC needs, while balancing the priorities of contributing terminology providers
- A method to allow stakeholders and developers of additional services using B2TERM to suggest modifications or specify emerging needs.

These aspects are addressed by:

- Establishing an international **KOINE Governance and Operations Committee (OC)**, initially composed of members of this project's consortium, to formalise representation and decision making mechanisms. The OC tasks include decisions on which tools and procedures will be used to link experts to developers to extend terminologies within the EOSC Service Catalogue as well as which workflows will be used to bridge stakeholder requests to the management of participating terminology services. Furthermore, mechanisms to ensure the quality and interoperation of individual resources and to upgrade the system with new technologies without compromising stability are determined.
- Establishing, through engagement with the Stakeholder Forum (WP5), an international **KOINE Scientific Advisory Board** (SAB, T6.2), primarily composed of members outside the consortium. The SAB's primary task will be to advise the OC by reviewing the system and ensuring it is current and well-connected with external initiatives such as additional services provided by stakeholders. The SAB will seek members from underrepresented disciplines (e.g., the life sciences) to increase the B2TERM coverage and its relevance to EOSC users.
- Establishing a GO-FAIR Implementation Network on Vocabulary Services that will serve as an open forum for discussing architectures, software, and specifications developed by KOINE with the wider global community. As thematic GO-FAIR INs (e.g., metabolomics, Food Systems) are in the process of being established, we will identify or create additional INs so as to promote collaboration and sustainability relevant to KOINE's activities and partners beyond the funding horizon for KOINE.

For international visibility and recognition, the KOINE governance engages actively in important advisory bodies (e.g., GO-FAIR Implementation Network, Research Data Alliance (RDA) working groups, other EOSC Implementation Roadmap projects) to execute and further refine the functions developed in the model above. This engagement with international stakeholders ensures the long-term visibility, recognition, sustainability, scalability and trust of users and stakeholders of the service. As such, the Research Data Alliance is of particular importance as there have been relevant previous developments in the semantic sector that will be considered in KOINE. This includes for example, the Basic Vocabulary of Foundational Terminology Query Tool[44], and the Data Type Model and Registry[45], as well as the outcomes of the Agrisemantics Working Group[46].

For users and stakeholders alike, trust in project outcomes, their services and their long-term sustainability is a key issue for any project. Within KOINE, this is approached by the development of a robust sustainability model to allow governance and operations mechanisms to persist beyond the project, adhering to core principles and standard operating procedures.

Gender dimension

The gender dimension has been taken into account in the planning of the project and in defining the proposal priorities, concept and approaches. WP5 is give special attention to requirements of female scientists, which will thus be reflected in WP2 design and WP3 implementation. WP5 and WP7 will promote gender equality in workshops, training, and outreach events. Particular attention will be given to diversity measures among speaker invitations, discussions and event contributions of any kind. Out of the 21 funded partners, 7 have identified female key personnel who will take leadership roles in the project. The coordinator will further encourage gender balance in the additional recruitment and assignment of staff to the project.

[44] https://rd-alliance.org/group/data-foundation-and-terminology-wg/outcomes/data-foundation-and-terminology
[45] https://rd-alliance.org/group/data-type-registries-wg/outcomes/data-type-registries
[46] https://www.rd-alliance.org/groups/agrisemantics-wg.html

1.4 Ambition

The high-impact ambition of the KOINE project is to make an essential and concrete contribution to increase EOSC data interoperability by proposing a sustained, long-term terminology service – B2TERM – that is integrated in operational (inter)disciplinary research data infrastructures. Through interoperability, KOINE directly increases the semantic findability and re-usability of research data produced and consumed within the EOSC. For the very first time since the introduction of the FAIR data principles and the development of FAIR-related services, a project will focus on enabling Semantic Interoperability within the EOSC Catalogue of Services as a whole.

1.4.1 Advancing FAIRness beyond the state-of-the-art

FAIR data is both a challenge and an opportunity. In the European research landscape, the uptake of the Open Research Data Pilot and its extension to cover all areas of Horizon 2020 already result in more data being accessible (the "A" in FAIR). However, being accessible is not a sufficient goal in itself - the data should also be interoperable in order to be re-used (the "I" and "R") in order to generate value throughout the research lifecycle. **With KOINE, we argue that key to research data interoperability is the use of FAIR terminologies and the transition from ambiguous labels to describe data semantics to unambiguous and persistently identified terms.**

Generic and discipline specific data discovery services continue to rely on user-based search input in the form of ambiguous keywords such as name, discipline, publication year for metadata-driven search, and list search results in a more or less standardized manner. The precision of search results thus largely depends on the quality and extent of the metadata offered by data providers while the quantity returned (recall) depends on the coverage offered by repositories. A search for research data by users who need to locate and evaluate the usefulness of data is thus often tedious and a futile exercise.

Semantically-rich data is a fundamental requirement for efficient interdisciplinary data discovery and exploitation. Data need to self-describe their correct interpretation and thus semantics, unambiguously for both humans and machines. **The state-of-the-art is that the meaning of research data is generally ambiguous and implicit. To machines, data semantics are mostly entirely inaccessible.** To advance FAIRness beyond the state-of-the-art, the project will research, design, develop, evaluate and promote the **novel service B2TERM** as an element of the EOSC Service Catalogue. As a central part of B2TERM, the proposed **Terminology Interoperability Framework (TIF)** harmonizes access to terminologies, develops a **Common Semantic Model** for the description of (essential) variables, and facilitates the exchange and use of terminology required to describe research data semantics. As such, the TIF will determine a set of **Compliance Standards** for the connection of FAIR terminologies in the EOSC.

Thus, KOINE will provide practices, policies, instruments and services that close the loop between the challenge of semantic interoperability in the production and publication of research and its actual use and re-use in the broader science-society-industry environment. KOINE will set the European (and possibly global) standard for the programmatic, persistent access to and use of terminologies, starting off with established terminologies, integrating such resources in operational (inter)disciplinary research data infrastructures, and extending terminologies to generic research data discoverability resources such as B2FIND.

KOINE greatly emphasises the integration of the B2TERM service in operational (inter)disciplinary research data infrastructures, which is key **to bring the adoption of harmonized terminologies into research communities. KOINE excels at this, both at the disciplinary and at the interdisciplinary level**, by showcasing this with 23 project partners as key stakeholders representing major European research infrastructures. Furthermore, KOINE emphasises the strong integration of research communities and other stakeholders (1) in gathering requirements for B2TERM and the TIF as well as (2) in training to ensure that the use of unambiguous terminology becomes the new normal in research.

With the goal of generalizing to any scientific area and offering generic services for EOSC, by the end of the project our objective is to establish concrete links – via our large and multidisciplinary consortium – to other existing projects inside or outside the EOSC Implementation Roadmap and the ESFRI landscape. The EOSC is the right framework in which to establish the B2TERM multidisciplinary service. It is only through the links in this European context that a terminology service in a multidisciplinary framework makes sense in the

first place. The EOSC also makes it possible to integrate relevant contributions from interest groups beyond the scope of the project in terms of time and topic and to keep B2TERM up to date.

1.4.2 Closing the loop around the perception of semantic interoperability and FAIRness

In the KOINE project, FAIR is considered principally under the perspective of harmonizing available terminologies and terminology services rather than creating new ones. This includes not only the research perspective but also the management and social interaction around this perspective.

In contrast to the initial perception that FAIR data would flow almost freely among research communities and other relevant stakeholders in the EU, there is a growing understanding that this is not the general case. The lack of data interoperability is most importantly explained by the fact that published data generally make little or no use of unambiguous terminology. To allow true data interoperability in the context of FAIR, KOINE will go one step beyond the provision of harmonized access to terminologies: KOINE will study and analyse the generic and discipline-specific terminology landscape (WP2), integrate terminology services and build the B2TERM service (WP3) with rich features including notification for terminology updates and the mediation of requests for updates to terminologies, and integrate B2TERM in technical (WP4) and social (WP5) infrastructures to ensure interoperable (FAIR) data. KOINE will also develop sustainability models and outline possible business opportunities (WP6).

The ambition of KOINE is to effectively showcase the major advantages of a future EOSC with semantic data interoperability by use of FAIR terminologies and make the B2TERM service available to all European researchers using the EOSC Service Catalogue.

To power this ambition, KOINE will launch a B2TERM Stakeholder Forum involving a wide range of scientific domains and communities to gather requirements for how to harmonize and publish terminology in the EOSC (WP5), offer capacity building and training to ensure unambiguous terminologies are used in research data (WP5) and launch an Ambassador Programme (WP7) that will create a first group of B2TERM promoters. These processes will exploit the use of terminologies to catalyse new products and services for industry and society at large, extending the FAIR data chain beyond research. Another pillar to support innovation and input to B2TERM services at a smaller scale is the development and distribution of Open Educational Resources and the delivery of user workshops (“Terminology Carpentries”, WP5). Both tasks will enable individual researchers and interested third parties from various backgrounds to continuously evaluate the service and contribute their own ideas.

1.4.3 Assessment of Technology Readiness Levels

The KOINE project builds its service, B2TERM, on existing technologies and providers that are already at TRL 6-7, for example AgroPortal, CLARIN-TS and NERC-VS, or higher, for example ACTRIS Data Portal, PANGAEA, B2FIND. The high TRL of these platforms directly serves the creation of B2TERM as a lightweight new service harmonizing their contents for efficient and systematic use. The TRL of B2TERM will progress during the project with its integration in disciplinary operational data research infrastructures (e.g., ACTRIS, SIOS, eLTER) and interdisciplinary operational data discovery services (e.g., B2FIND). At the same time, these integrations will contribute to lifting the TRL of some of the existing technologies and providers from TRL 6-8. A full list of existing technologies and providers that B2TERM builds on is given in Table 9.

Table 9: TRL of operational technologies and providers that will serve as foundational services and multipliers for the B2TERM service and the KOINE project.

Productive level	TRL 9	• B2FIND • PANGAEA • 4TU.ResearchData • Languages for knowledge representation (e.g., RDFS, OWL, SKOS) • Environment Ontology (ENVO) • SDGIO for the UN Sustainable Development Goals
	TRL 8	• LTER DEIMS-SDR • ACTRIS Data Portal • LifeWatch Italy Data Portal • Data Inra Institutional repository

		• ESGF-DKRZ • ARCHE - A Resource Centre for the HumanitiEs (and other CLARIN centres)
	TRL 7	• SIOS Svalbard • NERC-VS • CLARIN Concept Registry • CLAVAS - CLARIN Vocabulary Service • ACDH Vocabs
Pre-production level	TRL 6	• EnvThes • AgroPortal (relying on NCBO BioPortal TRL 9 core technology) • EuroArgo Data Subscription Service (DSS) • LifeWatch Thesauri

2. Impact

2.1 Expected impacts

The KOINE project will impact a diverse array of stakeholders in the European research area and beyond, including public researchers, commercial/industrial users, and policy analysts in search of meaningful data. Among these, KOINE's initial impact will be focused on the researchers and research infrastructures using EOSC data services. KOINE will provide its stakeholders with improved machine-readable and –actionable terminological descriptors to support handling of research data, thus improving (inter)disciplinary data findability, interoperability, reusability and potential exploitation through improved data integration. To do so, KOINE will align and boost interoperation between existing and internationally adopted terminology providers, broadening its impact to all users of their services. In close collaboration with the producers of research data, KOINE will improve automated, machine-driven understanding of what research data is about, creating new possibilities for more automated data mobilisation and analysis, across the data lifecycle. KOINE's outcomes will contribute to the reduction of costs and increases in the efficiency of data handling across EOSC services.

KOINE will be delivered by a consortium that has extensive experience in all phases of the research data interoperability lifecycle and has strong links to significant national, pan-European, and global (e.g., EuroArgo, GOOS, Research Data Alliance, FAIRMetrics Group) infrastructures, initiatives and organizations. The consortium's experience and activities provide it with a clear understanding of the current challenges facing data interoperability. In this proposal, its members have translated these into a set of priorities and activities which will achieve a substantial, broad and sustainable impact. Each partner in the KOINE consortium brings strong motivation and solid experience to secure the project's success. A point-by-point impact achievement plan is detailed in Table 10.

KOINE will have significant impact on several stakeholder groups:

- Using the B2TERM service, **(inter)disciplinary operational research data infrastructures** will be able to persistently describe their research data with unambiguous, machine-readable, and interoperable terminology.
- **Projects, organisations, initiatives, and networks generating data and developing or using terminologies** will have the ability to describe their data with proven, well-adopted, and FAIR-compliant terminologies which they will be able to shape to their needs.
- Using B2TERM, **research communities** will be able to enhance the discoverability of their data, as well as their data-driven products and services, and ultimately the quality of their research.
- **Data policies initiatives, research funders and standards organisations** aiming to align and implement data description frameworks to respond to the requirements such as FAIR data principles, in particular to the principle of Interoperability.

KOINE has the potential to significantly impact the practically universal and often laborious tasks of data discovery, formatting and cleaning. Notoriously, these tasks consume as much as 80% of time invested in handling data prior to analysis, and present a significant bottleneck for efficient transformation of data into the information which powers knowledge-based economies. By making the terminology-based description of research data a) interoperable across existing terminology providers, b) easily applied to data sets, and c) accessible to machines, KOINE will furnish new possibilities in automated data handling. For example, algorithms capable of understanding the machine-readable data descriptions provided by KOINE's technological outputs would be able to automatically transform datasets such that their units are compatible. KOINE can thus contribute to significantly reduce the costs and efficiency in data analysis for a wide range of stakeholder groups, including data discovery services and the industrial data sector. While KOINE's impact on the direct users of open data will be most immediate, its most significant long-term impact will be on society at large. As data are more efficiently transformed into scientific information and knowledge, KOINE will help secure the long-term value of published research data, discoverable across research disciplines. By unlocking interoperability across disciplines, KOINE will aid in the development of new products and services that will aid society in multiple, and often unforeseeable ways.

2.1.1 Expected impacts listed in the work programme

The Horizon 2020 Work Programme 2018-2020 says that "*In order to realise an EOSC that truly supports interdisciplinary research and Open Science, a new pan-European model for research data and related services that is both scalable and flexible needs to be put in place, so that it can be adapted to the emerging needs of the scientific community and support the whole research data lifecycle*."

The challenge is to strengthen Europe's position as provider of information and data products and services that meet a high level of interoperability across disciplines. The overall goal of KOINE is to ensure that machine-actionable terminology- and semantic-based descriptions of data are able to boost data interoperability, discovery and exploitation across disciplines. The project will thus support user-centric scientific, industrial, and societal applications that maximize the impact of research data. The major expected impacts of KOINE and how they contribute to meeting the expected impact and targeted outcomes for projects under the work programme topic are shown in Table 10. As detailed in Table 11, the KOINE project delivers quantifiable and measurable results, through which its concrete impact and contribution to the related Key Performance Indicators can be evaluated.

Table 10: Impacts expected by INFRAEOSC-02-2019 and the method of achievement in KOINE.

Expected Impact	**KOINE Method of Achievement**
Integrate co-design into research and development of new services	The design of B2TERM will result from collaborative efforts involving stakeholders (e.g., data publishers and providers, research infrastructures) and existing terminology services operating in agronomy, the humanities, the ecological and the oceanographic domains. Through a Stakeholder Forum (SF) and formal decision making processes (WP6), users and developers will be involved in the design of machine-readable, accessible and user-focused integrated services and interfaces. Thus, the project will result in the improvement of B2FIND and DSS supporting the application of B2TERM by multiple stakeholders and leading to the improvement of existing infrastructures.
Better support scientific, industrial and societal applications	KOINE will align and orient well-adopted FAIR terminology services to maximize the impact of research data by improving cross-domain (meta)data interoperability, discovery and exploitation by scientific, industrial, and societal stakeholders.
Strong user orientation	KOINE will engage its wide range of user groups from the beginning of the project. Through the SF (WP5) and advisory groups (WP6), user dialogue will ensure that the service is developed and optimised for both data users and data providers (e.g., RIs and data publishers/metadata aggregators). The same mechanisms will also be used to engage an even broader range of users.

Support the objectives of Open Science	At its core, KOINE will provide access to controlled scientific terminologies that adhere to the FAIR principles, and provide tools to facilitate their integration into the research data lifecycle. As the use of formal terminologies for knowledge representation is a primary component of data Interoperability (FA**I**R), the project if deeply aligned will enable Open Science across disciplines. Further, all KOINE's software and the terminologies it uses will be published in open repositories such as GitHub.
Improve access to content and resources	By adhering to international standards (e.g., languages for knowledge representation) and by facilitating the use of standardised and well-adopted terminologies to annotate research data, KOINE will harmonize access to any well-annotated data or information products. Additionally, KOINE will provide common annotation schemas as a means to extend access through external services. An interoperability layer will be built between terminology providers and B2TERM.
Facilitate interdisciplinary collaborations	KOINE will align and bridge interdisciplinary terminologies which are used in fields such as Earth science, ecology, medicine, sustainable development, and agronomy. By developing a Terminology Interoperability Framework (TIF) for these terminologies, KOINE's deployment in EOSC will enable cross-domain data discovery and interoperability.
Open up the EOSC ecosystem of e-infrastructure service providers to new innovative actors	KOINE will facilitate this by: • Harmonising and aligning standardised terminologies and services built upon them, such that their wide user bases (e.g., NASA, NOAA, EBI, etc.) can better interface with the EOSC ecosystem. • Demonstrating the feasibility of true between terminologies across multiple scientific domains and sectors (e.g., engaging scientific instrument manufacturers), encouraging bold new actions towards data sharing and synthesis.

Table 11: KOINE Key Performance Indicators.

Number	Description	More than
KPI1: Increase in interoperability and FAIRness across disciplines in operational research e-infrastructures		
KPI1.1a	Number of active terminologies identified in current terminology services	200
KPI1.1b	The degree of FAIRness of terminologies earmarked for incorporation into B2TERM (will be FAIRness assessed manually before integration in B2TERM and after integration; the score difference will also be a relevant KPI)	100
KPI1.2	Number of research datasets annotated with B2TERM terminologies	100
KPI1.3	Number of terminologies served by B2TERM	30
KPI2: Increase in terminological interoperability and FAIRness in interdisciplinary operational data discovery services		
KPI2.1	Number of terminologies harmonized to serve the Essential Variables	30
KPI2.2	Level of increase in the precision and recall of dataset search in B2FIND, as a consequence of B2TERM application	+40%
KPI2.3	Number of registered subscriptions based on the Data Subscription Service (DSS) in B2FIND	150
KPI3: B2TERM engagement, availability and re-use		
KPI3.1	Number of terminology services joining the TIF and implementing its API (at least 4 in the project and more via partnership and outreach)	5
KPI3.2	Number of queries to B2TERM API	1000

KPI3.3	Number of data providers repositories integrating B2TERM	10
KPI3.4	Number of B2TERM pilot applications to support data integration and analytics	5
KPI4: Outreach, liaison & training activity results		
KPI4.1	Number of initiatives engaged with FAIR data management liaised with KOINE	10
KPI4.2	Number of participants in Stakeholder Forum	60
KPI4.3	Number of participants in Ambassador Programme	50
KPI4.4	Number of open access papers in journals or conference contributions based on KOINE outputs and activities	15
KPI4.5	Number of open source software applications of software applications incorporating KOINE outputs	10
KPI4.6	Number of KOINE recommendations for funders and policy makers	5
KPI4.7	Number of Open Educational Resources published	15
KPI4.8	Number of KOINE related events	15

2.1.2 Impact on the development of the EOSC as a whole

KOINE is expected to have impact on a wide range of different research communities and infrastructures who contribute to the EOSC, as well as those that plan to contribute. The EOSC provides a solid framework in which the established B2TERM service can greatly enhance harmonized access to multidisciplinary data in the European (and possibly global) context. Thus, the EOSC serves as a multiplier, maximizing the impact of contributions from KOINE's stakeholders and interest groups, well beyond the scope of KOINE as a project. KOINE's service, B2TERM, is instrumental for establishing a comprehensive, integrated European FAIR data space through structured terminological and semantic representation, terminology-based interoperability, accessibility and interlinking of research data. Additionally, B2TERM will provide a reliable foundation which existing and future components of the EOSC ecosystem can develop innovations in data exploitation based on terminological description. KOINE exemplifies this extensibility by enriching already existing services of the EOSC catalogue such as the B2FIND data search and discovery service with aligned, machine-readable terminologies semantics.

2.2 Measures to maximise impact

a) Dissemination and exploitation of results

Dissemination and exploitation of the B2TERM service and KOINE outputs are a major outcome of the KOINE project. KOINE will regularly promote the main achievements of the project to a broad range of stakeholders. The project aims to reach (inter)disciplinary operational data research infrastructures involved in the project as well as stakeholders interested in increasing the level of terminological and semantic interoperability in their data products and services. This includes various research communities, policy makers, data owners, technology providers, potential users of open data and data management solutions, and the general public.

To be successful, the activities planned in WP5 and WP7 will have to reach out to a large number of stakeholders potentially interested in using, evaluating and further refining B2TERM. This section lays out the dissemination plan. The KOINE dissemination and exploitation strategy consists of the following elements:

- KOINE Stakeholder Forum (from WP5)
- Implementation of an International KOINE Scientific Advisory Board (SAB)
- Exploitation plans of B2TERM early adopters (Ambassador Programme) and KOINE partners
- Community validation of B2TERM and B2FIND
- Knowledge management and intellectual property protection
- Capacity building and Training
- Data and Software management plan, open data and open code
- Scaling and sustainability plan within the EOSC
- Self-sustaining Business Model

KOINE Stakeholder Forum

The forum will be established at the very beginning of the project and will consist of terminology providers, users and data curators from different research domains. The stakeholders will come from discipline-specific communities (e.g., ecosystem and biodiversity, atmospheric and climate sciences, coastal engineering and linguistics), as well as from interdisciplinary groups (e.g., urban studies), composed by members of the KOINE consortium but also connected stakeholders (Figure 3). The KOINE Stakeholder Forum has the task to collect requirements, but also to validate KOINE's services. Finally, the KOINE Stakeholder Forum will help to establish KOINE's Scientific Advisory Board.

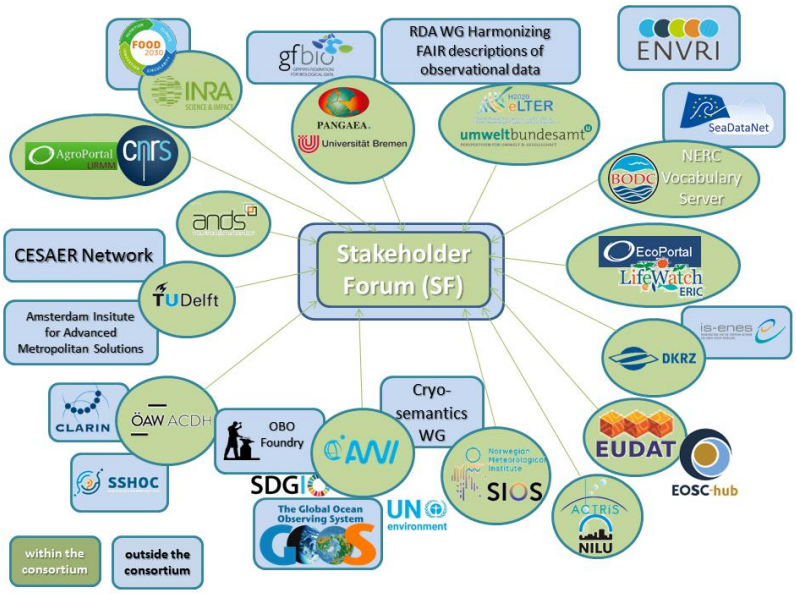


Figure 3: KOINE Stakeholder Forum (SF) with involved institutions, infrastructures and networks they represent.

International KOINE Scientific Advisory Board (SAB)

Engaging stakeholders is a major task of KOINE due to the fact that the adoption of terminological and semantic interoperability and FAIRness in the workflows of research communities and consequently, in their infrastructures often have a grassroots, bottom-up character. As a result, KOINE puts considerable effort in creating a basis of stakeholders to interact with and steer the project's course. Terminology providers and users from different research domains as well as data curators engaged through the SF (WP5) will be a source of members for KOINE's Scientific Advisory Board (SAB). The SAB will include (inter)disciplinary communities and will seek members from under-represented disciplines in KOINE (e.g., the life sciences). The experts on the SAB will ensure that KOINE is connected with on-going initiatives and programmes, such that its outputs will impact a broad audience by meeting current needs.

Communication and outreach with the KOINE Ambassador Programme

We will launch a KOINE Ambassador Programme comprising early adopters in the development and refinement of case studies and services. These Ambassadors will constitute the first promoters of project results (WP7), having a key role in raising awareness about the concerns addressed by the project and demonstrating the KOINE's services functionality to their communities. Ambassadors will also actively promote B2TERM integrations in infrastructures beyond those carried out by KOINE, at national, European and global scale. The KOINE members will equip the Ambassadors with tutorial and training materials and guidelines to provide consistency in their dissemination and impact.

Community validation of B2TERM and B2FIND

Iterative testing of the B2TERM and B2FIND services with research communities and other stakeholders will be a central aspect in KOINE. Indeed, the impact of the project and its promotion will rely on the thorough testing of its products. During the project, the validated test results will be communicated back to developers (WP2 and WP3) in order for them to effect changes and express any persistent needs to the

Operations Committee (WP6). After the project, a B2TERM Helpdesk provided by TIB will serve as the point of contact for further feedback and community-led development.

Knowledge management and intellectual property protection

The KOINE Consortium Agreement – based on official guidelines and recommendations in Horizon 2020 – will provide the framework for knowledge management and IPR issues in the project. In particular, the consortium partners will ensure that all knowledge that is required for the implementation of the B2TERM service will become available to all relevant partners. The KOINE consortium will ensure that publications are at least green open access and – in cases where excellent journals without open access options are targeted – the proposal budged allows for gold open access. IPRs of integrated third-party software components will be taken into account to avoid any infringement of third-party rights. This will ensure that the dissemination of KOINE outcomes will not be inhibited by improper rights management.

Capacity building and Training

In the KOINE project, both capacity building and training are essential tools to provide terminology providers and users from various research communities with the awareness and skills needed to take full advantage of B2TERM and B2FIND services (WP5). To maximize the impact of both, KOINE focuses on two key aspects: The development and broad distribution of Open Educational Resources (OERs) and the delivery of "Terminology Carpentry" workshops (see also Section 2.2(b)).

Data and Software management plan[47]

As a project within the EOSC framework, the KOINE project will adopt the principles of Open Science and develop a culture of Open Science practices and interdisciplinary research to foster scientific discoveries and innovation towards a broad range of disciplines in EU. A major expected outcome of KOINE is software source code that will be developed publicly, following established guidelines[48] (see also Table 12) and licensed as recommended by the Open Source Initiative[49]. This will permit all interested stakeholders and developers outside the consortium to be impacted through interaction with KOINE's development repositories, and broadly share, adapt, and potentially contribute to the project's progress.

Scaling and sustainability plan within the EOSC

KOINE partners will develop, evaluate and promote the B2TERM service as an element of the EOSC Service Catalogue (WP6). To scale up the project's outcomes, the KOINE Terminology Interoperability Framework (TIF) will determine, for the first time, a set of Compliance Standards for the connection of FAIR terminologies in the EOSC (WP2, WP6). These Compliance Standards will be used as a guide for new terminology providers to enable Terminological and Semantic Interoperability of their own systems as well as the connection to B2TERM. After KOINE, the sustainability of the B2TERM service and its role in aligning key terminology providers will be ensured by TIB in collaboration and under advice of the SAB. TIB will operate the service and offer B2TERM to interested stakeholders via the EOSC Service Catalogue (WP6).

Self-sustaining Business Model

A not-for-profit operational and self-sustainable business model for B2TERM will be developed in WP6. This allows B2TERM and its core components (e.g., the Terminology Interoperability Framework) will be operational at TRL 8-9, to be used and further developed by all interested stakeholders in the long-term, perpetuating its impact and permeation across communities. A Business Model Canvas (BMC) will be developed, maintained and regularly updated during the project.

[47] The Software Sustainability Institute. (2018, December 10). Checklist for a Software Management Plan (Version 1.0). Zenodo. http://doi.org/10.5281/zenodo.2159713

[48] Jiménez RC, Kuzak M, Alhamdoosh M et al. Four simple recommendations to encourage best practices in research software. F1000Research 2017, 6:876. http://doi.org/10.12688/f1000research.11407.1

[49] http://opencource.org/licenses

Two possible models for cost-recovery are envisioned:

- **Value-added service and support model:** While the basic (but fully functional) B2TERM service will be freely provided to interested stakeholders via the EOSC Service Catalogue, service-level agreements and value-added services (e.g., professional training, mass requests) will be offered against a fee.
- **Bespoke service model**: As above, the core B2TERM service will be provided at no cost, however, fees would be charged for provisioning of consulting, commercial support, maintenance, deployment/integration and custom extension development for stakeholders, who would want to implement B2TERM components on top of their own infrastructure.

b) Communication activities

Communication activities raising awareness about KOINE, its achievements and expected results will start early in the project, with the objective of establishing a fit-for-purpose B2TERM service in community dialogue and development. These channels will cover traditional dissemination means (flyers, posters, public web-pages, scientific/professional conference/events presence, press releases) as well as social network dissemination channels (e.g., social media and networks like Twitter, social software development solutions etc.). An action plan will be created for KOINE dissemination activities, during and beyond the lifetime of the project. This plan will be maintained and consistently updated throughout the project (WP7) and will also include communication strategies to secure funding opportunities. Table 12 adds further details on the scope of KOINE communication activities.

Among these, KOINE will place particular focus on building an online presence that speaks to the EOSC community, and allows us to:

- Disseminate information about the work and progress in KOINE and the development of the B2TERM service in particular
- Provide access to the Terminology Interoperability Framework (TIF) for community input
- Disseminate a Common Semantic Model for the description of variables, interoperable with the Essential Variables for Oceans, Biodiversity, Ecosystems, and Climate
- Attract additional data research infrastructures or e-Infrastructures and interdisciplinary operational data discovery services to integrate the B2TERM service
- Discuss important developments, both technical and social, and further progress the exchange and use of terminology required to describe research data
- Communicate about the set of Compliance Standards for the connection of FAIR terminologies in the EOSC and gather feedback from the research communities and other important stakeholders.

Table 12: Details on KOINE communication activities.

Activity	Description
Project website	A central hub for all relevant information about the project, which aggregates all dissemination, engagement and communication activities of KOINE into one conclusive online presence, including social media accounts (see below).
Logo and branding	Graphic design for project logo in line with the project's identity/brand, integration in the project website and templates.
Press releases	Press releases about KOINE and B2TERM will be made in national and European contexts with particular focus on future EOSC activities.
Dissemination material	Design, content creation and production of leaflets, poster templates, videos, and communication packages for KOINE services.
Social media	Multiply communication efforts by creating KOINE social media accounts, including Facebook, Twitter, YouTube, ResearchGate.
Open Source Software (OSS)	Developed products of KOINE will be published on code hosting platforms (e.g., GitHub.com or GitLab.com). This ensures automatic archival in

	SoftwareHeritage.org, while release versions will be made citable through Zenodo.org.
Open Educational Resources (OERs)	KOINE offers OERs combining teaching resources on semantic terminology and semantic technologies in cooperation with the international teaching framework "The Carpentries". The OERs' aim to increase the uptake of B2TERM. KOINE will incorporate OERs into the Carpentries curriculum and maintains them in order to reach out to a broad range of international research communities.
'Terminology Carpentry' Workshops	These workshops will show users how to integrate and utilise the B2TERM service developed by KOINE.
Public events	Midterm and final public events will trigger adoption, collect feedbacks and promote usage of KOINE services. These will include participation in EOSC meetings and other conferences (e.g., EUDAT) with booths and demonstrative sessions.
Publications	Dissemination via open access (Gold or Green) scientific articles, publication of KOINE presentations and posters on the KOINE website and content sharing platforms (e.g., SlideShare, FigShare).

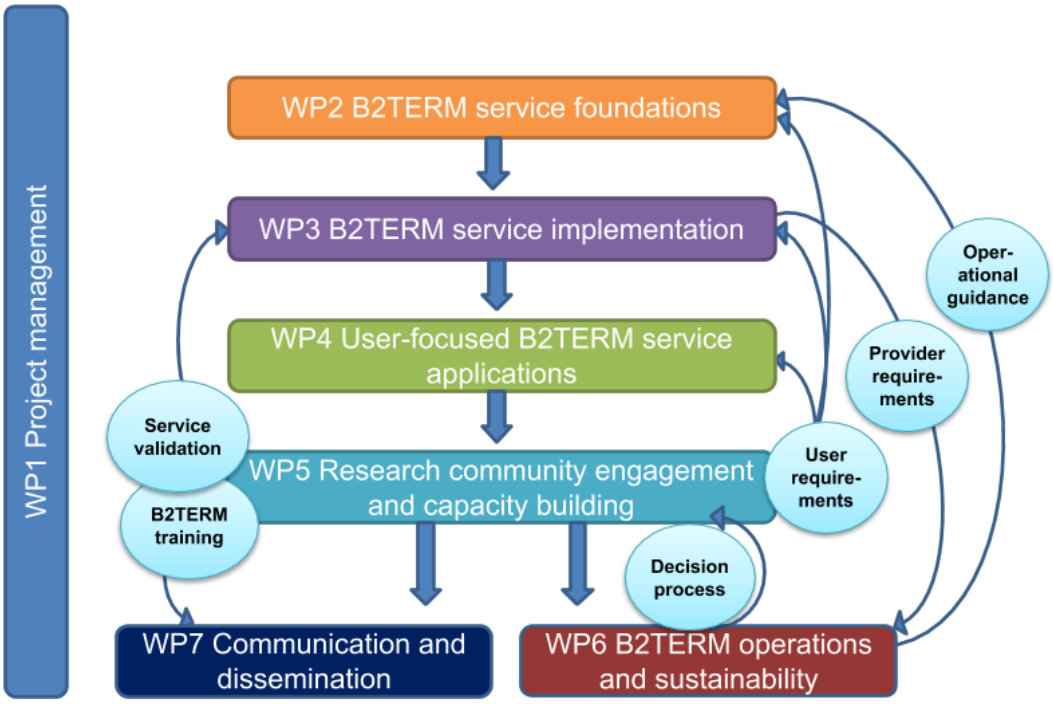


Figure 4: Dependencies among KOINE WPs.

3. Implementation

The KOINE work programme is structured in seven WPs (Figure 4 and Table 3.1a). WP1 is concerned with project management. WP2 will study and analyse the generic and discipline-specific terminology landscape and develop the specifications for the Terminology Interoperability Framework (TIF), B2TERM and its integrations in data centres. WP3 will integrate these results, thus creating the TIF and implement and deploy the B2TERM service. WP4 ensures that B2TERM is integrated in (inter)disciplinary operational research data infrastructures. The effort allocated to WP4 underscores KOINE's belief in integrating terminology in systems and services researchers use for data management. WP5 focuses on social integration and will create a B2TERM Stakeholder Forum. Through the Forum, a wide range of scientific domains and communities

will be involved to gather requirements for how to harmonize and publish terminology in the EOSC and offer capacity building and training to ensure unambiguous terminologies are used in research data. In WP6, the KOINE project will focus on governance structure, sustainability and cost-recovery of the developed B2TERM service, which are essential to ensure the long-term availability and service provision through the EOSC Service Catalogue. WP7 will look into the best ways to ensure dissemination and communication, which includes launching an Ambassador Programme (WP7) that will create a first group of B2TERM promoters. Each individual WP within the project has its own roadmap and objectives, and the consortium members will work towards these throughout the project. However, as the high-quality outputs generated by the individual WPs are essential to the success of the overall project, all WPs must strive for success in the overall work plan, requiring close integration and alignment between the executed activities. The time frame of the main tasks foreseen in the work plan is shown in Figure 5 as Gantt chart.

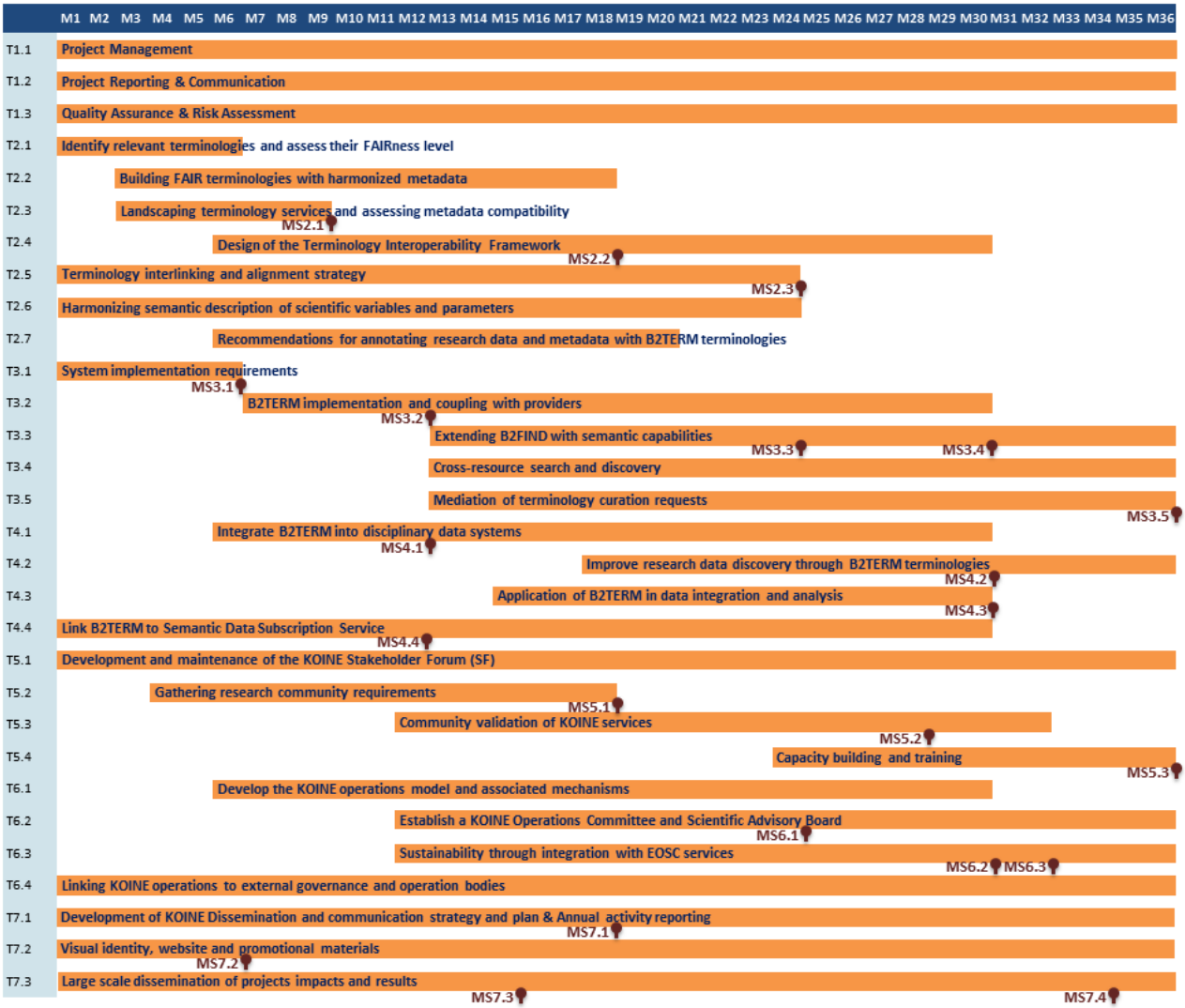


Figure 5: KOINE Gantt chart showing task durations and milestones.

3.1 Work plan — Work-packages, deliverables

Table 3.1a: List of work-packages

Work-package No	Work-package Title	Lead Participant No	Lead Participant Short Name	Person-Months	Start Month	End month
1	Project management	1	TIB	30	1	36
2	B2TERM service foundations	2	CNRS	97	1	30
3	B2TERM service implementation	7	UvA	120	1	36
4	User-focused B2TERM service applications	4	UHB	187	1	36
5	Research community engagement and capacity building	3	EAA	143	1	36
6	B2TERM operations and sustainability	5	AWI	48	1	36
7	Communication and dissemination	10	LW ERIC	34	1	36
				659		

Table 3.1b: Work-package description

Work-package number	1		**Lead beneficiary**		TIB		
Work-package title	Project management						
Participant number	1	2	3	4	5	7	10
Short name of participant	TIB	CNRS	EAA	UHB	AWI	UvA	LW ERIC
Person months per participant	24	1	1	1	1	1	1
Start month	1			**End month**			36

Objectives

The aim of this WP is to ensure the high-quality level of the project's results. It will do this via the continuous monitoring of the implementation and completion of the project tasks, activities, milestones and deliverables, safeguarding their proper and timely development according to the Description of Work (DoW) and the project's work-plan, while ensuring the successful collaboration among the partners.

Description of work

Task 1.1: Project Management [M1-M36]
Lead: TIB, Participants: CNRS, UvA, UHB, EAA, AWI, LW ERIC. Effort: 10 PM
This task deals with all necessary project management tools, mechanisms and structures for the high quality, efficient and timely administrative coordination of the project. It incorporates Administration Management activities, including procedures and guidelines for activity planning and monitoring, cost and time management, submission of periodic progress reports and cost statements, preparation of annual review reports, review presentations, and timely submission of high quality deliverables to the Commission. TIB will be responsible for the day-to-day coordination of project-related activities and tasks, as well as the administrative management of the project; contributions will be made by all the Core Consortium partners (participation in project meetings, teleconferences, taking over responsibilities, reporting, coordination of allocated WPs and tasks).

Task 1.2: Quality Assurance & Risk Assessment [M1-M36]
Lead: TIB, Participants: CNRS, UvA, UHB, EAA, AWI, LW ERIC. Effort 10 PM
The task focuses on defining and specifying the appropriate mechanisms and processes that will be established in order to maintain high quality. Additionally, T1.2 deals with the identification of potential project management risks and the respective monitoring of each risk profile as well as with the definition and timely application of contingency plans. TIB will organize the activities in this task, with input from the Work-package leaders, focusing on the preparation of periodical risk reports, the identification of challenges, the suggestion of remedial actions and the implementation of any corrective measures, when necessary.

Task 1.3: Project Reporting & Communication [M1-M36]
Lead: TIB, Participants: CNRS, UvA, UHB, EAA, AWI, LW ERIC. Effort 10 PM
The project coordinator will act as the point of contact for partners in communications with the Commission. The coordinator will ensure that the annual reporting to the EC, semi-annual technical internal reporting, milestone review and midterm review after 12 months will be implemented as planned in the DoW. TIB, as the project coordinator, will lead this task. All Work-package leaders will contribute to the relevant activities.

Deliverables

D1.1 Project Management Plan [M3]. The core management handbook and structure will provide guidelines for the project manager and the project members to follow during the full cycle of the project. The handbook will set out the governance, project calendar, cluster and work-package description and reporting roles and responsibilities for all partners. Moreover, it will define the communication mechanisms for the project's partners and the templates for reporting and formal communication of the project's activities and outcomes.

D1.2 Annual Public Report [M12, M24, M36]. Report on the project's progress, targeting the general public. The report will focus on the impact of the conducted work in terms of developing a novel service for interdisciplinary data interoperability and integrating the service in operational infrastructures (WP4) and research communities (WP5), thus providing an essential component for the realization of the EOSC vision.

D1.3 Quality Assurance & Risk Assessment Plan [M6]. The plan will incorporate details on the quality assurance processes adopted within KOINE. It will define all processes and instruments to be used for the regular quality monitoring and risk assessment in the form of a handbook for project partners.

D1.4 Data Management Plan [M6]. The report lists research data volume, access, licencing and integration according to the legal framework / GDPR. This report will also detail data characteristics, privacy/security plans, authorisations and thus answer data security and privacy questions, such as: Where will the data be physically processed? What are the physical security protection features? Which privacy protocols are implemented? Which guidelines for personal data processing will be enforced?

Work-package number	2		**Lead beneficiary**		CNRS						
Work-package title	B2TERM service foundations											
Participant number	1	2	3	4	5	7	8	10	11	19	20	22
Short name of participant	TIB	CNRS	EAA	UHB	AWI	UvA	UMAAS	LW ERIC	UKRI	INRA	TUD	ANDS
Person months per participant	5	24	4	3	12	5	4	12	20	4	4	0
Start month	1			**End month**				30				

Objectives

WP2 aims to investigate properties and aspects related to terminologies (accessibility, format, metadata description, exploitation, condition of use, alignment) and federate diverse terminology services (ontology repositories, vocabulary services, etc.). We argue that KOINE shall encourage and enable FAIR terminologies, a key feature of B2TERM. Therefore, this WP hosts research activities necessary to inform and build the foundations of the project, addressing the technical challenges of WP3 and WP4. The WP objectives are:

- To landscape the terminologies and terminology services that will be harvested by B2TERM, starting with those that are of immediate relevance to KOINE partners (WP4) and extending to a wider spectrum, to enhance exploitation and adoption of B2TERM in the wider community (WP5).
- To assess the FAIRness of terminologies inside the terminology services using a scoring system based on harmonized and enriched terminology metadata description to serve this information to B2TERM.
- To design the architecture of the Terminology Interoperability Framework (TIF) and specify an application programming interface (API) that will be implemented by terminology services in order to provide content to B2TERM.
- To identify a Common Semantic Model for human and machine readable descriptions of scientific variables and parameters across multiple disciplines by building on existing and emerging standards where possible.
- To provide guidelines to users of B2TERM for semantically annotating research data and data sets with FAIR terminologies. The recommendations will be used by current and new users of B2TERM.

WP2 includes three of KOINE's terminology services providers (AgroPortal, EcoPortal and NERC-VS) as well as developers and stakeholders of important terminologies in KOINE's ecosystem (INRA, EAA, AWI, UHB). Other expertise is brought by UMAAS, TIB and TUD.

Description of work

Task 2.1: Identify relevant terminologies and assess their FAIRness level [M1-M6]
Lead: LW ERIC, Participants: CNRS, UKRI, EAA, INRA, AWI, ANDS. Effort: 4 PM

The key to data interoperability is their annotation with FAIR terminologies that are easily accessible from multiple domains. This task will landscape the terminologies (i.e., controlled vocabularies, taxonomies, thesauri and ontologies) that will be used by B2TERM and manually assess their FAIRness level (before integration). Initially, the task will undertake a review to identify important terminologies used to annotate and discover datasets by KOINE partners who represent multiple scientific domains. We will rationalise to key terminologies (e.g., units, parameters, instruments, taxons) and this review will be used as the basis in other tasks within this and other WPs, such as in T2.6 and WP4. Such landscaping has already been conducted in the biomedicine and agronomy domain and has proven to be essential in building terminology services for these areas. Our recent experience in building AgroPortal (CNRS-LIRMM) will support and accelerate the creation of EcoPortal. EcoPortal is a new terminology service within the ecology and biodiversity domain that is also engaged in the project. Redundancies between terminology services (e.g., a terminology being present in several services) will be reported and discussed with the Operations Committee (T6.2). Where an identified terminology is not available in a format that can be handled by one of the KOINE terminology services, this task will liaise with the developers of the service to adopt standard practices appropriate for B2TERM. This may potentially be required for some terminologies used by KOINE, such as WORMS and ITIS terminologies that are heavily used in PANGAEA.

Task 2.2: Building FAIR terminologies with harmonized metadata [M3-M18]
Lead: CNRS, Participants: UKRI, LW ERIC, EAA, INRA, AWI, UMAAS. Effort: 16 PM

With the adoption of the semantic web and linked data, the volume of terminologies has grown notably across disciplines. Consequently, the following questions arise: Which terminology should we reuse or trust? How to evaluate the quality or relevance of a terminology? With the rapid expansion of new and overlapping terminologies, properly describing them with sufficient and accurate metadata that enable users to evaluate, identify and select them is becoming increasingly important. Terminology services are ideally positioned to describe metadata about terminologies such as scientific scope, description, community, adopters, level of expressiveness, annotation scenario supported, relations. In this task, we will investigate the commonalities and differences between terminologies and terminology services identified respectively in T2.1 and T2.3 to

produce a standardised metadata profile to accurately describe terminologies. As encouraging FAIR terminologies is a core objective of KOINE, T2.2 will develop a set of indicators specific to terminologies, to assess their FAIRness level. The indicators will be based on the recently proposed FAIR metrics[50] and the standardised metadata profile produced in this task. This will support the development of the FAIRness Scorer (WP3) automatically computing the FAIRness score of terminologies in B2TERM. With WP5, we will consolidate the preliminary work started in 2018 by the RDA Vocabulary and Semantic Services Interest Group (VSSIG) on "ontology metadata" and deliver a RDA recommendation that will be adopted by each terminology service within KOINE as well as by B2TERM. Additionally, this task ensures that KOINE's strategy to describe terminologies is compatible with standard vocabulary approaches such as Google's Schema.org, W3C's Data Catalog Vocabulary (DCAT), Asset Description Metadata Schema (ADMS), etc.

Task 2.3: Landscaping terminology services and assessing metadata compatibility [M3-M9]
Lead: UvA, Participants: CNRS, UKRI, INRA, ANDS. Effort: 6 PM

While identifying key terminologies in T2.1 and their associated metadata in T2.2, we will review and compare the technologies between a wide range of existing portals, vocabulary and ontology repositories, or vocabulary services. These will include ontology repositories such as, NCBO BioPortal, AgroPortal, EBI OLS, AberOWL, OntoHub, LOV or vocabulary services such as, NERC, ANDS, Bartoc.org, EIONET or underlying technologies such as SKOSMOS, etc. We will evaluate the differences, specificities and functionalities with respect to terminology lifecycle management, alignment/interlinking and semantic description of terminologies. Results from this task will inform the design of the TIF architecture in T2.4. In this task, we will also assess the metadata richness of each reviewed terminology service or any indicator relevant for terminology services to participate into B2TERM. This information will contribute to the FAIRness Scorer (T2.2). Finally, we will deliver a state-of-the-art recommendation report that will be adopted for community engagement (WP5) and outreach (WP7).

Task 2.4: Design of the Terminology Interoperability Framework (TIF) [M6-M30]
Lead: CNRS, Participants: UKRI, LW ERIC, UvA, TIB, ANDS. Effort: 21 PM

This task will work with T5.2 (community user requirements) and T3.2 (implementation) for the specification and design of the TIF.

Subtask T2.4.1: Design KOINE's Terminology Interoperability Framework

We will establish the architecture of the TIF in terms of its components, their relations and design. The architecture specification will be implemented in the development of B2TERM in T3.2. We will rely on a distributed approach, consuming and producing web services built with state-of-the-art technologies (REST web service and JSON-LD content). This task will also ensure that the architecture can support the slice mechanism described previously in Section 1.3. Slices (an idea originally proposed in the NCBO BioPortal and also available in AgroPortal) currently allow users to interact (both via API or user interface) with a pre-defined subset of terminologies in AgroPortal or BioPortal, making browsing or the integration of terminologies into workflows more efficient and manageable. In addition, the terminology service features are customized to the chosen subset, enabling users to focus on their specific use cases. For instance, INRA researchers can access only vocabularies developed by their fellows on AgroPortal's slice http://lovinra.agroportal.lirmm.fr. In KOINE, we will incorporate and expand on this functionality, enabling slices in B2TERM and solutions that not only customize the list of terminologies, but also the activation of services (search, annotation, recommendation, communities syndication) in a slice. This functionality will ensure that B2TERM can be customised to suit the needs of specific sub-groups of terminology users or specific scientific communities.

Subtask T2.4.2: Specify a reference API for terminology services to participate into B2TERM

This subtask will specify the reference standard API that terminology services involved within KOINE should implement to provide content to B2TERM. This specification will be publicly available for other terminology services that would embrace the project vision and would also like to participate in B2TERM. We will use the OpenAPI (formally Swagger) for the specification and reference implementation. Note that B2TERM will itself re-expose its content following the same specification. Once the API is specified and benchmarked to work with the TIF architecture, it will be implemented in T3.2 by each KOINE's terminology services.

[50] Wilkinson, M. D., Sansone, S. A., Schultes, E., Doorn, P., da Silva Santos, L. O. B., & Dumontier, M. (2018). A design framework and exemplar metrics for FAIRness. Scientific data, 5. https://doi.org/10.1038/sdata.2018.118

Task 2.5: Terminology interlinking and alignment strategy [M1-M24]
Lead: CNRS, Participants: UKRI, LW ERIC, EAA. Effort: 9 PM

In this task, we will investigate, develop and implement the strategies necessary to deal with terminology interlinking and alignment at terminology services. Terminologies as designed independently, by different developers, and following diverse modelling principles and patterns can be both heterogeneous but also overlapping in terms of coverage. To achieve interoperability and integration, one solution is to identify/generate mappings (or correspondences) between different terminologies of the same domain. This process is known in the semantic web domain as ontology matching or ontology alignment. Building algorithms to identify these mappings is itself a scientific challenge, but when dealing with terminology services we also must address all issues related to mapping representation/description, generation, validation, merging, visualization, storage and retrieval. Furthermore, current tools to map terminologies are varied in their functionality. For example, some lack mapping quality or provenance. In addition, mappings are not uniformly described. Therefore, this task will enable KOINE's terminology services to work on manual, semi-manual or automatic alignment in order to deliver properly aligned terminologies to B2TERM, enabling end-users to leverage the semantics of interlinked terminologies through this service. Additionally, this task will setup the mechanisms (inside KOINE's terminology services) to handle B2TERM mediated queries related to terminology alignment. The results from this task will be used in T2.6 to facilitate the harmonization of parameters and in support of activities in WP5 and as guidance for the governance of new terminology services in B2TERM in WP6.

Task 2.6: Harmonizing semantic description of scientific variables and parameters [M1-M24]
Lead: UKRI, Participants: CNRS, LW ERIC, EAA, AWI, UHB, TIB. Effort: 26 PM

This task focuses on the terminologies necessary to express scientific variables and parameters unambiguously and with accuracy. In the real world of scientific monitoring and research, variables and parameters are complex concepts. They are often reported using simple terms (such as "temperature") or simple association of terms (for example "air temperature", "body temperature") that are chosen for their conciseness. However, such compact labelling is invariably ambiguous and dependent on context. In order to be described without ambiguity, any given measurement value must be associated with a term from a FAIR terminology that precisely links meaning with syntax. Clarity is further enhanced if terms adopt consistent and explicit syntax; for example: "Temperature of the air in the room" or "Concentration of total petroleum hydrocarbons per unit dry weight of sediment <63um." Re-using well established terminologies and conceptual models, our objectives are: a) to create a Common Semantic Model that will be used as part of the KOINE's TIF under T2.4.1 to conceptualise the descriptions of scientific variables and parameters; b) to feed the semantic model with FAIR terminologies; and c) to connect this to higher level discovery and aggregation concepts. Work will be organized in two subtasks:

Subtask 2.6.1: Deliver the Common Semantic Model

This sub-task will deliver the Common Semantic Model early in the project. It will be used in the design of KOINE's Terminology Interoperability Framework (T2.4.1) for implementation in WP3 (T3.2) and data annotation by WP4 (T4.1). We will use the study cases put forward by partners in T5.2.2 as a starting point and analyse them in the context of existing terminology models, ontologies and resources used to conceptualise variables and parameters in different disciplines. The emphasis will be on re-using existing models and concepts to minimise redundancy and optimise interoperability. As part of this sub-task we will select terminologies assessed under T2.1 to standardise the structured description of variables and parameters. This sub-task will also establish foundation for rules of governance in connection with WP6 and issue recommendations for user-centric, user-friendly downstream applications and service implementation in WP3 (with input from WP4 RIs and WP5 user communities).

Subtask 2.6.2: Terminology alignment through coordinated development

Using the alignment and interlinking strategies developed in T2.5, this subtask will align KOINE's Common Semantic Model with existing conceptual models (in particular, OGC's Observations and Measurements, O&M), ontologies (SOSA, OBOE, SERONTO, OBO ontologies) and keep it aligned with global standardisation efforts (such as Schema.org and Bioschemas). We will also seek interoperability with controlled vocabularies registered under the 2007/2/EC INSPIRE Directive[51]. It will contribute to and be informed by work on observational data parameter harmonisation taking place under the auspice of RDA

[51] http://inspire.ec.europa.eu/codelis, http://dd.eionet.europa.eu/vocabularies

VSSIG (WP5). In addition, we will extend, align and harmonise selected terminologies used in KOINE to specifically support internationally relevant multidisciplinary Essential Variables focusing on Climate (ECVs), Ocean (EOVs) and Biodiversity (EBVs).[52]

Task 2.7: Recommendations for annotating research data and metadata with B2TERM terminologies [M6-M20]

Lead: UKRI, Participants: CNRS, LW ERIC, UvA, UHB, AWI, TUD. Effort: 15pm

This task will deliver guidelines for annotating research (meta)data (observational, experimental, simulation) using the references to the unambiguous and persistently identified terms from B2TERM (e.g., http://vocab.nerc.ac.uk/collection/P01/current/TEMPPR01/) in order to facilitate data interoperability, discovery and exploitation within KOINE and beyond. The recommendations will draw on KOINE and research community user requirements gathered from partners in this task and T5.2, respectively. The guidelines will be used by B2TERM early adopters (e.g., PANGAEA, TUD 4TU.Centre for Research Data, B2FIND) to integrate terminologies into KOINE disciplinary data systems (RIs) as part of semantic annotation activities in T4.1. They will also be used to support capacity building and training workshops within research communities (T5.4) and as guidelines for other external and future users of B2TERM. Solutions will primarily focus on exposing annotations to users and machines, for example, on dataset landing pages, in published data files and including them in machine readable metadata by developing extensions to the metadata standards that are used by our stakeholders, such as Schema.org.

Deliverables

As WP2 has 7 different tasks that are important for the rest of the project, it has 5 deliverables that consist of synthetized recommendations or specifications. One deliverable is mutualized with T3.1 and described in WP3. Another one is the product of two tasks (T2.5 & T2.6).

D2.1 Report on identification and listing of terminologies relevant for KOINE partners [M6]. This deliverable will thus build from results of T2.1 and T2.3

D2.2 Recommendation for FAIR terminologies metadata description [M18]. This deliverable will thus build from results of T2.2.

D2.3 Description of TIF architecture and API specification for B2TERM [M12, M18]. This deliverable will thus build from results of T2.4.

D2.4 Common Semantic Model to describe scientific variables and parameters across disciplines at M9 and final report on parameter harmonisation at M24 [M9, M24]. This deliverable will thus build from results of T2.5 and T2.6.

D2.5 Recommendation on annotating research data with B2TERM terminologies [M15]. This deliverable will thus build from results of T2.7.

Work-package number	3	**Lead beneficiary**		UvA		
Work-package title	B2TERM service implementation					
Participant number	1	2	7	10	11	15	21
Short name of participant	TIB	CNRS	UvA	LW ERIC	UKRI	DKRZ	OEAW
Person months per participant	16	17	35	8	13	22	9
Start month	1			**End month**		36	

[52] Pereira, H. M., Ferrier, S., Walters, M., Geller, G. N., Jongman, R. H. G., Scholes, R. J., ... & Coops, N. C. (2013). Essential biodiversity variables. Science, 339(6117), 277-278. https://doi.org/10.1126/science.1229931

Objectives

This WP implements the B2TERM service and the Terminology Interoperability Framework, on which it relies, coupling existing terminology services to B2TERM and implementing the API recommendations specified by T2.4 so that B2TERM (and external systems committed to interoperation via T6.4) can more easily harvest terminologies and their descriptions. B2TERM is the basis then for improving search and discovery of terms by researchers and other stakeholders and it encourages consistent use of terms to enhance the interoperability of research data. The objectives of this WP are therefore to:

- Implement the formal specifications of T2.4 to produce interoperability in the registration, interlinking, and service of terminologies within EOSC by independent content providers.
- Provide a common terminology service for researchers, B2TERM, and perform integration of B2TERM into the existing B2FIND service.
- Apply KOINE's Terminology Interoperability Framework to selected existing terminology services: AgroPortal, EcoPortal, NERC-VS and CLARIN Terminology Services.
- Build additional facilities for cross-resource search, discovery and mediation of request related to terminologies upon the B2TERM service, including support for the construction of community-specific slices of terminologies to better promote standard practices and provide tailored guidance to researchers in specific research domains.

All development in this WP is guided by the design specifications and recommendations in WP2 and requirements of researchers and research infrastructures as gathered in WP5 for application in WP4, and is based on existing technologies, developed by bodies such as EUDAT and NCBO using expertise provided by the partners in the WP.

Description of work

Task 3.1: System implementation requirements [M1–M6]
Lead: UvA, Participants: TIB, DKRZ, CNRS. Effort: 7 PM

Establish the key architectural and implementation requirements of B2TERM (T3.2) and the extended B2FIND (T3.3) along with the requirements of additional services for search and mediation (T3.4 and T3.5). This includes some review of the technologies provided by NCBO, EUDAT and individual partners (all TRL 6 or greater) identified prior to the initiation of the project against the latest developments so as to make final recommendations based on the most contemporary information available. This task will:

- Finalise the technology requirements for the implementation of B2TERM, extension to B2FIND and the implementation of search and mediation services built upon B2TERM.
- Confirm technology choices for implementing the B2TERM service in terms of suitability and maturity.
- Compile a detailed review of existing technologies to be disseminated both within the project consortium and to the wider community to ensure awareness of and alignment with selected technologies.

The output of this task will be a public deliverable.

Task 3.2: B2TERM implementation and coupling with providers [M7–M30]
Lead: UvA, Participants: TIB, UKRI, LW ERIC, CNRS, OEAW. Effort: 62 PM

Construct B2TERM, deploy it as a live production service and work with participating terminology services to couple their existing resources with the service. B2TERM will act as a lightweight federator for existing terminology services that implement the APIs defined in T2.4.2, providing content to the sub-services developed in Tasks 3.3–3.5. This task will implement the service based on the architectural specification defined in T2.4.1 and requirements defined in WP5, and it will also implement any necessary internal indexing or caching mechanisms required to support search, discovery and mediation without usurping the established role of existing providers, as well as support any clients or other applications deployed in WP4. This task will:

- Implement the B2TERM service and its interfaces using the standard protocols established in WP2, acting as a base for implementation and integration activities in WPs 3 and 4.
- Build upon technologies already implemented at the terminology service level (e.g., in AgroPortal and EcoPortal) to expose their functionalities at a federal level and provide cross-provider functionality without having to rebuild it from scratch.

- Enable the coupling of the B2TERM service with KOINE's terminology services (AgroPortal, EcoPortal, NERC-VS, CLARIN-TS) and provide the means to couple other terminology services developed outside of the project that offer their content following the API and specification discussed in WP2.
- Develop a system for evaluating FAIRness of terminologies served by B2TERM. This FAIRness assessment capability will be based on the accepted published FAIR guidelines and the specification of T2.2; it will consume harmonized metadata information about the terminologies originating from the source repositories.

The outputs of this task will be demonstrated via an implemented B2TERM service running as a production service.

Task 3.3: Extending B2FIND with semantic capabilities [M13-36]
Lead: DKRZ, Participants: UvA. Effort: 22 PM

Extend the existing B2FIND service to leverage B2TERM capabilities. The B2FIND service provides an endpoint to discover, localize and access EOSC research data and will be enhanced, in conjunction with B2TERM, with greater semantic functionalities. These advanced capabilities include semantic search for data by filtering using the Common Semantic Model and validating of metadata versus specific terminologies. This gives users the ability to see whether a metadata set is compatible with a given vocabulary or ontology and enhances the interoperability of research data. This entails augmentation of B2FIND with:

- The ability to search by terminological element (e.g. vocabulary or ontology term), with basic support for alternative labels, synonyms and related terms.
- B2TERM functionalities incorporated into the B2FIND service, allowing B2FIND to directly leverage terminology semantics in various contexts.
- Validation of research dataset metadata against terminologies available in B2TERM (e.g., check if a terminology annotated to a keyword or another facet in a metadata record is compliant. (Concrete example: The keyword 'Temperature' is assigned to a ontology about 'Body temperature', but B2FIND-B2TERM validation shows, that 'Climate temperature' is meant).

The outputs of this task will be demonstrated via an augmented B2FIND service running in production with accompanying documentation.

Task 3.4: Cross-resource search and discovery [M13–M36]
Lead: UvA, Participants: TIB, CNRS. Effort: 19 PM

Develop a framework for cross-resource search and discovery that is both reliable and scalable with large numbers of distributed terminologies, and which can also intelligently filter large numbers of highly similar or replicated entries. In order to ensure consistent application of terminologies within scientific disciplines, prevent confusing proliferation of near-synonymous terms and support communities of practice, the 'slicing' model advocated by WP2 will be adopted to define community/domain-specific views across all terminologies. This task will:

- Develop a mechanism to produce community/domain-specific 'slices' of indexed terminologies to support best semantic practices within communities.
- Extend the existing (RESTful) terminology index and lookup API in place in AgroPortal / BioPortal / EcoPortal to support search and discovery across terminologies within a given slice that can be invoked directly by both terminology services and clients (e.g. those developed in WP4 in Task 4.1).
- Implement a publish/subscribe mechanism for harmonising terminology data between registered terminologies via a central facility (to enable federated search and mediation via a single gateway or replicated set of gateways).
- Develop a slice administrator mechanism allowing a restricted admin group to select terminologies, and components in their community specific slice.

The outputs of this task will be demonstrated via a search portal service, both for 'full-scope' search and for 'sliced' browsing, plus documentation for how to invoke via other clients.

Task 3.5: Mediation of terminology curation requests [M13–M36]
Lead: TIB, Participants: UvA, CNRS, UKRI, LW ERIC. Effort: 10 PM

Develop the framework for mediation of requests for terminology curation by consumers to producers. In addition to finding and accessing terms (in a sliced manner), B2TERM will also offer a feature for brokering requests to curate and augment terminologies, e.g., to request new terms, declare equivalencies/mappings between terms, or any other issues, removing the need for users to know the specific contact points to make

requests for specific terminologies. The system developed should encourage end users' participation and feedback. This activity will be supported by a helpdesk after the project completes; this task will be focused on technical aspects. It will:

- Define the necessary mediation framework for making curation requests.
- Implement the framework within the context of the B2TERM service.

The output of this task will be demonstrated via a mediation interface, supported by documentation of the mediation process from both the user and terminology services' perspectives.

Deliverables

D3.1 State of the Art review of semantic technologies [M6]. A review of technologies applicable to the implementation of B2TERM and KOINE in general.

D3.2 B2TERM implementation [M24]. A reference implementation of B2TERM, a service for accessing and using FAIR terminologies in EOSC.

D3.3 Extension of B2FIND with B2TERM semantic capabilities [M36]. A reference implementation of B2FIND with B2TERM integration for EOSC.

D3.4 Seamless search and discovery of terminological content by domain [M36]. The implemented terminological search and discovery service for KOINE.

D3.5 Mediation framework for terminology curation and augmentation [M36]. The implemented terminological curation mediation service for KOINE.

Work-package number	4			**Lead beneficiary**				UHB							
Work-package title	User-focused B2TERM service applications														
Participant number	1	2	3	4	6	9	10	12	13	15	17	18	19	20	21
Short name of participant	TIB	CNRS	EAA	UHB	NILU	MET	LW ERIC	FMI	CNR	DKRZ	CSC	IFREMER	INRA	TUD	OEAW
Person months per participant	2	12	12	30	5	15	22	5	5	23	16	6	15	15	4
Start month	1			**End month**						36					

Objectives

The main objective of this WP is to integrate B2TERM (developed in WP3) into (inter)disciplinary data infrastructures in order to improve research data interoperability, discovery and exploitation, within and beyond the EOSC. The early adopters of B2TERM are data providers (e.g., data publishers, portals and catalogues, research infrastructures (RI)s, and e-infrastructures) as well as data users (e.g., individual scientists). The table below lists involved data providers and their repositories. The specific objectives are:

- Leverage B2TERM terminologies to improve the annotation of (meta)data within (inter)disciplinary data infrastructures, targeting the provision of rich and meaningful (meta)data to a broader user community.
- Apply B2TERM terminologies to improve research data discovery on data repositories portals.
- Demonstrate the added values of B2TERM by integrating the service with user-focused data exploration and analysis services.

Data Repository & Portal	**Data Domain**	**Data Provider**
PANGAEA Data Portal (https://www.pangaea.de)	Earth sciences	UHB
LTER DEIMS-SDR (https://deims.org)	Biodiversity and ecosystems	EAA
TUDelft 4TU.ResearchData Portal (https://data.4tu.nl)	Engineering	TUD

ACTRIS Data Portal (http://actris.nilu.no)	Atmospheric	NILU, FMI, CNRS, CNR
LifeWatch Italy Data Portal (https://www.lifewatch.eu/catalogue-of-data)	Biodiversity and ecosystems	LW ERIC
Data Inra Institutional repository (http://data.inra.fr)	Agriculture	INRA
SIOS Svalbard (https://www.sios-svalbard.org/metadata_search)	Arctic & Atmosphere	MET
ESGF (https://esgf-data.dkrz.de)	Climate	DKRZ
ARCHE (https://arche.acdh.oeaw.ac.at/)	Humanities	OEAW

Description of work

Task 4.1: Integrate B2TERM into disciplinary data systems [M6-30]
Lead: UHB, Participants: EAA, NILU, CNR, FMI, CNRS, LW ERIC, MET, TUD, INRA, DKRZ, OEAW. Effort: 90 PM

This task focuses on the integration of B2TERM terminologies into disciplinary data systems in order to provide rich and meaningful (meta)data to users. The implementation will follow the recommendations based on the semantic model and annotation guidelines outlined in WP2 (T2.6 and T2.7). Annotations will be disseminated through standardised metadata services (e.g., Schema.org, OGC Catalogue Services for the Web (CSW)[53] and OAI-PMH implementation)[54].

Subtask 4.1.1: Annotate meta(data) with B2TERM terminologies. [M6-30]
Lead: EAA, Participants: UHB, NILU, CNR, FMI, CNRS, LW ERIC, MET, TUD, INRA, DKRZ, OEAW

In this subtask, we (data providers listed in Table 4.1) will plan and implement the integration of terminologies served by B2TERM in our data systems. We will develop client applications (e.g., REST client) to retrieve content from relevant terminologies (i.e., terms and their relations) in B2TERM programmatically. We will improve our annotation tools or employ third party tools to enable annotation of meta(data), e.g., observed properties, features-of-interest and research topics, with B2TERM terminologies. Annotation activities will be carried out manually (e.g., through data curators and individual researchers), semi-automatically or automatically, depending on the data providers' requirements.

Subtask 4.1.2: Share meta(data) enriched with semantic annotations. [M15-30]
Lead: TUD, Participants: UHB, NILU, CNR, FMI, CNRS, LW ERIC, MET, EAA, INRA, DKRZ, OEAW

In order to facilitate data sharing and reuse across community boundaries, we will disseminate the metadata enriched through semantic annotations (generated in T4.1.1) as follows:

- We will enhance our existing metadata provision services including related schemas to disseminate enriched metadata to other metadata catalogues (e.g., B2FIND).
- We will improve our data systems to incorporate the annotations in data files (e.g., delimited text files, and NetCDF), such that they are available to data users.
- We will demonstrate how the annotations may be embedded in data landing pages (e.g., using Schema.org) to improve data visibility in Search Engine Results Page (SERP).

Task 4.2: Improve research data discovery through B2TERM terminologies [M18-36]
Lead: UHB, Participants: NILU, CNR, FMI, CNRS, LW ERIC, MET, EAA, TUD, INRA, DKRZ. Effort: 42PM

We will leverage terminologies provided by B2TERM to enable semantic searches on data portals (Table 4.1). For example, we will utilize B2TERM enriched metadata (from T4.1) and implement semantic query expansion (e.g., by using synonyms and related terms) to produce relevant data search results on the portals. We will also implement features such as semantic-based faceted search and taxonomy-based autocomplete through the portals' search engine (e.g., Elasticsearch).

[53] https://www.opengeospatial.org/standards/cat
[54] http://www.openarchives.org/OAI/openarchivesprotocol.html

Task 4.3: Application of B2TERM in data integration and analysis [M15-30]

Lead: LW ERIC, Participants: CNRS, DKRZ, UHB, CSC, TIB. Effort: 28PM

This task describes how B2TERM will be applied to support data integration (T4.3.1), data exploration and analysis (T4.3.2), and semantic enablement of scholarly literature (T4.3.3).

Subtask 4.3.1: Semantic integration of functional biogeography data. [M15-30]

Lead: LW ERIC, Participants: CNRS

We will demonstrate how B2TERM terminologies may be applied to support semantic integration of functional biogeography data, with the aim to understand and generalize Trait-Environment-Relationships (TER) across Mediterranean Basin. We will combine and link community (terrestrial plants and phytoplankton) composition, trait, climate and soil data using the semantic annotations available from T4.1.1. Through this integration, we will evaluate the response of functional traits, such as biomass for plants and biovolume for phytoplankton to environmental factors, including major components of ongoing changes affecting ecosystems (e.g., temperature, light intensity). This will allow us to run statistical analyses in a systematic and unified way and test the generality of TER along temperature and light gradients.

Subtask 4.3.2: Leverage B2TERM to expedite and facilitate data exploration and analysis. [M15-30]

Lead: DKRZ, Participants: UHB, CSC

We will demonstrate through the following use-cases how terminologies provided by B2TERM facilitate data exploration and analysis within computational notebooks.

- The first use case addresses exploratory analysis of cross-domain data facilitation by B2TERM (i.e., terrestrial and marine). Using a Jupyter notebook, we will demonstrate how users can discover relevant datasets from B2FIND, thereby making use of terminologies provided by B2TERM to harmonize the heterogeneous nomenclature of observation variables within the search results. Based on the meaningful descriptions of retrieved datasets, the users will filter and merge datasets for further data analysis and visualization, e.g., examination of patterns and correlations. EGI Jupyter notebooks provided at the EOSC-hub will be used to implement and disseminate the technology use case within the EOSC environment.
- The second use case focuses on the calculation of climate indices and the visualization of parameters of the IPCC AR6 data. The calculation is defined and initiated by individual scientists. The automatic recalculation of the indices based on data subscriptions follows. The data subscription service (T4.4.1) triggers the (re-)calculation of climate indices when new data-of-interest is found. The calculation is submitted automatically to Jupyter notebooks service at CSC Pouta cloud platform and/or EOSC provided Jupyter notebooks service. The resulting data is made available with sufficient metadata and provenance information in order to support further re-use. This workflow will be implemented in such a way that it can be configured and executed through a transparent interface by the end user.

Subtask 4.3.3: Leverage B2TERM in scholarly communication. [M15-30]

Lead: TIB, Participants: UHB

This task will prototype the integration of B2TERM in the Open Research Knowledge Graph[55], an existing project led and sustained by TIB that aims at increased machine actionability of scholarly communication. B2TERM terminologies will be used in machine actionable representations of information published in scholarly articles. For example, we will apply B2TERM terminologies to represent datasets (i.e., table headers) and statistical hypothesis tests included as results in scholarly articles.

Task 4.4: Link B2TERM to Semantic Data Subscription Service (DSS). [M1-M30]

Lead: CSC, Participants: DKRZ, IFREMER. Effort: 27PM

In the context of the ENVRIplus[56] project, a data subscription service (DSS) was prototyped to enable researchers to register a subscription for data matching their specific requirements. The DSS recognises new matching data, activates defined pre-processing of the data if requested, and informs the subscriber through email that new data is available with a link to access it. In this task, DSS will be further developed through its integration with B2TERM (T4.4.1) and B2FIND (T4.4.2). The application of this integration will be demonstrated in T4.4.3.

[55] https://orkg.org
[56] http://www.envriplus.eu/

Subtask 4.4.1: Further development of DSS. [M1-M24]

Lead: CSC, Participants: DKRZ

This task will further develop the DSS prototyped in ENVRIplus to an independent service component. The focus will be on programmable interfaces to register, manage and process subscriptions to enable use of DSS in interdisciplinary data exploration, including integration with B2TERM. The tasks T4.4.2 and T4.4.3 build on this development.

Subtask 4.4.2: Integration between DSS and B2FIND. [M12-M30]

Lead: DKRZ, Participants: CSC

This task will integrate the data subscription service with B2FIND. This enables the data user to register a subscription based on the search in B2FIND, and also notify users about changes to subscriptions managed by DSS. The integration will enable notifications based on the metadata catalogued at B2FIND service and thus allow for interdisciplinary data discovery from multiple science fields.

Subtask 4.4.3: Application of DSS in the context of Euro-Argo. [M12-M30]

Lead: IFREMER, Participants: CSC

This task will demonstrate the application of DSS to support semantically enhanced data discovery, exploration and access. IFREMER will implement OceanWorks data discovery API (a collaboration with NASA-JPL), which will enable users to define data subscriptions through the integration between B2TERM and DSS, as addressed in task T4.4.1. The data will come from Argo and DBCP GDACs (Global Data Assembly Centres for the global array of 20000 profiling floats and the global array of 10000 drifting buoys). Through the data subscriptions, users will be notified when their datasets-of-interest are available.

Deliverables

D4.1.1 Implementation Strategy for (Meta)data Annotation [M18]. The report provides an overview of the implementation strategies for the meta(data) annotation by different disciplinary data systems.

D4.1.2 Metadata Provision Services Enhanced by Data Providers [M24]. Examples of web services which expose annotated metadata from data providers.

D4.2 KOINE-B2TERM in Practice [M36]. The last report of WP4 which summarizes the concepts and results on B2TERM integration with disciplinary data systems to improve data discovery.

D4.3 Integration of B2TERM in Pilot Applications [M30]. B2TERM pilot applications on supporting research data integration and analysis.

D4.4 Application of Data Subscription Service (DSS) [M30]. The report on the integration of DSS with B2TERM and B2FIND, and its application in the context of oceanographic data discovery.

Work-package number	5		**Lead beneficiary**				EAA									
Work-package title	Research community engagement and capacity building															
Participant number	1	2	3	4	5	8	9	10	11	14	15	19	20	21	22	23
Short name of participant	TIB	CNRS	EAA	UHB	AWI	UMAAS	MET	LW ERIC	UKRI	BSC	DKRZ	INRA	TUD	OEAW	ANDS	SISTEMA
Person months per participant	3	8	28	5	7	8	10	14	8	3	5	5	30	4	0	5
Start month	1						**End month**			36						

Objectives

The main objective of WP5 is to integrate B2TERM within social infrastructures, i.e., to engage them with research communities in order to support the development, testing and validation of KOINE services. User-centric, agile development and feedback will inform and respond to the activities performed in WP2, WP3 and WP4. In addition, WP5 will provide important input for the development of the long-term sustainability

and maintenance strategy of B2TERM (WP6). WP5 will also provide training on the use of B2TERM, B2FIND and DSS services. WP5 is a relatively resource-intensive WP. This is intentional and reflects the importance of community engagement and the fact that some of the main barriers to data sharing are cultural, not technical[57,58,59]. The WP aims to co-locate events to reduce KOINE's carbon footprint.

Description of work

Task 5.1: Development and maintenance of the KOINE Stakeholder Forum (SF) [M1-M36]
Lead: TUD, Participants: EAA, LW ERIC, UKRI, CNRS, UHB, MET, BSC, DKRZ, AWI, UMAAS, INRA, OEAW, TIB. Effort: 30 PM

The forum will consist of terminology providers, users and data curators from different research domains. The stakeholders will come from discipline-specific communities (e.g., ecosystem and biodiversity, atmosphere and climate, marine and cryosphere research, coastal engineering and linguistics), as well as from interdisciplinary groups (e.g., urban studies), which collect data from various disciplines using different terminologies. Project members already have access to these research communities: TUD is an active member of the CESAER[60] network of science and technology universities (engineering disciplines) and is also one of the founding members of the Amsterdam Institute for Advanced Metropolitan Solutions[61] (interdisciplinary urban research); EAA is connected with the LTER-Europe[62] network and many environmental research communities involved in cluster projects like ENVRI-FAIR[63] and the RDA Vocabulary Service and Semantic Interest Group[64] where EAA leads together with UHB the emerging Working Group on Harmonizing FAIR descriptions of observational data. AWI partner is part of the UNESCO/IOC GOOS EOV[65] panels as well as Cryosemantics working group working on harmonising cryosphere terminologies. In addition, to ensure diversity of stakeholders and disciplines involved, ACDH-OEAW[66] in its role as national coordinator of CLARIN-ERIC[67], the European Research Infrastructure for Language Resources and Technology, will convey access to humanities researchers working with digital language resources. For the ecosystem and biodiversity domain this task will help to create a community of terminology providers and users which will use EcoPortal[68] as a registry for terminologies and a platform for terminology provision and exchange. The KOINE Stakeholder Forum will also help identify KOINE Ambassadors who will advocate for KOINE services within their respective communities (T7.3).

Task 5.2: Gathering research community requirements [M4-M18]
Lead: EAA, Participants: TUD, LW ERIC, UKRI, CNRS, UHB, MET, BSC, DKRZ, AWI, UMAAS, INRA, OEAW, TIB, ANDS. Effort: 39 PM

This task will gather research community requirements for B2TERM. It is closely aligned with T2.4 (design) and T3.2 (implementation). The requirements will be gathered during focus group meetings with semi-structured discussions involving two different types of research communities:

- Disciplinary communities - primary group where the basic API functionalities of B2TERM will be determined (ecosystem and biodiversity community and coastal engineering community, including the OpenEarth consortium[69]).

[57] http://blogs.lse.ac.uk/impactofsocialsciences/2018/11/14/the-main-obstacles-to-better-research-data-management-and-sharing-are-cultural-but-change-is-in-our-hands/
[58] https://doi.org/10.1177/2515245917751886
[59] https://doi.org/10.6084/m9.figshare.4055448.v1
[60] https://www.cesaer.org/
[61] https://www.ams-institute.org/
[62] http://www.lter-europe.net/
[63] http://envri.eu/envri-fair/
[64] https://www.rd-alliance.org/groups/vocabulary-services-interest-group.html
[65] http://www.goosocean.org/index.php?option=com_content&view=article&id=14&Itemid=114
[66] http://www.oeaw.ac.at/acdh/
[67] https://www.clarin.eu/
[68] http://ceur-ws.org/Vol-1933/poster-paper-13.pdf
[69] https://www.openearth.nl/

- Interdisciplinary research communities - secondary group, which will provide the requirements for B2TERM in an interdisciplinary research setting (urban studies).

Through the engagement of the coastal engineering research community and Deltares institute[70], Dutch SMEs (such as Evides[71]) will also be involved in the requirements gathering. This task will also help to define which workflows are needed for an efficient mediation between terminology providers and users. In order to address KOINE's objective to harmonize terminologies needed to describe observational research data, this task will also gather specific requirements for the Terminology Interoperability Framework supporting mainly the activities of T2.6, but also of T2.5 and T2.7. The main stakeholders involved will be environmental, ecosystem and biodiversity as well as earth sciences. The requirements will be gathered through a couple of workshops and hackathons (using already existing formats like the S4BioDiv[72] and the LifeWatch semantic workshops[73]) focusing on:

- Essential Biodiversity Variables (e.g., community ecology for plants and phytoplankton)
- Taxonomic information
- Essential Climate Variables

The outcomes of this activity will also be integrated in the prospective RDA WG on harmonisation of research observation semantics.

Task 5.3: Community validation of KOINE services [M12-M32]
Lead: LW ERIC, Participants: TUD, EAA, CNRS, MET, DKRZ, AWI, UMAAS, INRA, OEAW, UKRI, BSC, SISTEMA, ANDS. Effort: 36 PM

The objective of this task is to perform user-centric, iterative testing of B2TERM and B2FIND services with the research communities. MiLE+[74] methodology will be used for the evaluation. The results of the validation exercise will be communicated back to WP2 and WP3. The EcoPortal community will provide a particularly valuable case study of B2TERM and B2FIND implementations, given it will be a newly established semantic repository within an important disciplinary community of KOINE's consortium. LifeWatch will provide the platform to store data annotations for scientists not involved in any Research Infrastructure. Through the engagement with the coastal engineering communities, Dutch SMEs, such as Evides, will also be involved in KOINE services validation. The validation will be conducted in three main stages:

- Stage 1: B2TERM and B2FIND sandbox evaluation in a facilitated workshop.
- Stage 2: B2TERM and B2FIND beta version testing in a hands-on workshop.
- Stage 3: FAIRness assessment of annotated data (hands-on workshop). Data providers will learn about the FAIR principles and metrics and will perform a FAIRness self-assessment using a web-based software tool such as FAIRshake.

Task 5.4: Capacity building and training [M24-M36]
Lead: TUD, Participants: EAA, LW ERIC, UKRI, CNRS, UHB, MET, BSC, DKRZ, AWI, UMAAS, INRA, OEAW, TIB, ANDS. Effort: 33 PM

This task will focus on providing training needed by research communities to take full advantage of B2TERM and B2FIND services. The target audience are terminology providers and terminology users.

Subtask 5.4.1: Development and Distribution of Open Educational Resources (OERs)

Open educational resources (OERs) will be developed which will combine the extensive teaching experience around semantic technologies within our consortium with the internationally acclaimed peer-to-peer teaching framework "The Carpentries". The Carpentries already recognised[75] that there is a gap in ontology usage. Our OERs aim to fill this gap by increasing the uptake of B2TERM among diverse research communities. We will offer to on-board our OERs to the Carpentries curriculum and maintain them in that broad, public

[70] https://www.deltares.nl/en/
[71] https://www.evides.nl/english
[72] http://fusion.cs.uni-jena.de/s4biodiv2017
[73] https://www.lifewatch.eu/ontology-semantic-web-for-biodiversity-ecosystem-research
[74] Triacca, L., et al. (2004). MiLE: Systematic Usability Evaluation for E-learning Web Applications. In L. Cantoni & C. McLoughlin (Eds.), Proceedings of ED-MEDIA 2004--World Conference on Educational Multimedia, Hypermedia & Telecommunications (pp. 4398-4405). https://www.learntechlib.org/primary/p/11709/
[75] https://doi.org/10.1371/journal.pcbi.1005510

context in order to leverage the international Carpentries network to reach out to a broad range of research communities world-wide.

Subtask 5.4.2: Delivery of "Terminology Carpentry" workshops
Our "Terminology Carpentry" OERs will be initially delivered to institutions from within the KOINE Stakeholder Forum. During these Carpentries-like, effective workshops the participants will be invited to bring their own data (approach adapted from "Bring Your Own Data - BYOD"[76] workshops) and will be set up with all necessary tools on their own computers to immediately utilise B2TERM and to continue applying the skills they were taught in their day-to-day practices.

Deliverables

D5.1 User-driven KOINE service requirements [M9, M18]. A report summarising the requirements for defining the Terminology Interoperability Framework of B2TERM gathered through consultation with various communities identified through the KOINE Stakeholder Forum.

D5.2 KOINE service validation [M28]. This deliverable will summarise the outcomes of the validation of B2TERM and B2FIND by the research communities.

D5.3 Terminology Carpentry OERs [M36]. Carpentry-style workshop available as Open Educational Resources introducing the concepts and usage workflows of B2TERM.

Work-package number	6			**Lead beneficiary**		AWI			
Work-package title	B2TERM operations and sustainability								
Participant number	1	2	3	5	8	9	10	11	16
Short name of participant	TIB	CNRS	EAA	AWI	UMAAS	MET	LW- ERIC	UKRI	EUDAT LTD.
Person months per participant	18	3	3	12	3	3	3	3	0
Start month	1			**End month**		36			

Objectives

KOINE is committed to harmonizing distributed and well-adopted terminologies to create a service that provides interoperable data descriptors for EOSC users. Aside from technical integration, alignment of vision and operational leadership must be secured across all services (including external services) involved, as well as with the EOSC leadership and Executive Board. This WP seeks to address this challenge and has the following objectives:

- Design and implement an inclusive operations and decision-making model to efficiently maintain and evolve KOINE services. This will feature:
  - A formal operations model to specify and periodically update the interaction between components of the Terminology Interoperability Framework (TIF; developed in T2.4) and B2TERM to a) support harmonization and interoperability across services and b) persistently align the development of the TIF and B2TERM within and beyond the project.
  - A periodic review process for the TIF to ensure it effectively bridges B2TERM (in the context of the EOSC service catalogue) and the external terminologies upon which B2TERM relies.
  - A process - built on T5.1 and T5.2 - through which stakeholder input will be used to set development priorities in order to address emerging local user needs, while balancing the priorities of contributing terminology services.

[76] https://www.dtls.nl/fair-data/byod/

- Establish a KOINE Operations Committee and engage advisory bodies (e.g., the GO-FAIR initiative, RDA working groups, other EOSC Implementation Roadmap projects) to execute and validate the functions developed in the model above.

Develop a robust sustainability model to allow operation mechanisms to persist beyond the project, adhering to core principles and standard operating procedures.

Description of work

Drawing heavily from the outcomes of WPs 2, 4 and 5, this WP will progressively consolidate and convert the mechanisms used to manage the project into a forward-looking operations and decision-making structure to sustain the services developed and integrated within KOINE. The key features of this work are described below.

Task 6.1: Develop the KOINE operations model and associated mechanisms [M6-M30]
Lead: AWI, Participants: TIB, CNRS, UKRI, LW ERIC, EUDAT. Effort: 15 PM

In this task, an operational model will be developed, specifying inclusive decision-making and implementation pathways in the context of KOINE's multi-stakeholder federation of cooperating resources. This model will be a synthesis of experience gained through the management of KOINE's components (especially in WP1, WP5), and clarify how KOINE's outcomes can be best integrated into the EOSC Service Catalogue. In particular, this task will:

- Review existing operational models and best practices used to manage multi-terminology services in order to compose a model tailored to the EOSC use case. Such tailoring will draw from participation in the EOSC Stakeholder Fora and associated RDA Working Groups. This will be closely linked with T2.3, T3.1 and T5.1 that will, respectively, review, benchmark, and build stable links to such services.
- Building on the above, identify the elements of the KOINE service in need of periodic review by an Operations Committee to ensure currency and interoperability (e.g., the import and/or integration of terminologies into the system, the interface between terminology services and B2TERM functionality).
- Define the type and role of members included in decision-making and operational mechanisms, allowing for contributions from expert developers, users, and other stakeholders.
- Define coordination mechanisms with governance structures of terminology services external to the EOSC ecosystem (e.g., those managed by RDA vocabulary working groups, the OBO Foundry, Earth Science Information Partners Semantic Technologies Committee, European Bioinformatics Institute, Marine Metadata Interoperability) to ensure interoperability. This task will also work with the landscaping activities and compatibility assessments in T2.3.
- Define mechanisms to ensure compliance with the TIF (WP2) guidelines of the European Interoperability Framework (EIF), in particular the Semantic Interoperability layer (Section 3.5, Recommendations 30-32).

T6.2: Establish a KOINE Operations Committee and Scientific Advisory Board [M12-M36]
Lead: AWI, Participants: TIB, UKRI, EAA, CNRS, UMAAS. Effort: 10 PM

This task will implement, evaluate, and refine the model specified in T6.1 through the creation of an Operations Committee (OC) and an external Scientific Advisory Board (SAB). The proceedings of these entities will demonstrate how best to manage the KOINE service and its requisite collaborations. Thus, this task will:

- Establish an international KOINE OC, initially composed of members of the project consortium, to formalise representation and decision making mechanisms. Key subtasks include:
    - Formalising the procedures which will be used to link domain experts to developers in order to extend terminologies within the EOSC framework, in close collaboration with T3.5.
    - Formalising, in line with T3.5, which procedures will be used to transfer KOINE requests to the governing structures of its participating terminology services; this includes brokering the creation and update of terms in participating terminologies.
    - Establishing standard operating procedures to ensure that the quality and FAIRness of individual resources are stable and well-maintained, in collaboration with the FAIR principles (UMAAS)
    - Establishing mechanisms to assess the need for updates to the system and trigger these as

new technologies and terminologies arise, without compromising stability and the core user experience.

- Establish, through engagement with the Stakeholder Forum (WP5), an international KOINE Scientific Advisory Board (SAB), primarily composed of members outside the consortium. The SAB's primary task will be to advise the OC by reviewing the system and ensuring it is current and well-connected with external initiatives. Notably, the SAB will seek members from underrepresented disciplines (e.g., the life sciences) to increase the service's coverage and relevance to EOSC users.
- Establish a GO-FAIR Implementation Network on Vocabulary Services that will serve as an open forum for discussing architectures, software, and specifications developed by KOINE with the wider global community. As thematic GO-FAIR INs (e.g., metabolomics, Food Systems) are in the process of being established, we will identify or create additional INs so as to promote collaboration and sustainability relevant to KOINE's activities and partners beyond the funding horizon for KOINE.

Task 6.3: Sustainability through integration with EOSC services [M12-M36]
Lead: TIB, Participants: AWI, EUDAT. Effort: 15 PM

KOINE aims to offer the project's outputs in a sustainable form and to have the B2TERM service included in the EOSC Service Catalogue as part of the EUDAT Service Portfolio. To achieve this aim, KOINE, in collaboration with EUDAT, will develop a sustainability plan for the maintenance and future development of B2TERM beyond the project duration. This task will also ensure that the service complies with the EOSC Rules of Participation, necessary for listing in the EOSC Service Catalogue. Thus, subtasks include:

- Development of a Business Model Canvas (BMC) which will focus on the financial aspects in order to sustain KOINE's modular services past the lifetime of the project. Development of the BMC will begin as the KOINE services arise, and it will be progressively updated during the project. This will include the development of an appropriate, not-for-profit cost recovery model that allows the use of B2TERM and its core components (e.g., the Terminology Interoperability Framework) to be further developed by all interested stakeholders. As such, the core services of KOINE (such as B2TERM) will be included in service portfolio of TIB[77] and thus maintained for at least a midterm timeframe. This will include the development and implementation of a cost model that aims to make the use of services free-of-charge to end users.
- For the cost recovery of on top services such as consulting, co-development and integration support, two possible strategies are envisioned:
    - Provisioning of the B2TERM service to interested stakeholders via the EOSC Service Catalogue. While use of the B2TERM service will be free, service-level agreements and value-added services (e.g., professional training, mass requests) will be offered against a fee.
    - Provisioning of consulting, commercial support, maintenance, deployment/integration and custom extension development for interested stakeholders, who wish to implement B2TERM components on top of their own infrastructure.
- Development of an Exit strategy that will enable the transfer of the B2TERM service to suitable partners (as described in WP3). Such a strategy is a standard procedure and will only be applied should TIB not be able to permanently sustain the B2TERM service.
- Develop the necessary documentation (e.g. Terms and Conditions) to offer the service as well as guidelines for future development of B2TERM. This is closely linked to the fulfilment of the necessary legal requirements associated with providing B2TERM and its components (e.g. TIF), its integration within the EOSC service catalogue. Further, this documentation process will ensure compliance with privacy and data protection regulations. For the legal assessment of the services, the Project Coordination Office (Section 3.2) will initiate an award procedure (open call) in order to assign this task to suitable subcontractor which is also familiar with EOSC requirements.
- Perform the EOSC onboarding process to register B2TERM in the EOSC Service Catalogue, assessing compliance of B2TERM with the EOSC Rules of Participation.

Task 6.4: Linking KOINE operations to external governance and operation bodies [M1-M36]
Lead: AWI, Participants: UKRI, EAA, MET. Effort: 8 PM

When dealing with and developing FAIR-compliant technologies, it is essential to coordinate across similar resources in pursuit of ever-greater alignment of distributed systems. This task will link the activities of the

[77] https://www.tib.eu/en/publishing-archiving/

KOINE Operations Committee to those of governing bodies internationally, along the model developed in T6.1.3. This is essential in realising a truly interoperable future. In particular, this task will:

- Build on the external coordination recommendations in T6.1 and establish Memoranda of Understanding with other semantic and terminology services (e.g., the OBO Foundry and Library and the Earth Science Information Partnership, and the World Meteorological Organisation vocabularies). This will serve to extend alignment beyond EOSC stakeholders and open up new opportunities for Open Science.
- Establish EOSC as a stakeholder in external terminology integration systems and individual terminologies, thus ensuring EOSC user needs are transmitted and considered widely.

Deliverables

D6.1 KOINE operations model: core principles and standard operating procedures [M30]. This document will be the outcome of T6.1 and comprise the definition of the Operations Committee's mandate composition, core principles, standard operating procedures (SOPs), and its relation to the EIF. This will include a formalisation of the TIF Compliance Standards, the API definition, and the minimum information required to support interoperation between terminologies and terminology services. Its SOPs will address the guidelines for the addition, review, and removal of terminologies in the EOSC space, synthesizing recommendations from WPs 2, 4, 5.

D6.2 Proceedings of the KOINE Operations Committee kick-off meeting [M25]. As the first working drafts of D6.1 are completed, the first appointed committee will meet and begin proceedings, "dry-running" the SOPs defined in D6.1 and composing amendments as needed. This will include assessing the sustainability of the processes and structures set in D6.1. The first Proceedings of the Committee will be delivered, which will include an evaluation of the initial drafts of D6.1.

D6.3 Coordination agreements and strategy: interlinking KOINE with external terminology services [M30]. This report will include and describe the memoranda of understanding and related agreements between the KOINE service and external terminology integration and applications systems worldwide. This will complement the technical interoperation implemented in WP2 and consolidate the outcomes of WP5 into a strategic document, reviewed by the Operations Committee.

D6.4 Sustainability plan and linkage to EOSC [M30]. This deliverable comprises a plan to address the sustainability of the KOINE operations model beyond the project duration. The plan will note dependencies, potential risks, and the contingency plans / mechanisms to address them. This includes:

- A Sustainability plan including a Business Model Canvas;
- An Exit Strategy which will only be applied should TIB not be able to permanently sustain the B2TERM service;
- Complete Terms and Conditions for the service which fulfils the necessary legal requirements.

Work-package number	7	**Lead beneficiary**		LW ERIC	
Work-package title	Communication and dissemination				
Participant number	1	2	4	10	11
Short name of participant	TIB	CNRS	UHB	LW ERIC	UKRI
Person months per participant	8	3	2	18	3
Start month	1			**End month**	36

Objectives

This WP designs and implements the communication and dissemination of KOINE, with regard to all project related activities, research outputs and products including their scientific impacts and commercialisation possibilities for the wide spectrum of stakeholder groups by means of using a vast range of communication

and dissemination channels, tools and techniques. In this sense, WP7 supports all other WPs by providing them with adapted communication materials to reach out to their target groups. In particular, WP7 closely supports WP5 in the activities devoted to the engagement of the research community, the establishment of the necessary collaboration channels with relevant parties to understand and analyse their business cases, needs and requirements, and actively engage the community. KOINE will further explore and establish cross-domain synergies that will enhance substantially the visibility, outreach, competitiveness and adoption of the project's output. More specifically, the objectives of this WP are to:

- Disseminate regularly and consistently information about the project and its activities, as well as the impact on the explored research fields, via a wide spectrum of communication and dissemination channels, including social media, webinars, press releases and blog posts;
- Promote the use and adoption of project outputs, via a wide spectrum of communication and dissemination channels, including social media, webinars, press releases and blog posts;
- Exploit the outputs during and beyond the project's duration, ensuring their sustainability and the commercial adoption of the developed solutions.

In order to ensure the impactful outreach of the project, synergies will be established with related projects at European and global level, e.g., via the joint organisation and collocation of workshops and conferences, including establishing a global Ambassador Programme as well as a Network of Associated Partners, among other synergies.

Description of work

LW ERIC will lead this WP, with the main support of TIB, and the contribution of the leading partners of WP 1, 2, 4 and 5, to ensure the coordination and effectiveness of dissemination and communication activities.

Task 7.1: Development of KOINE Dissemination and communication strategy and plan & Annual activity reporting [M1-M36]

Lead: LW ERIC, Participants: TIB, UKRI, CNRS, UHB. Effort: 6pm

This task is directed at the development of the communication strategy for KOINE, detailing its mission and vision, targeting all identified target and stakeholder groups in a customised fashion and implementing operative engaging approaches for each community, by means of digital and printed media. Furthermore, in this task, KOINE's dissemination and communication strategy will be defined and implemented by means of the exact media mix, thus maximising the project's impact towards a vast range of audiences. In the same context, risk evaluation and measurement tools will be appropriately taken in consideration to monitor and assess the effectiveness of the communication strategy and plan. The operative communication plan will be composed, with its initiation from the draft plan for the dissemination and exploitation of results, and will be further developed in its definitive version in M4, as well as revised and updated at mid-term of project life cycle. The plan will outline all scheduled and implemented tasks in this context, e.g., attendance by all partners of relevant meetings, workshops and conferences, dissemination of project results, webinars and social media; as well via the regular publishing of website blog posts, among other related activities. On an annual basis, Dissemination and Communication Activity reports will be drafted and released focusing on the progresses and intermediate results, and updated plans for the following period.

Task 7.2: Visual identity, website and promotional materials [M1-M36]

Lead: LW ERIC, Participants: TIB, CNRS, UHB. Effort: 16pm

This task is dedicated to the design and implementation of project communication tools and materials, in accordance to Communication Strategy and Plan. This task includes the development of:

- The visual identity of the project (logo and template for documents and presentation);
- A set of digital tools (website, newsletters, social media accounts, videos, digital leaflets);
- Printed materials (brochures, reports, leaflets, booklets, stickers etc.);
- Development of KOINE services communication packages (specific tools dedicated to main deliverables of the project to promote specific results towards specific target groups).

Digital tools and printed materials will be updated taking into account project advancements.

Task 7.3: Large scale dissemination of projects impacts and results [M1-M36]

Lead: TIB, Participants: LW ERIC, UKRI, CNRS, UHB. Effort: 12pm

This task is devoted to promote KOINE towards the identified scientific communities, with the aim of engaging them and increasing users for the developed services through the presentation of project use cases.

The task will be carried out through:

- Creation of an Ambassador Programme (e.g., early adopters, committed researchers, etc.) is foreseen involving early adopters in the development and final tuning if case studies and services, aiming at making this first group of users the first promoters of project results
- Media relations, media advisory, press releases and news and contents for relevant websites, blogs and media will be issued during the project, on the occasion of milestone accomplishments. These will be localised and distributed by the project partners using their existing communication channels and systems.
- Dissemination of open access (Gold or Green) scientific articles produced on different project's outcomes and activities.
- Events
    - The organisation of two public meetings, a mid-term and a conclusive one, the first to trigger adoption of the services and collect feedback on them, the latter with the aim of presenting the final releases of the services and use cases developed during the project life, addressing the scientific community and relevant policy makers
    - Participation to EOSC related events and main conferences (e.g., RDA) with exhibition booths featuring demonstrative sessions.

Deliverables

D7.1 Dissemination & Communication Strategy and Plan [M4]. A detailed plan for the dissemination of the project's results. This will include an indicative list of events that partners plan to attend, webinars, blogs posts etc.

D7.2 Annual Dissemination & Communication Report [M12, M24, M36]. Annual reports on the dissemination activities undertaken and plans for the following period.

D7.3 Project Logo, Website & Promotional Materials [M6]. Project logo will be designed together with website layout and leaflet, constituting the first set of communication materials of the project

D7.4 Report on Large scale dissemination of projects impacts and results [M12, M24, M36]. A report on the community engagement activities carried out during the reporting period. Three versions of the deliverable will be produced, at the end of each year of the project.

Table 3.1c: List of Deliverables

Deliverable (number)	Deliverable name	Work-package number	Short name of lead participant	Type	Dissemination level	Delivery date (in months)
1.1	Project Management Plan	1	TIB	R	CO	3
1.2	Annual Public Report	1	TIB	R	PU	12, 24, 36
1.3	Quality Assurance & Risk Assessment Plan	1	TIB	R	PU	6
1.4	Data Management Plan	1	TIB	R	PU	6
2.1	Report on identification and listing of terminologies relevant for KOINE partners	2	LW ERIC	R	PU	6
2.2	Recommendation for FAIR terminologies metadata description	2	CNRS	R	PU	18
2.3	Description of TIF architecture and API specification for B2TERM	2	CNRS	R	PU	12, 18
2.4	Common Semantic Model to describe scientific variables and parameters across disciplines at M9 and final	2	UKRI	R	PU	9, 24

	report on parameter harmonisation at M24					
2.5	Recommendation on annotating research data with B2TERM terminologies	2	UKRI	R	PU	15
3.1	State of the Art review of semantic technologies	3	UvA	R	PU	6
3.2	B2TERM implementation	3	UvA	DEM	PU	24
3.3	Extension of B2FIND with B2TERM semantic capabilities	3	DKRZ	DEM	PU	36
3.4	Seamless search and discovery of terminological content by domain	3	UvA	DEM	PU	36
3.5	Mediation framework for terminology curation and augmentation	3	TIB	DEM	PU	36
4.1.1	Implementation Strategy for (Meta)data Annotation	4	EAA	R	PU	18
4.1.2	Metadata Provision Services Enhanced by Data Providers	4	TUD	DEM	PU	24
4.2	KOINE-B2TERM in Practice	4	UHB	R	PU	36
4.3	Integration of B2TERM in Pilot Applications	4	LW ERIC	DEM	PU	30
4.4	Application of Data Subscription Service (DSS)	4	CSC	R	PU	30
5.1	User-driven KOINE service requirements	5	EAA	R	PU	9, 18
5.2	KOINE service validation	5	LW ERIC	R	PU	28
5.3	Terminology Carpentry OERs	5	TUD	R	PU	36
6.1	KOINE operations model: core principles and standard operating procedures	6	AWI	R	PU	30
6.2	Proceedings of the KOINE Operations Committee kick-off meeting	6	AWI	R	PU	25
6.3	Coordination agreements and strategy: interlinking KOINE with external terminology services	6	AWI	R	PU	30
6.4	Sustainability plan and linkage to EOSC	6	TIB	R	PU	30
7.1	Dissemination & Communication Strategy and Plan	7	LW ERIC	R	PU	4
7.2	Annual Dissemination & Communication Report	7	LW ERIC	R	PU	12, 24, 36
7.3	Project Logo, Website & Promotional Materials	7	LW ERIC	DEC	PU	6
7.4	Report on Large scale dissemination of projects impacts and results	7	TIB	R	PU	12, 24, 36

3.2 Management structure, milestones and procedures

The project management strategy of KOINE is tailored to the project's scope, aims, and expected impacts, as well as to the number of partners and stakeholder groups, in order to provide an effective operation of the consortium and its ecosystem of collaborations.

The management strategy ensures that all innovation and impact-related objectives of the project are achieved within time, cost, and resource constraints. It capitalizes on tried-and-tested project management structures and procedures, as well as on state-of-the art technological support; both leverage the comprehensive project management experience of the organizations involved in the project, and of the project coordinator in particular. This most crucially concerns the development, implementation and long-term sustainability of the **B2TERM service together with its integration in (inter)disciplinary data infrastructures and research communities**, which represents the major outcome of the KOINE project and will serve as a multiplier to increase Semantic interoperability within the EOSC Catalogue of Services as a whole.

The KOINE project management strategy is driven by the following principles:

- ***Quality:*** The B2TERM service and its integrations developed in the scope of the project reaches the high-quality standards required by the participating research communities, operational infrastructures and the EOSC Service Catalogue;
- ***Impact:*** KOINE increases EOSC data interoperability in the form of a sustained, long-term terminology service (B2TERM) that is integrated in (inter)disciplinary research data infrastructures. Through interoperability, KOINE directly increases the semantic findability and re-usability of research data produced and consumed within the EOSC;
- ***Balance:*** Consensual, balanced strategic and operational decisions are needed to guarantee that project partners, including the disciplinary operational research data infrastructures that are part of the initial development of the Terminology Interoperability Framework (TIF), and the interdisciplinary providers of operational data discovery services which serve as primary B2TERM integrators, contribute as one to the coherent KOINE vision, and work together to achieve the project objectives;
- ***Resilience:*** The management structure and procedures set in place by the consortium guarantees that the project is able to proactively face risks and problems of different nature (technical, financial, administrative, etc.).

The following chapters introduce the management structure of the KOINE project, the main management boards and roles, and the procedures to be applied to support a smooth operation of the project on a day-to-day basis.

3.2.1 Management structure and roles

The KOINE Consortium consists of 23 partners, of which 2 are participating without direct funding (EUDAT Ltd. and ANDS). The management structure has the aim to be efficient and transparent in translating decisions into action at the Consortium level, and in liaising with the European Commission. The management has the following structures (Figure 6).

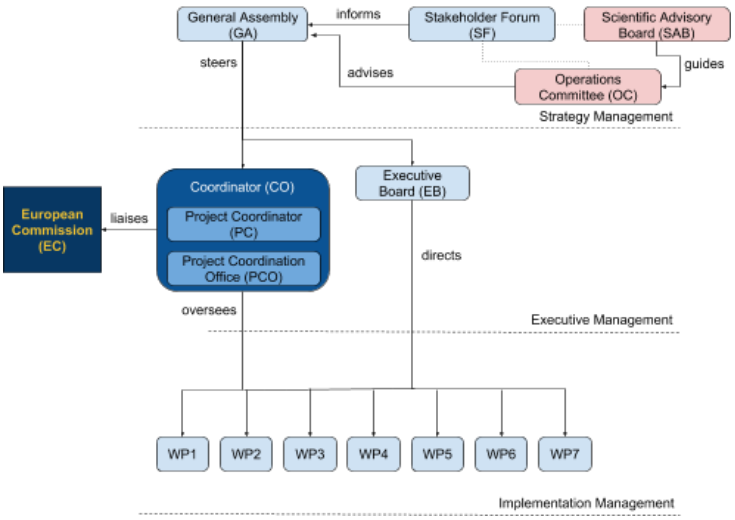


Figure 6: KOINE management and decision structures. The KOINE GA is informed by the KOINE Stakeholder Forum. The KOINE Scientific Advisory Board will guide the Operations Committee which will advise the GA in a later stage of the project. The GA steers both the KOINE Coordinator and KOINE Executive Board, which guide and oversee WP operations.

Project strategy will be managed by the **General Assembly (GA)**, which is the ultimate decision-making body of KOINE Consortium. It consists of one Delegate per Partner, including associated partners, who may appoint a substitute to attend and vote at any meeting (one vote per Delegate). The GA is chaired by the Coordinator and is informed by the Stakeholder Forum (SF). In a later stage of the project the KOINE Scientific Advisory Board (SAB) guides the Operations Committee (OC), which will advise the GA. The GA takes decisions on any change to the Annex I of the Grant Agreement or to the budget, on changes in the composition of the Consortium (entry of new parties or withdrawal of existing ones), declaration of a Partner being a Defaulting Party and establishment of corrective measures or termination of the defaulting party's participation, re-allocation of tasks and/or resources as a function of overbooking or underbooking. It approves the project implementation plan and budget.

The GA appoints the **Executive Board (EB)**, the WP Leaders, and all other members of *ad hoc* committees. The EB is the supervisory body for the execution of the project and shall report and be accountable to the GA. In order to guarantee a smooth and effective day-to-day implementation of the project, the GA can delegate specific tasks and responsibilities to the EB and to the Coordinator. Scope and limits of the delegated powers are to be defined in the CA.

The consortium-wide executive level management is implemented by the **Coordinator (CO)** and the EB. The CO is the contracting partner of the EU and is responsible for the adherence to the contractual obligations of all activities of the project, for the prompt delivery of all reports and deliverables required by the EC and by the CA. At all levels of KOINE's organisational structure the CO has the right of participation and veto in case of decisions conflicting with the EC Grant Agreement. The CO will appoint an individual as **Project Coordinator (PC)** who will chair the EB and lead WP1. The CO will provide additional managerial and administrative support to the PC as required, to form a project coordination office led by the **Consortium Project Manager (CPM)**.

The EB is the main managing, coordinating and supervisory body for the project execution. It is composed by the PC (Chair) and all WP Leads to secure proper representation of all WPs at EB meetings. The EB is responsible for the proper execution and implementation of the decisions of the GA, for the monitoring of the effective and efficient implementation of the project, the WPs status, the timeliness and quality control of deliverables and achievement of milestones, the early identification of possible technical and managerial problems, risks and trouble-shooting.

The **WP Leaders (WPLs)** are responsible for coordinating, supervising, and contributing to the specific research, implementation, networking activities. The WP leaders report on the WP status to the EB, and are responsible for each WP's progress set against the planned deliverables and milestones, and assess risks and propose contingency plans. They are members of the EB where the interaction and transfer of knowledge between the activities of KOINE is ensured.

The **Project Coordination Office (PCO)** will establish and monitor quality management procedures for the review and sign off of project deliverables. It will be the responsibility of WPLs with the oversight of PCO and EB to ensure that deliverables are of the appropriate high quality and consistent across the project.

3.2.2 Management procedures

The management and communication procedures of KOINE are designed to ensure that the project components are properly coordinated and comply with contractual, legal, ethical, financial and EOSC-specific regulations. The management of the project ensures that the project's aims and objectives are achieved on schedule and within budget through:

- Establishing management structures (as outlined in the previous section);
- Planning, monitoring and coordinating the activities of the partners;
- Organising consortium meetings;
- Providing administration support;
- Liaising with the Commission and their reviewers; and
- Managing finances.

The consortium is well-balanced from a management point of view. The appointed coordinator has comprehensive experience in managing European collaborative research projects, and the project uses suitable management structures (Section 3.2.1) and procedures (this section). The consortium builds upon existing collaborations between individual project partners, which have repeatedly proved to be successful in several areas of activity, from research projects to exploitation work and the provisioning of scientific events.

Decisions will be taken as described above with the **GA** being the highest decision making body with the exclusive right to make material decisions such as substantial changes to the work plan, budget or major amendments of the contractual agreements. Decisions are taken by simple majority of the votes unless otherwise specified in the Consortium Agreement (**CA**). The **GA** meets at least once per calendar year. For all other decisions, the subsidiary principle is executed, meaning that each decision shall be made as close as possible to the level of execution. In the event of risks occurring, the Consortium is prepared to adjust the work plan to avoid a failure of activities. In case no decision can be taken, the matter will be addressed at the superior executive level (e.g., from WPLs to PCO).

The **EB** meets (largely by video-conferencing) at least monthly, or upon request of the Chair or of at least three of its members. Business arising from the different activities will be considered, including the ratification of results of any round of the call for proposals having a significant financial or strategic impact on the project. When urgent matters arise between the monthly meetings these will, when possible, be dealt with by the **EB** per email, or using other communication channels set up within the project. **WPLs** will set up regular meetings with their WP teams to coordinate activities to ensure delivery. The **SAB** will be convened annually to review the system and ensure it is well-connected with external initiatives; their assessments will feed into the proceedings of the OC when it is established (WP6).

For issues regarding the necessary legal requirements of the planned B2TERM service, its components (e.g., TIF), its integration within the EOSC Service Catalogue and to ensure external privacy and data protection, the PCO will initiate an award procedure for a suitable subcontractor.

KOINE will use email, online conferences, and popular Web-based tools to allow for an effective and lean execution of the work plan as follows:

- A Web-based project-specific information space for communication among the project members and with the European Commission. Access will be granted on an invitation-only basis for all confidential information, though the general aim of the project is to run its activities openly and transparently. This environment will include information about the KOINE project and associated procedures and templates, the applications for funding, evaluation reports, meeting minutes, the tabling of consortium proposals, their rankings and final decisions taken by the PCO. Reports of the results of the funding will also be filed including a record of funds actually used. Further on, there will be an online shared folder to maintain all documents relevant to the project, including contractual information, information about the work plan, deliverables in various versions, including embedded internal reviews.
- A project-wide email distribution list will be set up for the preparation of the work, where partners exchange information, ideas and opinions. The list may also be used as a bi-lateral communication tool between partners, keeping the other partners informed of the progress in project tasks.

- Periodic web-based conferences, with a periodicity to be set, where consortium members discuss project issues and project progress.

3.2.3 Milestones

The Milestones mark the completion of the major phases of the project. There are milestones for the completion of the design, prototype, deployment, integrations and evaluation phases. The Milestones are described in Table 3.2a.

Table 3.2a: List of milestones

Milestone number	Milestone name	Related work-package(s)	Due date (in month)	Means of verification
2.1	Landscaping of FAIR terminologies and terminology services completed	2	9	Demonstrator
2.2	API specified and implemented	2, 3	18	Demonstrator
2.3	KOINE recommendations completed	2	24	D2.2, D2.3, D2.4, D2.5
3.1	Technology selection finalised	3	6	D3.1
3.2	Internal prototype of B2TERM	3	12	Demonstrator
3.3	Public release of B2TERM, V1	3	24	Demonstrator
3.4	Deployment of enhanced B2FIND service	3	30	Demonstrator
3.5	Public release of B2TERM, V2	3	36	Demonstrator
4.1	B2TERM and data centres integration started	4	12	The outline of D4.1.1
4.2	Semantically-enabled data portals online	4	30	D4.2
4.3	B2TERM pilot applications demonstrated	4	30	D4.3
4.4	Documented DSS interfaces available	4	12	D4.4
5.1	User-driven requirements gathering completed	5	18	D5.1
5.2	KOINE services validates by the research communities	5	28	D5.2
5.3	Stakeholder training completed	5	36	D5.3
6.1	KOINE Governance and Operations Committee Kick-Off	6	24	D6.2
6.2	Legal assessment of B2TERM service completed	6	30	D6.4
6.3	B2TERM service included in EOSC Service Catalogue	6	32	D6.4
7.1	Midterm review of “Dissemination & Communication Strategy and Plan” and Midterm KOINE public event	7	18	Events held
7.2	Release of KOINE logo, website and communication materials	7	6	Website operational and Materials available
7.3	Release of B2TERM communication packages	7	15	Materials available
7.4	Final KOINE public event	7	34	Event held

3.2.4 Risk management

The following risks have been identified as critical to the success of the project, and mitigation measures have been planned. Risks will be followed in the project according to a standard risk management plan implemented in WP1. A general risk is connected to the timely hiring of personnel for performing the work

plan, according to scientific excellence and gender balance. It is somewhat mitigated by the strong permanent staff of some partners. Specific competences might require long advertisement of the hiring and selection time. Risks connected to innovation and software development will be managed by the EB based on reports by WP leaders, of extra incurring costs or delays in the deliverables. This is mitigated by a consortium involving experienced project managers and innovation teams.

Table 3.2b: Critical risks for implementation

Risk No.	Description of Risk	Likelihood / Impact	WP No	Proposed risk-mitigation measures
Management Risks				
R1	Withdrawal of consortium member from the project	low / high	All	Consortium members are all highly committed to the KOINE project. In the unlikely event that a partner leaves the project, if possible, the consortium will find a suitable replacement. If this would not be possible, the tasks allocated to the dropout, will be re-assigned to the rest of the consortium members.
R2	Withdrawal/Unavailability of a key staff in the project	medium /high	All	Each consortium member is responsible for the personnel management of its staff. Therefore, the affected consortium member should and will handle the substitution of the participating staff.
R3	Key milestones or deliverables are delayed	high / high	All	The management team will monitor all effort spent during the project, identifying and dealing with any miscalculations as early as possible. The various milestone cycles will also act as measures to ensure that the project is on track. In the unlikely event of having delays in finalising key deliverables, the management team together with the EC will adjust the work plan.
R4	The dissemination of the project results is not sufficient to create impact	high / high		Strategy for dissemination is defined with clear responsibilities. Detailed dissemination plan is created already. All activities will be reviewed regularly during the full project duration. If a review of the dissemination activities establishes that the impact is not sufficient, a choice of remedial measures will be proposed by the WP leader(s), and the Management Board.
R5	Budget underestimated or project objectives not fully achieved	high / high		The consortium would put any effort required to achieve the objectives. The organisation of the work allows to achieve key results with strong impact even if objectives are not fully met.

Technical Risks				
R6	Failure to meet user requirements. The software functionalities do not meet user requirements.	Medium / high	All	In order to avoid misspecification of software functionalities, KOINE will follow an iterative development process and involve stakeholders in all stages of the development.
R7	Risk of poor data quality. The risk in carrying out this kind of project, which has to work with existing, but exemplary data, lies in the quality of data used as basis for the project.	Medium / high	All	The project members are aware of the quality of the raw data and took this into account when preparing the proposal.
R8	Risk of losing thematic relevance during the runtime of the project, issues could appear which will raise the need to re-orientate parts of the project to a different direction than stated in the proposal.	Low / high	All	Communication and objectives adjustment will be accomplished in the project by monitoring thematic-related activities i.e. other EU-funded projects, national projects and other important influences on the market and research environment. Furthermore, the end users and dissemination leaders will monitor the research and development process during the complete lifetime of the project.
R9	Risk of not achieving sufficient scalability.	Low / medium	All	Our approach to achieve scalability is based on distributing tasks in a cluster of commodity hardware. Hence, scalability can be simply increased by adding nodes to the cluster. Furthermore, the Coordinator's representatives have a proven track record on cutting edge research and development.

3.3 Consortium as a whole

The KOINE consortium brings together organisations that cover large research communities and have pioneered, promoted, and greatly shaped the adoption of the FAIR principles and data-driven technologies and services. The consortium is characterized by its network of multiplier organisations that are directly involved as project partners. With the help of these organisations, which are also scientific networks and digital service providers, KOINE engages European and global research infrastructure serving thousands of scientists.

KOINE involves network partners and multipliers from three broad research disciplines, namely the Earth and Environmental Sciences, including Food and Agronomy, the Engineering Sciences, and Humanities and Social Sciences, as well as the established EUDAT service B2FIND and the Data Subscription Service already prototyped in Marine Environmental Science (IFREMER). Moreover, disciplinary data publishers (e.g., PANGAEA, DKRZ, BODC, 4TU.Centre for Research Data, INRA) and data centres of research infrastructures (e.g., ACTRIS, eLTER, SIOS, LW ERIC, CLARIN) as well as interdisciplinary data discovery services (e.g., B2FIND) apply the B2TERM service and thus actively catalyse the transition in the current practice of using ambiguous labels to describe research data semantics toward using unambiguous, persistently identified and machine-readable terminology. Within data discovery services such as B2FIND, the adoption of terminology enables semantic data discovery and exploitation across disciplines for stakeholders including individual researchers, research communities, public sector, industry, and educators.

Through KOINE, these national and pan-European research infrastructures will not only ensure that terminology is integrated in their data curation workflows; they will additionally catalyse the adoption of the advanced practices in research communities, by the researchers themselves. Through research infrastructures, KOINE will actively work with researchers to ensure the delivered service and approaches are effectively addressing researcher's needs and are efficiently integrated in their scientific workflows.

The importance of B2TERM technical and social integrations in research and e-Infrastructures and their research communities is reflected in KOINE's overall effort shared by the corresponding WP4 (28%) and WP5 (22%), i.e. half of the total effort. Partners have thus adequate resources to implement these integrations.

Research infrastructures, data centres and publishers, and data discovery services are complemented by partners with world-class expertise in ICT, generally, and semantic technologies, specifically. ICT partners (e.g., TIB, CNRS-LIRMM, UvA, UMAAS, CSC) ensure that the Terminology Interoperability Framework, the B2TERM service, and B2TERM integrations are specified and implemented according to the highest standards in software engineering. All ICT partners have solid and proven expertise in collaborating with research communities to develop effective e-Infrastructure services for science. The specification, development and implementation of B2TERM receives approximately a third of the total effort. This is adequate for this proposal as KOINE primarily integrates existing software systems and the smaller share compared to integrations also reflects that the adoption of terminology in research is primarily a social rather than technical challenge.

Furthermore, KOINE includes two SMEs, namely SISTEMA and EUDAT Ltd., that are willing to contribute to the development or application of B2TERM and KOINE outputs. As a user, SISTEMA will evaluate the effectiveness of terminology-enhanced datasets to information mining, also for stakeholder groups outside academia.

These organisations are at the core of the FAIR data movement not only in their countries or at pan-European level, but also globally. They have an impressive track record in current research activities, the development of terminologies, the development and maintenance of research data infrastructures and in the management of research data as a whole (the research data lifecycle), all of these being central to this call. In a nutshell, the KOINE consortium has the network and means to reach out to thousands of researchers and their organisations aiming to increase data interoperability and FAIRness of research data.

The formation of the KOINE consortium was directed by the following guidelines:

- Establish a consortium that has the **skills, expertise, and motivation** in order to meet the ambitious objectives of the project, and to achieve the expected impact that will significantly increase the semantic findability and re-usability of research data produced and consumed within the EOSC through interoperability;
- **Balance the distribution of skills and know-how across the partners**, paying account to critical complementarities and positive overlaps, and ensure that all key stakeholders of the research lifecycle are present and actively involved in the implementation of the work plan;
- Bring together organisations that, through their ethos and goals, will **collaborate effectively** in achieving the mission of the project. In particular we strove to create a balanced mix of partners with experience in running large-scale European projects and organisations that are relatively new to this landscape and might bring in novel ideas and approaches to well-known problems;
- Include **research communities and infrastructures that can serve as early adopters and multipliers** ('best practice examples'), which drive to achieve a large societal impact within the present and future EOSC.

3.4 Resources to be committed

KOINE ensures sufficient resources are allocated to partners in order to implement project deliverables. The requested total contribution to KOINE is 5,987,332.50 Euro, including management. The allocation of resources is shared among WPs as shown in Figure 7. Costs for project management (WP1) and communication and dissemination (WP7) are kept at 4.6% and 5.2% of the overall budget, respectively. WP4 and WP5 together account for 50% of the overall budget, which reflects the emphasis KOINE sets on service integration in (inter)disciplinary operational infrastructures as well as ensuring that technical infrastructure is adopted by the social infrastructure, in particular research communities. Beneficiaries are located in 9 countries, plus Australia (represented by ANDS). The ACTRIS research infrastructure is represented by

CNRS, NILU, FMI, CNR, and BSC; these partners share the total effort required for ACTRIS to implement project deliverables. Staff effort allocated at the participants is distributed as compiled in Table 3.4a.

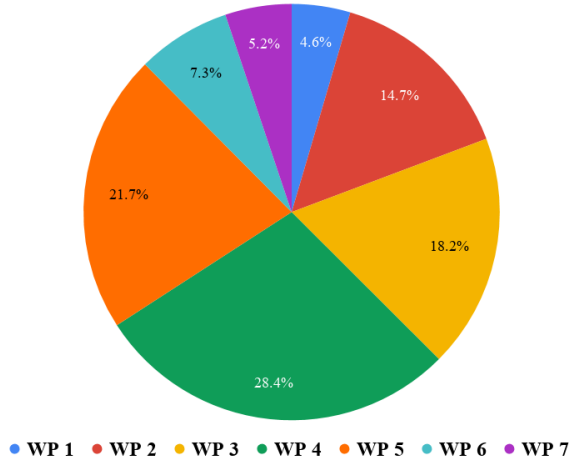


Figure 7: Allocation of resources among KOINE WPs.

Table 3.4a: Summary of staff effort

	WP1	WP2	WP3	WP4	WP5	WP6	WP7	Total Person-Months per Participant
1/TIB	24	5	16	2	3	18	8	76
2/CNRS	1	24	17	12	8	3	3	68
3/EAA	1	4	0	12	28	3	0	48
4/UHB	1	3	0	30	5	0	2	41
5/AWI	1	12	0	0	7	12	0	32
6/NILU	0	0	0	5	0	0	0	5
7/UvA	1	5	35	0	0	0	0	41
8/UMAAS	0	4	0	0	8	3	0	15
9/MET	0	0	0	15	10	3	0	28
10/LW ERIC	1	12	8	22	14	3	18	78
11/UKRI	0	20	13	0	8	3	3	47
12/FMI	0	0	0	5	0	0	0	5
13/CNR	0	0	0	5	0	0	0	5
14/BSC	0	0	0	0	3	0	0	3
15/DKRZ	0	0	22	23	5	0	0	50
16/EUDAT ltd.	0	0	0	0	0	0	0	0
17/CSC	0	0	0	16	0	0	0	16
18/IFREMER	0	0	0	6	0	0	0	6
19/INRA	0	4	0	15	5	0	0	24
20/TUD	0	4	0	15	30	0	0	49
21/OEAW	0	0	9	4	4	0	0	17
22/ANDS	0	0	0	0	0	0	0	0
23/SISTEMA	0	0	0	0	5	0	0	5
Total Person Months	30	97	120	187	143	48	34	659

Table 3.4b: 'Other direct cost' items (travel, equipment, other goods and services, large research infrastructure)

3/EAA	Cost (€)	Justification
Travel	13,000	Kick-off meeting, mid-term project meeting, mid-term review, final review, mid-term public event, final public event.

Other goods and services	40,000	Organisation of WP5 training events.
Total	53,000	

6/NILU	Cost (€)	Justification
Travel	8,000	Kick-off meeting, mid-term project meeting, mid-term review, final review, mid-term public event, final public event.
Total	8,000	

14/BSC	Cost (€)	Justification
Travel	2,000	Kick-off meeting, mid-term project meeting.
Total	2,000	

KOINE Glossary

B2FIND	Generic metadata and discovery service for research data sets in EOSC
B2TERM	Syndicated terminology service that consumes the content from terminology services part of the TIF. It offers user and application programming interfaces to access terminology, supports slicing and terminology mediation.
Common Semantic Model	Conceptualises the descriptions of scientific variables and parameters harmonising the exchanged information for both humans and machines.
Compliance Standards	Rules established by KOINE for serving terminologies to B2TERM and therefore make them more FAIR
Data Subscription Service (DSS)	Service for personalized data discovery to notify users when relevant new data are published or existing data sets are extended or updated.
KOINE	Greek term for a standard language that arises as a result of contact between two or more dialects of the same language. We use it as project name to catalyse the use of a common language for interdisciplinary data interoperability, discovery and exploitation in the EOSC.
Slice	Restricted view on all terminologies available in B2TERM. This mechanism (originally proposed in the NCBO BioPortal and available in AgroPortal) will allow a community to restrict B2TERM to community defined subsets of terminologies and features.
Terminology	Overarching name for any set of fixed denotations that are used to describe something with the goal to reduce ambiguity and facilitate interoperability. A terminology can range from a simple controlled vocabulary (a simple list of terms), thesaurus to a complex ontology (formal definitions of terms and their relation expressed in a machine readable way) or any other kind of knowledge organisation source.
Terminology Interoperability Framework (TIF)	A framework that enables the syndication of terminology services into B2TERM. It harmonizes programmatic access to terminologies as well as their metadata description. It relies on a Common Semantic Model and determines the Compliance Standards.
Terminology Service	Overarching expression used in this proposal to name any kind of web application serving terminology content. This includes ontology repositories (e.g., NCBO BioPortal, AgroPortal, EBI-OLS, EcoPortal), terminology or vocabulary service (e.g., ANDS, NERC-VS, CLARIN-VS). B2TERM is itself a terminology service which differs from other as it consumes content from other terminology services rather than by hosting directly terminologies.

KOINE: Interdisciplinary data interoperability, discovery and exploitation in the EOSC

Proposal addressing H2020 call INFRAEOSC-02-2019
Hannover, 28 January 2019

Contents

Section 4: Members of the consortium

4.1. Participants (applicants)

Technische Informationsbibliothek (TIB)

TIB is the German National Library for science and technology, Information Centre of the Leibniz Association as well as university library for Leibniz University Hannover. TIB represents and operates a national research infrastructure facility for the provision of scientific information covering engineering, information technology, mathematics, physics, chemistry and architecture. TIB is member of the German research organization Leibniz Association comprising over 90 leading institutes from all areas of science. TIB preserves and organizes information, data and knowledge in its target domains and provides direct access to these large-scale information spaces through digital services, irrespective of time and place. With its vast collections and innovative services, TIB aims to support the complete life-cycle of research and the digitization of science and technology in general. As the world's largest specialised information centre in its fields, TIB has outstanding expertise in developing, managing and preserving knowledge, particularly in key areas such as grey literature (e.g. conference and research reports), big research data, vocabularies and ontologies, films and 3D objects, as well as patents and standards. Some examples of key services include the TIB portal giving access to more than 100 Million documents and research artefacts, the audio-visual portal AV-Portal comprising more than 17,000 scientific videos, license negotiation, Open Access and Digital Preservation offerings, the research data management software Leibniz Data Manager or the collaborative OpenCourseWare authoring platform SlideWiki. In 2009 TIB founded the DataCite association with meanwhile more than 130 international member organizations and hosts the DataCite headquarter. In addition to serving the scientific communities in its fields, TIB is information service provider for major international corporations including Volkswagen, Siemens, Continental or Daimler. In close cooperation with L3S research center from Leibniz University Hannover, TIB performs world-class research aiming to advance information, data and knowledge sharing in the digital age for example with its Open Research Knowledge Graph and to facilitate the digitalization of science, industry and society at large.

Role in the project

As a relevant world class research infrastructure with an international dimension, TIB is a leading institution in science & technology information organization with a research track record demonstrating scientific excellence in developing and applying methods for data and system interoperability, in particular **knowledge-based methods for semantic interoperability** of exchanged data. **TIB coordinates the KOINE project** and will support the consortium in connecting researchers across scientific disciplines with interoperable infrastructures in the EOSC, with in-depth understanding of:

- All aspects of findable, accessible, interoperable and reusable data - in particular persistent identification, metadata standards, languages for knowledge representation, development of interoperable and reusable vocabularies, provenance, and licensing which will be core to the B2TERM service foundations (WP2) and development (WP3, WP4)
- Techniques underpinning next generation knowledge infrastructures - in particular information extraction from data, representation and curation of information in knowledge-based systems,

processing and visualization of information, as well as the integration of such techniques in scientific workflows exposed as services to research communities (WP5 and WP7).

As a service provider to a wide range of academic and industrial customers, TIB will be responsible for the transfer of KOINE project results, mainly the **B2TERM service, into a successful, sustained and long-term service**. As such, TIB will ensure the transition of B2TERM from the project prototype into a fully operational service far beyond the project's lifetime (WP6).

Key personnel

Dr. Markus Stocker [m] is head of the Knowledge Infrastructures research group leader at the German National Library of Science and Technology (TIB). Markus holds a PhD in Environmental Informatics from the University of Eastern Finland, a MSc in Environmental Science from the University of Eastern Finland, and a Diploma (MSc) in Informatics from the University of Zurich, Switzerland. He is author of 40 peer-reviewed journal and conference proceeding papers, with roughly 1000 citations. He has managed partner contribution and been involved in various H2020 projects, including THOR, ENVRIplus, OpenAIRE, FREYA, ENVRI-FAIR as well as nationally-funded projects in Finland and Germany. Prior to TIB, Markus held a postdoctoral research associate position at PANGAEA, the Data Publisher for Earth & Environmental Science, at the MARUM Center for Marine Environmental Sciences, University of Bremen, Germany. He continues to be affiliated with MARUM. As a member of the Research Data Alliance (RDA), Markus is involved in various groups, in particular the WG Persistent Identification of Instruments and the IG From Observational Data to Information. He has several years of professional experience in software development and semantic technologies, with positions at Hewlett Packard Labs, Bristol, UK and Clark & Parsia, Washington DC, USA.

Alexandra Garatzogianni [f] studied management, linguistics, media and communication. Her business interests focus on innovation and project management, international business development, software product management, startup consulting in physical and digital incubators and data-driven entrepreneurship. She has successfully lead European Consortia within the context of Horizon 2020 ICT Research and Innovation projects, feasibility studies at European level as well as prototype development for a major Chinese state-owned telecommunications operator and world's fourth-largest mobile service provider by subscriber base. She is currently working as Senior Research and Innovation Project Manager at the Leibniz Information Center for Science and Technology, University Library (TIB) and at the Leibniz University of Hannover. She considers herself a citizen of the world, and as such has the C2 certified level in five foreign languages, i.e. German, English, French, Spanish, European and Brazilian Portuguese and the B2 level in Russian.

Dr. Angelina Kraft [f] is a data scientist and head of the Research Data and Scientific Software team at TIB. The team develops quality standards and infrastructure services for the publication of research data and scientific software according to the FAIR (Findable, Accessible, Interoperable and Reuseable) principles. She co-facilitates the task force Open Science - Research Data management within the CESAER (Conference of European Schools for advanced engineering Education and Research) initiative and serves as reviewer for various conferences and journals, including the International Open Science Conference and MDPI (Multidisciplinary Digital Publishing Institute). Angelina Kraft has published more than 20 research papers at national and international conferences and more than 25 associated digital research data sets. She is member of the DataCite community engagement steering group (CESG) and responsible for the further development of services for digital data management at TIB such as the generic research data repository (RADAR).

Dr. Javad Chamanara [m] is a postdoctoral researcher at the L3S/TIB joint lab working on distributed semantic research data management systems. Javad holds a Ph.D. in Computer Science

from the University of Friedrich Schiller Jena, Germany, an MSc in Software Engineering from the University of Poly-technique Iran, and a BSc in Software Engineering from the University of Science and Culture from the University of Iran. He has authored many journal and conference papers on heterogeneous data processing, query languages, conceptual modeling of research data. Javad has invented a new query language (QUIS) for data retrieval from heterogeneous data sources. He also has introduced a new data life cycle (JenPlane) to mitigate the shortcomings of traditional waterfall lifecycles such as DataOne and USGS. Before joining L3S he was team lead and software architect of BExIS, a web-based research data management system used by many projects including iDiv, AquaDiva, Max Plank BGC, Biodiversity Exploratories, and ATTO. He has run a large national project to conceptualize and develop ontologies for 21 geo-related domains such as geochemistry, geophysics, exploration, and mining. Javad has attended many learning programs, e.g., an internship at the University of Michigan US and a big data analysis summer camp at the University of Lisbon Portugal. He is currently a member of the Research Data Alliance (RDA) and together with Barbara Magagna, Mark Schildhauser, and others have submitted a working group on variable harmonization to the 13th plenary meeting.

Dr. Gino Erkeling [m] is the EU research liaison officer at TIB with project administration expertise. He has worked for the German Aerospace Center (DLR) and the European Space Agency (ESA) and has over 10 years experience as a project manager. In particular, he has coordinated a Helmholtz alliance at Muenster University (Germany) between 2008 and 2012. Between 2012 and 2017, he was a team leader of the ESA mission Mars Express. Since 2015, he is Co-I of Mars Express. He is well familiar with German and European grant programmes, in particular with funding via the German Federal Ministry for Economic Affairs and Energy (BMWi), the German Federal Ministry of Education and Research (BMBF) and the European research and innovation framework programme Horizon 2020.

Prof. Dr. Sören Auer [m] is professor for Data Science and Digital Libraries at Leibniz University of Hannover and director of TIB German National Library of Science and Technology. Sören has made substantial contributions to semantic web technologies, knowledge engineering, software engineering, usability, as well as databases and information systems. Sören is author (respectively co-author) of over 100 peer-reviewed scientific publications, which attracted more than 10.000 citations and result in an H-index of 45. He received several awards, such as the ESWC 7-year best paper award or the OpenCourseware Innovation award. He led several large-scale collaborative research projects such as the European Union's H2020 flagship project BigDataEurope. Sören is co-founder of several high-impact research and community projects such as the Wikipedia semantification project DBpedia, the OpenCourseWare authoring platform SlideWiki.org or the spatial data integration platform LinkedGeoData. The technology Sören develops with his team fuels many industrial applications. He is organiser, programme or track co-chair of renowned conferences and workshops, including OKCON 2010, ESWC 2010, ICWE 2011, WWW 2012, European Data Forum. He serves as an expert for industry, the European Commission, the W3C and board member of the Open Knowledge Foundation.

Publications

- Stocker M, Paasonen P, Fiebig M, Zaidan MA, Hardisty A (2018) Curating Scientific Information in Knowledge Infrastructures. Data Science Journal, 17:21. https://doi.org/10.5334/dsj-2018-021
- Diepenbroek M, Schindler U, Huber R, Pesant S, Stocker M, Felden J, Buss M, Weinrebe M (2017) Terminology Supported Archiving and Publication of Environmental Science Data in PANGAEA. Journal of Biotechnology, 261:177-186. https://doi.org/10.1016/j.jbiotec.2017.07.016
- Auer S, Bizer C, Kobilarov G, Lehmann J, Cyganiak R, Ives, ZG (2007) DBpedia: A Nucleus

for a Web of Open Data. Int. Semantic Web Conference, ISWC/ASWC 2007: pp. 722-735.
- Attard J, Orlandi F, Auer S (2016) Data Value Networks: Enabling a New Data Ecosystem. 2016 IEEE/WIC/ACM International Conference on Web Intelligence (WI), Omaha, NE, pp. 453-456. https://doi.org/10.1109/WI.2016.0073
- Kraft A, Dreyer B, Löwe P, Ziedorn F (2017) 14 Years of PID Services at the German National Library of Science and Technology (TIB): Connected Frameworks, Research Data and Lessons Learned from a National Research Library Perspective. Data Science Journal, 16:36. https://doi.org/10.5334/dsj-2017-036

Relevant projects

- 2018-2020: **ENVRI-FAIR** is the connection of the ESFRI Cluster of Environmental Research Infrastructures (ENVRI) to the European Open Science Cloud (EOSC) (H2020). Participating research infrastructures (RI) of the environmental domain cover the subdomains Atmosphere, Marine, Solid Earth and Biodiversity / Ecosystems and thus the Earth system in its full complexity. The overarching goal is that at the end of the proposed project, all participating RIs have built a set of FAIR data services which enhances the efficiency and productivity of researchers, supports innovation, enables data- and knowledge-based decisions and connects the ENVRI Cluster to the EOSC.
- 2018-2020: Big Data Value Spaces for COmpetitiveness of European COnnected Smart FacTories 4.0 - **BOOST 4.0** (H2020): BOOST 4.0 addresses the need for development of large scale, data-driven "connected smart" Factories 4.0 model through 10 lighthouse factories. The BOOST 4.0 project demonstrates how European industry can build unique strategies and competitive advantages through big data across all phases of product and process lifecycle (engineering, planning, operation, production and after-market services) building upon the connected smart Factory 4.0 model to meet the Industry 4.0 challenges.
- 2016-2018: Large-scale pilots for collaborative OpenCourseWare authoring, multiplatform delivery and Learning Analytics - **SildeWiki** (H2020). SlideWiki is a collaborative platform that allows individuals and groups to create, import and share slide decks, allowing to fork and remix existing slides, integrate LaTeX/MathML etc. The SlideWiki EU project started 2016 with seventeen partners from Europe and Brazil using the award-winning open-source SlideWiki platform from Germany as a basis. All project goals aiming at creating a large scale accessible learning and teaching platform using educational technology, skill recognition and global collaboration.
- 2013-2016: Durable Architectural Knowledge - DuraArK (FP7). **DURAARK** was a European FP 7 project (Durable Architectural Knowledge) focussing enrichment and on long-time preservation of architectural building data. TIB acts as a content provider of architectural 3D models and is involved in the areas of metadata and persistent identifiers, contributing with experience and knowledge in metadata, persistent identifier technologies and scientific data registration.

Relevant infrastructure

TIB runs its own computing center with several dedicated cluster infrastructures comprising dozens of server nodes and more than a petabyte of overall storage space. The computing center has a NESTOR certification for long-term digital archiving. The TIB service infrastructure comprises a knowledge graph containing bibliographic metadata for more than 100M research artefacts, the DOI and DataCite registries, the AV-Video-Portal, the SlideWiki OpenCourseWare collaboration infrastructure and a large portfolio of open-source software technology platforms such as the Research Data Management framework Leibniz Data Manager, the Ontario federated knowledge graph management or the Open Research Knowledge Graph.

National Center for Scientific Research (CNRS)

The National Center for Scientific Research (CNRS) is a public organization under the responsibility of the French Ministry of Education and Research. As the largest fundamental research organization in Europe, CNRS carries out research in all fields of knowledge, through its ten scientific institutes. CNRS encourages collaboration between specialists from different disciplines in particular with universities thus opening new fields of enquiry to meet social and economic needs. CNRS heads 950 research units, a hundred of service units in France, as well as 33 international units around the world. In those, work 32,000 employees that conduct scientific studies or assist researchers in this task, with a budget over 3.3 billion euros a year. Regularly, the CNRS is the scientific organisation that receives the most ERC grant yearly.

Within KOINE, the CNRS involves three research units – LIRMM, CEFE, AERIS – and will be represented by 'Occitanie Est delegation.' The delegation is also strongly connected to the University of Montpellier (UM) a French research-intensive university where education and research cover most of the scientific and technological fields. Since 2017, university's research and education actions are driven by the *Montpellier University of Excellence* (MUSE) initiative – 17M€/year– a federated program to build an internationally recognized university notably for its impact in the fields of agriculture, environment and health. Multiple joined research units are also involved in the national Digital Agriculture Convergence Lab (#DigitAg) based in Montpellier, which gathers the main national actors (IRSTEA, INRA, SupAgro, INRIA and companies) working on digital agriculture to meet the food security and environmental sustainability challenge.

LIRMM – Laboratory of Informatics, Robotics, and Microelectronics of Montpellier (www.lirmm.fr) is a 350-person cross-faculty research entity of the University of Montpellier and the CNRS. LIRMM research activities cover a broad range of topics, including: design and verification of integrated, mobile and communicating systems, modeling of complex systems, research on algorithms, bioinformatics, human-machine interaction, robotics, database, distributed systems, AI, knowledge engineering and more. LIRMM's Informatics department counts 85 permanent researchers, and more than 70 PhD candidates. Since 2008, LIRMM has been involved in more than 40 EU projects (http://www.lirmm.fr/lirmm_eng/international/international-parnerships). LIRMM obtained A+ during the HCERES 2010 & 2014 evaluations. In addition, LIRMM has chosen health, agriculture and environmental sciences as priority domains of application for its research in Informatics and Robotics. Several research teams (ADVANSE, FADO, GRAPHIK, ZENITH, TEXTE) have good expertise in knowledge engineering, semantic Web, text mining, services and ontologies. GRAPHIK & ZENITH INRIA associated teams are involved in several EU projects related to agri-food including H2020 GloPack, H2020 NoAW, EU COST FoodMC. The FADO team (involved in ANR D2KAB, H2020 SIFRm, H2020 eRosa) has been created in 2017 on the scientific topic of open and linked data in the context of the semantic Web. Research addresses fundamental questions (e.g., semantics, ontologies, alignment, knowledge representation) and is also applied to multiple domains including agri-food. The team develops a set of knowledge engineering platforms, including YAM++ (ontology alignment), Legato (data linking), SIFR BioPortal and AgroPortal (ontology repositories), AgroLD (RDF knowledge base).

CEFE – Center for Functional and Evolutive Ecology (www.cefe.cnrs.fr), is a joint research unit (CNRS, Univ. of Montpellier, Univ. Paul Valéry Montpellier 3, EPHE, IRD) and one of the largest

research center in ecology and evolution in France. Its mission is to perform independent, basic scientific research on (i) the dynamics of biodiversity, (ii) the impacts of the different components of global changes on ecological systems, and (iii) sustainable development issues. Work is conducted in a large variety of systems worldwide, but particular expertise has been developed in Mediterranean and tropical ecosystems. One of the main objectives is to develop scenarios on the evolution of biological systems, as well as strategies for their conservation and restoration. CEFE currently hosts 148 permanent staff, 79 PhD candidates, and 93 non permanent staff, including 47 post doctoral researchers. CEFE is the administrative host of the "CNRS SémanDiv GDR" launched in 2017, a national working group devoted to biodiversity semantics lead by Dr. Eric Garnier.

AERIS – French Data center for Atmospheric research (www.aeris-data.fr) is appointed by eleven institutes and agencies, and has the objective to facilitate and enhance the use of atmospheric data, whether from satellite, aircraft, balloon, or ground observations, or from laboratory experiments. It generates advanced products and provides services to facilitate data use, to prepare campaigns, and to interface with modeling activities. It consists of four Centres for Data and Services (CDS): ICARE, which is mainly dedicated to aerosols and cloud satellite data, and constitutes one of the pillars of the ACTRIS-2 data center; ESPRI, which is dedicated to atmospheric chemistry data; SEDOO for data from large field campaigns; and SATMOS, which is dedicated to services for users of geostationary satellite data. AERIS has close relationships with different laboratories for transferring prototype products and expertise on data. It has strong expertise in data curation, storage, preservation and dissemination.

Role in the project

Within KOINE CNRS-LIRMM will be leading WP2 and be strongly involved in WP3. LIRMM provides expertise on working with terminologies and developing vocabulary and ontology repositories in biomedicine and agronomy. Since 2016, LIRMM designs and maintains AgroPortal, an ontology repository for agronomy and related domains (http://agroportal.lirmm.fr) in collaboration with INRA and other important agri-food actors (IRD, CIRAD, CGIAR). Within KOINE, LIRMM participates as both a terminology services for B2TERM but also an expert in semantic technologies, ontology metadata, semantic annotation and ontology repository interoperability. We will also collaborate with the National Center for Biomedical Ontology (NCBO) at Stanford University which develops BioPortal (health and biomedicine). The NCBO technology is reused within KOINE by 3 terminology services: SIFR BioPortal (French biomedicine), AgroPortal (agronomy) and EcoPortal (ecology & biodiversity – led by LifeWatch ERIC). LIRMM is also leading a national ANR supported 3,1M€ project to collaborate with 2 other actors of KOINE (CNRS-CEFE, INRA) on exploitation of semantics to transform data into knowledge in agronomy and biodiversity (www.d2kab.org). LIRMM is currently involved (as task leader) on a response proposal for the DT-SFS-26-2019 call to build the Food and Nutrition Security thematic cloud for EOSC. Led by O. Le Gall (INRA).

Within KOINE, CNRS-CEFE brings its expertise at the interface between ecology and semantics, and will be involved in parameter harmonization (WP2), the development of a functional biogeography use case in collaboration with LifeWatch ERIC (WP4), and in the testing and validation phases of the services provided (WP5).

Within KOINE, CNRS-AERIS brings its expertise in the atmospheric domain and will participate in user-focused service applications (WP4) through the ACTRIS Data center.

Key personnel

Dr. Clement Jonquet [m] is Assistant Professor (will be Associate Professor in 2019) at University of Montpellier, France since Sept. 2010 and was visiting at the Stanford Center for Biomedical Informatics Research (BMIR) from 2015 to 2018. He is a researcher at LIRMM within the FADO research group working on (biomedical/agronomical) ontologies, semantic data indexing and annotation, semantic Web, text mining, knowledge representation. Dr. Jonquet has 12 years expertise in semantic web and application to health and agronomy. In the context of the National Center for Biomedical Ontology (NCBO) project supported by NIH (U54-HG004028) he contributed actively to the design, evolution and development of the NCBO BioPortal the reference ontology repository in biomedicine. Since 2013, Dr Jonquet is the PI of the SIFR project (Semantic Indexing of French Biomedical Data Resources) mainly funded by the ANR Young Researcher program (ANR-12-JS02-01001) and recipient of the H2020 Marie Skłodowska-Curie program (701771). He is co-PI of PractiKPharma project (Practice-based evidences for actioning Knowledge in Pharmacogenomics – ANR-15-CE23-0028). He also leads the AgroPortal project, a repository of ontologies for agronomy. Dr. Jonquet is the (co)author of 80 publications cumulating more than 2200 citations (h-index 19 in Google Scholar as of Dec. 2018), including 23 international journals in multiple domains (biomedical informatics, semantic Web, distributed systems & AI), 6 as first author. He co-supervised 3 PhD students and 12 MSc students and will be "habilitated" (French HDR) in 2019. He chairs the Web Science Montpellier and AgroHackathon Meetup groups.

Dr. Konstantin Todorov [m] is Assistant Professor at the University of Montpellier, France and a researcher at the LIRMM laboratory within the FADO group since 2012. His research lies in the field of Artificial Intelligence with a particular interest in knowledge and data representation, integration and information retrieval. He currently works on problems related to data linking, knowledge graph profiling and recommendation, key discovery and ranking, the use of background knowledge in the matching and linking processes. He is involved in a number of national and international projects in the field of data lifting and linking (e.g., ANR DOREMUS, Datalyse, PHC LDCT between LIRMM and IST Innsbruck (Austria)). Dr. Todorov has co-supervised 3 PhD and 8 master students and has participated in the development of a number of data fusion and ontology matching tools, such as Legato, YAM++ online, CODA. He is a PC member of several international web data conferences including ESWC, ISWC, WebConf, ACM WebScience, SEMANTiCS, and other and reviews regularly for the Semantic Web Journal, the Journal of Web Semantics, Fuzzy Sets and Systems, etc.

Dr. Eric Garnier [m] is Research Director (DR1) at CNRS. He is a researcher in plant ecology, whose work on plant functional diversity combines conceptual, experimental and methodological approaches, contributing to the broad field of trait-based comparative ecology. He has taken part in the development of large trait databases and more recently, has developed interests in semantics as a key component of the emerging discipline of ecoinformatics. He is the head of the national working group of semantic for biodiversity (CNRS SémanDiv GDR), that will continue the development of the TOP Thesaurus and facilitate the adoption of semantic technology by ecologists. He is also involved in the French initiative for the design of a "National Focal Point for Biodiversity Data", launched in 2018 by the Ministry for Higher Education and Research.

Dr. Guillaume Brissebrat [m], Research Engineer, is responsible of the Observatoire Midi-Pyrénées data service (SEDOO) which is part of AERIS and he will be the technical director of AERIS (as of July 2019). Dr. Brissebrat has more than 12 years of experience in database development and management. He has been involved in international projects (AMMA, MISTRALS) as responsible of the database conception, metadata activities, and other services associated to data centers.

Vincent Douet [m] is CNRS Software Engineer at the ESPRI/AERIS datacenter at Institut Pierre-Simon-Laplace since September 2017. Before he was working at IPGP datacenter and worked on many projects such as RESIF (French national equipment for the observation and understanding of the solid Earth) and InSIGHT (NASA mission to explore Mars' deep interior, (May 2018)). His range of expertise is full-stack web development, database management, metadata management.

Publications

- Jonquet C, Toulet A, Arnaud E, Aubin S, Dzalé-Yeumo E, Emonet V, Graybeal J, Laporte M-A, Musen MA, Pesce V, Larmande P (2018) AgroPortal: an ontology repository for agronomy. Computers and Electronics in Agriculture, 144:126–143, see also: http://agroportal.lirmm.fr/
- Todorov K, Bellahsene Z, Achichi M, Ben Ellefi M Linking and Disambiguating Entities Across Heterogeneous RDF Graphs. Web Semantics, IN PRESS, 2019
- Jonquet C, Toulet A, Dutta B, Emonet V (2018) Harnessing the power of unified metadata in an ontology repository: the case of AgroPortal, Data Semantics, pp. 1-31, Elsevier.
- Noy NF, Shah NH, Whetzel PL, Dai B, Dorf M, Griffith NB, Jonquet C, Rubin DL, Storey M-A, Chute CG, Musen MA (2009) BioPortal: ontologies and integrated data resources at the click of a mouse, Nucleic Acids Research, Vol. 37 (web server), pp. 170-173
- Garnier E, Stahl U, Laporte M-A, Kattge J, Mougenot I, Kühn I, Laporte B, Amiaud B, Ahrestani FS, Bönisch G, Bunker DE, Cornelissen JHC, Díaz S, Enquist BJ, Gachet S, Jaureguiberry P, Kleyer M, Lavorel S, Maicher L, Pérez-Harguindeguy N, Poorter H, Schildhauer M, Shipley B, Violle C, Weiher E, Wirth C, Wright IJ, Klotz S (2017) Towards a thesaurus of plant characteristics: an ecological contribution. Journal of Ecology 105, 298–309.

Relevant projects

- 2013-2017: **SIFR** (http://www.lirmm.fr/sifr – ANR-12-JS02-0010 and H2020 MSCA #701771) Building ontology-based services to leverage biomedical ontologies and terminologies in indexing, mining and retrieval of French biomedical data.
- 2015-2018: **AgroPortal** (http://agroportal.lirmm.fr/) supported by ANR (Labex NUMEV, Labex Agro, IBC of Montpellier). Design and development of an advanced prototype ontology repository for agronomy and related domains. Led by C. Jonquet.
- 2019-2023: ANR **D2KAB** (www.d2kab.org – ANR-18-CE23-0017). Create a framework to turn agronomy and biodiversity data into FAIR knowledge in agronomy and biodiversity. The project will focus on ontology services and mappings and linked data exploitation. Led by C. Jonquet.
- 2017-2021: GDR # 2011 SémanDiv (Biodiversity semantics, funded by CNRS). Interactions between French partners from different fields (ecologists, librarians, computer scientists) concerned by semantic issues pertaining to the ecological component of biodiversity, from organism to ecosystem. Led by E. Garnier.
- 2017-2020: H2020 **Eurochamp-2020** (www.eurochamp.org – #730997) New Research Infrastructure which aims at further integrating the most advanced European atmospheric simulation chambers into a world-class infrastructure for research and innovation. AERIS is responsible for the Data Portal.
- 2017-2010: H2020 **ACTRIS-PPP** (www.actris.eu) project for the preparation of the ACTRIS economic and governing model. Now AERIS is a major partner of the EOSC ENVRI-FAIR cluster.

Relevant infrastructure

AgroPortal, is a community effort started by the Montpellier scientific community (LIRMM, IRD, CIRAD, INRA, Bioversity International) to build an ontology repository for agronomy and related domains. The portal features ontology hosting, search, versioning, visualization, comment, and recommendation; enables semantic annotation; stores and exploits ontology alignments; and enables interoperation with the semantic web. Our goal is to facilitate the adoption of metadata and semantics to facilitate open science. By enabling straightforward use of ontologies, we expect data managers and researchers to focus on their tasks, without requiring them to deal with the complex engineering work needed for ontology management. The platform currently hosts 102 vocabularies or ontologies with more than 2/3 of them not present in any similar ontology repository and 11 private ontologies. We have identified 80 other candidate ontologies that will be loaded in the future to complement this valuable resource. The platform already has more than 95 registered users and some vocabularies are visited more than 100 times per month. However, the current AgroPortal prototype only partially addresses the needs of the community: it is not multilingual, it is limited in terms of ontology alignment capabilities and does not provide semantic-search and retrieval of agri-food data.

AERIS has for objective to facilitate and enhance the use of atmospheric data whether from satellites, ground, airplanes or balloons. For this, AERIS generates products from observations, but also provide many support services for the use of data, help to conduct synergies, campaigns or interface with models. More than 3000 datasets are available on the AERIS data portal and thousands of users are registered.

Umweltbundesamt Gesellschaft mit Beschränkter Haftung (EAA)

EAA was established in 1985 and is the environmental specialist institution of the Austrian Federal Government. It provides expertise on the condition of the environment and environmental changes as well as on measures to avoid or reduce environmental pollution. In 1998 the Agency became a limited liability company owned by the Federation, in order to allow a more flexible management. Currently, the agency employs some 450 people. The Environment Agency Austria undertakes environmental monitoring, assessment and evaluation. It plays an important role in the EIONET, in that it is the National Focal Point (NFP) for the EEA as well as a partner in several European Topic Centres and it is the National Reference Centre of the EEA in several areas.

The EAA Department for Ecosystem Research & Environmental Information Management deals with long term ecosystem monitoring and collection of data, semantic data management, as well as the coordination of regional research activities and networking in the field of LTER. EAA currently coordinates the H2020 INFRAIA project eLTER (European Long Term Ecosystem Research). Furthermore, it chairs ILTER (International Long Term Ecosystem Research, LTER (Long Term Ecosystem Research) Europe and LTER Austria. EAA has been involved in a large number of pertinent international research and coordination projects. Altogether the agency has conducted almost 50 FP5/6/7 projects, five of which it coordinated.

EAA also provides hands-on experience by running the Austrian LTER Site Zöbelboden which is situated in the National Park Kalkalpen and part of a large forested area.

Role in the project

EAA will lead KOINE's WP5 (Research community engagement and capacity building) and be involved in WP 2, 4 and 6 where it will contribute its expertise regarding semantic integration and setup of community processes. Throughout the project, EAA is an important interface to the pertinent European and international policy frameworks and the eLTER RI as a major usecase for semantics. With the documentation of sites and datasets using DEIMS-SDR the use of semantics will play an important role in WP4.

Key personnel

Dr. Johannes Peterseil [m], PhD in ecology from the University of Vienna with special focus on landscape ecology, head of the department Ecosystem Research & Monitoring. He contributes to a number of European projects dealing with long term ecosystem observation and monitoring on European and national scale as well as management of long term observation data. Leader of the Expert Panel on Information Management of LTER Europe and member of the Information Management committee of ILTER.

Barbara Magagna [f], master degree in landscape planning and in geoinformatics with long experience in the field of GIS, landscape ecology modelling and database management for projects operating at different scales. She was involved in a number of European projects supporting in the field of ontology engineering and process facilitation. She supported the development process of the European ontology for ecological observations SERONTO and the underlying vocabulary EnvThes. Within ENVRI and ENVRIplus she contributed to the establishment of a reference model for environmental RIs as well as the implementation of provenance.

Doron Goldfarb [m], master's Degree in Computer Science from Vienna University of Technology (UT) and spent academic visits at DTU Lyngby/Denmark and University of Texas at Dallas/USA. He currently works at the EAA as software developer and researcher in the field of Semantic technologies, contributing to the EUDAT2020 and ENVRI*plus* projects in this regard. Before joining EAA, he worked at the Austrian National Library as WP1 lead of the EU-FP7 project DM2E and as researcher at the Vienna UT, in the context of Linked Data for the Digital Humanities.

Christoph Wohner [m], master's degree in Geospatial Technologies from the University of Technology Graz and is interested in web development, web mapping, GIS, metadata and how to connect them with environmental research. His work at the Environment Agency Austria includes the development and maintenance of DEIMS-SDR, an Ecological Research Site and Dataset Registry for long term observation and experimentation data and facilities, web semantics and INSPIRE. His language skills include German (mother tongue), English (fluent), French (basic) and Romanian (basic).

Publications

- Wohner C et al. (2019) DEIMS-SDR - a web portal to document research sites and their associated data. Ecological Informatics, forthcoming https://authors.elsevier.com/tracking/article/details.do?aid=920&jid=ECOINF&surname=Wohner
- Goldfarb D, Wohner C (2017) Enriching the EnvThes controlled vocabulary via aggregated semantic concepts. In: 20th EGU General Assembly, EGU2018, Proceedings from the conference held 4-13 April, 2018 in Vienna, Austria, p.19700
- Fiore N, Magagna B, Goldfarb D (2017) EcoPortal: a proposition for a semantic repository dedicated to ecology and biodiversity. In: Algergawy, A., Karam, N., Klan, F., Jonquet, C. (Eds.) Semantics for Biodiversity. Proceedings of the 2nd International Workshop on Semantics

for Biodiversity co-located with 16th International Semantic Web Conference (ISWC 2017). Vol-1933. pp. 6. http://ceur-ws.org/Vol-1933/poster-paper-13.pdf

- Goldfarb D, Le Franc Y (2017) Envhancing the Discoverability and Interoperability of Multi-Disciplinary Semantic Repositories. In: Algergawy, A., Karam, N., Klan, F., Jonquet, C. (Eds.) Semantics for Biodiversity. Proceedings of the 2nd International Workshop on Semantics for Biodiversity co-located with 16th International Semantic Web Conference (ISWC 2017). Vol-1933. pp. 11. http://ceur-ws.org/Vol-1933/paper-7.pdf
- Schentz H, Peterseil J, Bertrand N (2013) EnvThes - interlinked thesaurus for long term ecological research, monitoring, and experiments. Proceedings EnviroInfo 2013: Environmental Informatics and Renewable Energies. Shaker Verlag, Aachen.

Relevant projects

- ENVRIplus – Environmental Research Infrastructures Providing Shared Solutions For Science And Society (2015-2019) clusters research infrastructures (RIs) for Environmental and Earth System sciences, built around the ESFRI roadmap. www.envriplus.eu
- ECOPOTENTIAL – Improving Future Ecosystem Benefits through Earth Observations (2015-2019), is about reaping the benefits of Earth Observation (e.g. Copernicus Programme) for ecosystem management, including large forested areas. www.ecopotential-project.eu
- eLTER – European Long Term Ecosystem Research (2015-2019), is an infrastructure project to advance LTER Europe which contains many forested sites, broad experience as to infrastructure integration and management owing to co-ordinating role. www.lter-europe.net/elter
- Advance_eLTER – Advancing the European Long-Term Ecosystem, critical zone and socio-ecological Research Infrastructure towards ESFRI (2017), awarded to eLTER as an emerging ESFRI RI. The project has produced the eLTER ESFRI 2018 Roadmap application and conducted a large stakeholder event. www.lter-europe.net/elter-esfri/advance-elter

Relevant infrastructure

LTER Europe, ILTER, and eLTER RI – European and global network on Long Term Ecosystem Research, both chaired by EAA which thereby gained significant in-depth experience on large scale infrastructure integration and management (www.lter-europe.net and www.ilter.network). The EAA supports the eLTER RI building a sustainable research infrastructure on ecosystem and socio-ecological research in Europe (coordinated by UFZ Germany).

Universität Bremen (UHB)

The University of Bremen is a public university in Bremen Germany. University of Bremen (UHB) hosts the excellence cluster MARUM (Centre for Marine Environmental Research), a central research facility offering a number of technical and scientific services in the field of ocean research and furthermore developing innovative technology in support for scientific operations. Jointly with the Helmholtz Centre for Polar and Marine Research at Alfred Wegener Institute (AWI), MARUM hosts PANGAEA, an information system and publisher for geoscientific and environmental data which holds mandates from the World Meteorological Organization (WMO) (World Radiation Monitoring Center - WRMC).

Essential services supplied by PANGAEA are data curation, long-term data archiving and data publication. Data curation includes quality control of metadata and the development of ontologies

and vocabularies according to international protocols and standards. Metadata are extensive and each dataset can be cited using a universally unique DOI. A significant portion of published data sets include author links to ORCID. The system is operated in the sense of the Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities which is a follow up to the Budapest Open Access Initiative. PANGAEA hosts more than 380,000 datasets comprising around 13 billion individual measurements which have been collected during more than 240 European as well as international research projects.

The successful cooperation between PANGAEA and the publishing industry along with the correspondent technical implementation enables the cross-referencing of scientific publications and datasets archived as supplements to these publications. PANGAEA is the recommended data repository of several international scientific journals, such as "Scientific Data" by the Nature publishing group. It is also the recommended depository for data supplements to Elsevier publications in three subject areas, namely Agriculture and Biological Sciences, Earth & Planetary Sciences, Environmental Sciences. The "Guidelines for Authors" of about 500 Journals recommend depositing supplementary data at PANGAEA and emphasize the cross-linking application.

PANGAEA is laid out as a permanent facility, guaranteeing the long-term availability and accessibility of archived data and metadata in secure and machine readable formats.

Role in the project

As the WP4 leader, UHB will be coordinating the integration of B2TERM into disciplinary data infrastructures. UHB will provide actual settings in which KOINE's B2TERM will be applied to improve the curation and discovery of PANGAEA datasets. Through a pilot application, UHB will also demonstrate the value of B2TERM in facilitating data exploration analysis. UHB will also contribute to the development of semantic model and metadata annotation strategies in WP2, user validation of KOINE products in WP5 and finally outreach activities in WP7.

Key personnel

Dr. Michael Diepenbroek [m] (http://orcid.org/0000-0003-3096-6829) is a Geologist and Computer Scientist holding a PhD in Geology from the Free University of Berlin. He conceived and implemented the scientific information system PANGAEA. He is working at MARUM, where he is responsible for the operation of PANGAEA and he is the director of the ISCU World Data Center for Marine Environmental Sciences (WDC-MARE) and took a leading role in its initiation in 2001. He is a member of the Strategic Committee, later Scientific Committee of the ICSU World Data System.

Dr. Robert Huber [m] (http://orcid.org/0000-0003-3000-0020) is a Geologist and Information Specialist holding a PhD in Marine Geology. He worked several years as information system architect for the aerospace industry and the renewable energy industry. Since 2002 he is employed at the Centre for Marine Environmental Sciences (MARUM) at the University Bremen and responsible for projects in scientific data management and IT development at the PANGAEA working group. He was/is leading the data management work package for ESONET, FixO3 and COOPEUS and was/is involved in the related EU projects EMSODEV, HYPOX, SIOS and ENVRI and ENVRIplus.

Dr. Anusuriya Devaraju [f] (http://www.anusuriya.com, https://orcid.org/0000-0003-0870-3192) is a data scientist holding a PhD in Geoinformatics from the University of Muenster, with a specialization in semantic integration of geospatial information. She has a Master's degree in System Design for Internet Applications from the Newcastle University, UK. Currently, she is

working at the Centre for Marine Environmental Sciences (MARUM), University Bremen, where she is responsible for developing and implementing techniques based on data mining and machine learning to improve the discovery of PANGAEA datasets. Prior joining the research center, she worked at Commonwealth Scientific and Industrial Research Organisation (CSIRO) Australia and developed information models and tools to improve the discovery of CSIRO research assets. During her postdoctoral time at Forschungszentrum Jülich, she contributed significantly to TERENO data management and integration.

Publications

- Diepenbroek M, Schindler U, Huber R, Pesant S, Stocker M, Felden J, Buss M, Weinrebe M (2017) Terminology supported archiving and publication of environmental science data in PANGAEA, Journal of Biotechnology, Vol. 261, 2017, pp. 177-186, https://doi.org/10.1016/j.jbiotec.2017.07.016
- Klump J, Huber R (2017) 20 Years of Persistent Identifiers – Which Systems are Here to Stay?. Data Science Journal. 16(9), pp. 1-7, https://doi.org/10.5334/dsj-2017-009
- Klump J, Huber R, Diepenbroek M (2016) DOI for geoscience data – how early practices shape present perceptions, Earth Science Informatics, Vol 9(1), pp.123-136, https://doi.org/10.1007/s12145-015-0231-5
- Diepenbroek M, Grobe H, Reinke M, Schindler U, Schlitzer R, Sieger R, Wefer G (2002) PANGAEA-an information system for environmental sciences, Computers & Geosciences, 28 (10), pp. 1201-1210, https://doi.org/10.1016/S0098-3004(02)00039-0

Relevant projects

UHB is, or was involved in, a variety of EU projects such as THOR, FREYA, European Seas Observatory NETwork (ESONET), FixO3, COOPEUS and COOP+ (Cooperation EU/US), EMSO, HYPOX, the Svalbard Integrated Arctic Earth Observing System (SIOS) and Common Operations of Environmental Research infrastructures (ENVRI and ENVRIplus) focusing on research infrastructures and associated data management. Further, PANGAEA serves as long-term data archive for data which was collected during many European projects, such as the Atlantic Data Base for Exchange Processes at the Deep Sea Floor (ADEPD), Assessment of the Black Sea Sedimentary System since the last Glacial Extreme (ASSEMBLAGE), Biogas Transfer in Estuaries (BIOGEST), CARBOOCEAN, CoralFISH, DARCLIFE, EUR-OCEANS, EURO-BASIN, EPOCA, ERA-CLIM, ESONET, ESOP, HERMES, HERMIONE, HYPOC, MATER, METROL, ORFOIS, PROMESS, SINOPS, etc.

Relevant infrastructure

Hosted by MARUM, PANGAEA is the designated archive for the journal Earth System Science Data (ESSD) and recommended data repository of several international scientific journals such as "Scientific Data" by the Nature publishing group. Furthermore, PANGAEA co-chairs the ICSU World Data System (WDS) and Research Data Alliance (RDA) working groups on data publication and holds mandates from the WMO (World Radiation Monitoring Centre - WRMC). Essential services supplied by PANGAEA are data curation, long-term data archiving and data publication. Data curation includes quality control of metadata and the development of ontologies and vocabularies according to international protocols and standards. Metadata are extensive and each dataset can be cited using a universally unique Digital Object Identifier (DOI). The system is operated in the sense of the Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities which is a follow up to the Budapest Open Access Initiative.

Alfred-Wegener-Institut, Helmholtz-Zentrum für Polar- und Meeresforschung (AWI)

The Alfred-Wegener-Institute Helmholtz Centre for Polar and Marine Research is one of the national research centres of the Helmholtz Society in Germany. It has an annual budget of more than 150 million Euros and a staff of more than 1200 employees. The Institute is primarily active in the cold and temperate regions of the world. Working together with numerous national and international partners, we are actively involved in unravelling the complex processes at work in the Earth System. The institute, founded in 1980, is financed by the German Federal Ministry for Education, Science, Research and Technology (BMBF) and several German states (Länder). Its official mandate requires that the institute 1) participates in fundamental scientific research in marine regions, especially in the poles, 2) provides nation-wide co-operation of marine expeditions and related logistics and 3) contributes to international co-operative projects in polar and regional seas. AWI is embedded in a number of European Research initiatives, such as the European Polar Board (EPB), the European Marine Board (EMB), and the European Climate Research Alliance (ECRA). The institute has three departments: Geosciences, Biosciences and Climate Sciences. A data processing center, library and three research vessels support these departments.

Role in the project

AWI will lead KOINE's WP6 (Operations and Sustainability) and contribute to WP2 (FAIR terminologies: Development and Adoption of the Terminology Interoperability Framework) and WP5 (Research community engagement and capacity building). In WP6 the AWI partner will leverage experience from its contributions to the governance of the Open Biological and Biomedical Ontology (OBO) Foundry and Library (http://obofoundry.org). In WP2, they will extend a set of internationally adopted terminology resources which have been maintained by its personnel for ~8 years, in order to align them to KOINE service and Terminology Interoperability Framework. In WP5, AWI will leverage its involvement in multidisciplinary consortia in the information sciences to engage and align international activity to the KOINE vision.

Key personnel

Dr Pier Luigi Buttigieg [m], PhD, is an AWI data scientist with expertise in ontology development, bio/ecoinformatics, and marine microbial ecology. He is a member of the OBO Foundry Operations Committee, and leads/co-leads several reference ontology projects including the UN Sustainable Development Goals Interface Ontology (SDGIO, a collaboration with UN Environment) and the Environment Ontology (ENVO). He applies semantic technologies to support both the Essential Ocean Variable (EOVs) and the Essential Biodiversity Variables (EBVs). Further, he leads the development of the Microbial EOV as a member of the UNESCO/IOC Global Ocean Observation System (GOOS) Biology and Ecosystems Panel. In KOINE, he will a) propagate best practices in coordinating and sustaining terminology services from the OBO Foundry (WP6), b) contribute to the development of terminology solutions for the EOVs and EBVs, in close association with GOOS and the lead EBV developers (WP2), and c) engage the consortia and networks he contributes to in order to propagate the vision and outcomes of KOINE in the context of EOSC. He has previously worked on several EU projects in marine observation (e.g. MicroB3, DARCLIFE, AltantOS) and is active in international communities such as the Earth System Information Partnership (ESIP).

Publications

- Buttigieg PL, Pafilis E, Lewis SE, Schildhauer MP, Walls RL, Mungall CJ (2016) The environment ontology in 2016: bridging domains with increased scope, semantic density, and interoperation. J Biomed Semant 7, 57, https://doi.org/10.1186/s13326-016-0097-6
- Hardisty AR, Michener WK, Agosti D, García EA, Bastin L, Belbin L, et al. (2018) The Bari Manifesto: An interoperability framework for essential biodiversity variables. Ecol. Inform. https://doi.org/10.1016/j.ecoinf.2018.11.003
- Buttigieg PL, Fadeev E, Bienhold C, Hehemann L, Offre P, Boetius A (2018) Marine microbes in 4D — using time series observation to assess the dynamics of the ocean microbiome and its links to ocean health. Curr. Opin. Microbiol. 43, https://doi.org/10.1016/j.mib.2018.01.015
- Walls RL, Deck J, Guralnick R, Baskauf S, Beaman R, Blum S, et al. (2014) Semantics in Support of Biodiversity Knowledge Discovery: An Introduction to the Biological Collections Ontology and Related Ontologies. PLoS One 9, e89606, https://doi.org/10.1371/journal.pone.0089606

Relevant projects

- The UN Sustainable Development Goals Interface Ontology (SDGIO): Launched in 2014 in collaboration with UN Environment, this ontology seeks to link disciplinary activities and data streams to the UN Sustainable Development Agenda for 2030. The ontology has been deployed in UN Data Systems and is fully interoperable with the Open Biological Ontology Foundry and Library (OBO) ontologies used in KOINE. SDGIO is being included in the UN Semantic Interoperability Framework (UNSIF). This grants KOINE an opportunity to link EOSC data with UN initiatives and data stores.
- The UNESCO/IOC Ocean Best Practices System: Launched in 2017 as part of the H2020 AtlantOS project, the OBPS (www.oceanbestpractices.org) serves the entire Ocean Observing community (including stakeholders such as the World Meteorological Organization as well as consortia running national and international infrastructures). A set of ocean-focused ontologies and vocabularies are used to annotate methodological documents deposited in this system, and create a basis for advanced semantic search and discovery. The approaches used in this system to link information to semantic descriptors is highly relevant to KOINE's mission.
- The Food Ontology (FoodOn): Launched in 2017 and still actively supported (led by the British Columbia Centre for Disease Control), FoodOn provides a semantic interface for the LanguaL.org food indexing thesaurus, integrating it with OBO Foundry ontology terms from UBERON, ChEBI, and NCBITaxon and others. This provides a platform to interoperably link global food data stores across regional and national boundaries and is interoperable with the OBO resources used in KOINE.
- The Agronomy Ontology (AgrO): Deployed in 2017 and still actively supported, AgrO is an ontology of agronomic practices, equipment, and is led by the CGIAR (www.cgiar.org) as part of its mission to address global poverty, hunger, and environmental degradation. It is part of the Big Data for Agriculture programme and is fully interoperable with the OBO resources used in KOINE.
- The Environment Ontology (EnvO): Launched in 2013 and actively maintained, EnvO has become a widely-used and supported ontology for Earth and Environmental Science in its broadest sense, providing semantics for numerous projects in multiple fields including the Earth, life, medical, and development sciences. EnvO will be deployed within KOINE, as it has been selected by the Global Ocean Observing System as the focal point of Essential Ocean Variable development, and also serves the Essential Biodiversity Variable community.

Relevant infrastructure

The AWI hosts an extensive IT infrastructure including supercomputing facilities, sensor and data management systems, data portals, and close organisational association with long-term information archives such as the Earth science information publisher, PANGAEA. This infrastructure can both facilitate development/testing of the terminology and semantic interoperability services developed in KOINE, and greatly benefit from KOINE's outcomes. In particular, the institute's growing SensorWeb is well-suited to interface with KOINE's outcomes, which can FAIRly link sensor parameter metadata from the hundreds of sensors depoloyed in the field and on vessels (such as the German icebreaker RV POLARSTERN) to other EOSC services. The AWI's flagship, multidisciplinary observatories such as the Long-Tern Ecological Research Station (LTER) HAUSGARTEN and programmes such as FRAM (FRontiers in Arctic marine Monitoring) will provide multiple use cases to KOINE, aligning its outputs to operational scenarios featuring year-round surface-to-seafloor ocean observation.

NILU - Norsk institutt for luftforskning (NILU)

NILU was founded in 1969 and is an independent non-profit research institute specializing in climate and air pollution research with ca 180 employees. In the last years, NILU has been involved in more than 90 EU and ESA financed projects e.g. ACTRIS-1, ACTRIS-2, ACTRIS-PPP, ENVRI-FAIR, ENVRIplus, NextGEOSS, InGOS, MACC2, GEOMON, EUCAARI, EUSAAR, ACCENT, EARLINET-ASOS, SCOUTO3, MEGAPOLI and also ESFRI initiatives e.g. ICOS and SIOS, and ESA-CCIs. NILU is responsible for the ACTRIS Data Centre since start, and for data curation of all in situ data within ACTRIS. NILU is responsible for the widely used database infrastructure EBAS (http://ebas.nilu.no) hosting observation data of atmospheric chemical composition and physical properties. EBAS is used for ACTRIS in situ data. EBAS is nominated as a DCPC in the WMO-WIS system, and NILU has been involved in the WMO-WIGOS process. The institute is involved in developing INSPIRE guidelines, E-reporting and GEOMS (Generic Earth Observation Metadata Standard). NILU plays central roles in the WMO – GAW program, and is represented in 3 of the Scientific Advisory Groups (on Aerosols, Reactive Gases and Total Deposition respectively) as well as hosting the World Data Centre for Aerosols (GAW-WDCA) and the World Data Centre for Reactive Gases (WDCRG). NILU serves as the EMEP Chemical Coordination Centre (EMEP-CCC). The EMEP program comprise more than 40 Parties and forms the basis for UNECE CLRTAP abatement policies, as well as for the EU policies on Air Quality (www.emep.int). NILU is a member to the European Topic Centre for Air Pollution and Climate Change Mitigation (ETC/ACM). NILU hosts also the official databases of the Arctic Monitoring and Assessment Programme (AMAP) and the OsloParis Commission (OSPAR CAMP). NILU also hosts the Earth Observation Validation Data Centre (EVDC), a database developed and operated on contract by ESA. Furthermore, NILU operates state of the art observation facilities of atmospheric trace constituents in Norway and Polar Regions: the observatories Zeppelin at Svalbard, Birkenes in southern Norway, and the Troll station in Antarctica. All sites contribute to ACTRIS, Zeppelin and Birkenes contribute to ICOS, while Zeppelin is also contributing to SIOS.

Role in the project

NILU leads the ACTRIS data centre (DC), and runs the ACTRIS data portal in addition to the IN SITU topical DC unit. In KOINE, NILU will coordinate the ACTRIS DC units in collecting existing topical vocabulary and systemise it as a contribution to specifying B2TERM. In KOINE WP4, NILU will implement the B2TERM vocabulary in its portal. As ACTRIS DC co-ordinator, NILU will also work with the other ACTRIS DC partners in disseminating B2TERM vocabulary to further communities.

Key personnel

Dr. Cathrine Lund Myhre [f] is a senior scientist at NILU and has a PhD in spectroscopy and about 20 years of experience in understanding of atmospheric composition change, including data management and data quality requirements. Lund Myhre is author or co-author of more ca 40 papers in peer reviewed literature with more than 2300 citations, and contributed to more than 40 scientific reports including acting as contributing author to IPCC 5thAR, WG I, Chapter 2 (*Observations: Atmosphere and Surface).* She is an experienced project leader e.g. a number of Norwegian Research Council projects, the national monitoring of greenhouse gases, and leads the ACTRIS Data Center in ACTRIS-1 and ACTRIS-2. She will be highly involved in the new EU-project ENVRI-FAIR, and lead WP8: Implementation of Atmospheric subdomain.

Dr. Markus Fiebig [m] received his PhD in meteorology from the University of Munich, Germany, in 2001. He is a senior scientist at NILU and manages the GAW World Data Centre for Aerosol. In administering the underlying database infrastructure, he works on cross domain data archive interoperability and user interaction within the WMO Global Atmosphere Watch programme, and ENVRIplus, and the Norwegian Scientific Data Network (NorDataNet). He has authored or co-authored 40 peer reviewed publications and is and has been coordinating tasks and work packages in several international projects (ACTRIS2, ACTRIS1, EUSAAR, ENVRIplus). He will be task leader in ENVRI-FAIR, WP8.

Richard O Rud [m] is a system developer working at NILU with a background in informatics and industrial ecology. He is mainly working as a system developer on tools for data management and curation as well as machine-to-machine interfaces for data and metadata interoperability. In addition he works with data curation for the EBAS data center. At NILU he has worked on projects including ACTRIS-2, ENVRIplus, NextGEOSS and the EVDC project. He will contribute to ENVRI-FAIR, WP8.

Publications

- Dalsøren SD, Myhre G, Hodnebrog Ø, Lund Myhre C, Stohl A, Pisso I, Schwietzke S, Höglund-Isaksson L, Helmig D, Reimann S, Sauvage S, Schmidbauer N, Read KA, Carpenter LJ, Lewis AC, Punjabi S, Wallasch M (2018) Discrepancy between simulated and observed ethane and propane levels explained by underestimated fossil emissions, Nature Geoscience, v 11, 178–184, https://doi.org/10.1038/s41561-018-0073-0
- Myhre G, Aas W, Cherian R, Collins W, Faluvegi G, Flanner M, Forster P, Hodnebrog Ø, Klimont Z, Mülmenstädt J, Myhre CL, Olivié D, Prather M, Quaas J, Samset BH, Schnell JL, Schulz M, Shindell D, Skeie RB, Takemura T, Tsyro, S (2017) Multi-model simulations of aerosol and ozone radiative forcing for the period 1990–2015, Atmos. Chem. Phys.,17, 2709-2720, https://doi.org/10.5194/acp-17-2709-2017
- Myhre C et al. Large methane release from the Arctic seabed west of Svalbard, but small release to the atmosphere. Geophys. Res. Lett., 43, https://doi.org/10.1002/2016GL068999
- Collaud Coen M et al. (2013) Aerosol decadal trends – Part 1: In-situ optical measurements at

GAW and IMPROVE stations, Atmos. Chem. Phys., 13, 869-894, https://doi.org/10.5194/acp-13-869-2013

- Schmale J, et al. (2017) Collocated observations of cloud condensation nuclei number concentrations, particle number size distributions, and chemical composition, Scientific Data 4, Article number: 170003, https://doi.org/10.1038/sdata.2017.3

Relevant projects

- Participation in the H2020 project ACTRIS 2 (Aerosol, Clouds, and Trace gases Research Infrastructure) -Grant agreement n. 654109 (2015-2019).
- Participation in the H2020 project ENVRIPLUS (Environmental Research Infrastructures Providing Shared Solutions for Science and Society) - Grant agreement n. 654182 (2015- 2019).
- Participation in the H2020 ACTRIS Preparatory Phase Project – Grant agreement n. 739530 (2017-2019).
- Participation in the H2020 project ENVRIFAIR (ENVironmental Research Infrastructures building Fair services Accessible for society, Innovation and Research) - Grant agreement n. 824068 (2019 -2022).

Relevant infrastructure

- ACTRIS (Aerosol, Clouds, and Trace gases Research Infrastructure) www.actris.eu, Co-ordinating data centre, hosting ACTRIS data portal and IN SITU topical data centre unit.
- EBAS, data archive infrastructure for observations related to atmospheric composition from surface in situ stations, serving frameworks such as EMEP, GAW, and AMAP in addition to ACTRI.

Universiteit van Amsterdam (UvA)

UvA is the largest university in the Netherlands and a core member of the League of European Research Universities (LERU). The Faculty of Science at UvA embeds a concentration of e-Science development and application groups, benefiting from its central position in Amsterdam Science Park. Science Park holds the main European Internet Exchange (AMS-IX), one of the SURFnet core locations (which includes the optical network exchange NetherLight), the national supercomputer and Grid centre (Surf-SARA) and a number of Dutch Science Organisation institutes. The university contributes to innovation in computer science with its Institute of Informatics and Institute for High Performance Computing and Advanced Network. These institutes collaborate with the national BigGrid initiative with respect to advancing the state-of-the-art for advanced virtual infrastructures. UvA has extensive experience in participating in EU projects. The System and Network Engineering research group in UvA has a special research interest in software tools for quality-critical applications on Clouds (H2020 SWITCH), programmable infrastructure (FP7 GEYSES), advanced networks (GigaPort), software-defined networking (FP7 NOVI), the semantic model for describing network and virtual infrastructure (CineGrid, FP7 NOVI) and semantics-driven optimisation in research data infrastructures (FP7 ENVRI, H2020 ENVRIplus, H2020 VRE4EIC, H2020 ENVRI FAIR). UvA will contribute this expertise to the design and development of FAIR vocabulary services for EOSC, along with the development of advanced supporting knowledge infrastructure.

Role in the project
UvA is responsible for the technical implementation of the B2TERM service, and contributes to other activities in WP3, as well as having a role in the API specification activity in WP2.
Key personnel
Dr. Zhiming Zhao [m] obtained his Ph.D in computer science in 2004 from University of Amsterdam, and he is currently a senior researcher on advanced network, Software Defined Networking, distributed workflow systems, and big data infrastructure. He was the scientific coordinator of the H2020 SWITCH project and the leader of the 'Data for Science' theme within the H2020 ENVRIplus project, and leads the main technical coordination activity in H2020 ENVRI FAIR. He participated in the H2020 VRE4EIC, and FP7 ENVRI and GEYSES projects. **Dr. Paul Martin [m]** obtained his PhD in Informatics in 2011 from the University of Edinburgh with an interest in semantic modelling, distributed artificial intelligence and argumentation. After working within the Data Intensive Research group at the University of Edinburgh for four years, he joined the System and Network Engineering group at the University of Amsterdam in 2015. Since 2011 he has worked in a number of software and infrastructure-related EU projects: the FP7 projects ADMIRE, ENVRI and VERCE, and the H2020 projects SWITCH, ENVRIplus and VRE4EIC.
Publications
• Martin P et al. (2017) Computational Challenges in Global Environmental Research Infrastructures. In Chabbi, A. and Loescher, H. W., editors, Terrestrial Ecosystem Research Infrastructures: Challenges and Opportunities, chapter 12, pages 305–340. CRC Press. • Wang J et al. (2017) Planning virtual infrastructures for time critical applications with multiple deadline constraints. Future Generation Computer Systems. https://doi.org/10.1016/j.future.2017.02.001 • Koulouzis S et al. (2017) Seamless Infrastructure Customisation and Performance Optimisation for Time-critical Services in Data Infrastructures, in proceedings of the Eighth International Workshop on Data-Intensive Computing in the Clouds, in the context of Supercomputing, Denver, U.S., ACM SIGHPC. • Mork R, Martin P, Zhao Z (2015) Contemporary Challenges for Data-intensive Scientific Workflow Management Systems, International workshop on Workflows in support of large-scale science (WORKS 15), in the context of IEEE Supercomputing 2015. https://doi.org/10.1145/2822332.2822336 • Zhao et al. (2015) Reference Model Guided System Design and Implementation for Interoperable Environmental Research Infrastructures, In the proceedings of IEEE eScience Munich Germany, p583- 588. https://doi.org/10.1109/eScience.2015.41
Relevant projects
EU H2020 ENVRI-FAIR, http://envri.eu/envri-fair/, 2019–present. Implementation of FAIR services for environmental research infrastructures in Europe. UvA coordinates the common implementation and support work package, providing development expertise to the different environmental sub-domains in the project. EU H2020 ENVRIplus, http://www.envriplus.eu, 2015–present. Implementation of common solutions for a cluster of ESFRI infrastructures in the field of environmental science. UvA coordinates the 'Data for Science' theme within the project and leads development of semantic linking and optimisation solutions for research infrastructures. This project builds upon EU FP7 ENVRI, www.envri.eu, 2011–2014, which UvA coordinated. EU H2020 SWITCH, http://www.switch-project.eu, 2015–2018. A software workbench for time-

critical, self-adaptive cloud applications. UvA coordinates the project and leads development of the real-time infrastructure planning and provisioning component.
EU H2020 VRE4EIC, http://www.vre4eic.eu, 2015–2018. Development of a European interoperable virtual research environment for multidisciplinary research and innovation. UvA contributes to the requirements analysis and dissemination via its links to other projects such as ENVRIplus.

Relevant infrastructure

OpenLab and ExoGENI. OpenLab is a facility that provides an open experimentation environment for Software Defined Networking (SDN) research, hosted by the SNE group at the University of Amsterdam. The OpenLab currently consists of an ExoGENI rack, a cluster with 1 head node, 8 worker nodes, and 1 storage node with 6TB of storage space. It also has 700GB of fast Flash SSD storage space. All the machines connected using 10G Ethernet infrastructure with both traditional and OpenFlow links.

Maastricht University (UMAAS)

Maastricht University is a top-ranked international university, renowned for the quality of the research and teaching. Maastricht University hosts 16.800 students in international BA, MA and graduate programs, and 3.500 (FTE) employees. The Institute of Data Science (IDS) at Maastricht University is an interfaculty program to foster collaboration for inter- and multidisciplinary research and training founded on accurate, reproducible, multiscale, distributed, and efficient computation.

Role in the project

In WP2, UMAAS will work with the consortium to build FAIR terminologies with harmonized metadata (T2.2), and lend its expertise to FAIRness assessment of relevant terminologies (T2.1); In WP5, it will co-lead a hands on workshop to undertake FAIRness assessment of annotated data (T5.3), provide input to community requirements (T5.2), participate in the validation of KOINE services (T5.3), and contribute to capacity building and training (T5.4). In WP6, it will participate in the establishment of a KOINE Operations Committee, with a focus on standard operating procedures in relation to FAIR and in coordination with GO-FAIR.

Key personnel

Prof. Dr. Michel Dumontier [m] is a Distinguished Professor of Data Science at Maastricht University, and a co-founder of the FAIR principles. His research focuses on the development and application of computational methods for scalable and responsible data science. Formerly a professor in Biomedical Informatics at Stanford University (2013-2016), and at Carleton University (2006-2013), he obtained his Bachelor of Science in Biochemistry at the University of Manitoba (1999) and his PhD in Bioinformatics at the University of Toronto (2004). Dr. Dumontier is now one of four Distinguished Professors at Maastricht University, and one of two dozen in the Netherlands. His group combines semantic web technologies, such as ontologies and Linked Data, with machine learning and network analysis to tackle problems in the areas of computational drug discovery and precision medicine.

Publications

- Wilkinson M, Dumontier M, et al, and B Mons (2016) The FAIR Guiding Principles for scientific data management and stewardship. Nature Scientific Data, 3:160018, https://doi.org/10.1038/sdata.2016.18
- Wilkinson MD, Sansone SA, Schultes E, Doorn P, Bonino da Silva Santos LO, Dumontier M. (2018) A design framework and exemplar metrics for FAIRness. Sci Data, 26;5:180118, https://doi.org/10.1038/sdata.2018.118
- Wilkinson MD, et al, Dumontier M (2017) Interoperability and FAIRness through a novel combination of Web technologies. PeerJ Computer Science, 3:e110, https://doi.org/10.7287/peerj.preprints.2522v2
- van Soest J, Sun C, Mussmann O, Puts M, van den Berg B, Malic A, van Oppen C, Towend D, Dekker A, Dumontier M (2018) Using the Personal Health Train for Automated and Privacy-Preserving Analytics on Vertically Partitioned Data. Stud Health Technol Inform. 247:581-585, http://www.persistent-identifier.nl/urn:nbn:nl:ui:27-fc553d50-254f-46b1-b09f-0e0978addba
- MA Musen, C Bean, KH Cheung, M Dumontier, KA Durante, O Gevaert, A Gonzalez-Beltran, P Khatri, SH Kleinstein, MJ O'Connor, Y Pouliot, P Rocca-Serra, S-A Sansone, JA Wiser (2015) The center for expanded data annotation and retrieval. J Am Med Inform Assoc. 22(6):1148-52.

Relevant projects

- **EOSC-Life**: Providing an open collaborative space for digital biology in Europe (Horizon 2020), 2019-2022. **Bio2Vec**: Smart Analytics Infrastructure for the Life Sciences (KAUST), 2018-2021, http://bio2vec.net
- **Biomedical Data Translator** (National Institutes of Health), 2016-2019, https://ncats.nih.gov/translator
- **Data Commons Pilot Phase Consortium** (National Institutes of Health), 2017-2021, https://nihdatacommons.us/
- **BioCADDIE**: Biomedical and Healthcare Data Discovery Index Ecosystem (NIH), 2016-2017. https://datamed.org/
- **CEDAR**: Center for Expanded Data Annotation and Retrieval, NIH, 2014-2018, https://metadatacenter.org/

Relevant infrastructure

A dedicated, cloud-based research data storage is available and supported 24/7 by the ICT department using an industry standard service level. IDS has access to the Surf Sara Life Science Grid which consists of 5600 Xeon 2.2-2.6GHz cores running Linux CentOS 6.x 64bit with a total of 41 TB of memory, and a network backbone of 160Gbit/s to the Grid storage facility with a disk storage capacity of 8000 TB and 18000 TB of stored data.

Norwegian Meteorological Institute (MET)

The Norwegian Meteorological Institute (METNO) is the national meteorological service in Norway. METNO represents Norway in ECMWF, EUMETSAT, EUMETNET, WMO and other international forums, and takes part in international projects funded by EU and other bodies on climate, atmospheric and marine research including the application of remote sensing

techniques, and air pollution research. METNO employs about 400 persons, among them 80 scientists doing research within numerical weather prediction, ocean modelling, remote sensing, air pollution, product development, instrumentation, climatology and climate and data management. METNO has extensive experience in developing methods and operational applications that have led to innovation and value addition in both the private and public sector. METNO operates an extensive operational meteorological observation network that covers parts of the Arctic, including manned stations on the islands Jan Mayen, Bjørnøya and Hopen, as well as a network on the Svalbard archipelago. METNO supports and implements a free and open data policy. Official data and products from the institute are regarded as public sector information and are freely available to the public for use, distribution and processing. Channels for public distribution of weather forecasts include the website and the mobile application yr.no (with 5-10 million unique users weekly – the largest weather website in the world outside the US) as well as various lower level download services. METNO is connected to WMO data exchange, both through WMO Information System and through the Global Telecommunication System. Data management is based on a metadata driven approach where datasets are documented at both discovery and use level including the interfaces for accessing them. In addition to close linkages to national High Performance Computing and Storage systems, METNO has also local HPC and HPS systems that scales for future needs. Within these systems central data management nodes of Svalbard Integrated Arctic Earth Observation System (SIOS), Norwegian Scientific Data network (NordataNet), WMO Global Cryosphere Watch (GCW) and WMO Year of Polar Prediction (YOPP) are operated.

Role in the project

METNO leads the implementation of the SIOS Data Management System and the Norwegian Scientific Data Network. As a partner METNO will contribute to WPs 4, 5 and 6, working in the areas of terminology mapping and implementation of mappings in discovery interfaces as well as evaluation of B2TERM. METNO is further represented in WP5 as one of the partners of the ACTRIS infrastructure and will perform testing and evaluation of B2TERM within the AeroCom model evaluation database that is hosted by METNO.

Key personnel

Dr. Øystein Godøy [m] is Head of the division for Remote Sensing and Data Management within the Research and Development Department. He is a climate and remote sensing scientist with extensive experience in distributed data management. He has been part of or led design and implementation of data management in a number of projects funded by EU, EUMETSAT, ESA and RCN, is active in WMO data management efforts (WIGOS and WIS), is member of the joint SAON/IASC Committee on Information and Data Services representing WMO and of the GCW Steering Group and was active in the data management of the International Polar Year. Currently his main data management activities include development of the Norwegian Scientific Data Network (project leader), GCW Information System (leading development and implementation), YOPP Data Management (leading development and implementation) and SIOS Data Management System (Data manager and leading implementation). He is also co-charing the Polar Semantics Working Group, a joint working group between SAON/IASC Arctic Data Committee and IARPC in the US.

Dr. Lara Ferrighi [f] is a research scientist within the Remote sensing and data management department. She holds a PhD in Theoretical chemistry and has academic experience within Computational Solid States and Material Sciences. She is now involved in the development and implementations of tools and services for users on several web portals, including Norwegian Scientific Data Network, Svalbard Integrated Arctic Earth Observing System, National Ground Segment for Satellite Data and Global Cryosphere Watch.

Prof. Dr. Michael Schulz [m]: Deputy Head of MetNo's Climate Modelling and Air Pollution Section with 25 researchers, as well as Professor II (20%) at University of Oslo. He assumes responsibility for atmospheric chemistry, aerosols, climate and earth system modelling at MetNo. Earlier career achieving PhD at university of Hamburg (1993), Group leader at LSCE, France, as guest scientist at NASA Goddard, and supervising more than 20 graduate and PhD students. He is the leader of AeroCom aerosol since 2002, with internationally leading contributions to the understanding of aerosol-climate interactions. PI in multiple EU and national projects in Germany, France and Norway. Lead and Contributing Author of IPCC AR4 and AR5. Now co-chair of AerChemMIP contributing to CMIP6. Co-coordinator of the large Norwegian Infrastructure project INES supporting the NorESM climate model development, starting in July 2018. *Peer-reviewed publications: >120; H-index: 56 ; citations: 13000; ranked as Highly Cited Researcher by Thomson Reuters since 2014; ORCID: 0000-0003-4493-4158*

Dr. Jonas Gliß [m] is an early-career postdoc within the Climate Modelling and Air Pollution Section and is specialised in the evaluation of climate and aerosol models against observations. He is lead developer of the pyaerocom software (https://pyaerocom.met.no/) for model evaluation and has extensive experience in mapping of different metadata vocabularies from various databases and research infrastructures.

Publications

- Godøy Ø, Saadatnejad B (2017) ACCESS climate data management, AMBIO, https://doi.org/10.1007/s13280-017-0963-1
- Koffi B, Schulz M, Breon FM, Dentener F, Steensen BM, Griesfeller J, Winker D, Balkanski Y, Bauer, SE, Bellouin N, Berntsen T, Bian HS, Chin M, Diehl T, Easter R, Ghan S, Hauglustaine, DA, Iversen T, Kirkevag A, Liu XH, Lohmann U, Myhre G, Rasch P, Seland O, Skeie RB, Steenrod SD, Stier P, Tackett J, Takemura T, Tsigaridis K, Vuolo MR, Yoon J, Zhang K (2016) Evaluation of the aerosol vertical distribution in global aerosol models through comparison against CALIOP measurements: AeroCom phase II results. Journal of Geophysical Research-Atmospheres 2016, 121 (12), 7254-7283, https://doi.org/10.1002/2015JD024639
- Gliß J, Bobrowski N, Vogel L, Pöhler D, Platt U (2015) OClO and BrO observations in the volcanic plume of Mt. Etna - implications on the chemistry of chlorine and bromine species in volcanic plumes. Atmospheric Chemistry and Physics, https://10.5194/acp-15-5659-2015
- Pulsifer PL, Yarmey L, Godøy Ø, Friddell J, Parsons M, Vincent WV, de Bruin T, Manley W, Gaylord A, Hayes A, Nickels S, Tweedie C, Larsen JR, Huck J (2014) Towards an International Polar Data Coordination Network. Data Science Journal. 13, pp.PDA94–PDA102, http://doi.org/10.2481/dsj.IFPDA-16
- Parsons MA, Godøy Ø, LeDrew E, de Bruin TF, Danis B, Tomlinson S, Carlson D (2011) A conceptual framework for managing very diverse data for complex, interdisciplinary science, Journal of Information Science, 2011, pp. 1–15 http://jis.sagepub.com/content/early/2011/10/20/0165551511412705

Relevant projects

- Coordination of data management in Svalbard Integrated Arctic Earth Observing System (SIOS), https://www.sios-svalbard.org/
- Participation as work package leader for data management in H2020 project APPLICATE (Advanced predictions in polar regions and beyond)
- Participation as work package leader for data management in H2020 project INTERACT (International Network for Terrestrial Research and Monitoring in the Arctic)
- Participation as work package leader for data management in EU FP 7 project ACCESS (Arctic

Climate Change, Economy and Society)
- Participation as work package leader for data management in EU FP 6 project DAMOCLES (Developing Arctic Modeling and Observing Capabilities for Long- term Environmental Studies)
- Aerosol Model Comparison Project (AeroCom): coordination, since 2003 (http://aerocom.met.no/)
- Participation in ACTRIS-2 Infrastructure Activities in H2020 (model evaluation and aerosol trends, https://aerocom-trends.met.no/)

Relevant infrastructure

- Svalbard Integrated Arctic Earth Observing System (SIOS), https://www.sios-svalbard.org/
- WMO Global Cryosphere Watch (GCW) Data and Information System, https://globalcryospherewatch.org
- Aerosol, Clouds and Trace Gases Research Infrastructure (ACTRIS) - https://www.actris.eu/

LifeWatch ERIC (LW ERIC)

LifeWatch ERIC (http://www.lifewatch.eu), the e-Science European Research Infrastructure for Biodiversity and Ecosystem Research, is a distributed Research e-Infrastructure to advance biodiversity research by providing major science-based knowledge in order to address the big environmental challenges, including knowledge-based solutions to environmental managers. This goal is achieved by providing access through a single infrastructure to a multitude of sets of data, services and tools enabling the construction and operation of Virtual Research Environments (VREs), which allow enhanced capture of data with new innovative technologies and knowledge-based decision making-support for the sustainable management of biodiversity and ecosystems. LifeWatch entered into operational phase in 2016, legally becoming an ERIC at the beginning of 2017, which includes six countries as full members: Belgium, Greece, Italy, The Netherlands, Slovenia and Spain, although Portugal is expected to join the ERIC this year.

As an e-Infrastructure of distributed nature, LifeWatch is composed by Common Facilities, located in Spain (Statutory Seat and the ICT e-Infrastructure Technical Offices), Italy (Service Centre) and The Netherlands (Virtual Laboratories and Innovations Centre). The Statutory Seat and the ICT e-Infrastructure Technical Offices jointly assist to the coordination and management of the day-to-day institutional relationships, administrative, legal, and financial issues. The Service Centre provides the frontend interface with the Biodiversity Scientific Community, identifies the needs of the multiple users groups from different domains and areas of interest, and coordinates the development and operation of the relevant Services. Finally, the Virtual Laboratories and Innovations Centre coordinates and manages the requirements and needs analysis, design and implementation of the scientific case studies and production of the LifeWatch Virtual Laboratories. The Distributed Centres, developed in Belgium, Greece, Italy and Slovenia, are responsible for the design, construction and operation of a plethora of datasets and web services, which are essential to the national networks but also to the international scientific community and stakeholders.

LifeWatch ERIC envisions becoming the biggest and more complete research infrastructure for the study of biodiversity and ecosystems in Europe, connecting scientists and citizens to biodiversity issues, and fostering liaisons with other environmental research infrastructures. Likewise, it

pretends to contribute to support training, outreach and environmental awareness programs.
For the KOINE project LifeWatch ERIC will rely, also, on the support of:

- the Applied Software Engineering (ISA) research group at University of Seville (US), member of the JRU LifeWatch Spain. The ISA group is formed by 18 Ph.D researchers, 2 granted PhD students and 3 engineers as technical staff. The research of the group spans several research areas including search-based software engineering, software testing, business process management, service-oriented computing, software engineering and methodologies, software product lines, semantics, and blockchain technologies. In the last five years, the group has participated in 7 research projects (including 2 funded by the EC) and 7 technology transfer contracts, and has stable collaborations with more than 25 international research groups.
- the CNR-IRET pertains to the CNR Department of "Earth System Science and Technology for the Environment". Its scientific mission is to carry out the activities of basic and applied research on the thematics: interactions between organisms and the environment; anthropogenic effects on the ecological balance; biological and evolutionary processes and mechanisms in terrestrial organisms in relation to the environment; and cophysiological mechanisms and productivity of agricultural and forest plants. Within KOINE, CNR-IRET brings its expertise on semantics for biodiversity and ecosystem research, and will be involved in the identification and assessment of terminology resources, parameters harmonization, annotation of meta(data), development of a functional biogeography use case in collaboration with CNRS-CEFE.

Role in the project

LifeWatch ERIC will lead KOINE‘s WP7 (Communication and Dissemination) and be strongly involved in WP2, WP3, WP4 and WP5. LifeWatch ERIC has established a Communication Unit within its Service Centre with expertise in communication and dissemination, their strategic planning and implementation, as well as organisation and promotion of international scientific events. These skills and competence will allow a smooth coordination of KOINE WP7 activities, as well as the direct implementation of Task 7.1 and 7.2. LifeWatch ERIC expertise in developing vocabularies and ontologies in biodiversity and ecosystem domain will contribute to WP2, WP4 and WP5. In particular, we will be involved in the identification, assessment and alignment of terminology resources (WP2); in the application of B2TERM in data integration and analysis through the development of a functional biogeography use case in collaboration with CNRS-CEFE (WP4); and in the testing and validation of KOINE services (WP5). LifeWatch ERIC is also offering to the project its semantic repository focused on ecology and related domains - ECOPORTAL. LifeWatch ERIC will develop a connector to allows B2Term to be interoperable with ECOPORTAL (WP3).

Key personnel

Prof. Dr. Jesús Miguel Santamaría [m] is the interim CEO of LifeWatch ERIC. From 2013 to 2017 he has been the Spanish representative of the “Environment Strategic Working Group” of ESFRI. He leads the Integrated Laboratory for Environmental Quality at the University of Navarra. He has a long and outstanding experience in the research of environmental pollution and its effects on ecosystems. An important part of his investigations are conducted within the framework of the Long Range Transboundary Air Pollution (LRTAP) Convention.

Prof. Alberto Basset [m] is the interim Head of the LifeWatch ERIC Service Centre and the Manager of the JRU LifeWatch-ITA. He is full professor of Ecology at the University of Salento, with main research interests in biodiversity organisation and ecosystem functioning with a particular focus on aquatic ecosystems. Since 2015, he is the pro tempore President of the European Ecological Federation and of the Euro- Mediterranean Federation of research networks on lagoon ecosystems and he is member of the board of editors of different international journals.

Dr. Juan Miguel González Aranda [m] Dr. Eng. Telecommunications and European Doctorate & Master on Industrial Organization & Management. [2012-2018]: Spanish Ministry HoU and Delegate for e—Science & e-Infrastructures in tightly collaboration with European Commission (DG R&I and DG CONNECT): e-IRGwww.e-irg.eu, European Open Science Cloud-(EOSC), EuroHPC Spanish "Sherpa", Group European Experts Data (GEDE) at Research Data Alliance Europa. LifeWatch ERIC Chief Technology Officer CTO www.lifewatch.eu also supporting other Environmental ESFRIs (ACTRIS, DANUBIUS-RI, EMSO, EPOS, ICOS, etc.). Involved at start of KIC EIT; [2004-2012]: Research Technologist Spanish Council for Scientific Research (CSIC)

Dr. Nicola Fiore [m] is the interim ICT Coordinator of the LifeWatch ERIC Service Centre. He is Dr. Eng. Computer Science and PhD in Computer Engineering. He has 17 years of working experience in IT, in the field of the design and development of Information Systems in both the private and public sectors, and 7 years of accredited professional experience in the area of Biodiversity and Ecosystem Bioinformatics research. In particular, his activity involves the definition of common policies, models and e-infrastructure to optimise technological implementation; definition of workflows; and coordination, harmonisation, integration and interoperability of data, applications and other services between the ESFRI and other Research Infrastructure initiatives in the environment thematic area. He is one of the member of the Vocabulary Services Interest Group leading the Ontology Market Place working group in the framework of the Research Data Alliance.

Prof. Dr. Antonio Ruiz-Cortés [m] is full professor at the University of Seville, member of the JRU LifeWatch Spain, and one of the guarantors of the Unit of Excellence of the Institute of Computer Engineering of the University of Seville, of which he is a founding member. He leads the research group "Applied Software Engineering" since its foundation as well as the Spanish Network of Excellence "Services Science and Engineering" since 2015. From September 2018 Antonio Ruiz is the President of the Spanish Society on Software Engineering (SISTEDES). Throughout his 27 years of professional career, the first 7 in the private Company, he has led 13 research projects and published more than 50 papers in high-impact journals, 90% of which are in the top 25% of most cited papers worldwide, in addition to more than 40 contributions to top conferences, 13 supervised theses and international patents.

Dr. José María García [m] is an accredited associate professor at University of Seville, within the JRU LifeWatch Spain and the "Applied Software Engineering" research group. He received a Ph.D. on Computer Science from the US, with a thesis on improving Semantic Web Services discovery and ranking using a lightweight and integrated approach. He has participated in several national and European research projects (e.g., MSEE, SOA4All, SUPPORT, S-Cube, BYTE and PRELIDA), including public-private partnerships. He has published more than 40 articles as conference papers, workshop papers, technical reports and journal articles. He has also co-organised national conferences and has been member of several conference and workshop program committees, as well as reviewer for international journals and funding agencies. His research is currently focused on cloud computing services and the application of semantics and blockchain technologies to different domains, such as service engineering and biodiversity.

Mr. Antonio Jose Sáenz Albanés [m] is the ICT Core e-Infrastructure Operation Coordinator of LifeWatch ERIC. He received a Diploma of Advanced Studies of the Official Postgraduate Program in Engineering and Technology of the Information of the University of Seville. He has 29 years of experience in IT in both the private and public sectors. He was for ten years project manager of the "Centro de Gestión Avanzada" (CGA) for IT-supported educational centres in Andalusia. From its inception and planning to fully operational capabilities (hundreds of thousands of distributed computers, thousands of distributed servers and centres, and more than a million daily users.)

Deploying the complete ITIL services stack, standards-based security operations, large-scale continuous integration and deployment of software, services and hardware, and implementing DevOps operations. He was for ten years CTO of a private company, responsible of all the ICT teams (~250 in-house and world-wide distributed) managing every aspect of production and operations such as recruiting, training, strategic selection of technologies and partners, development and services methodologies, (ISO 9001:2000 using RUP-agile, ITIL, ISO 27000, CMMi) and the management of the projects portfolio. Also he coordinated the R+D+I and defined the strategy and partner selection. He deployed several virtual research environments: nation-wide electric grid planning, improvement and adaptation of the biological station of Doñana to LifeWatch, and scenario-based water innovation and research laboratory (continental waters management). He was also Associated Lecturer in the Languages and Information Systems Department of the University of Seville.

Ms Sara Valeria Montinaro [f] is the Interim Chief Communication Officer of LifeWatch ERIC. Graduated in Communication Studies in 2007, she has a ten year professional experience in the field of communications, in particular digital one, having worked for European Institutions, such as the Committee of the Regions and European Parliament, international organisations and research infrastructures. Since 2014, she has been coordinating LifeWatch Italy communication activities and since 2018 she is in charge to coordinate LifeWatch ERIC communication office and dissemination activities. She has been involved in a number of European projects related to biodiversity and ecosystem research, contributing to the dissemination of their results, such as LLP "The European Scientific Research Game Project", or H2020 ENVRI-FAIR and ECOPOTENTIAL.

Dr. Ilaria Rosati [m] is a research technologist at the Research Institute on Terrestrial Ecosystems (IRET) of the National Research Council of Italy. Ilaria holds a PhD in Fundamental Ecology from University of Salento, and a MSc in Biological Science at the University of Bologna. She has a background on biodiversity organization, conservation and management in aquatic ecosystems. Currently, her research focus is on e-biodiversity and, in particular, she is working in the field of semantic technologies for data integration and interoperability. Ilaria, as a member of the LifeWatch ERIC's working group on "Metadata, Controlled Vocabularies and Ontologies", is responsible of the implementation and curation of thesauri and ontologies for the biodiversity and ecosystems domain. As a member of the Research Data Alliance (RDA), Ilaria is involved in various groups, in particular the WG "FAIRSharing Registry: connecting data policies, standards & databases" and the IG "Biodiversity Data Integration". She has participated in several research programs funded at Regional, National and European level and she is currently involved in two H2020 projects, ECOPOTENTIAL and ENVRI-FAIR.

Publications

- Kissling WD, Walls R, Bowser A, Jones MO, Kattge J, Agosti D, Amengual J, Basset A et al. (2018) Towards global data products of Essential Biodiversity Variables on species traits. Nature ecology & evolution 2:1531-1540, https://doi.org/10.1038/s41559-018-0667-3
- Fiore N et al. (2017) EcoPortal: a proposition for a semantic repository dedicated to ecology and biodiversity. In: Algergawy, A., Karam, N., Klan, F., Jonquet, C. (Eds.) Semantics for Biodiversity. Proceedings of the 2nd International Workshop on Semantics for Biodiversity co-located with 16th International Semantic Web Conference (ISWC 2017). Vol-1933. pp. 6, http://ceur-ws.org/Vol-1933/poster-paper-13.pdf
- Rosati I, Bergami C, Stanca E, Roselli L, Tagliolato P, Oggioni A, Fiore N, Pugnetti A, Zingone A, Boggero A, Basset A (2017) A thesaurus for phytoplankton trait-based approaches: development and applicability. Ecological Informatics 42:129-138, https://doi.org/10.1016/j.ecoinf.2017.10.01
- Rosati I, Bergami C, Fiore N, Oggioni A, Tagliolato P (2017) LifeWatch Italy Thesauri

Documentation. Version 1.0. Roma, CNR Pubblicazioni 2017, pp.18, ISBN 978-88-8080-249-5

- Basset A (2016) e-Science perspectives for the conservation of transitional and coastal wetlands. Aquatic Conservation-Marine and freshwater ecosystems 26, 411-415, https://doi.org/10.1080/11263504.2012.740091

Relevant projects

LifeWatch ERIC is involved in, a variety of European projects such as

- ERIC Forum (https://www.eric-forum.eu/),
- ENVRI-Plus (http://www.envriplus.eu/),
- ENVRI-FAIR, EcoPotential (http://www.ecopotential-project.eu/),
- Globis-B (http://www.globis-b.eu/), and
- Jerico-Next (http://www.jerico-ri.eu/)

focusing on research infrastructures and associated data management.

Relevant infrastructure

Ecoportal is an initiative driven by the research infrastructure LifeWatch ERIC. Its goal is to design and develop a repository for the semantic resources in the ecological domain. It will support the community in the management and alignment of their semantics. This will improve the discovery, integration and re-usability of data. LifeWatch Eric has included Ecoportal as priority for 2019. The first release will be published for the end of February 2019 to the address http://ecoportal.lifewatchitaly.eu and will be based on the NCBO technology. In the first realise all the effort done in the semantic field by the LifeWatch network will be collected and published: LifeWatch Italy Thesauri, LifeWatch Core Ontology, etc.

LifeWatch ERIC **Catalogue of Data** has for objective to facilitate and enhance the use of biodiversity primary data, species traits data and in general all the biotic and abiotic data useful in the biological field. Different types of datasets are available on the LifeWatch Catalogue of Data and hundreds of users are registered.

UK Research and Innovation (UKRI)

National Oceanography Centre - British Oceanographic Data Centre (NOC-BODC)

The British Oceanographic Data Centre (BODC) is a component of the UK Natural Environment Research Council's Environmental (NERC), one of the UKRI research councils. It is part of NERC's Data Centre network with designated responsibility for marine data. It is a component of the National Oceanography Centre (NOC), NERC's centre of excellence for oceanographic sciences with a remit to provide leadership and national capability in marine sciences. BODC's mission is to develop, coordinate and provide specialist data services for the UK and international marine science communities; to enable innovative use and re-use of data; to ensure long-term curation of valuable and unique marine data resources; and to champion Open Data. Its 50 staff include a wide range of research scientists who have direct experience of marine data collection and analysis. They work alongside information technology specialists with particular skills in controlled vocabularies, ontologies, standards and interoperability. This combination of skills brings a pioneering approach to the development of marine data management tools. Nationally, BODC manages data from the UK Argo programme and the National Tide Gauge Network. It also

manages the Marine Environment Monitoring and Assessment National (MERMAN) database, which holds the data required to meet the UK's commitment to European directives, including its mandatory monitoring requirements under the Oslo and Paris Convention (OSPAR). BODC is accredited as the oceanography Data Archive Centre (DAC) for the UK Marine Environmental Data and Information Network (MEDIN). It participates in the Intergovernmental Oceanographic Commission's International Oceanographic Data and Information Exchange (IOC-IODE) programme, acting as the UK's National Oceanographic Data Centre and national contact point. BODC collaborates actively with other oceanographic data centres in Europe, and in particular, coordinates metadata directories and leads on standards development, especially controlled vocabulary development. BODC has also been engaged in many current and recent European projects, including SeaDataCloud/SeaDataNet-II/SeaDataNet, MyOcean/MyOcean2, EUROFLEETS, NETMAR, CLIP-C, ODIP/ODIP2, FixO3, Celtic Seas Partnership (Life+), SenseOcean, ENVRI+, AtlantOS, and several EMODnet projects.

Role in the project

The primary goal and role of UKRI in the proposal is as a vocabulary service provider for the marine domain in B2TERM contributing to WP2 and WP3. Our expertise in developing standards, vocabularies, ontologies and data management interoperability best practices will contribute to WP2, WP3, WP5 and WP7. The NERC-VS uses a strong governance model to maintain terminologies and this knowledge will contribute to WP6.

Key personnel

Dr Louise Darroch [f] is a senior data scientist at BODC. She is a project manager of technical and software projects that deliver new capability to BODC, such as the OGC's Sensor Web Enablement standards and NOAA's ERDDAP (EU ENVRI+). She is currently leading the data management of one of the NERC's major research programmes (Shelf Sea Biogeochemistry) and is a marine vocabulary expert, specialising in the standardised terms of instruments and sensors and providing them to national and international users. Additionally, she has provided cross-discipline vocabulary expertise and products for several other EU-funded projects including AtlantOS, EUROFLEETSv2, BRIDGES and SenseOCEAN. Dr Darroch is also a member of several international expert groups where she provides vocabulary and data management best practice, including the IOC's Global Ocean Surface Underway Data (GOSUD) Project Team, OceanSITES Data Management Team and SWE Marine Profiles. She has recently co-authored a white paper (democratisation, standardisation and integrity of sensor data), accepted for inclusion in the Oceanobs'19 conference.

Dr Gwenaelle Moncoiffe [f] is a senior data scientist and the head of the Vocabulary Management Group at BODC. She has a PhD in biological oceanography and post-doctoral research experience in marine biogeochemistry. Gwen is an experienced marine scientist who has developed expertise in scientific data mark-up and semantic modelisation of reported variables. She combines this with a strong interest in cross-domain interoperability and in the use of controlled vocabularies by the scientific community. As a data scientist and data project manager, she has worked alongside research scientists to improve data management best practices in large multidisciplinary research programmes, co-authoring guidelines and best practice documents for e.g. SCOR-IGBP sponsored programmes (IMBER, SOLAS) and the European project EPOCA (Ocean Acidification). She was chair of the IOC-IODE Group of Experts in Biological and Chemical Data Management and Exchange Practices from 2007 to 2011. In this role, and in subsequent national and international collaborations including those developed as part of the European SeaDataCloud and EMODnet chemistry projects, she has contributed to the development of vocabulary and data standards to facilitate the exchange of physical, chemical and biological data from a variety of environments (ocean, atmosphere, terrestrial, benthic).

Dr Alexandra Kokkinaki [f] is a senior semantic web scientist in the British Oceanographic Data Centre (BODC). She is the technical lead of the NERC Vocabulary Server (NVS) and coordinated the development of NVS-related tools. She participated in ODIPI and ODIPII and collaborated with external partners to deliver the innovative Rosetta Stone translation service that provides dynamic vocabulary translations. She also worked with the SWE Marine Profiles group to deliver the Common SensorML profiles and a list of vocabularies specialized to sensorML. She was also the lead developer of BODC's OGC sensor infrastructure which delivers standardised sensor data and metadata. She is currently working in Oceanids UK project, which will provide a unified and consistent infrastructure to control the NERC fleet of unmanned long range vehicles and has the technical lead in the SeaDataCloud project in regards to the vocabulary developments. She has recently co-authored a white paper (democratisation, standardisation and integrity of sensor data), accepted for inclusion in the Oceanobs'19 conference.

Publications

- Leadbetter AM, Lowry RK, Clements DO (2014) Putting Meaning into Netmar – The Open Service Network for Marine Environmental Data. International Journal of Digital Earth 7 (10): 811–828, https://doi.org/10.1080/17538947.2013.781243
- Schaap D, Lowry R (2010) SeaDataNet – Pan-European infrastructure for marine and ocean data management: unified access to distributed data sets, International Journal of Digital Earth, vol. 3, no. 1, pp. 50-69.
- Wiebe PH, Allison D, Kennedy M, Moncoiffé G (2014) A vocabulary for the configuration of net tows for collecting plankton and micronekton. J. Plank. Res. 37(1):21-27, https://doi.org/10.1093/plankt/fbu101
- Kokkinaki A, Darroch L, Buck J, Jirka S (2016) Semantically Enhancing SensorML with Controlled Vocabularies in the Marine Domain. In Proceedings of GSWC 2016, Geospatial Sensor Webs Conference, Munster, Germany, August 29-31, 2016. CEUR Workshop Proceedings, ISSN 1613-0073, http://ceur-ws.org/Vol-1762/Kokkinaki.pdf
- De Pooter D, Appeltans W, Bailly N, Bristol S, Deneudt K, Eliezer M, Fujioka E, Giorgetti A, Goldstein P, Lewis M, Lipizer M, Mackay K, Marin M, Moncoiffé G, Nikolopoulou S, Provoost P, Rauch S, Roubicek A, Torres C, van de Putte A, Vandepitte L, Vanhoorne B, Vinci M, Wambiji N, Watts D, Klein Salas E, Hernandez F (2017) Toward a new data standard for combined marine biological and environmental datasets - expanding OBIS beyond species occurrences. Biodiversity Data Journal 5: e10989, https://doi.org/10.3897/BDJ.5.e10989

Relevant projects

- **SeaDataNet I, SeaDataNet II & SeaDataCloud (EU FP6/FP7, H2020) [2006-2020]** Contribute to and involved in the development of a pan-European infrastructure for marine and ocean data by a consortium of over 100 marine data centres and research institutes from 34 countries including interfacing to EUDAT services. SDN Common Vocabularies are curated and served by NERC-VS.
- **AtlantOS (EU H2020) [2015-2019]** An initiative to develop a sustainable and integrated Atlantic Ocean observing system, comprising over 15 international in-situ observing networks and 62 partners. BODC developed a vocabulary model that harmonised the data flow from observations to users, based on the globally important, Essential Ocean and Climate Variables (EOVs, ECVs). The vocabularies developed and used are curated and served by NERC-VS.
- **SenseOCEAN (EU FP7 Oceans of Tomorrow) [2014-2017]** Part of the Oceans of Tomorrow collaboration, SenseOCEAN drew on a consortium of European sensor developers and data curators to develop an innovative and integrated, multiparameter sensor package to monitor the health of the oceans. BODC developed 'plug and play' capability using OGC and W3C global semantic sensor web standards. BODC developed new vocabularies to underpin the

interoperability of these standards in the marine domain. The vocabularies developed and used are curated and served by NERC-VS.

- **EUROFLEETSv2 (EU FP7) [2013-2017]** A consortium of 31 European partners from 20 countries developing services and providing access to modern Research Vessels and Equipment for scientific excellence. BODC were responsible for metadata curation in support of service discovery using vocabularies curated and served by NERC-VS.
- **BRIDGES (EH H2020) [2015-2019]** Delivers innovative, deep ocean exploration, autonomous glider platforms, in a consortium of technicians, software developers and data curators. BODC provided data curation, ontology and vocabulary expertise towards the standardization of software and hardware components, allowing open access, commercialization and versatile data distribution. The vocabularies used in the standardization are curated and served by NERC-VS.

Relevant infrastructure

BODC hosts the NERC Vocabulary Server version 2.0 (NERC-VS) (Leadbetter et al., 2014) which, in addition to its predecessors, have successfully served controlled vocabularies to the marine domain for more than 10 years. This server provides access to lists of standardized terms that cover a broad spectrum of disciplines relevant to the oceanographic and wider community. NERC-VS makes use of the World Wide Web Consortium's Simple Knowledge Organization System (SKOS) to represent knowledge in a format understandable by computers. In SKOS, vocabularies are modelled as collections and terms are modelled as concepts. Collections and concepts have unique URIs that are resolvable through a RESTful interface to either HTML or RDF documents through content negotiation. Collections are also accessible through SOAP Web Services and a SPARQL endpoint. All of the vocabularies are fully versioned and assured (governed) by a dedicated group of experts before publication. The largest of these vocabularies is the 'BODC Parameter Usage Vocabulary' (P01, http://vocab.nerc.ac.uk/collection/P01/current/) which is used to annotate the fields of data files with the physical properties that the numbers represent. P01 concepts are built from a rigorous, exposed, semantic model. The use of the NERC-VS has been pivotal within the European Union SeaDataNet I/SeaDataNet II/SeaDataCloud (Schaap and Lowry, 2010), EUROFLEETSv2 and EMODnet projects. The infrastructure has also been used to underpin interoperability in an international Atlantic ocean observing system (EU H2020 AtlantOS) (Darroch and Buck, 2017), global OGC and W3C sensor standards in the marine domain (Marine SWE Profiles, EU Oceans of Tomorrow SenseOCEAN) (Kokkinaki et al., 2016), marine biological communities (De Pooter et al., 2017, Wiebe et al., 2014) and a prototype within the Ocean Data Interoperability Platform (ODIP) collaboration by linking EU, US and Australian research cruise programs through providing cruise information at an international level.

Finnish Meteorological Institute (FMI)

The Finnish Meteorological Institute (FMI, www.fmi.fi/en) functions under the Ministry of Transport and Communications and has the mandate of producing weather, atmosphere, climate and marine related services required by Finnish society, with the aim of promoting safety and serving the needs of the public, industry and commerce. FMI conducts scientific research including observations of the physical state of the atmosphere, its chemical composition, and electromagnetic phenomena, and develops and applies numerical models in order to analyse and forecast atmospheric physical and chemical processes. FMI employs about 680 people, about 350 of which are involved in research.

ILMATIETEEN LAITOS
METEOROLOGISKA INSTITUTET
FINNISH METEOROLOGICAL INSTITUTE

FMI has coordinated and participated in numerous national, European and international projects, and is a partner of many EU environmental research infrastructures: ACTRIS, ICOS, EISCAT_3D, EPOS, EURO-ARGO, ICOS, JERICO-NEXT; and their cluster activities: ENVRI+, ENVRI-FAIR. FMI is coordinating ACTRIS PPP, tasked with the implementation of ACTRIS as a research infrastructure. FMI contributes to numerous international measurement networks including GAW, EMEP, International Arctic Systems for Observing the Atmosphere (IASOA), and the Copernicus Atmosphere Monitoring Service (CAMS). FMI has demonstrated experience in data processing, management and curation, implementing open-data initiatives and provision of satellite data. FMIs Arctic Research Centre (FMI-ARC) hosts the National Satellite Data Centre (NSDC) focusing on fast delivery remote sensing product generation, providing free and open data for national and international partners (ESA, EUMETSAT, NASA, NOAA, European Union and national agencies and research institutes).and customers. This includes the hosting of a Copernicus Sentinel Collaborative Ground Station for Sentinel satellite data distribution and archiving, cloud services through public private-partnership arrangements, and maintaining external processing chains in virtual environments.

Role in the project

FMI is implementing the data infrastructure node for cloud remote sensing within ACTRIS. As a partner, FMI will contribute primarily to WP4 to integrate B2TERM into ACTRIS Data Centre and support WP2 service specification and WP5 community engagement.

Key personnel

Dr. Ewan J. O'Connor [m] is a Tenure-track Professor at FMI with 18 years of experience in ground-based active remote sensing of the atmosphere. He develops and uses new radar and lidar techniques to retrieve cloud, aerosol and turbulent parameters for understanding cloud and dynamical processes, and for the evaluation and improvement of numerical forecast and climate models. He has participated in numerous national and international research projects, and is responsible for managing Cloudnet, (providing standardised methods for deriving cloud parameters from cloud radar and lidar) and its integration within ACTRIS as the cloud remote sensing node. He is responsible for the design and implementation of the Finnish Doppler lidar network. He has 60+ peer-reviewed publications as author or co-author.

Dr Simo Tukiainen [m] received his PhD from the University of Helsinki in 2016 and has a strong background in retrievals of atmospheric profiles from satellite and ground-based remote sensing data using Bayesian inverse methods. He has 10+ years of experience in scientific computing, atmospheric retrievals and radiative transfer modelling. He is the lead developer of the GOMOS bright limb ozone product, FMIs OSIRIS O3/NO2 products, and FMIs ground-based FTIR CH4 profile data. He is involved in designing, developing and implementing the infrastructure for the ACTRIS cloud remote sensing node.

Publications

- Illingworth AJ, Hogan RJ, O'Connor EJ, Bouniol D, Delanoë J, Pelon J, Protat A, Brooks ME, Gaussiat N, Wilson DR, Donovan DP, Baltink HK, van Zadelhoff G, Eastment DJ, Goddard JW, Wrench CL, Haeffelin M, Krasnov OA, Russchenberg HW, Piriou J, Vinit F, Seifert A, Tompkins AM, Willén U (2007) Cloudnet – continuous evaluation of cloud profiles in seven operational models using ground-based observations, Bull. Amer. Meteor. Soc., 88, 883–898, https://doi.org/10.1175/BAMS-88-6-883
- Hirsikko A, O'Connor EJ, et al. (2014) Observing wind, aerosol particles, cloud and precipitation: Finland's new ground-based remote-sensing network, Atmos. Meas. Tech., 7, 1351–1375
- Uttal T et al. (2016) International Arctic Systems for Observing the Atmosphere: An

International Polar Year Legacy Consortium. Bull. Amer. Meteor. Soc., 97, 1033–1056, https://doi.org/10.1175/BAMS-D-14-00145.1

Relevant projects

- Coordinating the H2020 ACTRIS Preparatory Phase Project – Grant agreement n. 739530 (2017-2019).
- Participation in the H2020 project ACTRIS 2 (Aerosol, Clouds, and Trace gases Research Infrastructure) - Grant agreement n. 654109 (2015-2019).
- Participation in the H2020 project ENVRI-FAIR (ENVironmental Research Infrastructures building Fair services Accessible for society, Innovation and Research) - Grant agreement n. 824068 (2019-2022).
- Participation in the H2020 project ENVRIPLUS (Environmental Research Infrastructures Providing Shared Solutions for Science and Society) - Grant agreement n. 654182 (2015- 2019).
- Participation in the H2020 project RINGO (Readiness of ICOS for Necessities of Integrated Global Observations) - Grant agreement n. 730944 (2017- 2021).

Relevant infrastructure

ACTRIS (Aerosol, Clouds, and Trace gases Research Infrastructure) www.actris.eu

National Research Council (CNR)

The National Research Council (CNR) is the largest public research institution in Italy, and the only one under the Research Ministry performing multidisciplinary activities. Established on 18 November 1923, CNR's stated mission is to carry out research in its own Institutes, to promote innovation and competitiveness of the national industrial system, to foster the internationalization of the national research system, to develop technologies and solutions to emerging public and private needs, to advice Government and other public bodies, and to contribute to the qualification of human resources. Within CNR this proposal is carried on by the Institute of Methodologies for Environmental Analysis (IMAA). CNR-IMAA has a proven track-record in the management of research infrastructures and in the implementation of strategies to improve and harmonize operational activities of in-situ observation networks (e.g. standardization of techniques and procedures for data quality assurance, automation of data collection, processing and transfer, etc.).

CNR-IMAA has a leadership role in the remote sensing community, characterized by a high scientific productivity and by the proven capability to provide support and technological transfer to end users. CNR-IMAA has coordinated and participated in a large number of national, European and international projects. At present, CNR-IMAA is coordinating the ACTRIS-2 H2020 research infrastructure project on aerosol, clouds and trace gases, following on ACTRIS (FP7) built on the EARLINET-ASOS FP6 and EUSAAR FP6 projects; the first two were also coordinated by the CNR-IMAA. Apart ACTRIS-2, CNR-IMAA is currently participating in several H2020 projects: ACTRIS-PPP, DustClim, EUNADICS-AV, ECARS, ENVRIPLUS, . In the past, CNR-IMAA has been a partner in several FP6, FP7, and H2020 projects, like GEOMON, WEZARD, ITARS, EARLINET-ASOS, GAIA-CLIMand involved in or coordinated several ESA projects like ESA-VALID, ESA-CALIPSO and ESA-LIVAS. The CNR-IMAA is also currently coordinating BARON4C3S, a Copernicus project for rationalising, harmonising and improving access to open

and free observational records and data streams from selected in-situ GCOS-relevant Baseline and Reference observing networks. CNR-IMAA is furthermore involved in several international networks and initiatives like AERONET, CLOUDNET, GALION, GRUAN, the Scientific Advisory Group for Aerosols of the Global Atmosphere Watch (GAW) aerosol program of WMO, the Regional Steering Group of the SDS-WAS (Sand and Dust Storm Warning Advisory and Assessment System) of the WMO, the InDust COST action, theWMO-Lidar ad hoc Task team for volcanic ash detection and the Scientific Advisory Group on Volcanic Ash (VA-SAG) of WMO.

CNR-IMAA is currently leading activities at EU and international levels to foster the integration of observations across many key international networks engaged in systematic long-term observations of the atmosphere and to ensure implementation of common observing techniques, data and metadata strategies. Specific expertise at CNR-IMAA relates to improving the comparability of data streams, creating uniform data quality standards, sharing of QA/QC approaches, and realising observational program synergies.

CNR-IMAA has a leadership role in the remote sensing community, characterized by a high scientific productivity and by the proven capability to provide support and technological transfer to end users. CNR-IMAA has coordinated and participated in a large number of national, European and international projects. In particular, CNR-IMAA is the main responsible for the design, testing and development of the Single Calculus Chain (SCC), the common standardized automatic analysis software developed within EARLINET. As for the data management and provision, CNR is currently responsible of the EARLINET database for the data collection, data curation and data provision to internal and external users.

Role in the project

CNR-IMAA leads the ACTRIS Aerosol remote sensing data centre unit and as a partner CNR-IMAA will contribute primarily to WP4 to integrate B2TERM into ACTRIS Data Centre and support WP2 service specification and WP5 community engagement.

Key personnel

Dr. Lucia Mona [f] – Researcher. Dr Mona has a researcher profile that combines expertise on developments of lidar systems, instruments integration/combination, analysis methodologies, exploitation of state-of-the-art measurements for different application fields and integrated studies with models. She is responsible of the EARLINET (European Aerosol Research Lidar NETwork) database and is working on exploitation of EARLINET database for comparison/integration with other ground-based and satellite measurements and models and model evaluation/integration studies for peculiar long-range transport cases and for multi-year observation. The high level of Dr. Mona's research to date has been recognized at international level by the large number of participations in working groups as a key scientist and by the number of oral presentations and citations to her papers. Dr Lucia Mona is author or co-author of more than 40 papers in peer reviewed literature with more than 1700 citations. She is the responsible for the EARLINET database and its link to the ACTRIS data portal within the ACTRIS2 (H2020) project. She is the WP leader of the Observation Infrastructure WP within the EUNADICS-AV EU H2020 project, coordinating activities for the collection, cross-validation and exploitation of ground based and satellite based measurements relevant for aviation purposes. She is also leading the EC-ACTS (Earlinet and Cloudnet - Aerosol and Clouds Teams for Sentinel-5P Validation) ESA validation project. She is participating in international current projects: ACTRIS PPP (H2020), EUNADICS-AV (H2020), ECARS (H2020), ACTRIS2 (H2020), ENVRIplus (H2020) and GAIA-CLIM (H2020). She is leading the AEROSAT (International Satellite Aerosol Science Network) Working Group on Aerosol Typing. She is member of the Regional Steering Group of the SDS-WAS (Sand and Dust Storm Warning Advisory and Assessment System) of the WMO.

Dr. Giuseppe D'Amico [m] - Researcher. He has over 14 years' research experience in the field of aerosol atmospheric remote sensing studies with lidar techniques. Dr. Giuseppe D'Amico is the main developer of the EARLINET Single Calculus Chain (SCC) the standard and traceable ACTRIS tool to retrieve aerosol optical properties from raw lidar measurements. He has long experience in developing algorithm for the retrieval of aerosol vertical profiles, standardized quality control procedures and data traceability and interoperability. Dr. Giuseppe D'Amico is also involved in the designing and the implementation of ACTRIS aerosol remote sensing node infrastructure.

Francesco Amato [m] - Researcher. He is working since 2015 on EARLINET/ACTRIS database. Dr. Francesco Amato is the main developer of the EARLINET database procedures for automatic data quality control and tools for automatic monitoring of data access and user statistics. He is also involved in the designing and the implementation of ACTRIS aerosol remote sensing node infrastructure.

Publications

- The EARLINET publishing group 2000-2015;EARLINET All 2000-2015. World Data Center for Climate (WDCC) at DKRZ, https://doi.org/10.1594/WDCC/EARLINET_All_2000-2015
- Papagiannopoulos N, Mona L, Amodeo A, D'Amico G, Gumà Claramunt P, Pappalardo G, Alados-Arboledas L, Guerrero-Rascado JL, Amiridis V, Kokkalis P, Apituley A, Baars H, Schwarz A, Wandinger U, Binietoglou I, Nicolae D, Bortoli D, Comerón A, Rodríguez-Gómez A, Sicard M, Papayannis A, Wiegner M (2018) An automatic observation-based aerosol typing method for EARLINET, Atmos. Chem. Phys., 18, 15879-15901, https://doi.org/10.5194/acp-18-15879-2018
- D'Amico G, Amodeo A, Baars H, Binietoglou I, Freudenthaler V, Mattis I, Wandinger U, Pappalardo G (2015) EARLINET Single Calculus Chain – overview on methodology and strategy, Atmos. Meas. Tech., 8, 4891-4916, https://doi.org/10.5194/amt-8-4891-2015
- Mona L, Papagiannopoulos N, Basart S, Baldasano J, Binietoglou I, Cornacchia C, Pappalardo G (2014) EARLINET dust observations vs. BSC-DREAM8b modeled profiles: 12-year-long systematic comparison at Potenza, Italy, Atmos. Chem. Phys., 14, 8781–8793, https://doi.org/10.5194/acp-14-8781-2014
- Pappalardo G, Mona L, D'Amico G, Wandinger U, Adam M., Amodeo A, Ansmann A, Apituley A, Alados Arboledas L, Balis D, Boselli A, et al. (2013) Four-dimensional distribution of the 2010 Eyjafjallajökull volcanic cloud over Europe observed by EARLINET, Atmos. Chem. Phys., 13, 4429-4450, https://doi.org/10.5194/acp-13-4429-2013

Relevant projects

- Coordination of the H2020 project ACTRIS 2 (Aerosol, Clouds, and Trace gases Research Infrastructure) -Grant agreement n. 654109 (2015-2019).
- Participation in the H2020 project ENVRIFAIR (ENVironmental Research Infrastructures building Fair services Accessible for society, Innovation and Research) - Grant agreement n. 824068 (2019 -2022).
- Participation in the H2020 ACTRIS Preparatory Phase Project – Grant agreement n. 739530 (2017-2019).
- Participation in the H2020 project ENVRIPLUS (Environmental Research Infrastructures Providing Shared Solutions for Science and Society) - Grant agreement n. 654182 (2015- 2019).
- Participation in the H2020 project EUNADICS-AV (European Natural Airborne Disaster Information and Coordination System for Aviation) - grant agreement n. 723986 (2015-2019)

Relevant infrastructure

ACTRIS (Aerosol, Clouds, and Trace gases Research Infrastructure) www.actris.eu

Barcelona Supercomputing Center (BSC)

The Barcelona Supercomputing Center – Centro Nacional de Supercomputacion (short named as BSC), created in 2005, is the leading supercomputing centre in Spain. It specialises in High-Performance Computing and its mission is twofold: to offer supercomputing facilities and services to Spanish and European scientists and to create knowledge and technology to be transferred to society. At the BSC, more than 500 people from 40 different countries perform and facilitate research into Computer Sciences, Life Sciences, Earth Sciences and Computational Applications in Science and Engineering. This multi-disciplinary approach and the combination of world-leading researchers and experts in HPC (High-Performing Computing) with state-of-the-art supercomputing resources make BSC unique. The BSC is one of the first eight Spanish 'Severo Ochoa Centre of Excellence" awarded by the Spanish Government, it is managing the Spanish Supercomputing Network, as well as one of the four hosting members of the European PRACE Research Infrastructure. The BSC hosts MareNostrum supercomputer, a Tier-0 PRACE system currently ranked as the #3 most powerful supercomputer in Europe (#13 in the world) with 13.7Pflop/s capacity. In addition, the BSC hosts other High-Performance Computing (HPC) resources, among which it is worth mentioning Minotauro, one of the most efficient supercomputers in the world (#35 in the last ranking of the top500 green list).

The Earth Sciences Department of the BSC (ES-BSC) was established with the objective of carrying out research in Earth system modelling and focuses its activity on emissions, air quality, mineral dust and global and regional climate modelling and prediction. ES-BSC is organized around four closely interacting groups (Atmospheric Composition, Climate Prediction, Computational Earth Sciences, and Earth System Services) comprising ~80 employees, including scientific, technical, and support staff. The department is an active member of the EC-Earth consortium, whose Earth System Model is widely used at ES-BSC for research and teaching purposes. During last 5 years (2013-2017), BSC-ES was granted 9 EU H2020 projects, 5 EU FP7 projects, 5 EU Copernicus projects, 10 national projects, 2 projects funded by the European Space Agency, 1 project funded by the French Ministry of Sciences, 1 project funded by the Flanders Research Foundation, 1 project from ERA-NET, 3 from ERA4CS and 1 ERC Consolidator Grant. During that same period, BSC-ES also participated in 21 RES and 4 PRACE projects. The BSC-ES international activity includes the coordination of the two World Meteorological Organisation (WMO) regional centres specialised in sand and dust warning and forecasting, as well as the participation in climate services initiatives like the Climate Services Partnership (CSP). Members of the BSC-ES participate in committees of the World Climate Research Programme (WCRP), such as the CLIVAR Scientific Steering Group or the Working Group on Seasonal to Interannual Prediction (WGSIP). The Atmospheric Composition group at ES-BSC aims at better understanding and predicting the spatiotemporal variations of atmospheric pollutants along with their effects on air quality, weather and climate and contributes to a variety of forecasting activities. The Atmospheric Composition group at ES-BSC aims at better understanding and predicting the spatiotemporal variations of atmospheric pollutants along with their effects on air quality, weather and climate and contributes to a variety of forecasting activities.

Role in the project
BSC is part of the ACTRIS Data Center consortium and as a partner BSC will participate to "WP5: Social Integrations: Integrate the EOSC-VS in research communities" taking part of the KOINE Stakeholder Forum and gathering research community requirements.
Key personnel
Dr. Carlos PEREZ GARCIA-PANDO [m] is AXA Professor, Ramon y Cajal Researcher and Head of the Atmospheric Composition Group at BSC. He holds an AXA Chair on Dust Storms at BSC. His research group is composed of ~20 people including senior researchers, postdocs, PhD students and technical support staff. His research focuses on understanding the physical and chemical processes controlling atmospheric aerosols, and evaluating their effects on climate, ocean biogeochemistry, air quality and health. His core area of expertise is atmospheric mineral dust. He is also a model developer with a large experience in supercomputers. Previously he has held research positions at the NOAA/National Centers for Environmental Prediction, the International Research Institute for Climate and Society, the NASA Goddard Institute for Space Studies and Columbia University. He has participated in ~30 international and national projects (in 7 of them as PI or co-PI). In the US, he has been PI and co-PI of competitive projects funded by the Department of Energy, NASA and NOAA. In 2017, he was awarded an ERC Consolidator Grant entitled FRAGMENT, which has started in October 2018. I was also awarded the Agustín de Betancourt y Molina prize for young researchers by the Spanish Royal Academy of Engineering. His work has resulted in ~60 peer-reviewed publications, 20 chapters in books, proceedings and reports, more than 150 contributions to conferences/workshops/seminars (26 as invited speaker) and the edition of a book of proceedings (Google Scholar; citations 3766; h-index 31). **Dr. Sara BASART [f]** received her PhD degree in Engineering Environmental at Technical University of Catalonia (UPC) in 2012 while doing her research at different research centres (Centro de Investigacion Atmosferica de Izana, Spain, and Laboratoire des Sciences du Climat et de l'Environnement, France). Her main research background covers mineral dust modelling, air quality and aerosols. At present, Dr Basart is a researcher in the Barcelona Supercomputing Center (BSC). She is the scientist in charge of the WMO Sand and Dust Storm Warning Advisory and Assessment System (SDS-WAS) Regional Center for Northern Africa, Middle East and Europe, and the Barcelona Dust Forecast Center (BDFC), hosted in BSC. She also participates in international projects like the International Cooperative on Aerosol Prediction (ICAP) initiative, H2020 (SOLWATT, ACTRIS and ACTRIS-2) and Copernicus (CAMS-50, CAMS-84, CAMS-95). She is Leader Project Investigator of the EU ERA4CS project DustClim. Recently, she was elected as a Chair of the COST Action InDust (CA16202). She has authored or co-authored more than 35 peer-reviewed publications in international journals and book chapters (Google Scholar: citations 1448; h-index: 19). Furthermore, she has participated in capacity building and transfer of knowledge activities associated with private contracts, European Commission and United Nations programmes. **MSc. Francesco BENINCASA [m]** is software designer and developer. He holds a Master of Science in Software Engineering with a thesis on Optimization Algorithms ("Heuristic Algorithms for Bin Packing Problem"). He is GNU/Linux and Python language expert and author of various software applications, web and stand-alone, for web communities, for data managing, automatic shopping, network data upload/download, data processing, formatting and visualizing. After working in the private sector on web and databases development and automatic shop applications, he started his experience in supercomputing and data manipulation (SCS SuperComputing Solutions, a CINECA spin-off - Bologna - Italy) oriented to biomedical simulations, participating to several FP6 and FP7 European projects. Since 2010 he is at the Barcelona Supercomputing Center (BSC) working on air quality data management, processing and visualization. Currently is co-leader

of the Data and Diagnostics Team inside the Earth Sciences Department. He is in charge of data processing and management and web development and maintenance of the WMO SDS-WAS NAME-E Regional Center (SDS) and the Barcelona Dust Forecast Center (BDFC), both projects operated by a consortium of BSC and AEMET (Spanish Meteorological Agency) under the umbrella of the World Meteorological Organization (WMO) with the goals to improve the understanding of sand and dust storms phenomena through air quality models comparison and evaluation (SDS) and to provide an operational daily dust forecast over the Mediterranean area (BDFC). Both previous projects are also involved in an EUDAT user pilot implementing storage and staging services. He is also co-chairman of the "weather, climate and air quality Interest Group" inside the Research Data Alliance (RDA).

Publications

- Benedetti A, Reid JS, Knippertz P, Marsham JH, Di Giuseppe F, Rémy S, Basart S, Boucher O, Brooks IM, Menut L, Mona L, Laj P, Pappalardo G, Wiedensohler A, Baklanov A, Brooks M, Colarco PR, Cuevas E, da Silva A, Escribano J, Flemming J, Huneeus N, Jorba O, Kazadzis S, Kinne S, Popp T, Quinn PK, Sekiyama TT, Tanaka T, Terradellas E (2018) Status and future of numerical atmospheric aerosol prediction with a focus on data requirements. Atmospheric Chemistry and Physics, 18, 10615-10643, https://doi.org/10.5194/acp-18-10615-2018
- Ansmann A, Rittmeister F, Engelmann R, Basart S, Jorba O, Spyrou C, Remy S, Skupin A, Baars H, Seifert P, Senf F, Kanitz T (2017) Profiling of Saharan dust from the Caribbean to western Africa – Part 2: Shipborne lidar measurements versus forecasts Atmospheric Chemistry and Physics, 17, 14987-15006, https://doi.org/10.5194/acp-17-14987-2017
- Huneeus N, Basart S, Fiedler S, Morcrette JJ, Benedetti A, Mulcahy J, Terradellas E, Pérez García-Pando C, Pejanovic G, Nickovic S, Arsenovic P, Schulz M, Cuevas E, Baldasano JM, Pey J, Remy S, Cvetkovic B (2016) Forecasting the northern African dust outbreak towards Europe in April 2011: a model intercomparison, Atmospheric Chemistry and Physics, 16, 4967-4986, https://doi.org/10.5194/acp-16-4967-2016
- Pérez García-Pando C, Miller RL, Perlwitz JP, Rodríguez S, Prospero JM (2016) Predicting the mineral composition of dust aerosols: Insights from elemental composition measured at the Izana Observatory. Geophysical Research Letters, 43(19).
- Binietoglou I, Basart S, Alados-Arboledas L, Amiridis V, Argyrouli A, Baars H, Baldasano JM, Balis D, Belegante L, Bravo-Aranda JA, Burlizzi P, Carrasco V, Chaikovsky A, Comerón A, D'Amico G, Filioglou M et al. (2015) A methodology for investigating dust model performance using synergistic EARLINET/AERONET dust concentration retrievals. Atmospheric Measurement Techniques, 8, 3577-3600, https://doi.org/10.5194/amt-8-3577-2015

Relevant projects

- IS-ENES3-Infrastructure for the European Network for Earth System modelling - Phase 3 (GA-824084). This project will deliver the third phase of the distributed e-infrastructure of the European Network for Earth System Modelling (ENES). It is initiated as the European climate modelling community faces the challenges of contributing to the next assessment report of the Intergovernmental Panel on Climate Change through the 6th phase of the CMIP.
- ACTRIS-2 Aerosols, Clouds, and Trace gases Research InfraStructure (GA- 654109). It addresses the scope of integrating state-of-the-art European ground-based stations for long-term observations of aerosols, clouds and short-lived gases. It consolidates and improves services offered within FP7 funded Integrated Infrastructures Initiative ACTRIS (2011-2015).
- ACTRIS-PPP Aerosols, Clouds and Trace gases Preparatory Phase Project (GA-739530). It is the pan-European RI that consolidates activities amongst European partners for observations of aerosols, clouds, and trace gases and for the understanding of the related atmospheric processes, to provide RI services to wide user groups.

Relevant infrastructure

BSC is a key element of and coordinates the Spanish Supercomputing Network, which is the main framework for granting competitive HPC time to Spanish research institutions. Furthermore, BSC is one of six hosting nodes in France, Germany, Italy and Spain that form the core of the Partnership for Advanced Computing in Europe (PRACE) network. PRACE provides competitive computing time on world-class supercomputers to researchers in the 25 European member countries. BSC operates MareNostrum, the most powerful supercomputer in Spain since its inception In March 2004.

The latest version, MareNostrum 4 (since July 2017) has a performance capacity of 13,7 Petaflop/s and is composed of two distinct parts. The general-purpose element, provided by Lenovo, has 48 racks with more than 3,400 nodes with next-generation Intel Xeon processors and a central memory of 390 Terabytes. Its peak power is over 11 Petaflop/s, i.e. it is able to perform more than 11,000 trillion operations per second, ten times more than MareNostrum 3 despite costing only a 30% increase in energy consumption. The second element of MareNostrum 4 is formed of clusters of three different technologies that will be added and updated as they become available. These are technologies currently being developed in the USA and Japan to accelerate the arrival of the new generation of pre-exascale supercomputers. MareNostrum 4 will have a disk storage capacity exceeding 10 Petabytes and will be connected to the Big Data infrastructures of BSC, which have a total capacity of 24.6 Petabytes. BSC has also other cutting-edge computing infrastructure based on latest available technology like FPGA boards, small clusters based on ARM SoCs, GPUs, etc.

Deutsches Klimarechenzentrum (DKRZ)

The German Climate Computing Centre, DKRZ, is a national German facility, providing state-of-the-art super-computing, data and other associated services to the German and also the international scientific community to conduct top of the line Earth System and Climate Modelling. DKRZ operates a fully scalable supercomputing system designed for and dedicated to earth system modelling including mass storage system to a capacity of more than 500 PBs. It provides its more than 1000 scientific users with the technical infrastructure needed for the processing and analysis of huge amounts of data from climate simulations, also covering training and support for related application software and data processing. DKRZ participates in many national and international projects aiming to improve the research infrastructure for climate modelling. DKRZ is operating with its ICSU World Data Center Climate (WDCC) a certified, community specific long-term data archive. Connected with its WDCC, DKRZ provides best practise examples in scientific data life cycle management for the Earth system research community (federated data infrastructures, long-term archiving service, grid-based data processing workflows). Through its research group on scientific computing DKRZ is linked to the Department of Informatics of the University of Hamburg. DKRZ is partner in IS-ENES (European Network for Earth System Modelling) and is one of the founding partners of EUDAT-CDI. In cooperation with the Centro Euro-Mediterraneo sui Cambiamenti Climatici (CMCC), it contributes the "ENES Climate Analytic Service" (ECAS) to EOSC-hub, which enables end users to carry out data analyses on large amounts of climate data on the basis of a PID-capable and server-side approach. DKRZ developed and maintains the EOSC-hub services B2FIND and B2HANDLE; in addition it leads the 'Discovery and Access' task in EOSC-hub.

Role in the project
DKRZ participates with different roles, both as a Community representative and as an EOSC-hub service provider. Due to its key knowledge of B2FIND software and workflow it will extent the search portal with semantic capabilities of B2TERM, thus engaging in WP3 with the lead of Task 3.3. In WP4 DKRZ is engaged in all Tasks (4.1, 4.2, 4.3 and 4.4), for integrating B2TERM into disciplinary data infrastructures (Earth System Grid Federation) as well as for developing use cases for B2FIND/B2TERM interaction that serve facilitated data exploration and integrating a Data Subscription Service (DSS). DKRZ will lead the subtask 4.3.2 and 4.4.2. As a representative of Climate Modelling Community DKRZ will test and validate B2FIND and B2TERM services.
Key personnel
Hannes Thiemann [m] is the Deputy Department Head of the Data Management Department at DKRZ. He leads the long-term archiving group which is responsible for the operation and further development of the ICSU World Data Center Climate at DKRZ. Following his degree as geophysicist he joined DKRZ in 1990 and specialized in data management soon after. He was involved in many national and international projects, most recently EUDAT2020, SeaDataCloud and IS-ENES3. In EUDAT he was the product manager of the metadata service B2FIND for many years. In EOSC-hub he is elected member of the project management board. **Heinrich Widmann [m]**, Diploma in physics, is a scientific programmer at DKRZ. Following his diploma, he worked within compute intensive research projects and afterwards in software industry. Since 2003 he worked in different positions as a scientific programmer at the Max-Planck-Institute for Meteorology and DKRZ. He has 10 years+ experience in computing and research data management in the environment of earth system modelling. He has very good skills in data management tools, high volume and performance storage systems and data portal technologies. He was product manager and one of the main developers of EUDAT service 'B2FIND'and is now engaged in EOSC-hub as leader of the task 'Data discovery and access'. **Claudia Martens [f]** has (besides others) a background in library and information science and is the current responsible lead for the B2FIND service. She worked for EUDAT service integration of B2SHARE and B2SAFE as well as for Research Infrastructure integration (e.g. ENVRIPlus) with B2FIND. Within EOSC-hub project she is responsible for representing a wide range of sientific outcome from divergent Communities which includes a strong experience in metadata management, both on community specific level as on an interdisciplinary one. She is also interacting with other EOSC-hub infrastructures (e.g. EGI services as Datahub) and EOSC projects (e.g. PaNOSC) in terms of service development and integration.
Publications
• Brase J, Lautenschlager M, Sens I (2015) The Tenth Anniversary of Assigning DOI Names to Scientific Data and a Five Year History of DataCite. D-Lib Magazine Vol. 21, No. 1-2, https://doi.org/10.1045/january2015-brase • Klump J, Bertelmann R, Brase J, Diepenbroek M, Grobe H, Höck H, et al. (2006) Data publication in the open access initiative. Data Science Journal 5:79–83, http://doi.org/10.2481/dsj.5.79 • Weigel T, Lautenschlager M, Toussaint F, Kindermann S (2013) A Framework for Extended Persistent Identification of Scientific Assets. Data Science Journal, Vol. 12, pp 10-22, https://doi.org/10.2481/dsj.12-036

Relevant projects

- EUDAT (EU FP7, H2020), European Data Infrastructure
- IS-ENES/IS-ENES2 (EU FP7), IS-ENES3 (EU H2020), Infrastructure for the European Network for Earth System Modelling
- SeaDataCloud (EU H2020), Further developing the pan-European infrastructure for marine and ocean data management
- EOSC-hub (EU H2020), Integrating and managing services for the European Open Science Cloud

Relevant infrastructure

DKRZ is running one of Germany's most powerful high performance computers and world class data storage and archiving hardware. The current system is a BULL system with 80000 cores and 3.2 PetaFLOPS. The file system is based on lustre and has a net storage capacity of 50 PetaBytes. For long term archiving of data DKRZ operates a tape library with the capacity to store more than 500 PetaByte of data.

Services: DKRZ supports complex and compute intensive national and international collaborative projects: simulations are carried out and resulting data are managed and made available through the World Data Center Climate (WDCC) operated by DKRZ. Data nodes and data portal of the Earth System Grid Federation (ESGF) are run by DKRZ.

EUDAT Ltd (EUDAT)

EUDAT ltd is a not-for-profit service company and value-added reseller of data management services, incorporated in Finland and formally established on February 2018. It is the legal voice of the well-established EUDAT Collaborative Data Infrastructure (CDI) and coordinates a partner network of 26 European organisations. EUDAT's goal is to provide the "digital data layer" to underpin the whole European research landscape and support the policy goals of frictionless Open Science across Europe. EUDAT ltd delivers comprehensive research data management solutions for European researchers and research infrastructures through its partner network, following a model of tailoring solutions in close collaboration with end-users.

Role in the project

EUDAT Ltd. contributes to the sustainability plan, the operation model and the provisioning of the KOINE service as an EOSC service through the EOSC Service Catalogue. EUDAT Ltd. activities are in tasks T6.1 and T6.3.

Key personnel

Dr. Damien Lecarpentier [m] is a Programme Director for research data at CSC and has been managing the EUDAT initiative since its inception in 2011 as Project Manager of EUDAT (2011-2015), Project Director of EUDAT2020 (2015-2018), and Head of the EUDAT Collaborative Data Infrastructure (CDI) Secretariat as of September 2016. Damien has an MA in Political Science and a PhD in Social Sciences from the Advanced School of Social Sciences in Paris. He joined CSC in 2009 as coordinator for international activities and was involved as work package and task leader in

several FP7-funded projects in the areas of grids (EGI_DS), HPC (DEISA2, PRACE, EESI) and e-Infrastructure policy (e-IRG e-Infranet), before taking responsibility of EUDAT. Damien has also contributed to the establishment of the Research Data Alliance (RDA) in Europe, through the iCORDI and the RDA support projects, and is currently actively involved in activities related to the European Open Science Cloud via the EOSCpilot, the EOSC-Hub, and the EOSCsecretariat.eu projects.

Mark van de Sanden [m] is the Technical Coordinator of EUDAT Ltd. and member of the EUDAT CDI Secretariat and has been the work package leader of the service building activities within EUDAT and the EUDAT2020 projects. Mark is also working as a System Architect at SURFsara within the data preservation services group. He has more than 20 years of experience in managing computers, supercomputers and large scale data infrastructures and 10 years in coordinating the SURFsara data services. He has been member of the management board of the EPIC consortium, and involved in the WLCG NL-T1, LOFAR Long Term Archive and PRACE data challenges as data and storage expert. He has a BSc in Computer Engineering.

Publications

- Riedel M, Wittenburg P, Reetz J, van de Sanden M, Rybicki J, von St. Vieth B, Fiameni G, Mariani G, Michelini A, Cacciari C, Elbers W, Broeder D, Verkerk R, Erastova E, Lautenschlaeger M, Budig R, Thielmann H, Coveney P, Zasada S, Haidar A, Buechner O, Manzano C, Memon S, Memon S, Helin H, Suhonen J, Lecarpentier D, Koski K, Lippert Th (2013) A Data Infrastructure Reference Model with Applications: Towards Realization of a ScienceTube Vision with a Data Replication Service'. Journal of Internet Services and Applications, 4:1, https://doi.org/10.1186/1869-0238-4-1
- Gentzsch W, Lecarpentier D, Wittenburg P (2014) Big Data in Science and the EUDAT Project. Global, Conference (SRII), 2014 Annual SRII Proceedings, April 2014, p. 191 - 194, https://doi.org/10.1109/SRII.2014.34
- Lecarpentier D, Michelini A, Wittenburg P (2013) The building of the EUDAT Cross-Disciplinary Data Infrastructure in EGU General Assembly Conference Abstracts, p. 7202, Vol. 15, 2013
- de Witt S, Lecarpentier D, van de Sanden M, Reetz J (2017) EUDAT - A Pan-European Perspective on Data Management, Conference: 2017 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC), https://doi.org/10.1109/NSSMIC.2017.8533053
- Berenji Ardestani S, Hakansson CJ, Laure E, van de Sanden M et al. (2015) B2SHARE: An Open eScience Data Sharing Platform, Conference: 2015 IEEE 11th International Conference on e-Science (e-Science), https://doi.org/10.1109/eScience.2015.44

Relevant projects

- **EUDAT(2011 – 2015) – EUDAT2020 (2015 – 2018**) aimed to establish the European data e-infrastructure by providing an integrated, cost-effective, and sustainable pan-European solution for sharing, preserving, accessing and performing computations with primary and secondary research data.
- **EOSC-hub** is an EU H2020 funded project. EOSC-hub creates the integration and management system of the future European Open Science Cloud that delivers a catalogue of services, software and data from the EGI Federation, EUDAT CDI, INDIGO-DataCloud and major research e-infrastructures. This integration and management system (the Hub) build on mature processes, policies and tools from the leading European federated e-Infrastructures to cover the whole life-cycle of services, from planning to delivery.
- **SeaDataCloud** is an EU H2020 funded project which aims at advancing the SeaDataNet services and increasing their usage, adopting cloud and high-performance computing technology for better performance. EUDAT is responsible for advancing the SeaDataNet

infrastructure with advancing the data infrastructure and integrating the data infrastructure with planned Virtual Research Environment.

Relevant infrastructure

The EUDAT CDI is the European e-infrastructure consisting of 26 partners delivering the collaborative data infrastructure providing integrated data services (b2share, b2safe, b2drop, b2handle, b2note, b2find and b2access) and resources to support research. This infrastructure and its services have been developed in close collaboration with over 50 research communities spanning across many different scientific disciplines and involved at all stage of the design process.

CSC - IT Center for Science Ltd. (CSC)

CSC is a Finnish center of expertise in ICT, providing services at an internationally high quality, for research, education, culture, public administration and enterprises, to help them thrive and benefit the society at large.

CSC is owned by the Finnish Government and the Higher Education Institutions. Established in 1971, CSC currently has over 320 employees, and supports a European-wide customer base of thousands of researchers in disciplines such as biosciences, environmental science, linguistics, physics, chemistry and mathematical modelling, as well as social sciences and humanities. CSC is listed on the Finnish Research Infrastructure Roadmap, hosting four national RIs: the CSC RI, including the national research network Funet, the energy efficient data centre Kajaani, and the HPC and Cloud computing facility; the Finnish node of the biomedical computing infrastructure ELIXIR; PRACE Finland; and a National Research Data Platform.

The CSC RI provides Finland's widest selection of scientific software and databases, training and expert support, and storage and data services. CSC has been in central role in many research data related European projects acting as initiator, work package leader and coordinator in several projects, such as EUDAT and RDA Europe. At the moment CSC coordinates the EUDAT Collaborative Data Infrastructure. CSC is a partner in the EOSCpilot project. It leads the governance work package, and contributes to the services, policy and engagement and dissemination work packages. CSC is also a key partner (providing Project Director and leading WP) in the EOSC-hub project (2018-2020) that creates the integration and management system of the future European Open Science Cloud that delivers a catalogue of services, software and data from the EGI Federation, EUDAT CDI, INDIGO-DataCloud and major research e-infrastructures. CSC is a partner in the environmental science cluster projects ENVRI (FP7) and ENVRIPlus (H2020).

On national level CSC has been in a central role in building the FAIR compliant research data management services, as well as supporting several national ESFRI research infrastructure communities in the area of environmental science, bioscience and social sciences.

Role in the project
CSC develops and integrates the data subscription service as well as supports its use in use cases within WP4. Task leader for T4.4.
Key personnel
Mr. Antti Pursula [m], M.Sc. in Applied Mathematics, has worked at CSC since 2001, currently as Program Director in Research Infrastructures unit, leading programs for research with sensitive data & environmental science e-infrastructures. His earlier work experience at CSC includes development and management of scientific software, working as Development manager for software solutions group (2007-2010) and as Director for application services unit (2010-2013). Currently (since 2014) he serves as the Project director of the Nordic 6M€ Tryggve project, which develops and integrates IT services for sensitive data in biomedicine. He has managed CSC's participation in ENVRI and ENVRIPlus projects (2011-2019), and is a member of the Executive board of ENVRIPlus. In addition Mr. Pursula participates in developing the European Open Science Cloud data policies and services for sensitive data in the EOSC-hub project as well as in the EUDAT Collaborative Data Infrastructure. He has wide experience on project and task management in international setting. **Mr. Christopher Ariyo [m]**, Lic.Sc. in Metallurgy, has worked at CSC since 1999, currently as Service Manager for EUDAT services in Research Data Services group of Research Infrastructures unit, leading the EUDAT team working in SeadDataCloud, EOSC-hub (data discovery, access and metadata management), EOSC Competence Centers (Marine, ELIXIR, EISCAT_3D), and various other projects (ENVRIplus, Marinet). As part of this activity Mr. Ariyo has contributed to designing the EUDAT data subscription service and steered this development effort. His earlier work experience at CSC includes Service Area Manager for EUDAT for Data Access and Re-use (2016-2018), leading EUDAT partners team in SeaDataCloud (2016- 2018), development manager position for a storage platforms for online and offline storage including backup and restore (2010-2013) and have System Specialist from 1999-2013 and from 2013- Senior System Specialist in Storage management (online, offline and backup). Have had tremendous experience in system, storage, project and team management through years experience in working in the different fields.
Publications
• Koulouzis S, Carval T, Martin P, Grenier B, Chen Y, Heikkinen J, Zhao Z (2018) Dynamic Optimization for Time-critical Data Services: A Case Study in Euro-Argo Research Infrastructure, 20th EGU General Assembly, EGU2018, Proceedings from the conference held 4-13 April, 2018 in Vienna, Austria, p.16012 • Asmi AJ, Konijn J, Pursula A (2014) European environmental research infrastructures are going for common 30 years strategy, Geophysical Research Abstracts, Vol. 16, EGU2014-5743, 2014, EGU General Assembly 2014, Full report: https://doi.org/10.6084/m9.figshare.2067537.v1 • Gentzsch W, Lecarpentier D, Wittenburg P (2014) Big Data in Science and the EUDAT Project, Global, Conference (SRII), 2014 Annual SRII Proceedings, April 2014, p. 191 - 194, http://doi.org/10.1109/SRII.2014.34 • Lecarpentier D, Michelini A, Wittenburg P (2013) The building of the EUDAT Cross-Disciplinary Data Infrastructure in EGU General Assembly Conference Abstracts, p. 7202, Vol. 15 • Lecarpentier D, Reetz J, Wittenburg P (2013) EUDAT: Toward a Pan-European Collaborative Data Infrastructure (Case Study), T Critchlow and K. Kleese van Dam editors, Data-Intensive Science, Chapman & Hall/CRC Computational Science

Relevant projects

- Work package leader in the EOSCPilot project (2017-2018).
- Key partner with Project director and WP leader in EOSC-hub project (2018-2020)
- Participant and task leader on ENVRI (2011-2014) & participant and Executive Board member in ENVRIPlus (2015-2019) projects
- Coordinator of EUDAT (2012-2015) and EUDAT2020 (2015-2018) projects
- Work package leader in the FAIRsFAIR project (2019-)

Relevant infrastructure

CSC operates the national HPC Centre; a energy efficient datacentre facility; the national research and education network Funet; a national research data platform and services; and a national Cloud resource for research. CSC is the Finnish node of the biomedical research infrastructure ELIXIR, and a technical centre for the language research infrastructure Fin-CLARIN. CSC's cloud service ePouta is a cloud computing environment (infrastructure as a service, IaaS) designed for processing sensitive data. CSC is initiator of the EUDAT collaboration and coordinator of the EUDAT Collaborative Data Infrastructure, and the associated EC projects EUDAT and EUDAT2020, with a suite of five integrated services - B2SHARE (store and share research data), B2DROP (synchronize and exchange research data), B2FIND (find collections of scientific data quickly and easily), B2SAFE (manage data and optimize access), and B2STAGE (transfer data to HPC workspaces for processing).

Ifremer (Ifremer)

Ifremer, through its research work and expert advice, contributes to knowledge of the oceans and their resources, to monitoring of marine and coastal environments and to the sustainable development of marine activities. To these ends, Ifremer conceives and operates tools for observation, experimentation and monitoring, and manage oceanographic databases. It also operates a great part of the French ocean research fleet, including all underwater systems and large-scale mobile facilities and equipment. Ifremer performs targeted applied research to address the questions posed by society (climate change effects, marine biodiversity, pollution prevention, seafood quality, etc.). Results include scientific knowledge, technological innovations, and systems for ocean observation and exploration.

The department of "Research Infrastructures and Information Systems" provides services and carries out projects in order to create or improve the marine research infrastructures such as the French Oceanographic Fleet, Argo observing system, High Performance Computers for data processing and ocean modelling. It also promotes research on the infrastructure under its care, to relevant national and international funding agencies. The “Information Systems and Oceanographic Databases” team develops and operates large Information systems and databases for national, EU-funded projects (DG research H2020, DG Mare, DG Growth with Copernicus) and international initiatives (ARGO) on its own e-infrastructures and shared European e-infrastructure in the scope of SeaDataCloud and ENVRIplus projects.

Ifremer is coordinating the Copernicus Marine In Situ TAC and is in charge of the operation of the Global component of the In Situ TAC. Ifremer is also involved in the System evolution activities for the T&S global and regional product assessment and the tools enhancement to connect to the Central Information System of the CIS (Copernicus Marine Central Information System).

Role in the project
Ifremer integrates B2TERM service in the Argo data and metadata discovery graphic user interface in WP4 Task 4.4.3.
Key personnel
Thierry Carval [m], head of the Coriolis Data centre, involved in the data management committee for main JCOMM networks (Argo, OceanSITES, EGO, GTSPP). He has been coordinating the development of the Copernicus Marine Environment Monitoring Service in situ TAC distributed architecture and its interoperability with existing international initiative with project such as ENVRIPLUS or AtlantOS. **Jérôme Detoc [m]** is an IT engineer, specialist on Cloud and NoSQL information systems. He is contributing to H2020 EOSC-hub Marine Competence Centre and ENVRIPLUS data management workpackage. He is leading the development of Ifremer sensors quality information system.
Publications
• Wong A. et al. (2018) Argo Quality Control Manual for CTD and Trajectory Data. http://doi.org/10.13155/33951 • Rannou P et al. (2018) Coriolis Argo floats data processing chain. SEANOE. http://doi.org/10.17882/45589 • Coppola L et al (2016) Dyfamed observatory data. SEANOE. http://doi.org/10.17882/43749 • Charria G et al (2014) PREVIMER: A contribution to in situ coastal observing systems. Mercator Ocean - Quaterly Newsletter, 49, 9-20. Open Access version : http://archimer.ifremer.fr/doc/00197/30785/ • Carval T et al. (2016) IQuOD 4th workshop in Tokyo: Copernicus Marine in situ TAC, a service for operational oceanography. The 4th Annual IQuOD Workshop - International Quality controlled Ocean Database. http://archimer.ifremer.fr/doc/00353/46410/
Relevant projects
• SeaDataCloud (H2020 INFRAIA-01-2016, GA no. 730960, 2016-2020, 10 M€, 56 partners, 29 countries): Further developing the pan-European infrastructure for marine and ocean data management. IFREMER coordinates the project and especially contributes to the development of virtual research environment in WP10, together with CSC, DKRZ and CINECA. • ENVRIplus (H2020 INFRADEV-4-2014-2015, GA no. 654182, 2015-2019, 15 M€, 38 partners, 13 countries): Environmental Research Infrastructures Providing Shared Solutions for Science and Society. IFREMER represents several Marine Research Infrastructures and is in charge of task 8.2 on cataloguing in theme 2 “data for science”. In addition, IFREMER is responsible for 2 uses cases in work package 9, for sensor registry and euro-ARGO data subscription which is used as a foundation for the platform proposed in this project. • ODIP (FP7-INFRA-2012-3.2, GA no. 312492, 2012-2015, 0.7 M€, 10 partners, 6 countries): Establishing and operating an Ocean Data Interoperability Platform. The project aims at fostering collaboration between USA, Australia and Europe on Marine e-infrastructure. Digital playground and Virtual Research Environment are especially prototyped in the project. IFREMER coordinate the impact assessment work package and contributes to the development and animation of the project. • Copernicus MEMS, DG Growth, Copernicus Service operated by Mercator Ocean. IFREMER coordinates the consortium in charge of collection and qualification of marine in-situ observations as sub-contractor of Mercator-Ocean.

Relevant infrastructure
Euro-ARGO ERIC: Ifremer hosts the ERIC and develops and operates the database used to manage one of the two Global Data Assembly Centres. IFREMER is also represented in the science steering team and very active in this field.

Institut National de la Recherche Agronomique (INRA)
INRA, with around 10,000 permanent or contractual staff, is a French public research institution for FNS: agriculture (including forests), food and food security, environment and land management, with a particular emphasis on sustainable development. The missions of INRA are to: (i) serve the public interest by maintaining a balance between excellence of research and the demands of society; (ii) produce and disseminate scientific knowledge and innovation, particularly for FNS; (iii) contribute to the expertise, training, promotion of scientific and technical culture and science/society debate. INRA is also participating actively in the open science movement with an incentive open access policy and involvement in citizen science projects. INRA has a long-standing engagement and expertise in Open Science (http://2025.inra.fr/openscience_en) and has created platforms for training and resources for Open Science: data centers, software forges, etc. The data resources of INRA (such as those gathered by the portal http://data.inra.fr and complying to the policy disseminated through the Website http://datapartage.inra.fr) will not all be listed here but they cover the entire FNS domain, ranging from the environment of food production to health and including agriculture, food processing, nutrition, etc. The Department of Scientific Information (DIST) works transversely to support INRA's scientific strategy and provide innovative services within the areas of data, information and knowledge management and sharing. INRA-DIST develops tools and services to support research activities and promotes open access to scientific and technical information –both publications and data–. INRA-DIST members are also involved in several working groups of the Research Data Alliance (RDA, https://www.rd-alliance.org/) as co-chair (Agricultural data interest group, Wheat data interoperability working group, Agrisemantics working group) or participant. INRA-DIST is a partner of AgInfra+ ("Accelerating user-driven e-Infrastructure innovation in Food & Agriculture", http://plus.aginfra.eu/, 2017-2019, H2020 grant #731001), a project to exploit core e-Infrastructures such as EGI.eu, OpenAIRE, EUDAT and D4Science, towards the evolution of a unified data infrastructure, so as to provide a sustainable channel addressing adjacent but not fully connected user communities around Agriculture and Food. More recently, INRA-DIST has coordinated e-Rosa ("e-Infrastructure Roadmap for Open Science in Agricultural and Food Sciences", http://www.erosa.aginfra.eu/, 2017-2018, H2020 grant #730988) to deliver said roadmap in the summer 2018. INRA is also leading a GO FAIR IN to improve interoperability in Food systems. It takes part in the project primarily through the Department of Scientific Information (DIST) which works transversely to support INRA's scientific strategy and provide innovative services within the areas of data, information and knowledge management and sharing. DIST develops tools and services to support research activities, and promote open access to scientific and technical information, both publications and data. DIST members are also involved in many groups of the research data alliance as co-chair (Agricultural data interest group, Wheat data interoperability working group, Agrisemantics working group) or participants.
Role in the project
In WP2, INRA will participate in the landscaping activities on both relevant terminologies and

repositories. Outputs of the Agrisemantics Working Group will be considered. Based on its previous collaboration with LIRMM on AgroPortal, INRA will also contribute to the task 2.2 dedicated to standardized metadata. INRA will put major efforts on WP4 in interconnecting the institutional data repository Data Inra (https://data.inra.fr), based on the Dataverse technology to B2TERM. INRA will enrich metadata of datasets (and possibly datasets themselves) recorded in Data Inra using terminology services from B2TERM. Relevant results will be sent to B2FIND. INRA will finally implement and test improved features for data discovery for datasets recorded in Data Inra.

Key personnel

Ms. Ir. Odile Hologne [f] is the head of the Department of Scientific Information at the National Institute for Agricultural Research (INRA-DIST). For many years, she held positions related to IT development, as well as information management and publishing in public institutions under the French Ministry of Agriculture. In her current position, she is involved in various national and international working groups on e-science and open science (open access to publications or to research data). She was one of the members of the senior team of the RDA Europe 3rd phase. She was the coordinator of the e-ROSA H2020 project "e-Infrastructure roadmap for open science in agriculture".

Ms. Sophie Aubin [f] (https://orcid.org/0000-0003-0870-3192) is a linguistics engineer at the Scientific and Technical Information Department of Inra, with a specialization on Semantics and Text Mining. She has a Master degree in Computational Linguistics from INALCO, France. Currently, she is responsible for developing vocabulary management and information analysis services for researchers. She has participated in the H2020 eRosa project which produced a shared vision of a future sustainable e-infrastructure for research and education in agriculture. Sophie is co-chairing the Agrisemantics Working Group in the Research Data Alliance which has just released its recommendations for a larger adoption of semantic approaches to improve data discoverability and interoperability. She will contribute to WP2 landscaping and standard metadata activities.

Ms. Ir. Esther Dzale Yeumo [f] is chair of the data competence centre to support INRA's data management and sharing policy and services for research units. Esther is also co-chair of the RDA Agricultural data interest group and Wheat Data Interoperability working group which delivered framework for describing, representing and linking Wheat data. She is involved in the development of Agroportal, an ontology repository dedicated to the agronomic and plant domains.

Dimitri Szabo [m] (https://www.linkedin.com/in/dimitriszabo/) is an Agronomist and Information Technologies specialist. He has a Master's degree in Agronomy, with a specialty in Information and Communication Technologies at Montpellier SupAgro. He had experiences in Data Management Solutions for Biology and Agronomy Applied Research. He is currently working at the Institut National de la Recherche Agronomique (INRA) on the data repository of the institute: https://data.inra.fr. He will lead INRA's contribution to WP4.

Publications

- Dzale Yeumo WE, Alaux M, Arnaud E, Aubin S, Baumann U, Buche P, Cooper L, Ćwiek-Kupczyńska H, Davey RP, Fulss RA, Jonquet C, et al. (2017) Developing data interoperability using standards : a wheat community use case in F1000Research, Vol 6 : 1843, https://doi.org/10.12688/f1000research.12234.1
- Jonquet C, Toulet A, Arnaud E, Aubin S, Dzalé Yeumo E, Emonet V, Graybeal J, Laporte M-A, Musen MA, Pesce V, Larmande P (2018) AgroPortal: an ontology repository for agronomy. Computers and Electronics in Agriculture, 144:126–143,

https://doi.org/10.1016/j.compag.2017.10.012

- Caracciolo C, Aubin S, Whitehead B, Zervas P (2018) Semantics for Data in Agriculture: a Community-based Wish List, https://doi.org/10.31220/osf.io/eapdv
- Zervas P, Manouselis N, Karampiperis P, Hologne O, Janssen S, et al. (2018) e-ROSA D3.7 Foresight Roadmap Paper, https://doi.org/10.5281/ZENODO.1479659
- Aubin S, et al. Reports of the RDA agrisemantics working group for data interoperability https://rd-alliance.org/groups/agrisemantics-wg.html

Relevant projects

- INRA is a partner of AgInfra+ ("Accelerating user-driven e-Infrastructure innovation in Food & Agriculture", http://plus.aginfra.eu/, 2017-2019, H2020 grant #731001), a project to exploit core e-Infrastructures such as EGI.eu, OpenAIRE, EUDAT and D4Science, towards the evolution of a unified data infrastructure, so as to provide a sustainable channel addressing adjacent but not fully connected user communities around Agriculture and Food.
- INRA was the coordinator of e-ROSA ("e-Infrastructure Roadmap for Open Science in Agricultural and Food Sciences", http://www.erosa.aginfra.eu/, 2017-2018, H2020 grant #730988), listed in the current call as one of the "relevant projects" on which the Food Cloud Demonstrator should build. e-ROSA delivered its roadmap in the summer 2018.
- INRA was a partner of OpenMinTeD ("Open Mining Infrastructure for Text and Data", http://openminted.eu/, 2015-2018, grant #654021), a project to create an open mining infrastructure for scientific publications and data, in order to ensure the availability of FAIR new or existing mining tools and platforms.
- INRA is a partner of ANR D2KAB (www.d2kab.org – ANR-18-CE23-0017). To create a framework to turn agronomy and biodiversity data into FAIR knowledge in agronomy and biodiversity. The project will focus on ontology services and mappings and linked data exploitation. Led by C. Jonquet (LIRMM Univ Montpellier).
- Inra is a partner of VisaTM supported by French MESRI minister. A partner project of the H2020 OpenMinTed project (http://openminted.eu) which builds a shared infrastructure for text and data mining. Led by C. Nédellec (INRA).

Relevant infrastructure

- Inra data portal - http://data.inra.fr
- Inra information portal for disseminating best practices for data management and sharing - http://datapartage.inra.fr
- Map of the Agri-food data ecosystem, output of the e-ROSA H2020 project - http://map.aginfra.eu/

TU Delft

Delft University of Technology (TUD) is an applied science and engineering institution of higher education and research in the Netherlands. According to International Reputation Rankings, TU Delft is currently the highest ranked Dutch university and one of the worldwide leading universities. TU Delft hosts over 18,000 students (undergraduate and postgraduate), more than 3,000 scientists and more than 2,000 people in the support and management staff.

TU Delft is a leading Dutch university in Open Science, embedding open access and publishing, FAIR data management and data stewardship, open education and open source software. It led the creation of the Dutch National Plan for Open Science, which is now being rolled out as a platform across the country. TU Delft is also part of the OpenAIRE project consortium.
TU Delft Library also hosts the 4TU.Centre for Research Data, the data archive for the technical sciences in the Netherlands. It hosts nearly 8,000 datasets from scientists in and outside the Netherlands, many in the NetCDF format. TU Delft has also recently published its Research Data Framework Policy (autumn 2018 http://bit.ly/TU-Delft-Research-Data-Framework-Policy), clarifying high-level roles and responsibilities for the handling of data, and the related training; faculty-based policies will follow in 2019. TU Delft is also one of the first universities (nationally or internationally) to embed a sophisticated network of Data Stewards to provide discipline-based support for researchers (http://bit.ly/DataStewardship) The Data Stewards provide ongoing advice and training, and run an ongoing survey on attitudes towards research data (https://openworking.wordpress.com/?s=tableau).

Role in the project

TU Delft is the co-leader of WP5, which is focused on requirements gathering and capacity building. TU Delft will leverage its extensive networks within the engineering communities: the CESAER association of universities of science and technology in Europe (where TU Delft is part of the task force on Open Science), the connection with the coastal engineering communities (and access to the Dutch SMEs through the Deltares independent research institute) and the partnership with the Amsterdam Metropolitan Solutions Institute (interdisciplinary urban studies). In WP5 TU Delft will first lead the task of the KOINE Stakeholder Forum development and maintenance. Through the forum, it will help with the requirements gathering for the KOINE services and also services validation. Finally, TU Delft will take the lead in the capacity building and training task: leading the development of open educational training resources and the delivery of carpentry-style workshops tailored for the research community.

TU Delft is also heavily involved in WP4, where it will take part in integrating KOINE services into its data archive (4TU Centre for Research Data) for technical sciences. In addition, TU Delft will annotate metadata in 4TU.Centre for Research Data with B2TERM terminologies and will lead the task on sharing metadata enriched with semantic annotations (leveraging our experience and expertise working with NetCDF disciplinary data format). TU Delft will also participate in the task focusing on improving research data discoverability through the use of KOINE services.

Finally, TU Delft is also involved in developing recommendations for annotating research data and metadata with B2TERM terminologies (part of WP2).

Key personnel

Alastair Dunning [m] is Head of Research Data Services at TU Delft, and also Head of the 4TU.Centre for Research Data. While at TU Delft, Alastair has overseen one of the most far-reaching institutional programmes for research data management (see description above). He is on the advisory board of the National Coordination Point for RDM, and also of the National Platform for Open Science. Previously he has run several European projects relating tp cultural heritage and technical infrastructure. This has included the Europeana Cloud project and the Europeana Libraries project. He was also Work Package leader for the task to assemble and deliver over 12m pages of digitised historical papers as part of Europeana Newspapers.

Marta Teperek [f] is the Data Stewardship Coordinator at TU Delft. She oversees the data stewardship project and the network of subject-specific data stewards at TU Delft. Before joining TU Delft, Marta led the creation and development of the Research Data Management Facility at the

University of Cambridge to support researchers at the University of Cambridge in good management and sharing of research data. Marta serves on the Editorial Board of the Data Science Journal and Co-Chairs the Research Data Alliance Libraries for Research Data Interest Group. She is a frequent speaker at international conferences and events related to Open Science and to Research Data Management, regularly publishes blog posts and peer-reviewed publications on these topics, and is professionally active on Twitter. Marta is a scientist by training and completed her PhD in molecular biology at the University of Cambridge.

Publications

- Dunning A, de Smaele M, Bohmer J (2017) Are the FAIR Data Principles fair? International Journal of Digital Curation, Vol 12 No 2, https://doi.org/10.2218/ijdc.v12i2.567
- Dunning A, Bernardou A (2017) From Europeana Cloud to Europeana Research: Tools, Users and Methods. Cultural Heritage Infrastructure in Digital Humanities, http://eprints.gla.ac.uk/158066/
- Teperek M, Cruz MJ, Verbakel E, Böhmer JK, Dunning A (2018) Data Stewardship – addressing disciplinary data management needs, https://doi.org/10.31219/osf.io/5w9pj
- Cruz MJ, Böhmer JK, Gramsbergen E, Teperek M, de Smaele M, Dunning A (2018) From Passive to Active, From Generic to Focused: How Can an Institutional Data Archive Remain Relevant in a Rapidly Evolving Landscape?, https://doi.org/10.31219/osf.io/jgrkb
- TU Delft Data Stewards, EPFL Library Research Data Team, Krause J, Lambeng N, Andrews H, Boehmer J, Teperek M (2018) Quantitative assessment of research data management practice (Version 3) [Dataset]. Zenodo. http://doi.org/10.5281/zenodo.1168800

Relevant projects

- Cruz MJ, Gramsbergen E (2018, April 30). Adding Value and Facilitating Data Reuse: the Case of the 4TU.Centre for Research Data. https://doi.org/10.31219/osf.io/rvfs2
- Teperek M (2018) Research Data Management in Engineering disciplines. http://doi.org/10.5281/zenodo.1478516
- Cruz, MJ, Kurapati S, Turkyilmaz-van der Velden Y, (2018, October). The Role of Data Stewardship in Software Sustainability and Reproducibility. Zenodo. http://doi.org/10.5281/zenodo.1472945
- den Heijer k (2018, October). FAIR science following the onion model. Zenodo. http://doi.org/10.5281/zenodo.1472284
- TU Delft is also part of the Open AIRE consortium (https://www.openaire.eu/tudelft)

Relevant infrastructure

As mentioned above, TU Delft has an advanced institutional embedding of research data management and is host of the 4TU.Centre for Research Data, the Dutch data archive for the technical sciences.

OEAW

The Austrian Academy of Sciences (OeAW) is Austria's largest non-university research facility with 1600 employees looking back on a history of over 170 years. The Academy's 28 institutes perform applied and basic research in many different disciplines ranging from archaeology, cultural sciences, linguistics, and literary studies to quantum physics and molecular

biology at the highest international level. The Academy is represented in the project by one of its institutes, the Austrian Centre for Digital Humanities (ACDH-OeAW), which was founded to foster the change towards digital paradigm in the humanities. The ACDH-OeAW pursues a dual agenda of conducting digitally-enabled research and providing technical expertise and support to the research communities at the Academy, as well as on the national and international level. Conducting numerous outreach activities, it promotes knowledge sharing and the build-up of social infrastructures. The ACDH-OeAW is the coordinating institution in Austria of both ESFRI research infrastructure consortia in the humanities (CLARIN and DARIAH). It is involved in numerous national and international projects (e.g., PARTHENOS, ELEXIS, SSHOC).

Role in the project

Within KOINE, ACDH-OeAW will be the representative for CLARIN and the SSH domain in general.

Key personnel

Matej Ďurčo [m] is head of the ACDH's technical working group "Tools, Services & Systems" and was one of the key figures in founding the institute. Since 2002, while still studying computer science at the Technical University Vienna, he has been a team member of the ACDH's institutional predecessors, the Austrian Academy Corpus (AAC) and the Institute for Corpus Linguistics and Text Technology (ICLTT), engaging primarily in corpus linguistics and development of text technological applications. Since 2009, he has been part of the Austrian research infrastructures core group, contributing substantially to the development of key technical components such as the Component Metadata Infrastructure (CMDI), the Federated Content Search (FCS) and the Vocabulary Repository, both on the local and on the European levels. Currently, he acts as coordinator of the Metadata Curation Task Force in CLARIN and is co-head of the Virtual Competence Centre I (e-Infrastructure) in DARIAH. For his achievements in the build-up of CLARIN, he was awarded the first CLARIN Young Scientist Award in 2014 in Soesterberg (Netherlands). In his role as head of the ACDH's technical group, he has been coordinating the development of applications and the provision of services for the numerous projects of the institute and its cooperation partners. His focus lies on building and providing a consolidated portfolio of technical solutions, striving to offer the most appropriate tools for specific requirements of individual projects. His specific interests are semantic technologies, information visualisation and resource discovery.

Publications

- Ďurčo M, Lorenzini M, Sugimoto G (2018) Something will be connected - Semantic mapping from CMDI to Parthenos Entities. CLARIN. Selected papers from the CLARIN Annual Conference 2017. Budapest: Linköping University Electronic Press.
- Trognitz M, Ďurčo M (2018) One Schema to Rule them All. The Inner Workings of the Digital Archive ARCHE. Mitteilungen der Vereinigung Österreichischer Bibliothekarinnen und Bibliothekare, Bd. 71 (1), S. 217-231, https://doi.org/10.31263/voebm.v71i1.1979
- Ostojic D, Sugimoto G, Ďurčo M (2017) The Curation Module and Statistical Analysis on VLO Metadata Quality. In: CLARIN, (Hrsg.), Selected papers from the CLARIN Annual Conference 2016 Aix-en-Provence 26–28 October 2016 (CLARIN Annual Conference 2016 Aix-en-Provence 26–28 October 2016); Aix-en-Provence: Linköping University Electronic Press, S. 12.
- King M, Ostojic D, Ďurčo M, Sugimoto G (2016) Variability of the Facet Values in the VLO–a Case for Metadata Curation. In: CLARIN (Hrsg.), Selected Papers from the CLARIN Annual Conference 2015 (CLARIN Annual Conference 2015, October 14–16, 2015, Wroclaw, Poland); Wroclaw: Linköping University Electronic Press, S. 25–44.
- Wissik T, Ďurčo M (2016) Research Data Workflows: From Research Data Lifecycle Models to Institutional Solutions. In Selected Papers from the CLARIN Annual Conference 2015, October

14–16, 2015, Wroclaw, Poland, ed. K. De Smedt, 94-107. Linköping University Electronic Press, Linköpings Universitet.

Relevant projects

- 09.2015 – 12.2018, Humanities at Scale: Evolving the DARIAH-ERIC, improving DARIAH in fostering new and sustaining existing knowledge in digitally enabled research in the arts and humanities. Role ACDH: Definition of assessment criteria for in-kind contributions of partners to the consortium
- 09.2015 – 12.2017, CLARIN-PLUS: Following the recommendations of the 2013 ESFRI aiming to implement and strengthen CLARIN in the following areas: central (technical) hub, central office, partnerships with other infrastructures, outreach, governance. Role ACDH: Development of the Curation module and application for quality assessment of CLARIN metadata
- 05.2015 – 04.2019, H2020, Parthenos: Strengthening the cohesion of research in the broad sector of Linguistic Studies, Humanities, Cultural Heritage, History, Archaeology and related fields through a thematic cluster of European Research Infrastructures, integrating initiatives, and building bridges between different fields. Role ACDH: mapping of CLARIN metadata to common semantic framework, integration of specialized NLP services into the common e-Infrastructure platform, task lead on semantically assisted resource discovery services
- 01.2019 – 04.2022, H2020, SSHOC – Social Sciences and Humanities Open Cloud, a European open cloud ecosystem for social sciences and humanities (SSH) contributing to the Open Science agenda and the realisation of the European Open Science Cloud (EOSC). Role ACDH: contributing to task on making language resources FAIR, task lead on developing training material for fostering communities, empowering users and building expertise, task lead on implementation of SSHOC Marketplace - the central discovery service

Relevant infrastructure

- ACDH-OeAW, in cooperation with the computing centre of the Academy, runs numerous web applications, many custom-tailored for individual cooperation projects, but also a few applications that serve a specific task to the community outside of specific projects. Two such applications of central importance are ACDH Vocabs and ARCHE.
- ACDH Vocabs https://vocabs.acdh.oeaw.ac.at/en/ is a vocabulary repository for publishing controlled vocabularies used across various humanities projects.
- ARCHE https://arche.acdh.oeaw.ac.at/ is a service aimed at offering stable and persistent hosting as well as dissemination of digital research data and resources for the Austrian humanities community.

Australian National Data Service (ANDS)

Note: The Australian National Data Service is in the process of being merged and transitioned into the Australian Research Data Commons (ARDC). The ARDC is not yet a legal entity - once it is established legally this application will be updated. The relevant and current information on ANDS is included in the meantime.

Monash University (Australian National Data Service) (soon to be ARDC) is a leading Australian public research university based in Melbourne. Monash is home to major research facilities, including the Australian Synchrotron, the Monash Science Technology Research and Innovation

Precinct (STRIP), the Australian Stem Cell Centre, 100 research centres and 17 co-operative research centres. Monash University is the lead institution and legal entity for the Australian National Data Service (ANDS), which is the body participating in KOINE. The Australian government recognises the need for Australian researchers to have unprecedented access to data, enabling more efficient and more new research to be conducted in a richer data environment which can address significant research challenges. To enable this, the Australian National Data Service (ANDS) was established focusing on bringing about four transformations to data – from unmanaged to managed, from disconnected to connected, from invisible to findable and from single-use to reusable – that will enable Australia's research data as a whole to become a national strategic resource.
To this end ANDS is partnering with research institutions across Australia to create and build the Australian Research Data Commons (ARDC), a cohesive national collection of research resources to provide:

- A set of data collections that are shareable
- Descriptions of those collections
- An infrastructure that enables populating and exploiting the commons
- Connections between data, researchers, research, instruments and institutions

The ARDC is a meeting place for researchers and data, designed to make better use of Australia's research outputs; enable Australian researchers to publish easily, discover, access and use data; and enable new and more efficient research.
ANDS has initiated and supported a number of significant national services to enable researchers and research institutions to improve their research data management, leading to routine publication of their data with persistent identifiers into data stores that feed information to the ANDS collections registry. In addition, researchers will be able to find a wide variety of datasets through a variety of discovery paths. Most importantly ANDS will have engaged the research community to the extent that researchers see publishing their research data as their default practice.
The Australian National Data Service was established in January 2009. ANDS is supported by the Australian Government. The ANDS Project partners are Monash University (lead institution), the Commonwealth Scientific and Industrial Research Organisation (CSIRO) and the Australian National University (ANU).

Role in the project

ANDS (tbka ARDC) operates a standard vocabulary discovery and re-use infrastructure called "Research Vocabularies Australia". We will contribute to the KOINE project by:

- Testing standards-based interoperability across int'l services
- Validating EU approaches from an international perspective
- Exchanging best practices
- Reviewing materials

Key personnel

Dr. Adrian Burton [m] (http://orcid.org/0000-0002-8099-7538) is the Director, Services at ANDS. His portfolio responsibilities span ANDS IT services, ANDS skills agenda, and ANDS policy agenda. Part of this role is to oversee both the ANDS Online Services and National Service partnerships which provides standard vocabulary services for Australian eResearch groups. Adrian's research interest formerly included South Asian Linguistics.

ARDC service operations and development staff with expertise and experience in knowledge organisation infrastructure and services will contribute to the project on needs basis.

Publications

- Burton A, Aryani A, Koers H, Manghi P, La Bruzzo S, Stocker M, Diepenbroek M, Schindler U, Fenner M (2017) The Scholix Framework for Interoperability in Data-Literature Information Exchange. D-Lib Magazine 23(1/2), https://doi.org/10.1045/january2017-burton
- Brownlee R, Burton A, Frazier J and Walker R. (2016). Collaborative development of a multi-disciplinary research infrastructure for vocabulary creation, management, publication, discovery, access and re-use. VALA2016. http://www.vala.org.au/vala2016-proceedings/vala2016-session-11-brownlee
- Burton A, Koers H, Manghi P, La Bruzzo S, Aryani A, Diepenbroek M, Schindler U (2015) On Bridging Data Centers and Publishers: The Data-Literature Interlinking Service. MTSR 2015: 324-335, https://doi.org/10.1007/978-3-319-24129-6_28
- Burton A, Groenewegen D, Love C, Treloar AE, Wilkinson R (2012) Making Research Data Available in Australia. IEEE Intelligent Systems 27(3): 40-43 (2012), https://doi.org/10.1109/MIS.2012.57
- Burton A, Treloar AE (2009) Designing for Discovery and Re-Use: the 'ANDS Data Sharing Verbs' Approach to Service Decomposition. IJDC 4(3): 44-56, https://doi.org/10.2218/ijdc.v4i3.124

Relevant projects

ANDS has worked with research institutions and other partners on a series of major program streams incorporating hundreds of individual projects. Examples:

- Trusted Research Outputs (2016 - 2017): The Trusted Research Output projects are working with six of the NeCTAR VLs to capture and expose trust-enhancing information for a range of research outputs.
- Trusted Data Repositories (2016 - 2017): Implementing a number of trusted data repository services to understand what that involves for different disciplines and institutions. The intention is to improve the trustworthiness of data repository services for these institutions and disciplines, and to learn what is needed to move beyond simple data storage services. A number of the projects will also use the RDA/WDS Trusted Data Repository checklist to inform their development activity and assess the resulting services.
- High Value Collections (2016 - 2017): The High Value Collections (HVC) program builds on open data collections where ANDS will partner with 25 institutions to form collections of strategic significance to them, which are tied into national initiatives and other relevant NCRIS facilities. Most of these collections will be openly available, enabling the discovery of the rich data being produced and making optimal use of national facilities. The highlight of this program will be building capability at the institutions and forming collections that exploit Australia's research data infrastructure.
- Collection Enhancement Partnerships (2016 - 2017): ANDS is continuing to work on increasing the value of existing collections; through sharing and making data open, but also through other transformations, particularly enhancing the usage of data. ANDS is engaging in partnerships with selected institutions and contributes a combination of effort and funds, through Collection Enhancement Partnerships. The resulting enhanced research data collections will enable research over richer data and promote the data assets of the institution.
- Institutional Research Data Capabilities (2015 - 2017): Through the Institutional Research Data Capability Program ANDS has reached out to research organizations in areas that had not worked on research data through ANDS funded projects in the past. The organizations targeted in this program are the medical research institutes and the CRCs. Through this program ANDS has funded four exemplar projects addressing questions around research data typical for these organisations and these areas.

Relevant infrastructure
Research Vocabularies Australia (RVA) (https://www.ands.org.au/online-services/research-vocabularies-australia) is a service portal that makes it easy to find and use controlled vocabularies used in research. It also makes it possible for Australian research organisations to publish, re-purpose, create, and manage their own controlled vocabularies. Vocabularies change over time, so the service enables management of new versions while retaining superseded versions. Research Data Australia (RDA) is a web portal for discovering data collections produced by, or relevant to, Australian researchers. It helps to find, access, and reuse data for research from over one hundred Australian research organizations, government agencies and cultural institutions.

SISTEMA GmbH (SISTEMA)
SISTEMA GmbH is a privately-held company born in 2009 as extension of commercial and R&D activities of the Meteorological and Environmental Earth Observation (MEEO) company. SISTEMA is the official distributor of MEEO products, and it is also responsible for realizing specific research and development activities of these products. Strict contacts with University and Research Institutes ensure the maximum level of knowledge transfer and on-the-edge technologies availability: this is the philosophy that SISTEMA pursues any time the improvement of the existing products and services or the development of a new one is needed.
Role in the project
As a consumer, SISTEMA will evaluate B2TERM and its integration in research data infrastructures to determine if semantically annotated data help in information mining. Of particular relevance is information tailored to user groups other than scientists, such as policymakers, educators, journalists, citizen scientists.
Key personnel
Dr. Stefano Natali [m] holds an academic degree in Atmospheric Physics, and European Space Agency certifications in Space Component Engineering, Software Engineering Standards and Technologies, and Quality assurance. He has over 18 years' experience covering different roles (Managing director, Project Manager, Senior Analyst) in national and international projects. He has gained knowledge in developing advanced environmental services and algorithm. Moreover, he has a consolidated experience on access processing services for heterogeneous data (ground-measurement, numerical model, EO products, etc.) based on interoperability standards offered by OGC services (WFS, WMS, WCS).
Publications
• Baumann P, Mazzetti P, Ungar J, Barbera R, Barboni D, Beccati A, Bigagli L, Boldrini E, Bruno R, Calanducci A, Campalani P, Clements O, Dumitru A, Grant M, Herzig P, Kakaletrisk G, Laxton J, Koltsida P, Lipskoch K, Mahdiraji AR et al. (2015) Big data analytics for earth sciences: the EarthServer approach. International Journal of Digital Earth, 1-27. • Folegani M, Mantovani S, Natali S, Veratelli MG (2013) WHERE - World HEritage monitoring by Remote sEnsing: the role of the Microclimate processing chain for monitoring the UNESCO sites. Conference programme for '4th Workshop on Cultural and Natural Heritage', 06 Jun 2013 • Natali S, Beccati A, D'Elia S, Veratelli MG, Campalani P, Folegani M, Mantovani S (2011)

Multitemporal data management and exploitation infrastructure, in Analysis of Multi-temporal Remote Sensing Images (Multi-Temp), 2011 6th International Workshop on the, pp. 217 –220

Relevant projects

- HERACLES – "HEritage Resilience Against CLimate Events on Site": Lead by the National Research Council (Italy) HERACLES a 3-year Horizon2020 project whose main objective is to support effective resilience policies of Cultural Heritages against climate change effects. The role of SISTEMA is to provide meteo-climatological information as input to the monitoring and warning system.
- TAMP - Technology and Atmospheric Mission Platform and "Virtual Exploitation Environment Demonstration for Atmospheric Missions" – (VEEDAM): Lead by SISTEMA, with the scope to demonstrate the feasibility of the implementation of an exploitation platform for the Atmospheric Science community. The TAMP project started in Dec. 2014, and continued with VEEDAM ultil April 2018, involving the most relevant the most important European entities in Air Quality monitoring such as ECMWF, DLR, Meteo France, BIRA, KNMI, the Austrian meteorological Administration.
- PollenMON - Improvement of phenological models by means of time series of EO data for numerical pollen forecast (PollenMON): Co-funded by the Austrian Space Promotion Program and lead by the Austrian Meteorological Service, the project aims at improving the performance of Pollens forecast models integrating satellite-based meteorological information (temperature) and vegetation information. SISTEMA is responsible for the collection, pre/processing and data fusion of temperature maps collected from satellite data and ground measurements with model data.
- Earth Observation for Sustainable development (EO4SD) projects:
    - Climate Resilience (SISTEMA subcontractor of GMV, Spain)
    - Fragile and Conflict States (SISTEMA subcontractor of CLS, France)

  For both projects SISTEMA is providing and customising the eodataservice platform to manage the entire data portfolio of the two projects (more than 100 collections each)

Relevant infrastructure

SISTEMA can provide a data center facility called MEEO-DAF, that is an infrastructure created to support high computing demand of geospatial web platforms to manage geospatial data and services. Initially designed for Earth Observation data services it has been extended to provide computational resources and hosting to every kind of geospatial data. The facility gives academic, government and commercial users access to internally hosted climate and EO Data alongside a wide range of associated Services for data processing, visualization and analysis.

Platform:
ADAM (https://adamplatform.eu) is a cross-domain platform developed by the SISTEMA's mother company MEEO, for which SISTEMA owns all rights of use, modification, and distribution. ADAM provides access, processing and visualization services to a wide range of environmental data. In the framework of B2TERM it will be verified whether ADAM can benefit from the data description approach.

4.2. Third parties involved in the project (including use of third party resources)

The coordinator TIB intends to subcontract some work in WP6, as outlined below. CNRS intends to subcontract Stanford-BMIR and will also involve University of Montpellier as described.

Does the participant plan to subcontract certain tasks (please note that core tasks of the project should not be sub-contracted)	Y
TIB The subtask of the legal assessment associated to the provision and long-term sustainability of the B2TERM service will be sub-contracted (see also WP6, T6.3). In order to fulfil the necessary legal requirements of the planned B2TERM service, its components (e.g., TIF), its integration within the EOSC Service Catalogue and to ensure external privacy and data protection, the Project Coordination Office (section 3.2) will initiate an award procedure (open call) in order to assign this task to suitable subcontractor which is also familiar with EOSC legal requirements. **CNRS** Within our ontology repository development activity, LIRMM-CNRS partners with the US NIH funded National Center for Biomedical Ontology (NCBO) project which develops BioPortal, the reference ontology and terminology repository in biomedical sciences. NCBO technology is domain-independent and open source. In partnership with Stanford University (during C. Jonquet H2020 MSCA SIFRm project), CNRS-LIRMM designed, developed two ontology repositories for French biomedical terminologies (the SIFR BioPortal) and for agronomy and related sciences (the AgroPortal). Within KOINE, we plan to build atop that collaboration on the shared NCBO technology, therefore we envision that certain development tasks shall involve the Stanford Center for Biomedical Informatics Research (Stanford-BMIR). Stanford-BMIR headed by Pr. Musen (**See Letter of Support**), is home to world class scientists managing knowledge and data related to health, biomedical sciences and semantic Web. The research group is worldwide recognized for its work on ontologies (Protégé), biomedical ontologies (BioPortal) and metadata (CEDAR). Stanford-BMIR and CNRS-LIRMM have a long-standing collaboration (10 years) related to semantic annotation, ontology alignment and ontology repositories. This collaboration with Stanford-BMIR shall also be reinforced for 2019-2022 by a CNRS funded International Associated Lab, currently being setup between CNRS and Stanford. Within KOINE, we will continue to collaborate on the research and development questions related to interoperation of the NCBO technology with B2TERM as the technology is a pillar of different terminology services for B2TERM. CNRS-LIRMM requests 20K€ of subcontracting to Stanford-BMIR to participate over the span of the project (5-10K€) to NCBO's future triple store backend licensed software (AllegroGraph); and to cover (10-15K€) implementation costs for specific features that will engage Stanford-BMIR significantly.	
Does the participant envisage that part of its work is performed by linked third parties[1]	Y
CNRS LIRMM is a Joint Research Unit (JRU 5506) owned by the CNRS and the Université	

de Montpellier (UM). These organisms are linked by a contract attesting the relationship between the CNRS (beneficiary) and the linked third party UM and establishing the JRU. Only the costs of permanent staff will be declared by the linked third party: 2 permanent researchers (Drs. Jonquet & Todorov).	
Does the participant envisage the use of contributions in kind provided by third parties (Articles 11 and 12 of the General Model Grant Agreement)	N
If yes, please describe the third party and their contributions	

Section 5: Ethics and Security

5.1 Ethics

All partners in KOINE will adhere to the Charter of Fundamental Rights of the European Union and to data protection legislation, as well as, overarching ethical guidance such as the European Code of Conduct for Research Integrity. The consortium also expects project researchers to adhere to the ethical commitments contained their professional codes of conduct.

The main ethical risks associated with the project are related to the collection, handling, storage and analysis of personal data, e.g., when users register within the B2TERM service or participate within the Stakeholder forum. Under the European Union law, personal data is defined as "any information relating to an identified or identifiable natural person". The collection, use and disclosure of personal data at a European level are regulated by the following directives and regulation:

- Directive 95/46/EC on protection of personal data (Data Protection Directive)
- Directive 2002/58/EC on privacy and electronic communications (e-Privacy Directive)
- Directive 2009/136/EC (Cookie Directive)
- Regulation 2016/679/EC (repealing Directive 95/46/ EC)
- Directive 2016/680/EC

According to the Regulation 2016/679/EC, personal data: "*means any information relating to an identified or identifiable natural person ('data subject'); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person* (art. 4.1)."
Specific attention will be paid towards addressing the ethics of research throughout different stages of the project since the will be handling certain types of personal data. All activities will be carried out ensuring the ethical principles and in compliance with Directive 95/46/EC of the European Parliament and subsequent Regulation 2016/679/EC and Directive 2016/680/EC, about the protection of individuals with regard to the processing of personal data and on the free movement of such data, as well as Directive 2002/58/EC, concerning the processing of personal data and the protection of privacy in the electronic communications sector, as modified by Directive 2009/136/EC. The national data protection and privacy laws of the countries involved in the research will be also followed. The consortium guarantees that all personal data collected during the project will be kept secure and unreachable by unauthorized persons. The data will be handled with appropriate confidentiality, accessibility controls, and technical security. The partners will use state-of-the-art security policies and rules to ensure the integrity and availability of electronic information captured, stored, maintained, and used. In detail, the consortium will adopt the following principles when dealing with personal data:

- Confidentiality and anonymity – confidentiality will be guaranteed whenever possible. This may also be important dealing with industrial stakeholders. The only exception can be in some cases for the researcher directly interacting with a group of participants (e.g., focus group). Our research team will not make publicly accessible any personal data. Anonymity will be granted through generalization and pseudonymisation. Furthermore, provisions will be taken to avoid the possibility of information linkage.

- Informed consent – the informed consent policy requires that each participant will provide his/her informed consent prior to the start of any activity involving him/her. All people involved in the project's face-to-face research and evaluation activities (e.g. interviews, workshops) will be asked to read and approve an Informed Consent Form explaining how personal data will be collected, managed and stored. For the remaining data collection activities (collection of publicly available patent data and companies' data) informed consent procedure will not be applied.
- Circulation of the information limited to the minimum required to achieve a limited circulation of the information, the database containing in anonymous form the data collected from the users will be distributed to the partners, if needed at all, through protected and encrypted Internet connections; the raw data will only be shared if it is required for the development. The project partners will never pass on or publish the data without first protecting participants' identities. No irrelevant information will be collected; at all times, the gathering of private information will follow the principle of proportionality by which only the information strictly required to achieve the project objectives will be collected.

The project team will store research data on secure repositories protected by password at the head offices of partner organizations. File names will not refer to any personal information. Information that might enable data to be linked to individuals, such as the file linking participants' names to their respective code/pseudonym, will be password protected and encrypted so that access will be restricted to only those with the requisite credentials.

The partner organisations will be responsible for ensuring all ethical principles relating to their country and institutional context are adhered to. In some cases this will require additional ethical permission to be sought. This consortium involves partners who are experienced academics and experts with a history of conducting research in the kinds of settings described in this research project. The project partners each have a senior researcher lead who will supervise the data collection and analysis process. Special care will be taken to ensure that no documents or outputs in any of the partner countries contain personal identifiers. In the case of dissemination activities that may include identifiers, consent will be gained beforehand. Any ethical issues that arise during the project will be reviewed by the project coordinator and necessary actions taken.

5.2 Security

Please indicate if your project will involve:

- activities or results raising security issues: NO
- 'EU-classified information' as background or results: NO

3980 Broadway Suite 103-185 · Boulder, CO 80304 · www.esipfed.org

TIB – Leibniz Information Centre for Science and Technology
Welfengarten 1 B
30167 Hannover
Germany

Boulder, Colorado, 25 January 2019

Letter of Support to the H2020 KOINE Project (INFRAEOSC-02-2019 Call)

Dear Prof. Auer,

The Earth Science Information Partners (ESIP) is a community of Earth science data and information technology practitioners who facilitate the distribution of Earth science data, and provide products and services to decision makers and researchers in public and private settings. ESIP's strength comes from its depth of partner organizations, which now number ~110. Among these are all of the Earth observing data centers at the National Oceanographic and Atmospheric Administration (NOAA), National Aeronautics and Space Administration (NASA), and United States Geological Survey (USGS), as well as government research laboratories, research universities, modelers, education resource providers, technology developers, nonprofits and commercial enterprises. It is no underestimation that ESIP continues to have global impact on making science data matter.

ESIP's core mission is to support the networking and data dissemination needs of its members and the global Earth science data community by linking the functional sectors of Earth science: observation, research, application, education and use. Within ESIP there are a number of *collaborative committees*, *working groups*, and *clusters* of interested persons, organized around a variety of areas of interest within the Earth science discipline. Interest groups work on issues that range from providing technology support for Earth science research and data management, to addressing the societal and cultural implications of climate change. The ESIP Semantic Technology Committee (STC) is one such committee, formed to encourage and promote research and development of semantic technologies in support of Earth science data management, data discovery, data dissemination, and data analysis. The committee has broad collaborations with other ESIP members, working groups, clusters, and committees, as well as external organizations such the Research Data Alliance, the Open Biological and Biomedical Ontology (OBO) Foundry and Library, the American Geophysical Union, Open Geospatial Consortium, the World Wide Web Consortium, and others to identify semantic methods and tools that support the adoption of semantic technologies within ESIP and across its member organizations.

The KOINE project aims to develop B2TERM as a novel and innovative EOSC service for terminology, and integrate B2TERM into (inter)disciplinary operational infrastructures and diverse communities in research, industry or society at large. B2TERM is the EOSC cornerstone that enables semantic data interoperability, discovery and exploitation across disciplines and thus supports user-centric scientific, industrial and societal applications that maximize the scientific and societal impact of research data.

ESIP's STC oversees stewardship and open source community governance of the SWEET (Semantic Web for Earth and Environmental Terminology) ontology suite. SWEET provides an ontological

3980 Broadway Suite 103-185 · Boulder, CO 80304 · www.esipfed.org

foundation for, amongst other things, the Global Change Master Directory, a comprehensive directory of information about Earth science data, including a widely used controlled vocabulary of keywords that are integral to the provision of consistent metadata records used by NASA's Earthdata Search. This in turn provides data discovery and access services for over 17 petabytes of Earth science data, representing 100s of 1000s of different measurements of Earth system phenomena. The STC has been working collaboratively with to harmonize SWEET with the Environment Ontology (ENVO), W3C's Data Catalog Vocabulary (DCAT) and Semantic Sensor Network (SSN)/Sensor, Observation, Sample, and Actuator (SOSA) resources.

As H2020 KOINE Project's B2TERM annotation service will allow users to annotate distributed services with multiple, interoperable terminologies, services that rely on ENVO terms for their annotations, will be testably interoperable with services that rely on ontologies such as SWEET that are aligned with ENVO. This would be an enormous advantage for the ESIP community member services that use SWEET/GCMD terms for their annotations. Recently, ESIP's interest in addressing the topic of "*improving semantic understanding across communities*" was formally documented in ESIP's submission in response to the National Science Foundation's CI 2030 request for information. B2TERM will play a key role in the future of Earth science cyberinfrastructure and in ESIP's efforts to provide support for the networking and data dissemination needs of its members and the global Earth science data community. We enthusiastically endorse this project.

With best wishes.

Yours sincerely,

Erin Robinson, Executive Director
ESIP

Beth Huffer, Out-going Chair
ESIP Semantic Technology Committee

Lewis McGibbney, In-coming Chair
ESIP Semantic Technology Committee

LEIBNIZ INFORMATION CENTRE
FOR SCIENCE AND TECHNOLOGY
UNIVERSITY LIBRARY

TIB // P.O. Box 60 80 // 30060 Hannover, Germany

KOINE Consortium

Hannover, 21. January 2019

Intent to sustainably operate B2TERM

Dear KOINE Consortium,

TIB is the German National Library for science and technology, Information Centre of the Leibniz Association as well as university library for Leibniz University Hannover. TIB represents and operates a national research infrastructure facility for the provision of scientific information covering engineering, information technology, mathematics, physics, chemistry and architecture. Among the key services, TIB hosts the DataCite headquarter.

The KOINE project aims to develop B2TERM as a novel and innovative EOSC service for terminology, and integrate B2TERM into (inter)disciplinary operational infrastructures and diverse communities in research, industry or society at large.

With this letter I express TIB's intent to support the development of B2TERM as well as the sustainable operation of the service beyond the project's lifetime.

Yours sincerely,

Prof. Dr. Sören Auer
Director TIB

Director
Prof. Dr. Sören Auer
T +49 511 762-2531
F +49 511 762-2686
auer@tib.eu

Technische
Informationsbibliothek (TIB)
German National Library
of Science and Technology
Welfengarten 1 B
30167 Hannover, Germany
www.tib.eu

Your letter dated:

Your Ref.:
Our Ref.:

VAT Reg No:
DE214931803
Tax.-No. 25/202/26441

TIB
Nord/LB Hannover
IBAN: DE47 2505
0000 0101 4287 53
BIC: NOLADE2HXXX

UB
Nord/LB Hannover
IBAN: DE57 2505
0000 0106 0276 34
BIC: NOLADE2HXXX

TIB is a member of

Leibniz Association

AMSTERDAM INSTITUTE FOR
ADVANCED METROPOLITAN SOLUTIONS

Subject: Letter of Support to the H2020 KOINE Project (Call INFRAEOSC-02-2019: Prototyping new innovative services)

Amsterdam, January 18th, 2019

Dear Sir/Madam,

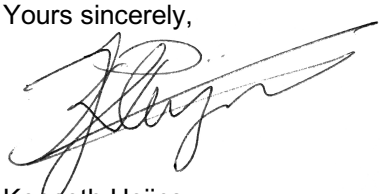
With this letter I declare on behalf of Amsterdam Institute for Advanced Metropolitan Solutions (AMS Institute) our interest in and support for the Horizon 2020 KOINE project proposal in the "Prototyping new innovative services - INFRAEOSC-02-2019' call from the European Commission.

AMS Institute is a public-private institute, initiated and supported by the city of Amsterdam. It is enabled and created by a unique consortium of public and private partners that cooperate on metropolitan innovation. At AMS Institute, metropolitan solutions are responses to the urban challenges of sustainability and quality of life, including energy, resource and food security, mobility and logistics, water and waste management, and its effects on health and wellbeing. AMS Institute currently participates in more than 80 research projects. Our projects are by definition multidisciplinary. The city of Amsterdam is our home and living lab to develop and test innovative solutions to urban challenges.

The interdisciplinary character of our research implies combining datasets and sources from various disciplines and from diverse stakeholders ranging from researchers, through policy makes to practitioners. In this regard, the lack of common terminologies and metadata poses a challenge for a greater data integration, data interoperability and increased research efficiency. The KOINE project aims to develop a novel EOSC service for terminology, B2TERM, that is to be integrated into (inter)disciplinary operational infrastructures and diverse communities in research, industry or society at large. B2TERM is the EOSC cornerstone that enables semantic data interoperability, discovery and exploitation across disciplines and thus supports user-centric scientific, industrial and societal applications that maximize the scientific and societal impact of research data.

AMS Institute is interested in this proposal for its relevance for improving the impact and replicability of interdisciplinary research currently conducted at our institute. In case of its positive evaluation, we would be glad to contribute with our point of views on the issues contemplated by the consortium.

Yours sincerely,



Kenneth Heijns
Managing Director
AMS Institute

Amsterdam Institute for Advanced Metropolitan Solutions
Marineterrein Amsterdam
Kattenburgerstraat 5,
Building 027W,
1018 JA Amsterdam
T +31 (0)20 6651350
www.ams-institute.org
office@ams-institute.org

Chris Mungall, PhD
Department Head, Molecular Ecosystems Biology
Environmental Genomics and Systems Biology
Lawrence Berkeley National Laboratory
Department of Molecular Ecosystems Biology
1 Cyclotron Road Mail Stop—977
Berkeley, CA 94720

TIB – Leibniz Information Centre for Science and Technology
Welfengarten 1 B
30167 Hannover
Germany

Jan 27, 2019

Re: Letter of Support to the H2020 KOINE Project (INFRAEOSC-02-2019 Call)

Dear Professor Auer,

I am writing to express my support for your proposal. I am a scientist at Lawrence Berkeley National Laboratory, a US Department of Energy National Laboratory. At the Lab, we forge multidisciplinary teams to solve national challenges in energy, environment, health issues, and technology development, as well as advance the engineering of biological systems for sustainable manufacturing.

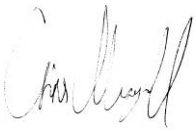
The KOINE project aims to develop B2TERM as a novel and innovative EOSC service for terminology, and integrate B2TERM into (inter)disciplinary operational infrastructures and diverse communities in research, industry or society at large. B2TERM is the EOSC cornerstone that enables semantic data interoperability, discovery and exploitation across disciplines and thus supports user-centric scientific, industrial and societal applications that maximize the scientific and societal impact of research data.

As the head of the Molecular Ecosystems Biology department in the Biosciences area at our institute, I am aware of the importance of terminologies and data interoperability in domains such as environmental science and genomics. Lawrence Berkeley Laboratory has been supportive of efforts such as the environment ontology EnvO. Our field is generating massive quantities of data from a variety of modalities and instrumentation types. Standards are vital to be able to organize and make use of this data. Your terminology services promise to deliver a crucial piece of the infrastructure required.

I am also heavily involved with and committed to the ontologies and standards community. I was one of the founders of the Open Biological Ontologies (OBO) Foundry, a project to coordinate the development of ontologies across the life sciences (and, increasingly, across intersecting domains such as earth, environmental sciences). I am one of the PIs on the OBO Services project, and also a PI on the Gene Ontology (GO) project, one of the most widely used

resources in the life sciences. I am pleased to see you are adopting existing standards such as EnvO and other ontologies from OBO, and allied efforts (e.g. ESIP, SWEET). Your plans to annotate scientific parameters using these existing terminologies will be necessary for the integration of diverse datatypes in environmental sciences and for integrating these with omics data.

Yours sincerely



Christopher J Mungall
Department head, Molecular Ecosystems Biology
TEL: 510.486.4170
CJMungall@lbl.gov

CESAER

CESAER Secretariat
Kasteelpark Arenberg 1 Box 2200
3001 Leuven
Belgium
Tel: +32 16 32 16 87
E-mail: info@cesaer.org

Delft, 17 January 2019

Letter of Support to the H2020 KOINE Project (INFRAEOSC-02-2019 Call)

Dear Sir/Madam,

I would like to express my support for the KOINE project proposal as the CESAER's Co-chair of the Task Force on Open Science.

The KOINE project aims to develop B2TERM as a novel and innovative EOSC service for terminology, and integrate B2TERM into (inter)disciplinary operational infrastructures and diverse communities in research, industry or society at large. B2TERM is the EOSC cornerstone that enables semantic data interoperability, discovery and exploitation across disciplines and thus supports user-centric scientific, industrial and societal applications that maximize the scientific and societal impact of research data.

CESAER is the association of 50 universities of science and technology in Europe. In our experience, the lack of data interoperability is one of the key barriers to open science and to better re-use of research outputs. This proposal is therefore of high interest for us and in case of its positive evaluation, we would be glad to contribute with our point of views on the issues contemplated by the consortium. Furthermore, we will facilitate the consortium in the stimulation of focus groups, stakeholders and multipliers, in particular through our dedicated Open Science task force.

Yours sincerely,

i/o H. G. Heymans

Wilma van Wezenbeek
CESAER's Co-chair of the Task Force on Open Science

CLARIN ERIC
c/o Utrecht University
Drift 10
3512 BS Utrecht
The Netherlands

phone
+31 30 253 6484

email
f.m.g.dejong@uu.nl
clarin@clarin.eu

website
www.clarin.eu

TIB – Leibniz Information Centre for Science and Technology
Welfengarten 1 B
30167 Hannover
Germany

Subject	Our Ref.	Place/Date
Letter of Support to the H2020 KOINE Project	CE-2019-1383	Utrecht, 21 January 2019

Dear Professor Auer,

This letter is meant to express CLARIN ERIC's support for project proposal KOINE to be submitted under the INFRAEOSC-02-2019 call of Horizon 2020. CLARIN ERIC is the pan-euroepan research infrastructure aiming at providing access to language resources and tools to support and advance the relevant research communities in the humanties and social sciences and beyond.

The KOINE project aims to develop B2TERM as a novel and innovative EOSC service for terminology, and integrate B2TERM into (inter)disciplinary operational infrastructures and diverse communities in research, industry or society at large. B2TERM is the EOSC cornerstone that enables semantic data interoperability, discovery and exploitation across disciplines and thus supports user-centric scientific, industrial and societal applications that maximize the scientific and societal impact of research data.

CLARIN will contribute terminologies, terminology services and expertise, experience on terminology creation and curation with respect to Language Resources. It will extend existing terminology services with B2TERM-compliant API enabling their integration

Yours sincerely,

Prof.Dr. Franciska de Jong
executive director CLARIN ERIC

MINERAL RESOURCES
www.csiro.au

26 Dick Perry Avenue, Kensington WA 6151
PO Box 1130, Bentley WA 6102, Australia
T (08) 6436 8500 • **ABN** 41 687 119 230

10 January 2019

TIB – Leibniz Information Centre for Science and Technology
Welfengarten 1 B
30167 Hannover
Germany

Letter of Support to the H2020 KOINE Project (INFRAEOSC-02-2019 Call)

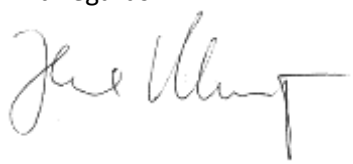
Dear Prof. Auer,

The Commonwealth Scientific and Industrial Research Organisation (CSIRO) serves governments, industries, business and communities across Australia. Through our Business Units Mineral Resources, Land & Water, Oceans & Atmosphere, Agriculture & Food, Data61 and Information Management & Technologies CSIRO contributes to the development of research data infrastructures and semantic interoperability.

The KOINE project aims to develop B2TERM as a novel and innovative EOSC service for terminology, and integrate B2TERM into (inter)disciplinary operational infrastructures and diverse communities in research, industry or society at large. B2TERM is the EOSC cornerstone that enables semantic data interoperability, discovery and exploitation across disciplines and thus supports user-centric scientific, industrial and societal applications that maximise the scientific and societal impact of research data.

CSIRO will support KOINE through the promotion of B2TERM to related groups that may adapt it in future in Australia. Among those groups are our own Business Units, our institutional and working links with the Australian Research Data Commons, and our involvement in relevant working groups of the Research Data Alliance.

Kind regards



Jens Klump
Science Leader Earth Science Informatics
jens.klump@csiro.au
(08) 6436 8828

www.cyverse.org
info@cyverse.org

TIB – Leibniz Information Centre for Science and Technology
Welfengarten 1 B
30167 Hannover
Germany

Jan. 27, 2018

re.: Letter of Support to the H2020 KOINE Project (INFRAEOSC-02-2019 Call)

Dear Prof. Auer:

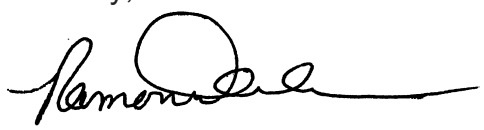
I am writing to express my support for the KOINE Project, which aims to develop B2TERM as a novel and innovative EOSC service for terminology, and integrate B2TERM into (inter)disciplinary operational infrastructures and diverse communities in research, industry or society at large. As Senior Scientific Analyst at CyVerse, I find very high value in the proposed infrastructure. CyVerse is the U.S. National Science Foundation's funded infrastructure for data intensive life sciences research. We provide data storage and computational capability to tens of thousands of researchers around the globe, currently hosting over 4PB of data. As the product champion of the CyVerse Data Commons, I am responsible for specifying features that support data management throughout the data life cycle, with a focus on producing and publishing FAIR data. Key among those features are tools for complying with metadata standards. For example, we provide metadata templates that allow users to specify terms from controlled vocabularies or ontologies. The existence of API standards and aligned terminologies resulting from the KIONE project will greatly facilitate the expansion of this service in CyVerse.

In additional, I am an active member of the Operations Committee of the Open Biological Ontologies (OBO) Foundry, which coordinates the development of ontologies across the life sciences and related domains such as Earth and Environmental Science. As the lead developer for several OBO Foundry ontologies, and I am pleased to see you are adopting existing OBO ontologies as standards in your project. KOINE's alignment of terminologies with OBO Foundry resources will build a bridge between EOSC and the various projects I work on, opening new doors. I am also the current chair of the Compliance and Interoperability Group (CIG) of the Genomics Standards Consortium, which is responsible for developing and maintaining the Minimum Information of any (x) Sequence (MIxS) standards. The CIG is in the midst of a major restructuring of how MIxS terms and checklists are stored and delivered, and a project such as KIONE will be highly beneficial to our efforts to promote the use of standard terminologies as values for MIxS fields. Finally, I am a member of the

Biodiversity Information Standards (TDWG) working group on terminologies, which promotes the use of controlled vocabularies for Darwin Core metadata. The terminologies served through KIONE can also benefit that effort.

Hoping for success on this proposal.

Sincerely,



Ramona Walls
Senior Scientific Analyst, CyVerse
Research Assistant Professor, Bio5 Institute
University of Arizona
Biosciences Research Laboratory
1601 East Helen St
Tucson, AZ 85721
telephone: 1-520-626-1489
email: rwalls@cyverse.org

STANFORD CENTER FOR
BIOMEDICAL INFORMATICS RESEARCH

MARK A. MUSEN, MD, PHD

Professor of Medicine and Biomedical Data Science
Chief, Biomedical Informatics Research
Phone (650) 725-3390 Fax (650) 725-7944
Email: musen@stanford.edu
http://bmir.stanford.edu

January 21, 2019

TIB – Leibniz Information Centre for Science and Technology
Welfengarten 1 B
30167 Hannover
Germany

RE: H2020 KOINE Project (INFRAEOSC-02-2019 Call)

Dear Prof. Auer:

I am delighted to write in support of your KOINE proposal. I head of the Stanford Center for Biomedical Informatics Research (BMIR) and I am the principal investigator of the United States *National Center for Biomedical Ontology* (NCBO; www.bioontology.org), one of the eight National Centers for Biomedical Computing founded under the NIH Roadmap as well as the *Center for Expanded Data Annotation and Retrieval* (CEDAR; https://metadatacenter.org), funded by the NIH BD2K program. I have performed research in the area of biomedical informatics and the use of ontologies for more than 30 years. My group developed the Protégé tool, the first platform used worldwide for ontology development, as well as the NCBO BioPortal, a repository of biomedical ontologies offering multiple services (search, annotation, indexing) used by thousands of people internationally.

I understand that the KOINE project plans to develop B2TERM as a novel and innovative service for accessing and using terminologies (and other semantic resources) within EOSC and to integrate B2TERM into multidisciplinary operational data centers and e-infrastructures. The scientific diversity and complementarity among the members of your consortium is exceptional and these elements are certainly key predictors for success and wide adoption. I am confident that B2TERM can become a cornerstone for EOSC that enables semantic data interoperability, discovery, and exploitation across disciplines and thus will maximize the scientific and societal impact of research data.

I am very enthusiastic about collaborating with your project and to explore the convergence between similar research questions in different scientific disciplines (biomedicine, environmental sciences, agri-food, oceanography, and so on). I strongly believe that the project's foreseen outcomes will have a strong impact on the advancement of data practices and adopting Open Science in Europe.

In case of approval and financing by the EU Commission, my group will be pleased to be associated (as we were during Dr. Jonquet's H2020 MSCA SIFRm project) to the AgroPortal and EcoPortal vocabulary and ontology repositories (both involved in KOINE) and to discuss the best synergies between our NBCO BioPortal and B2TERM.

Sincerely,

Mark A. Musen, M.D., Ph.D.
Professor of Medicine (Biomedical Informatics) and Biomedical Data Science
Chief, Biomedical Informatics Research

DataONE

TIB – Leibniz Information Centre for Science and Technology
Welfengarten 1 B
30167 Hannover
Germany

16 January 2019

Letter of Support to the H2020 KOINE Project (INFRAEOSC-02-2019 Call)

Dear Prof. Auer,

The KOINE project aims to develop B2TERM as a novel and innovative EOSC service for terminology, and integrate B2TERM into (inter)disciplinary operational infrastructures and diverse communities in research, industry or society at large. B2TERM is the EOSC cornerstone that enables semantic data interoperability, discovery and exploitation across disciplines and thus supports user-centric scientific, industrial and societal applications that maximize the scientific and societal impact of research data.

The US DataONE project (https://dataone.org) recognizes the importance of well-defined, shared vocabularies built upon appropriate semantic web technologies, for better realizing the FAIR principles for sharing of research data. We are pleased to indicate our enthusiasm for seeing this project move forward, and hope to work with our European colleagues in building and deploying these vocabulary services for improved global interoperability of mechanisms for discovering and re-using scientific information resources.

With best wishes.

Yours sincerely,

William Michener
Professor and Director of e-Science, University Libraries, University of New Mexico
Principal Investigator, DataONE (Observation Network for Earth)

THE UNIVERSITY OF NEW MEXICO

TIB – Leibniz Information Centre for Science and Technology
Welfengarten 1 B
30167 Hannover
Germany

Albuquerque, New Mexico USA, January 11, 2019

Letter of Support to the H2020 KOINE Project (INFRAEOSC-02-2019 Call)

Dear Prof. Auer,

On behalf of the Environmental Data Initiative (EDI), based at the University of New Mexico, USA, I am writing to express strong support for the proposal entitled *KOINE: Interdisciplinary data interoperability, discovery, and exploitation in the EOSC.* EDI is funded by the US National Science Foundation and operates a data repository with a staff of highly experienced informatics specialists and software developers. EDI faces the same data discovery and reuse challenges outlined in the KOINE proposal due to natural language ambiguity in research data documentation and machine-inaccessible semantic encoding.

The KOINE project aims to develop B2TERM as a novel and innovative EOSC service for terminology, and integrate B2TERM into (inter)disciplinary operational infrastructures and diverse communities in research, industry or society at large. B2TERM is the EOSC cornerstone that enables semantic data interoperability, discovery and exploitation across disciplines and thus supports user-centric scientific, industrial and societal applications that maximize the scientific and societal impact of research data.

EDI personnel are very interested in opportunities to collaborate with the KOINE project. We will pursue funding from US sources that would enable us to participate in KOINE research activities. The EDI Data Repository also houses over 49,000 environmental datasets, the majority of which are from the US Long Term Ecological Research Network (LTER). All EDI datasets are richly documented and openly accessible for use in KOINE semantics and terminology research.

With best wishes,

Kristin L. Vanderbilt

Kristin Vanderbilt, Ph.D.
Research Associate Professor
Environmental Data Initiative

Biology Department, MSC03 2020, 1 University of NM, Albuquerque, NM 87106, 505-277-3411

European Long-Term Ecosystem Research Network
Chair-person: Michael Mirtl
Email: michael.mirtl@umweltbundesamt.at
www.lter-europe.net

LTER-Europe Office

To whom it may concern
Prof. Sören Auer
TIB – Leibniz Information Centre for Science and Technology
Welfengarten 1 B
30167 Hannover
Germany

January 17, 2019

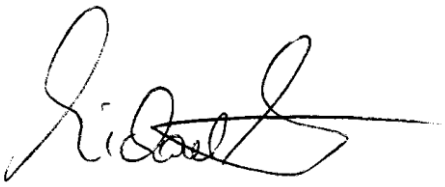
Subject: Letter of Support to the H2020 KOINE Project (INFRAEOSC-02-2019 Call)

Dear Prof. Auer,

The KOINE project aims to develop B2TERM as a novel and innovative EOSC service for terminology, and integrate B2TERM into (inter)disciplinary operational infrastructures and diverse communities in research, industry or society at large. B2TERM is the EOSC cornerstone that enables semantic data interoperability, discovery and exploitation across disciplines and thus supports user-centric scientific, industrial and societal applications that maximize the scientific and societal impact of research data.

The scope of KOINE touches the data- and information related activities and services of all current LTER activities in Europe, including the eLTER ESFRI process, various eLTER project engagements (incl. ENVRI fair, EuroGEOSS, eLTER H2020) and wider LTER-Europe Network´s activities. We´re looking forward to aligning with your team´s work to realize the project's goals and objectives.

Yours sincerely



Dr. Michael Mirtl
Chairman of LTER-Europe
eLTER ESFRI coordinator

GFBio e.V. - Campusring 1, 28759 Bremen

Dr. Michael Diepenbroek
Chairperson

GFBio e.V.
Campusring 1
28759 Bremen
Steuer-Nr. 60/147/14692
Finanzamt Bremen

mail info@gfbio.org
www.gfbio.org

TIB – Leibniz Information Centre for Science and Technology
Welfengarten 1 B
30167 Hannover
Germany

Bremen, 2019-01-17

Letter of Support to the H2020 KOINE Project (INFRAEOSC-02-2019 Call)

Dear Prof. Auer,

GFBio is a federated national data infrastructure for biodiversity operated by 20 partner institutions in Germany running now for around six years.

The KOINE project aims to develop B2TERM as a novel and innovative EOSC service for terminology, and integrate B2TERM into (inter)disciplinary operational infrastructures and diverse communities in research, industry or society at large. B2TERM is the EOSC cornerstone that enables semantic data interoperability, discovery and exploitation across disciplines and thus supports user-centric scientific, industrial and societal applications that maximize the scientific and societal impact of research data.

The main objective of GFBio is the long-term archiving and publication of biodiversity related data. A major challenge is the data discovery and integration of the heterogeneous data provided by environmental, collections, and genome data archives. Key in this respect is the usage of well-curated and persistently available terminologies. For this purpose GFBio operates a terminology service (https://terminologies.gfbio.org/) which is a ready to use component for the planned EOSC service layer B2TERM. In this sense GFBio appreciates and looks forward to corresponding services supplied by KOINE. They will help significantly to harmonize semantics of data descriptions between the various data archives in Europe and beyond.

With best wishes.

Yours sincerely,

Chairperson of GFBio e.V.

The Global Ocean Observing System

TIB – Leibniz Information Centre for Science and Technology
Welfengarten 1 B
30167 Hannover
Germany

Hobart, 25th of January 2019

Subject: Letter of Support to the H2020 KOINE Project (INFRAEOSC-02-2019 Call)

Dear Prof. Auer,

We would like to support the application of the KOINE Project within the H2020 Framework.

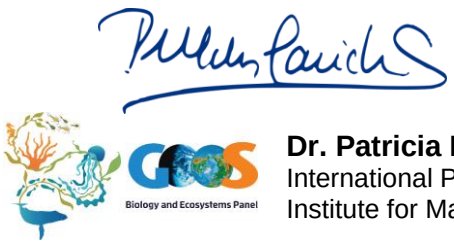
The Global Ocean Observing System is a program under the Intergovernmental Oceanographic Commission (IOC) of the United Nations Educational, Scientific and Cultural Organization (UNESCO) aimed in improving forecast capacity and management of marine resources in the face of climate change, and also to help mitigate the effects of natural disasters for a better use and protection of the ocean and coastal zones. Through three different expert panels in physics, biogeochemistry, and biology and ecosystems, GOOS supports ocean services, climate and ocean health.

The GOOS Biology and Ecosystems Panel (GOOS BioEco) is working towards developing a globally coordinated and sustained ocean observing system to systematically assess the status of the ocean's biodiversity and ecosystems and how these are responding to increasing resource use, including coastal development under long-term climate change scenarios.

The KOINE project aims to develop B2TERM as a novel and innovative EOSC service for terminology, and integrate B2TERM into (inter)disciplinary operational infrastructures and diverse communities in research, industry or society at large. B2TERM is the EOSC cornerstone that enables semantic data interoperability, discovery and exploitation across disciplines and thus supports user-centric scientific, industrial and societal applications that maximize the scientific and societal impact of research data.

We hope that the KOINE Project is approved under the H2020 framework as it aligns with BioEco's collaborative international approach to support and facilitate the implementation of best practices, develop new technologies, strengthen data sharing and interoperability, and enhance capacity building and technology transfer. These developments are needed to increase the availability and impact of scientific advice to decision makers.

Yours sincerely,



Dr. Patricia Miloslavich
International Project Officer – GOOS Biology and Ecosystems Panel
Institute for Marine and Antarctic Studies – University of Tasmania

IOC WMO UNEP ICSU

The Global Ocean Observing System

Ref: KOINE_28012019

28 January 2019

TIB – Leibniz Information Centre for Science and Technology
Welfengarten 1 B
30167 Hannover
Germany

Paris, 28 January 2019

Subject: Letter of Support to the H2020 KOINE Project (INFRAEOSC-02-2019 Call)

Dear Prof. Auer,

The Global Ocean Observing System (GOOS) would like to note its support for the KOINE project.

The **Global Ocean Observing System (GOOS)** is a sustained collaborative system of ocean observations, encompassing in situ networks, satellite systems, governments, UN agencies and individual scientists. We are organized around a series of components undertaking requirements assessment, observing implementation, innovation through projects, and a core team. GOOS is a programme executed by the Intergovernmental Oceanographic Commission (IOC) of the UNESCO, but its success relies on the coordinated contributions of people and organizations worldwide, and for this information flow is vital.

The development of natural language ontologies and their digital connection is of great benefit in connecting dispersed stakeholders to the information that they need, this is demonstrated through the current work on ocean ontologies within the Ocean Best Practice System (project) and in easing the flow of glider data in Europe (OceanGliders and BODC).

EOSC service for terminology, and integrate B2TERM into (inter)disciplinary operational

infrastructures and diverse communities in research, industry or society at large. B2TERM is the EOSC cornerstone that enables semantic data interoperability, discovery and exploitation across disciplines and thus supports user-centric scientific, industrial and societal applications that maximize the scientific and societal impact of research data.

The KOINE Project, under the H2020 framework, aligns with some key tenets of the GOOS approach to supporting the development of a fit-for-purpose global ocean observing system, through facilitating information flow, accessibility to best practices, and strengthens data sharing and interoperability. These increase the availability and impact of ocean information and data for decision makers. In particular it should multiply any usage across European Open Science Cloud (EOSC) for the GOOS Panels and their work in defining the Essential Ocean Variables, which help connect diverse stakeholders to data and information in a more reliable and efficient way. GOOS therefore supports the development of the KOINE project and looks forward to using its outcomes.

Yours sincerely,

Albert Fischer

Head, Ocean Observations and Services Section

To
Prof. Aue
Leibniz Information Centre for Science and Technology
Welfengarten 1 B
30167 Hannover
Germany

Dr. habil. Werner Leo Kutsch
Director General
Integrated Carbon Observation System
(ICOS ERIC)
Email: werner.kutsch@icos-ri.eu

Helsinki, 23 January 2019

Letter of Support your "KOINE" proposal

Dear Prof. Dr. Auer,

in my role as the Director General of the Integrated Carbon Observation System European Research Infrastructure Consortium (ICOS ERIC), I want to express my support to the "KOINE: Interdisciplinary data interoperability, discovery, and exploitation in the EOSC" project to be submitted under your coordination responding to the H2020 call INFRAEOSC-02-2019, Prototyping new innovative services.

The KOINE project aims to develop B2TERM as a novel and innovative EOSC service for terminology, and integrate B2TERM into (inter)disciplinary operational infrastructures and diverse communities in research, industry or society at large. B2TERM shall be the EOSC cornerstone that enables semantic data interoperability, discovery and exploitation across disciplines and thus supports user-centric scientific, industrial and societal applications that maximize the scientific and societal impact of research data. The expected results from KOINE will be of crucial importance for the further improvement of the data life cycle of ICOS and many other research infrastructures and its goal to establish a sustainable, community-driven service within the EOSC, co-designed with user communities is highly supported by ICOS ERIC.

I look forward to be working with you and your team to realize the project's goals and objectives.

Yours sincerely,

Dr. Werner Kutsch, Director General ICOS ERIC

THE UNIVERSITY OF ARIZONA
College of Agriculture and Life Sciences

Department of Biosystems Engineering

Shantz Bldg., B38, Room 403
1177 E. 4th Street
P.O. Box 210038
Tucson, AZ 85721-0038
Tel: (520) 621-1607
Fax: (520) 621-3963
http://cals.arizona.edu/abe

TIB – Leibniz Information Centre for Science and Technology
Welfengarten 1 B
30167 Hannover
Germany

Hurwitz Lab, University of Arizona, January 26th, 2019

Letter of Support to the H2020 KOINE Project (INFRAEOSC-02-2019 Call)

Dear Prof. Auer,

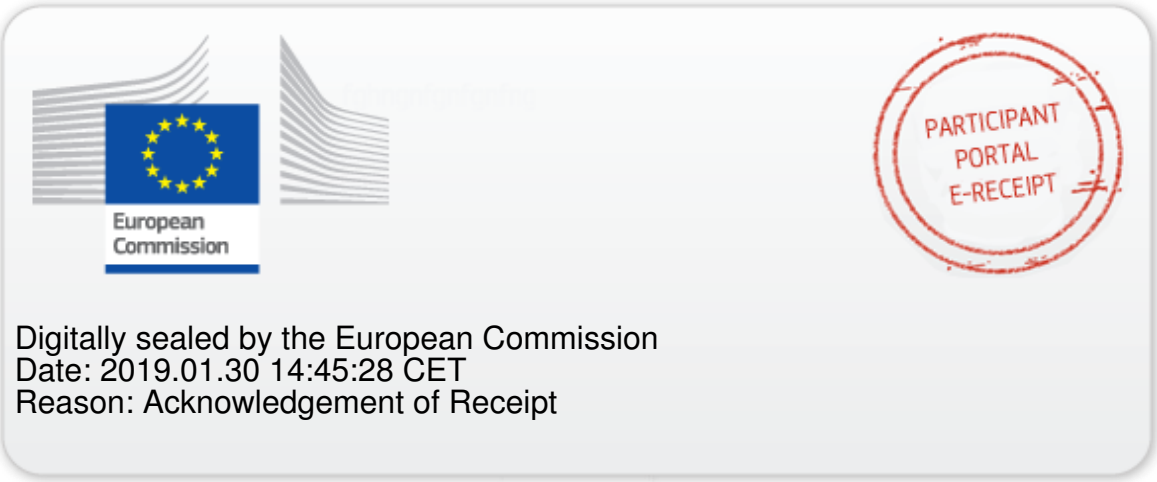
My lab at the University of Arizona is focused on developing computational methods for examining microbiome-host interactions in the environment. We use culture-free high-throughput metagenomics, big data analytics, and statistical analyses to address key questions in: (1) microbial community structure and its relationship to ecological factors; (2) functional and metabolic capacities in microbial communities; and (3) the potential cross-talk between viruses, their microbial hosts and ecosystem function. To promote this work, we strive to make data interoperable from public resources in a cyberinfrastructure platform we develop called iMicrobe. The KOINE project aims directly align with work we are pursuing in our lab, and we enthusiatcially endorse the project and offer our help and expertise.

The KOINE project aims to develop B2TERM as a novel and innovative EOSC service for terminology, and integrate B2TERM into (inter)disciplinary operational infrastructures and diverse communities in research, industry or society at large. B2TERM is the EOSC cornerstone that enables semantic data interoperability, discovery and exploitation across disciplines and thus supports user-centric scientific, industrial and societal applications that maximize the scientific and societal impact of research data.

The iMicrobe platform brings together analysis tools, environmental and microbiome data sets by leveraging National Science Foundation-supported cyberinfrastructure and computing resources from CyVerse, Agave, and XSEDE. The primary purpose of iMicrobe is to provide users with a freely available, web-based platform to: (1) maintain and share project data, metadata, and analysis products, (2) search for related public datasets, and (3) use and publish bioinformatics tools that run on highly-scalable computing resources. iMicrobe is not a data repository but rather a search engine that connects remote datastores to make data discoverable. As a result, projects like KOINE are fundamental to interconnecting datasets on the backend to allow users to compute on data derived from disparate repositories. Taken together, our collaboration with the KOINE project promotes work in the research community towards FAIR data principles.

Yours sincerely,

Bonnie Hurwitz, PhD
Assistant Professor of Agricultural and Biosystems Engineering; University of Arizona
bhurwitz@email.arizona.edu



Digitally sealed by the European Commission
Date: 2019.01.30 14:45:28 CET
Reason: Acknowledgement of Receipt

This electronic receipt is a digitally signed version of the document submitted by your organisation. Both the content of the document and a set of metadata have been digitally sealed.

This digital signature mechanism, using a public-private key pair mechanism, uniquely binds this eReceipt to the modules of the Participant Portal of the European Commission, to the transaction for which it was generated and ensures its full integrity. Therefore a complete digitally signed trail of the transaction is available both for your organisation and for the issuer of the eReceipt.

Any attempt to modify the content will lead to a break of the integrity of the electronic signature, which can be verified at any time by clicking on the eReceipt validation symbol.

More info about eReceipts can be found in the FAQ page of the Participant Portal. (http://ec.europa.eu/research/participants/portal/page/faq)