

## Horizon 2020

### Call: H2020-EO-2017 (Earth Observation)

### Topic: EO-2-2017

### Type of action: RIA (Research and Innovation action)

### Proposal number: 776252

### Proposal acronym: ETCETERA

### Deadline Id: H2020-EO-2017 Table of contents

Section	Title	Action
1	General information	
2	Participants & contacts	
3	Budget	
4	Ethics	
5	Call-specific questions	

#### *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 previous steps in the submission wizard.



Proposal ID **776252**

Acronym **ETCETERA**

## 1 - General information

Topic EO-2-2017

Call Identifier H2020-EO-2017

Type of Action RIA

Deadline Id H2020-EO-2017

Acronym ETCETERA

Proposal title\*

ExTensible and sCalable models EvaluaTion vs Earth obseRvations frAmework

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

Duration in months 30

Fixed keyword 1 Earth Observation / Services and applications

Add

Fixed keyword 2 Visual techniques / Visual analytics / Intelligent data understandi

Add

Remove

Fixed keyword 3 Space data exploitation

Add

Remove

Free keywords

Big Data Technologies, Diagnostics, Analytics, Scalability, Framework, Atmospheric Sciences



Proposal ID **776252**

Acronym **ETCETERA**

### Abstract

*ETCETERA will provide a powerful service framework for a user friendly access to both the whole Copernicus data & information infrastructure and to external data to different profiles of users, with a homogeneous interface the ability to retrieve atmospheric sciences related products such as observational data and numerical models outputs and to calculate on-demand analytics and diagnostics. ETCETERA will provide a fast and intuitive interface to understand the data meaning through a set of new tools exploiting the potential of the Big Data technologies. In other terms, ETCETERA will enable the execution of a set of processes already provided, or with the possibility to configure new ones to extract meaningful information like graphics, text tables or new datasets formatted according to user's needs. The platform to be designed and implemented will have these main features*

- 1) Complementarity with the future Copernicus Data and Information Access Services*
- 2) Capability to extend support to different storage sources*
- 3) Capability to retrieve input data from available data sources and calculate diagnostics and analytics on-demand*
- 4) Capability to configure dynamically different data diagnostics or analytics*
- 5) User friendly and uniform interface to access the data and output calculations. A pilot of the implemented platform will be distributed via a public repository with three use cases up and running on Air Quality topic. It is worth noting that users needs will be fully design with end users.*

*The final results are:*

- 1) To scientific community: a powerful tool to facilitate research by a comprehensive and easy access to data sources and capability to run data analysis.*
- 2) To SMEs: an opportunity to expand their business as simple user of ETCETERA or developing on it tailored services*
- 3) To policy makers, especially at a local level, an effective decision support tool to engage actionable insights*

Remaining characters

68

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



Proposal ID **776252**

Acronym **ETCETERA**

### Declarations

1) The coordinator declares to have the explicit consent of all applicants on their participation and on the content of this proposal.	<input checked="" type="checkbox"/>
2) The information contained in this proposal is correct and complete.	<input checked="" type="checkbox"/>
3) This proposal complies with ethical principles (including the highest standards of research integrity — as set out, for instance, in the <a href="#">European Code of Conduct for Research Integrity</a> — and including, in particular, avoiding fabrication, falsification, plagiarism or other research misconduct).	<input checked="" type="checkbox"/>
4) The coordinator confirms:	
- to have carried out the self-check of the financial capacity of the organisation on <a href="http://ec.europa.eu/research/participants/portal/desktop/en/organisations/lfv.html">http://ec.europa.eu/research/participants/portal/desktop/en/organisations/lfv.html</a> 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	<input checked="" type="radio"/>
- 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	<input type="radio"/>
- as sole participant in the proposal is exempt from the financial capacity check.	<input type="radio"/>
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	<input checked="" type="checkbox"/>
- they have the financial and operational capacity to carry out the proposed action.	<input checked="" type="checkbox"/>
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/her 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](#).





Proposal ID **776252**

Acronym **ETCETERA**

## List of participants

#	Participant Legal Name	Country
1	CAPGEMINI TECHNOLOGY SERVICES	France
2	BARCELONA SUPERCOMPUTING CENTER - CENTRO NACIONAL DE SUPERCOMPUTACION	Spain
3	KONINKLIJK NEDERLANDS METEOROLOGISCH INSTITUUT-KNMI	Netherlands
4	AGENZIA NAZIONALE PER LE NUOVE TECNOLOGIE, L'ENERGIA E LO SVILUPPO ECONOMICO SOSTENIBILE	Italy
5	RASDAMAN GMBH	Germany
6	Green City Solutions GmbH & Co. KG	Germany
7	FONDAZIONE CENTRO EURO-MEDITERRANEO SUI CAMBIAMENTI CLIMATICI	Italy

Proposal ID **776252**

Acronym

**ETCETERA**

Short name **CAPGEMINI TS**

## 2 - Administrative data of participating organisations

<b>PIC</b>	<b>Legal name</b>
947643150	CAPGEMINI TECHNOLOGY SERVICES

Short name: **CAPGEMINI TS**

### Address of the organisation

Street 5-7 RUE FREDERIC CLAVEL

Town SURESNES

Postcode 92287

Country France

Webpage <http://www.capgemini.com>

### Legal Status of your organisation

#### Research and Innovation legal statuses

Public body .....	no	Legal person .....	yes
Non-profit .....	no		
International organisation .....	no		
International organisation of European interest .....	no		
Secondary or Higher education establishment .....	no		
Research organisation .....	no		

#### Enterprise Data

SME self-declared status.....07/12/2004 - 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.



Proposal ID **776252**

Acronym

**ETCETERA**

Short name **CAPGEMINI TS**

### Department(s) carrying out the proposed work

#### Department 1

Department name

MU A&D

☐ not applicable

☐ Same as organisation address

Street

109 avenue Eisenhower

Town

Toulouse Cedex

Postcode

31036

Country

France

### Dependencies with other proposal participants

Character of dependence	Participant	
-------------------------	-------------	--

Proposal ID **776252**

Acronym

**ETCETERA**Short name **CAPGEMINI TS***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☒ FemaleFirst name **Carine**Last name **SAÜT**E-Mail **carine.saut@capgemini.com**

Position in org.

Business Developer

Department

MU A&amp;D

☐ Same as organisation☐ Same as organisation address

Street

109 avenue Eisenhower

Town

Toulouse Cedex

Post code

31036

Country

France

Website

Phone 1

+33531089143

Phone 2

+XXX XXXXXXXXXX

Fax

+33531088100

*Other contact persons*

First Name	Last Name	E-mail	Phone
Keith	Kelly	keith.kelly@capgemini.com	+33 6 47 42 84 07

Proposal ID **776252**

Acronym

**ETCETERA**

Short name **BSC**

**PIC**

999655520

**Legal name**

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

Legal person ..... yes

Non-profit .....yes

International organisation .....no

International organisation of European interest .....no

Secondary or Higher education establishment .....no

Research organisation .....yes

#### 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.**



Proposal ID **776252**

Acronym

**ETCETERA**

Short name **BSC**

### Department(s) carrying out the proposed work

#### Department 1

Department name

Earth Science Department

☐ not applicable

☐ Same as organisation address

Street

NEXUS II building, Jordi Girona 29

Town

Barcelona

Postcode

08034

Country

Spain

### Dependencies with other proposal participants

Character of dependence	Participant	
-------------------------	-------------	--

Proposal ID **776252**

Acronym

**ETCETERA**

Short name **BSC**

### 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 **Francesco**

Last name **Benincasa**

E-Mail **francesco.benincasa@bsc.es**

Position in org.

Research Support Engineer

Department

Earth Science

☐ Same as organisation

☐ Same as organisation address

Street

NEXUS II building, Jordi Girona 29

Town

Barcelona

Post code

08034

Country

Spain

Website

www.bsc.es

Phone 1

+34 934137581

Phone 2

+XXX XXXXXXXXXX

Fax

+XXX XXXXXXXXXX

### Other contact persons

First Name	Last Name	E-mail	Phone
Pierre Antoine	Bretonniere	pierre-antoine.bretonniere@bsc.es	+34934137612
Alicia	Sanchez	alicia.sanchez@bsc.es	
Mar	Rodriguez	mar.rodriguez@bsc.es	+34934137566



Proposal ID **776252**

Acronym

**ETCETERA**

Short name **KNMI**

<b>PIC</b>	<b>Legal name</b>
999518944	KONINKLIJK NEDERLANDS METEOROLOGISCH INSTITUUT-KNMI

*Short name: KNMI*

*Address of the organisation*

Street UTRECHTSEWEG 297

Town DE BILT

Postcode 3731 GA

Country Netherlands

Webpage www.knmi.nl

*Legal Status of your organisation*

#### Research and Innovation legal statuses

Public body .....yes

Legal person ..... yes

Non-profit .....yes

International organisation .....no

International organisation of European interest .....no

Secondary or Higher education establishment .....no

Research organisation .....yes

#### Enterprise Data

SME self-declared status ..... 15/05/2008 - no

SME self-assessment ..... unknown

SME validation sme..... 15/05/2008 - no

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





Proposal ID **776252**

Acronym

**ETCETERA**

Short name **KNMI**

### Department(s) carrying out the proposed work

#### Department 1

Department name

R&D Observation and Data Technologies

☐ not applicable

☒ Same as organisation address

Street

UTRECHTSEWEG 297

Town

DE BILT

Postcode

3731 GA

Country

Netherlands

### Dependencies with other proposal participants

Character of dependence	Participant	
-------------------------	-------------	--

Proposal ID **776252**

Acronym

**ETCETERA**Short name **KNMI***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☐ FemaleFirst name **Wim**Last name **Som de Cerff**E-Mail **wim.som.de.cerff@knmi.nl**

Position in org.

Researcher

Department

R&amp;D Observation and Data Technologies

☐ Same as organisation☒ Same as organisation address

Street

UTRECHTSEWEG 297

Town

DE BILT

Post code

3731 GA

Country

Netherlands

Website

www.knmi.nl

Phone 1

+31302206870

Phone 2

+XXX XXXXXXXXXX

Fax

+XXX XXXXXXXXXX

*Other contact persons*

First Name	Last Name	E-mail	Phone
Alessandro	Spinuso	alessandro.spinuso@knmi.nl	+31302206870
Andrej	Mihajlovski	andrej@knmi.nl	+31302206870

Proposal ID **776252**

Acronym

**ETCETERA**

Short name **ENEA**

**PIC**

999988521

**Legal name**

AGENZIA NAZIONALE PER LE NUOVE TECNOLOGIE, L'ENERGIA E LO SVILUPPO ECONOMICO

Short name: **ENEA**

Address of the organisation

Street Lungotevere Grande Ammiraglio Thaon di Reve

Town ROMA

Postcode 00196

Country Italy

Webpage <http://www.enea.it>

Legal Status of your organisation

### Research and Innovation legal statuses

Public body .....yes

Legal person ..... yes

Non-profit .....yes

International organisation .....no

International organisation of European interest .....no

Secondary or Higher education establishment .....no

Research organisation .....yes

### Enterprise Data

SME self-declared status ..... 03/09/2003 - no

SME self-assessment ..... unknown

SME validation sme..... 03/09/2003 - no

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



Proposal ID **776252**

Acronym

**ETCETERA**

Short name **ENEA**

### Department(s) carrying out the proposed work

#### Department 1

Department name

Department for Sustainability

☐ not applicable

☐ Same as organisation address

Street

Via Anguillarese 301

Town

Rome

Postcode

00123

Country

Italy

### Dependencies with other proposal participants

Character of dependence	Participant	
-------------------------	-------------	--

Proposal ID **776252**

Acronym

**ETCETERA**

Short name **ENEA**

### 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 **Matteo**

Last name **De Felice**

E-Mail **matteo.defelice@enea.it**

Position in org.

Staff Scientist

Department

Department for Sustainability

☐ Same as organisation

☐ Same as organisation address

Street

Via Anguillarese 301

Town

Rome

Post code

00123

Country

Italy

Website

Phone 1

0039051 6098 745

Phone 2

+XXX XXXXXXXXXX

Fax

+XXX XXXXXXXXXX

### Other contact persons

First Name	Last Name	E-mail	Phone
Pierluigi	Fanchin	pierluigi.fanchin@enea.it	0039051 6098194
Mario	Adani	mario.adani@enea.it	00390516098916
Massimo	D'Isidoro	massimo.disidoro@enea.it	00390516098905

Proposal ID **776252**

Acronym

**ETCETERA**

Short name **RASDAMAN**

**PIC**

972352251

**Legal name**

RASDAMAN GMBH

Short name: **RASDAMAN**

Address of the organisation

Street HANS HERMANN SIELING STRASSE 17

Town BREMEN

Postcode 28759

Country Germany

Webpage www.rasdaman.com

Legal Status of your organisation

#### Research and Innovation legal statuses

Public body .....no

Legal person ..... yes

Non-profit .....no

International organisation .....no

International organisation of European interest .....no

Secondary or Higher education establishment .....no

Research organisation .....no

#### Enterprise Data

SME self-declared status .....29/12/2010 - no

SME self-assessment ..... unknown

SME validation sme..... 29/12/2010 - yes

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



Proposal ID **776252**

Acronym

**ETCETERA**

Short name **RASDAMAN**

### Department(s) carrying out the proposed work

#### No department involved

Department name

☒ not applicable

☐ Same as organisation address

Street

Town

Postcode

Country

### Dependencies with other proposal participants

Character of dependence	Participant	
-------------------------	-------------	--

Proposal ID **776252**

Acronym

**ETCETERA**Short name **RASDAMAN***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☐ FemaleFirst name **Peter**Last name **Baumann**E-Mail **baumann@rasdaman.com**

Position in org.

CEO

Department

RASDAMAN GMBH

☒ Same as organisation☐ Same as organisation address

Street

HANS HERMANN SIELING STRASSE 17

Town

BREMEN

Post code

Country

Germany

Website

www.rasdaman.com

Phone 1

+491735837882

Phone 2

+XXX XXXXXXXXXX

Fax

+XXX XXXXXXXXXX

*Other contact persons*

First Name	Last Name	E-mail	Phone
Dimitra	Misev	misev@rasdaman.com	+491735837882





Proposal ID **776252**

Acronym

**ETCETERA**

Short name **Green City Solutions**

**PIC**

930164138

**Legal name**

Green City Solutions GmbH & Co. KG

*Short name: Green City Solutions*

*Address of the organisation*

Street Andreas-Schubert-Str. 23

Town Dresden

Postcode 01069

Country Germany

Webpage [www.greencitysolutions.de](http://www.greencitysolutions.de)

*Legal Status of your organisation*

**Research and Innovation legal statuses**

Public body .....no

Legal person ..... yes

Non-profit .....no

International organisation .....no

International organisation of European interest .....no

Secondary or Higher education establishment .....no

Research organisation .....no

**Enterprise Data**

SME self-declared status ..... 31/12/2014 - yes

SME self-assessment ..... 31/12/2014 - yes

SME validation sme..... unknown

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



Proposal ID **776252**

Acronym

**ETCETERA**

Short name **Green City Solutions**

### Department(s) carrying out the proposed work

#### No department involved

Department name

☒ not applicable

☐ Same as organisation address

Street

Town

Postcode

Country

### Dependencies with other proposal participants

Character of dependence	Participant	
-------------------------	-------------	--

Proposal ID **776252**

Acronym

**ETCETERA**

Short name **Green City Solutions**

### 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 **Liang**

Last name **Wu**

E-Mail **z.wu@mygcs.de**

Position in org.

CIO

Department

IT & Innovation

☐ Same as organisation

☐ Same as organisation address

Street

EUREF-Campus 7-8

Town

Berlin

Post code

10829

Country

Germany

Website

www.greencitysolutions.de

Phone 1

+ 49 176 23164521

Phone 2

+XXX XXXXXXXXXX

Fax

+XXX XXXXXXXXXX

### Other contact persons

First Name	Last Name	E-mail	Phone
Malgorzata	Olesiewicz	m.ole@mygcs.de	

Proposal ID **776252**

Acronym

**ETCETERA**

Short name **FONDAZIONE CENTRO EURO-MEDITERRA**

**PIC**

999419422

**Legal name**

FONDAZIONE CENTRO EURO-MEDITERRANEO SUI CAMBIAMENTI CLIMATICI

Short name: *FONDAZIONE CENTRO EURO-MEDITERRANEO SUI CAMBIAMENTI CLIMATICI*

*Address of the organisation*

Street VIA A IMPERATORE 16

Town LECCE

Postcode 73100

Country Italy

Webpage [www.cmcc.it](http://www.cmcc.it)

*Legal Status of your organisation*

#### Research and Innovation legal statuses

Public body .....no

Legal person ..... yes

Non-profit .....yes

International organisation .....no

International organisation of European interest .....no

Secondary or Higher education establishment .....no

Research organisation .....yes

#### Enterprise Data

SME self-declared status ..... 11/05/2005 - 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.**



Proposal ID **776252**

Acronym

**ETCETERA**

Short name

**FONDAZIONE CENTRO EURO-MEDITERRA**

### Department(s) carrying out the proposed work

#### Department 1

Department name

ASC Advanced Scientific Computing Division

☐ not applicable

☐ Same as organisation address

Street

Via per Monteroni

Town

Lecce

Postcode

73100

Country

Italy

### Dependencies with other proposal participants

Character of dependence	Participant	
-------------------------	-------------	--

Proposal ID **776252**

Acronym

**ETCETERA**

Short name **FONDAZIONE CENTRO EURO-MEDITERRANEA**

### 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 **Sandro**

Last name **Fiore**

E-Mail **sandro.fiore@cmcc.it**

Position in org.

Director of the ASC Advanced Scientific Computing Division

Department

ASC Advanced Scientific Computing Division

☐ Same as organisation

☐ Same as organisation address

Street

Via per Monteroni

Town

Lecce

Post code

73100

Country

Italy

Website

www.cmcc.it

Phone 1

+390832297332

Phone 2

+XXX XXXXXXXXXX

Fax

+390832277603

### Other contact persons

First Name	Last Name	E-mail	Phone
Giovanni	Aloisio	giovanni.aloisio@cmcc.it	+390832297221
Giulia	Galluccio	giulia.galluccio@cmcc.it	+390243986856

Proposal ID **776252**

Acronym **ETCETERA**

## 3 - Budget for the proposal

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	Capgemini Ts	FR	556625	13500	0	0	0	142531,25	0	712656,25	100	712656,25	712656,25
2	Bsc	ES	247500	23500	0	0	0	67750,00	0	338750,00	100	338750,00	338750,00
3	Knmi	NL	154000	8000	0	0	0	40500,00	0	202500,00	100	202500,00	202500,00
4	Enea	IT	92000	23000	30000	0	0	28750,00	0	173750,00	100	173750,00	173750,00
5	Rasdaman	DE	144241	8000	0	0	0	38060,25	0	190301,25	100	190301,25	190301,25
6	Green City Solutions	DE	132655	6000	0	0	0	34663,75	0	173318,75	100	173318,75	173318,75
7	Fondazione Centro Euro- mediterraneo	IT	159300	6000	0	0	0	41325,00	0	206625,00	100	206625,00	206625,00
	<b>Total</b>		1486321	88000	30000	0	0	393580,25	0	1997901,25		1997901,25	1997901,25

Proposal ID **776252**

Acronym **ETCETERA**

## 4 - Ethics issues table

<b>1. HUMAN EMBRYOS/FOETUSES</b>		Page
Does your research involve <a href="#">Human Embryonic Stem Cells (hESCs)</a> ?	<input type="radio"/> Yes <input checked="" type="radio"/> No	
Does your research involve the use of human embryos?	<input type="radio"/> Yes <input checked="" type="radio"/> No	
Does your research involve the use of human foetal tissues / cells?	<input type="radio"/> Yes <input checked="" type="radio"/> No	
<b>2. HUMANS</b>		Page
Does your research involve human participants?	<input type="radio"/> Yes <input checked="" type="radio"/> No	
Does your research involve physical interventions on the study participants?	<input type="radio"/> Yes <input checked="" type="radio"/> No	
<b>3. HUMAN CELLS / TISSUES</b>		Page
Does your research involve human cells or tissues (other than from Human Embryos/ Foetuses, i.e. section 1)?	<input type="radio"/> Yes <input checked="" type="radio"/> No	
<b>4. PERSONAL DATA</b>		Page
Does your research involve personal data collection and/or processing?	<input type="radio"/> Yes <input checked="" type="radio"/> No	
Does your research involve further processing of previously collected personal data (secondary use)?	<input type="radio"/> Yes <input checked="" type="radio"/> No	
<b>5. ANIMALS</b>		Page
Does your research involve animals?	<input type="radio"/> Yes <input checked="" type="radio"/> No	
<b>6. THIRD COUNTRIES</b>		Page
In case non-EU countries are involved, do the research related activities undertaken in these countries raise potential ethics issues?	<input type="radio"/> Yes <input checked="" type="radio"/> 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.)?	<input type="radio"/> Yes <input checked="" type="radio"/> No	
Do you plan to import any material - including personal data - from non-EU countries into the EU?	<input type="radio"/> Yes <input checked="" type="radio"/> No	
Do you plan to export any material - including personal data - from the EU to non-EU countries?	<input type="radio"/> Yes <input checked="" type="radio"/> No	
In case your research involves <a href="#">low and/or lower middle income countries</a> , are any benefits-sharing actions planned?	<input type="radio"/> Yes <input checked="" type="radio"/> No	
Could the situation in the country put the individuals taking part in the research at risk?	<input type="radio"/> Yes <input checked="" type="radio"/> No	



Proposal ID **776252**

Acronym **ETCETERA**

<b>7. ENVIRONMENT &amp; HEALTH and SAFETY</b>		Page
Does your research involve the use of elements that may cause harm to the environment, to animals or plants?	<input type="radio"/> Yes <input checked="" type="radio"/> No	
Does your research deal with endangered fauna and/or flora and/or protected areas?	<input type="radio"/> Yes <input checked="" type="radio"/> No	
Does your research involve the use of elements that may cause harm to humans, including research staff?	<input type="radio"/> Yes <input checked="" type="radio"/> No	
<b>8. DUAL USE</b>		Page
Does your research involve dual-use items in the sense of Regulation 428/2009, or other items for which an authorisation is required?	<input type="radio"/> Yes <input checked="" type="radio"/> No	
<b>9. EXCLUSIVE FOCUS ON CIVIL APPLICATIONS</b>		Page
Could your research raise concerns regarding the exclusive focus on civil applications?	<input type="radio"/> Yes <input checked="" type="radio"/> No	
<b>10. MISUSE</b>		Page
Does your research have the potential for misuse of research results?	<input type="radio"/> Yes <input checked="" type="radio"/> No	
<b>11. OTHER ETHICS ISSUES</b>		Page
Are there any other ethics issues that should be taken into consideration? Please specify	<input type="radio"/> Yes <input checked="" type="radio"/> 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](#)



Proposal ID **776252**

Acronym **ETCETERA**

## 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<sup>1</sup>](#), 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](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.

<sup>1</sup> 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.



# ETCETERA

**ExTensible and sCalable models EvaluaTion versus Earth obseRvations frAnework**

## List of participants

Participant No	Participant organisation name	Participant short name	Country
1 (Coordinator)	Capgemini Technology Services	CG	FR
2	Barcelona Supercomputing Center	BSC	SP
3	Het Koninklijk Nederlands Meteorologisch Instituut	KNMI	NL
4	Agenzia Nazionale per le Nuove Tecnologie, Energia e lo Sviluppo Economico Sostenibile	ENEA	IT
5	Rasdaman GmbH	RAS	DE
6	Green City Solutions	GCS	DE
7	Fondazione Centro Euro-Mediterraneo sui Cambiamenti Climatici	CMCC	IT

## Table of Content

1.	Excellence.....	3
1.1	Objectives.....	3
1.2	Relation to the work programme .....	8
1.3	Concept and Methodology .....	9
1.4	Ambition.....	21
2.	Impact .....	23
2.1	Expected impacts.....	23
2.2	Measures to maximise impact .....	27
3.	Implementation .....	36
3.1	Work plan.....	36
3.2	Management structure, milestones and procedures .....	38
3.3	Consortium as a whole .....	43
3.4	Resources to be committed.....	48
	WP1 - Management .....	48
	WP2 - Data Access .....	50
	WP3 - Diagnostics & Analytics.....	55
	WP4 - User Interface .....	58
	WP5 - Dissemination & Impact.....	60
	Table 3.1b: List of work packages.....	63
	Table 3.1c: List of Deliverables.....	63
	Table 3.2a: List of milestones.....	65

# 1. Excellence

## 1.1 Objectives

**In a world of Big Data, developers work on platform-centric approaches, which are a proven way to deal with volatility, emergence and adaptation to new needs:**

- Scientific needs have evolved towards more interactivity, more collaboration;
- Big Data approaches radically change scientific computing :
  - Complementarity: dealing with Copernicus<sup>1</sup> Data is already a step forward (20 TB/day)
  - Scalability and flexibility are key enablers for a platform-centric approach
- End-users ultimately open to help discover and/or test new ideas

**ETCETERA is an EO platform to foster the digital transformation of the Earth  
Observation ecosystem<sup>2</sup> using Big Data Technologies**

The Earth Observation (EO) ecosystem has a huge opportunity to positively impact people's lives across Europe and the planet, going well beyond the Space sector to almost every part of society and sector of business. However in order to make scientific discoveries or to invent new services for the consumer market, scientists currently have to address new challenges (see Figure 1):

- Deal with **unprecedented data volumes** coming from next-generation instruments and sensors, as well as simulation data that will be used to develop and refine next-generation processing chains. Several hundreds of petabytes will be common in the next decade for a single space mission. Moreover, to get more insights from space data, scientists will combine data from multiple space missions or instruments with additional, “common” data such as public sector information or social network sentiment feeds on key topics such as air quality.
- Design **new kinds of processing chains**, not limited to image processing at pixel level, but also using machine learning or statistical algorithms, much like “Big Data” companies are doing.
- The development lifecycle of these new processing chains will also significantly change, moving towards **more collaboration** and **more interactivity**, exploiting **real-time insights** and employing “**test & learn**” approaches involving automatic “artificial intelligence” analyses as well as interactive “human” ones.

---

<sup>1</sup> <http://www.copernicus.eu/>

<sup>2</sup> Earth Observation Ecosystem refers in this proposal to Copernicus EO and non EO data and services.

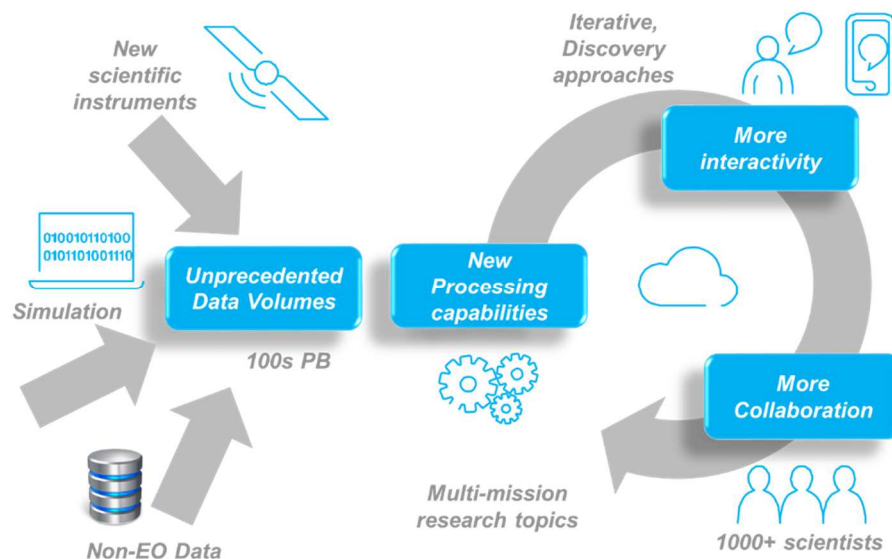


Figure 1: New challenges for scientists

ETCETERA, as outlined in this proposal, is part of a platform-centric approach to **facilitate the exploitation of Big Data EO ecosystem**, and to support the development of new EO services as well as non-EO services that are enriched with EO data. It is part of a vision for a future EO framework that includes current Copernicus initiatives such as Data & Information Access Services (DIAS) on which ETCETERA will build. The future ETCETERA framework will provide the scalability and flexibility required to boost creativity in the context of such endeavors, increasing business agility while controlling risks and reducing time-to-market. It will provide tools that support more interactive, collaborative and innovative ways of working, to publish and consume new generation EO services. Ultimately, the ETCETERA framework will help to discover and test new ideas and bring them to their target population more quickly and easily.

*A platform-centric approach brings disruptive capabilities and a more agile culture for building ecosystems and creating new services*

ETCETERA proposes then a flexible solution built on the following pillars:

- Enabling challenges of the Big Data context such as complementarity and heterogeneity of the appropriate tools and processes, allowing the combinations of the Copernicus data providers (Sentinel data providers: European Space Agency -ESA<sup>3</sup>- and European Organization for the Exploitation of Meteorological Satellites EUMETSAT<sup>4</sup>) along with Copernicus Information providers (Copernicus Core services: Copernicus Atmosphere Monitoring Service - CAMS- , Copernicus Climate Change Service -C3S- and Copernicus - Marine Environment Monitoring Service - CMEMS) and with other non-EO data sources to produce novel set of tools and services.
- Providing an extensible service framework able to orchestrate a variety of existing Earth Observation dissemination platforms affording an increasing volume of data of several petabytes. The ETCETERA

*ETCETERA addresses the challenges of the Copernicus program and is aligned with the future EO platform components and in particular with DIAS*

<sup>3</sup> ESA, <http://www.esa.int/ESA>

<sup>4</sup> EUMETSAT, <http://www.eumetsat.int/website/home/index.html>

solution will provide an **extensible and scalable service accessible via an homogeneous interface**, both as Application Program Interface (API) and Graphical User Interface (GUI), to retrieve EO data, numerical models outputs, analytics and diagnostics applied to Atmospheric Sciences and compute new ones.

- Orchestrating, via the proposed framework, the existing Earth System data dissemination platforms including Copernicus EO and model data and information within DIAS and from non-Copernicus sources such as ESA servers, non-European satellite data (MODIS<sup>5</sup>, CALIPSO<sup>6</sup>), in-situ observations (EARLINET<sup>7</sup>, AERONET<sup>8</sup>) and data outputs for, WMO GAW<sup>9</sup>, ESGF<sup>10</sup>, among others.

ETCETERA is therefore a demonstrative project and its outputs will produce the essential scientific basis for future EO monitoring and high-quality data transfer systems. The needs from the scientific community, the public authorities and private sector with a special focus on Small and Medium Enterprises (SME) are addressed via the new set of tools available from the ETCETERA framework. Free, full, open and proprietary data aspects are taken into consideration. In particular, ETCETERA will benefit directly from additional data on dust aerosols, particulate matters and water vapour from Sentinel 3 while Sentinel 4 and Sentinel 5 will bring complementary information on the chemical composition of the troposphere. **Public authorities may have legal obligation to monitor air quality in the future, expanding the market to Value Added Service companies, in urban air quality monitoring for instance.** The Copernicus context where the data are free and open access encourages to address this targeted audience.

Piloting an innovation project is fundamentally different from managing a delivery project. To conduct the project, the consortium has adopted the “Lean Startup” methodology<sup>11</sup>. Lean start-up aims to shorten product/service development cycles by adopting a combination of value-hypothesis-driven experimentation, iterative product/service releases, and validated learning. The former has become popular for piloting such innovative projects to success has become popular with start-up companies over the past few years. The methodology will be further explained in the section 1.3.

*Using Lean Startup methodology to ensure ETCETERA provides maximum impact*

The project, led by Capgemini Technology Services, brings together world-class research institutes in Europe with strong experience in climate and air quality impact modelling, parallel diagnostics and large scale scientific data management, such as the Barcelona Supercomputing Center (BSC), the Italian National Agency for New Technologies, Energy, and Sustainable Economic Development (ENEA), the Netherlands Meteorological Institute (KNMI) and Euro-Mediterranean Centre on Climate Change (CMCC), together with private companies specialized in air quality sector (Green City Solutions) and Big Data Technologies (Rasdaman) having a diversity of atmospheric-sciences-related issues.

---

<sup>5</sup> MODIS, <https://modis.gsfc.nasa.gov/>

<sup>6</sup> CALIPSO, <https://www-calipso.larc.nasa.gov/>

<sup>7</sup> EARLINET, <https://www.earlinet.org/>

<sup>8</sup> AERONET, <https://aeronet.gsfc.nasa.gov/>

<sup>9</sup> WMO GAW, [http://www.wmo.int/pages/prog/arep/gaw/gaw\\_home\\_en.html](http://www.wmo.int/pages/prog/arep/gaw/gaw_home_en.html)

<sup>10</sup> ESGF, <http://esgf.llnl.gov/>

<sup>11</sup> [https://fr.wikipedia.org/wiki/Lean\\_Startup](https://fr.wikipedia.org/wiki/Lean_Startup)

**Four main drivers** for the Big Data Shift of the Earth Observation ecosystem have been identified as success factors for this proposal:

1. **A downstream extension of the image processing value chain in the digital world**, opening new markets through new business models. The traditional, product-based, “remote sensing” Business to Business (B2B) market will evolve towards downstream digital applications for business, consumer and local authority markets. Completely new ways of consuming and leveraging imagery will generate demand and will have a huge economic impact: Google has published an Oxera report<sup>12</sup> that estimates the revenues from global Geo services at \$150 billion to \$270 billion per year. In the 2014 State of the Satellite Industry Report published by the Satellite Industry Association, the revenue from remote sensing activities is valued at \$1.5 billion. Rebooting the business model is a “common” practice when entering the digital world<sup>13</sup>. ETCETERA will include a real-life demonstrator of such a downstream service in the form of Green City Solutions’ air quality services.
2. **Newcomers among the key actors of imagery value stream**. Google<sup>14</sup> and Amazon<sup>15</sup> already have moved towards these huge opportunities of the future Earth Observation market. New imagery and more generally information coming from new sensors are becoming available all the time. For example, the European Commission (EC) via the Copernicus program with Sentinel satellites provides free and open imagery. The goal is to stimulate the emergence of new markets and actors, including start-ups that will disrupt the value that can be extracted from imagery and other open data.
3. **The use of platforms that bring scalability and flexibility to boost creativity, increasing business agility while controlling risks and reducing time-to-market**. These platforms combine all the required state-of-the-art capabilities (storage, processing, large-scale collaboration, mobile access, connectivity with sensors and end-users, etc) that have emerged from the digital transformation in the pioneer sectors. The same platform can be used for developing new services as well as for operating them, thus drastically reducing the time to market for new services.
4. **A more open, collaborative culture reshaping organizations and building ecosystems** with spatial agencies, satellite manufacturers, storage and processing platforms providers, research centres, scientific communities, start-ups and even end-users.

**Three main components** are the base of the foreseen solution (see Figure 2):

- **A Data Access component**: its key characteristics are to be scalable and open to both EO and non-EO data. It is capable of both providing data as inputs to analytics processing, as well as storing its outputs, which in turn can be used as inputs to the next steps of the processing chain.
- **An Analytics component**: that can take inputs from various sources via the Data Access layer and processe them to produce new data sets and actionable insights. The processing should support various types of scalable processing techniques and languages from the world of Big Data.
- **A User Interface component** for 2 distinct purposes:
  - Tools and services for service developers to efficiently leverage the Data Access and Analytics components. This will essentially be a packaging and integration of existing open source tools aligned with the underlying technologies of the Data Access and Analytics layers. This will be performed via an API.

---

<sup>12</sup> <http://www.oxera.com/Latest-Thinking/News/January-2013/Oxera-quantifies-the-benefits-of-Geo-services-to-g.aspx>

<sup>13</sup> <https://www.capgemini-consulting.com/rebooting-the-business-model-for-the-digital-age>

<sup>14</sup> <http://www.skyboximaging.com/>

<sup>15</sup> <https://aws.amazon.com/es/public-datasets/landsat/>



- Interactive Applications to hold the business logic and the user experience for new services. The objective is to go beyond simply displaying images (e.g. air quality visualized in a heat map), to help end-users make faster and more relevant decisions (e.g. combine images with information gathered from the field into a fully-featured air quality management application for public authority decision makers or business sector). This will be done via GUI.

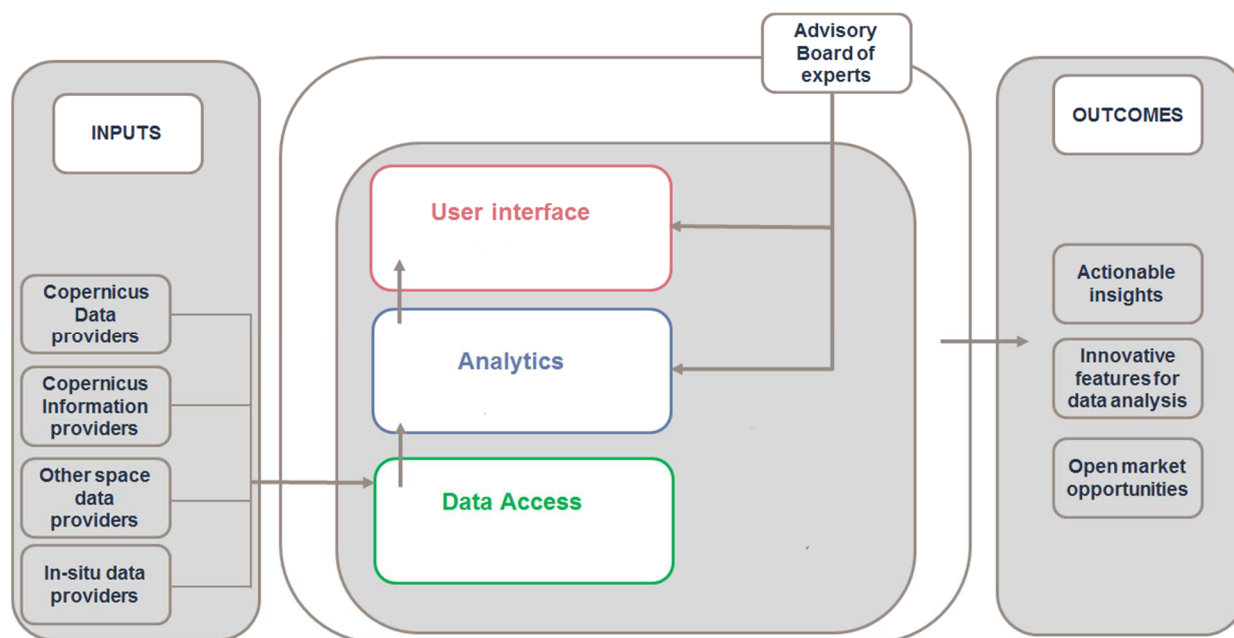


Figure 2: ETCETERA platform overview

In particular, ETCETERA will address two different kinds of objectives: the general objectives referred to the global aspects of ETCETERA and the specific objectives devoted to emphasize particular facets of the same project. The table below shows the global objectives of ETCETERA.

General Objectives (GO)	Description
<b>GO1</b>	Facilitate the access to very high resolution and disseminated data to extract scientific and political value of it.
<b>GO2</b>	Explore the scientific and technological frontiers of the present features in order to provide guidance for the development of future generation of monitoring and data-processing systems.
<b>GO3</b>	Engage with targeted end-user groups in key European economic sectors to strengthen their competitiveness, growth, resilience and ability by exploiting new scientific progress.

In addition, we have identified seven specific objectives which are summarized as well in the table below.

Specific Objectives	Description	Work packages involved to achieve the objective
<b>SO 1</b>	Adapt and integrate Big Data technologies focused on Earth Observations analysis to provide an effective and unified access to Copernicus Data & Information through a homogeneous interface.	WP2, WP3, WP4
<b>SO 2</b>	Create modular, scalable and complementary services.	WP2, WP3, WP4
<b>SO 3</b>	Produce new sets of tools such as innovative features for analyses and actionable insights in order to disseminate and promote the benefits of ETCETERA tools to their potential users.	WP3, WP5
<b>SO 4</b>	Design of a business model service-based for opening new business opportunities.	WP5, WP2, WP3, WP4
<b>SO 5</b>	Take into consideration and reach experts and non-experts users, including SMEs and stakeholders.	WP5
<b>SO 6</b>	Establish cooperation between science and policy actions at European and international level namely the support of the development of effective air quality policies, optimize public decision making and increase capability to manage high-level pollution risks.	WP5, WP4
<b>SO 7</b>	Demonstration of the service relevance and features via a use case on air quality.	WP3, WP4, WP5

## ***1.2 Relation to the work programme***

The Copernicus operational activities, as established in the annual work programmes, and which covers the time period 2014-2020, describe the actions to be undertaken to ensure funding and system evolution. In particular, the actions needed to implement Copernicus encompass evolving user needs and technological developments. As clearly explained in the guidance documents supporting the EO-2-2017 call, the present proposal aims at optimally aligning the relevance of the ETCETERA service framework to Copernicus and in particular to the on-going development of an Integrated Ground Segment (IGS) that enables the integration of Copernicus data and information in a Big Data context.

The ETCETERA proposal relates to the topic call identifier **EO-2-2017: EO Big Data Shift** within the work programme part **Leadership in Enabling and Industrial Technologies - Space**. This topic contributes to the

Big Data challenge by addressing the fundamental research problems related to the scalability and responsiveness of analytics capabilities, with a special focus on industry-validated, user-defined challenges.

The ETCETERA project will develop a framework demonstrating the use of new sets of tools based on new architecture, innovative algorithms and visualization tools to explore and exploit those data, discover new relationships, and enrich them with other remotely data sources by optimizing the data exploitation.

The ETCETERA project brings together an ecosystem with top academic and private players in the field of Big Data management and data analytics. The Copernicus complementarity and the flexibility aspects for the non-Copernicus data sources will be targeted under the Data access work package (WP2). The acceptance of the framework by traditional and non-traditional users will be tested in use cases under the Analytics and Diagnostics work package (WP3) and reflected via the integration of Users interface (WP4). The technology transfer will be optimized by being carried by a SME consortium member to expand and develop new market opportunities.

### ***1.3 Concept and Methodology***

#### **a) Concept**

In the last few years, the growth of Copernicus infrastructure, with the impressive quantity of data to be disseminated, represents a unique opportunity for the society to understand the Earth System from various points of view, and to take advantage of a massive access to that kind of data to **improve scientific research, foster economic growth and facilitate decisions taking at political level**.

To face these challenges, a unified Copernicus dissemination infrastructure is necessary to coordinate all different platforms operated by Member States and transnational institutions like ESA and EUMETSAT, and service operators. The development of the future DIAS, expected soon, as a unique access point to the whole Copernicus ecosystem, is a very important achievement in this context.

The ETCETERA proposal aims **to design and develop a framework** which will adopt Big Data and Machine Learning solutions to provide innovative services on top of Copernicus. The main objective is to allow a **full and homogeneous access** to Copernicus, to provide the possibility to extend the number of data sources to non-Copernicus platforms, and build above it a **Big Data analytics** service layer to extract meaningful information and provide this to the final users. As stated in previous section, the architecture is based on three components:

- **A distributed access component**, which will provide data retrievals from a potentially infinite variety of input sources: local or remote file systems, Copernicus EO data providers, etc. and include
  - a system of dynamic extensions that will give the capabilities to transform (on-the-fly or deferred depending on the data accessibility) external data in an internal format optimized for data analytics purposes. For each new type of data source, a new input adapter should be implemented to extend the system. According to needs and resources, some data can be cached to speed up the processes.
  - a database which stores all the information related to the data source, its features and metadata, etc.
- **A diagnostics/analytics component**, which will provide both core data manipulations and calculations capabilities (analytics) and orchestration support to run dataflows (diagnostics). In the

same way as storage, the main engine can be dynamically extended with as many types of outputs as required through output adapters according to user's needs.

- **User-friendly interfaces component**, both in terms of GUI and APIs.

Each component will provide its own API to communicate to other services and to allow an independent deployment. To address interoperability, all APIs developed within the project will adhere to officially recognized standards (see Standardisation section). **The three components described above should be modular and will have distributed systems capabilities such as federability.**

An extensive documentation (tutorial, step-by-step guides) will be provided to minimise the learning curve of the service. Furthermore, a set of test cases will be chosen to demonstrate how the proposed service would enhance the interoperability of the data-driven atmospheric research and, at the same time, maximise the reliability of the scientific workflow. The software related with the ETCETERA project will be released as free software on public repository such as GitHub so that all developers can write their own extensions. ETCETERA will design the tools to be offered in the service framework according to the users' needs as defined in WP3.

From the beginning of the project, **the aim is to engage with potential users to shape various aspects of the work** to be done. ETCETERA will place an important emphasis on the air quality demonstration project to be developed in collaboration with Green City Solutions, a SME consortium member. This will serve as a showcase to test our value-propositions and solutions with real-life users, both those using ETCETERA to develop the air quality service and end-users of the final air quality solution.

Our aim is to ensure that the innovations, technologies, methodologies and approaches promoted in ETCETERA are exploited in the EO value stream, showcasing its potential and enticing new service development. It is worth noting that the expected economic impact of the air quality sector, as reported in the Copernicus Market, states that small and micro companies operating in the air quality field interviewed for the study -such Plume Labs, which is a start-up based in France, offering information on air quality in the main cities- consider that around 7% of their revenues can be attributed to Copernicus. It improves the quality of the forecasts at the scale of city by up to 20% and can help taking opportunities in niche markets (tourism, etc.).

User engagement will be ensured taking advantage of external actors' collaboration but above all internal consortium members' experience. Hereafter, an overview of the users and stakeholders Advisory Board:

Scientific community	Business Sector	Public Organizations
BSC <sup>(*)</sup>	Green City Solutions <sup>(*)</sup>	INERIS
KNMI <sup>(*)</sup>		
ENEA <sup>(*)</sup>	ARIA Technology <sup>16</sup>	AIREAS <sup>17</sup>

<sup>16</sup> <http://www.aria.fr/>

<sup>17</sup> <http://www.aireas.com/welcome-to-aireas/>

CMCC <sup>(*)</sup>		
Max Planck Institute for Meteorology	Iseo Environnement <sup>18</sup> (+ international network of the Environnement SA group)	AEMET (Spanish Meteorological Agency) <sup>19</sup>

(\*) Consortium member

The final outcome of the project will be demonstrator of the designed framework, with the use cases on air quality as further explain below, implemented and up and running, but with the capability to be scalable and extended according to more diverse needs and types of data sources. **As research project, ETCETERA is clearly situated in the spectrum of “idea to application”, according to the Technology Readiness Level (TRL) number 5 (technology validated in relevant environment) of annex G of work programme.**

ETCETERA consortium, reinforced with the chosen methods and approaches such as “Lean Startup” or User Experience design (explained below in the “methodology” section), by their inter-disciplinarity, has all the skills, resources and tools required to meet these challenges. The inter-disciplinarity of the consortium, composed by computer and data scientists, climate and air quality researchers and project managers with experience in European projects, will help enabling innovation and achieving the intended impact.

### Synergy with current initiatives - Link with Copernicus

All European research projects on Atmospheric Sciences are potentially linked with ETCETERA, due to the fact that the scientific fields like climate and air quality base their studies on data analysis of Earth Observation and numerical models outputs. During the ETCETERA activities, a strong link to EU projects, community-driven organisations and Copernicus activities will be set up and ETCETERA will create synergy with current initiatives.

The following table describes the most relevant ones:

Initiative Name	Short description	Relevancy to ETCETERA
<b>Research Data Alliance<sup>20 21</sup></b>	RDA’s goal is to foster data sharing among different scientific communities by so-called Interest Groups (IG) and Working Groups (WG). Almost all members of the ETCETERA consortium are chairs or members of one or more	Possibility to discuss (online and during the plenary events) and disseminate project results with scientists and stakeholders; discussion about standardization. RDA groups’ outcomes will fertilize project developments at

<sup>18</sup> <http://www.iseo.fr/>

<sup>19</sup> <http://www.aemet.es/en/portada>

<sup>20</sup> <https://www.rd-alliance.org/node>

<sup>21</sup> CMCC has recently been granted by RDA Europe in the context of a Call for Collaboration project 2016 to adopt RDA recommendations for persistent identifiers management into their data analytics framework. <https://www.rd-alliance.org/rda-europe-call-collaboration-projects-2016-edition>

	of these groups: “Big Data”, “Geospatial”, “Array Database Assessment” and “Weather Climate and Air Quality”.	many levels: data management, standardization and use cases definitions.
<b>Open Geospatial Consortium (OGC)</b>	OGC is an international geospatial standards organisation.	OGC Web Services standards have been used as an API behind previous Copernicus and Non-Copernicus projects (such as CLIPC, ESGF, etc.); their use allows for a high level of interoperability for ETCETERA with existing and future projects.
<b>INSPIRE Directive</b>	INSPIRE aims to create EUs spatial data infrastructure for environmental and climate policies.	Description of ETCETERA services according to INSPIREs Catalog Service for the Web CSW standard allows for standardised access with maximum flexibility to all services generated by DIAS (including OGCs WCS, WMS, WPS, OpenDAP).
<b>Earth System Grid Federation (ESGF)</b>	The Earth System Grid Federation (ESGF) Peer-to-Peer (P2P) enterprise system is a collaboration that develops, deploys and maintains software infrastructure for the management, dissemination, and analysis of model output and observational data.	ETCETERA framework will be fully compatible to integrate ESGF system. In that way will be provided access to the datasets exposed by the ESGF nodes.
<b>EUDAT Collaborative Data Infrastructure<sup>22</sup></b>	EUDAT is a Service-oriented, Community driven, Sustainable and Integrated initiative, providing recommendations and software for the good use of data in different communities. Some partners of ETCETERA consortium are participating in EUDAT as infrastructure provider and/or user pilots.	Issues like data transfer and replica are very common to most communities sharing data, and even if file formats and organization are specific to the scientific domain, solutions could be easily extrapolated to other types of contexts. Above all, interoperability and integrations with this existing infrastructure will open the door for future

<sup>22</sup> <https://www.eudat.eu/eudat-cdi>

		extensions of the ETCETERA framework. ETCETERA software development and Data Management Plan will be inspired, among others, also by EUDAT solutions.
<b>Copernicus projects</b> (C3S_23a Software Infrastructure for the Climate Data Store, C3S_34a Global climate projections: data access, product generation and impact of front-line developments)	C3S-34a: delivers software to calculate and present metrics, statistics, time series, and tailored applications from climate model data.	ETCETERA software and tools are developed with interoperability, modularity and reusability with current Copernicus efforts.
<b>Copernicus QA4SEAS</b>	QA4Seas (Quality Assurance for Multi-model Seasonal Forecast Products) develops a strategy for the evaluation and quality control (EQC) of the multi-model seasonal forecasts provided by the Copernicus Climate Change Service (C3S).	The metadata conventions and propagation system defined in Copernicus in general and in the CDS in particular within QA4SEAS will be of interest for ETCETERA in the definition of the DMP following the global objective of making ETCETERA fit in the global Copernicus architecture.

## b) Methodology

As already mentioned, **the overall methodology will be based on Lean Start-up principles that are ideal for scenarios of innovation** and will be coordinated as part of the activities of WP1. Lean start-up is customer-centric methodology for developing products in contexts of high uncertainty, generally used by high-tech start-up companies. This methodology aims to shorten product development cycles and minimize risks of delivering products that are not relevant for their target population. The general principles are based on continuous deployment of improvements to a solution that includes an explicit feedback loop: Design, Build, Measure and Repeat.

We will use the lean start-up approach by adopting a combination of value-hypothesis-driven experimentation, iterative design and customer feedback. So, the first step of our methodology will be to understand users' needs according to several hypotheses and then develop a Minimum Viable Product (MVP). A MVP is a "version of a new product which allows a team to collect the maximum amount of validated learning about customers with the least effort".

Due to the variety of expertise needed for the execution of the work, when the phase of software development is integrated with the system administration, the overall project will be characterized by the so-called DevOps

approach<sup>23</sup>, a new tendency that breaks the traditional separation inside the Information Technology world between development and operations. It is the practice of operations and development engineers participating together in the entire service lifecycle, from design through the development process to production. This happens at technical level and integrates perfectly with the global lean methodology and with the particular methodology that each WP will adopt. **The lean start-up will be observed at each consortium project meetings in order to monitor the progress and ensure the solution within reach and avoid any dead end.**

ETCETERA is organised in five work packages. The work packages structure a set of tasks, deliverables and milestones associated. The list of work packages is:

- WP1: Management
- WP2: Data access
- WP3: Analytics and Diagnostics
- WP4: User interface
- WP5: Dissemination & Impact

**Each work package leader will adopt the suitable methodology to achieve its own specific objectives, as described hereafter in the following work package methodologies description:**

**WPI Management:** The main objective of this work package is to ensure an effective, smooth and high quality implementation of the project, with respect to general administration, management practices and financial management. The work package intends to provide effective daily project management that includes coordination, EC reporting, financial information, quality assurance procedure, risk management, data coordination, meeting organization in order to achieve project objectives on time, on budget and at the highest quality level.

**WP2 Data Access:** the goal of the work package is to set up the Big Data management and query platform, compatible with Copernicus data services and at the same time compatible with non-Copernicus data. This latter will be based on the pre-existing rasdaman scalable array analytics engine, plus further tools coupled to Rasdaman. The result is the service used by WP3. Therefore, a support task is defined to accommodate all needs of WP3, but also of external users of the ETCETERA data platform. The strategy is to have an early ramp-up phase where existing tools – in particular: rasdaman – get installed for a preliminary service; this also serves for a requirement elicitation. After this common task, work will spread out into individual enhancements.

Organizationally, three increments will deliver successively advanced services (basic/advanced/consolidated). This ensures that requirements expressed initially, but also requirements emerging, can be considered adequately. At every milestone, the status delivered will be discussed with stakeholders, and new steps for the next increment will be prioritized in a consensus process. A dedicated task ensures continued support and quality management.

**WP3 Analytics and Diagnostics:** The goal of WP3 is to implement a Big Data analytics and diagnostics framework on top of the access layer provided by WP2. The outcome will be a software system accessible through a fully standardized API that will be able to define, configure and execute data diagnostics and analytics computation over Atmospheric Sciences datasets. Results of queries will be inputs for the User

---

<sup>23</sup> <https://en.wikipedia.org/wiki/DevOps>



Interface layer provided by WP4. To support the development of the WP3 platform, an incremental and iterative approach will be adopted according the “Agile”<sup>24</sup> philosophy. In particular, among the Agile methodologies, “Scrum”<sup>25</sup> has been selected. Its incremental and iterative approach extends the traditional waterfall methodology (requirements analysis, design, coding, integration, testing, deployment) blending all its components into each iteration, adapting dynamically to discovered realities.

According to Scrum roles definitions, the WP leader (BSC) will take the role of Owner and Master. The iterations will have two weeks duration and one new release will be produced every 3 months (~6 iterations per release) with major releases adjusted to milestones and deliverables defined in the WP3 section. Use cases will have a fundamental role, as users will contribute to requirements analysis, participate actively to iterations writing “user stories” and to testing and validation phases about the required functionalities.

**WP4 User interface:** The MVP will be created with the principles of User Experience (UX) Design and quickly tested in real conditions with a panel of future users to improve the product with their feedback. More references are provided in Annex. The principle of the UX Design is to “make the user’s digital experience as positive as possible by thinking the experience before the product”. The aim is to remove all functional and emotional frictions that users may encounter. According to this philosophy, we will create our MVP’s user interface using Capgemini’s “Rapid Design Visualization” (RDV) methodology. The RDV method is a fast, collaborative and iterative process that is used to develop Users Interfaces (UI) that are adapted to the intended usage of the system. This method will include workshops with all the WP4 contributors and policy makers from the Advisory Board.

To begin, the WP participants will identify user needs, elaborate the first workflow of our solution and sketch the first draft of our UI in the form of “storyboards”. Then, they will create the wireframes of the user interfaces. This step will allow them to refine and animate the displays to illustrate and validate their interactions. It also allows them to refine information needed (including when and how) from the backend systems, notably from WP2 and WP3. Finally, the definitive visual displays will be created to make sense of the interfaces and animate them with a design prototype to be testing by real users.

**WP5 Dissemination & Impact:** The objectives of this work package are to ensure an appropriate dissemination and communication throughout the project supporting a wide-range of dissemination activities of the ETCETERA framework, keeping focused on the application on Big Data technologies for EO. One of the big challenges of this project is to strengthen the link between the EO/Climate and Big Data communities. In this way, communication and dissemination activities will also take place during events like scientific conferences and industrial workshops for both communities.

Given the involvement of many partners in the RDA activities, project activities will be discussed and disseminated in multiple working and interest groups (WG/IG), for example Big Data IG and Weather, Climate and Air Quality IG. Regarding the exploitation plan to be performed under the WP5, Green City Solutions would compare the data generated by the ETCETERA project with their own data and examine the accuracy and coherence of both datasets.

The focus will specifically be on air pollution values in hotspot areas in cities, especially regarding particulate matter and other harmful gases. If a correlation between the datasets can be observed and confirmed, the information concerning the air quality can be used for the development of the data driven business model.

---

<sup>24</sup> <http://agilemanifesto.org/>

<sup>25</sup> <http://scrummethodology.com/>

They will focus on the user impact of the use cases stretching from municipalities to companies and even citizens as existing customers of their services (WP5). Inquiries from governmental customers have already shown that analysis on the efficiency of existing measures to prevent air pollution are in a high demand especially as a Software as a Service (SaaS). Prediction monitoring of air pollution levels could be another potential service which would allow Green City Solutions to grow its business and expand.

Figure 3 below showcases the architecture overview within the methodology. Management and Dissemination activities are considered as activities covering the overall project and are not represented in the figure.

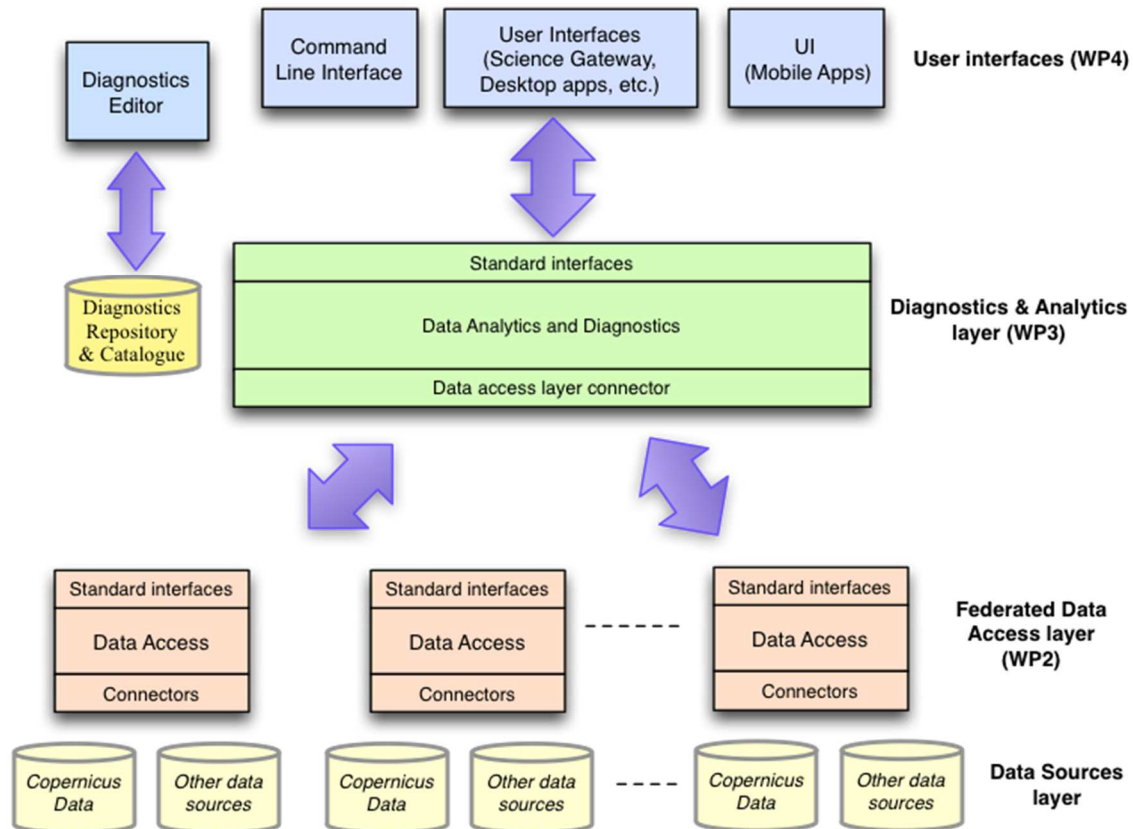


Figure 3: ETCETERA architecture within the methodology overview

### IT security: authentication

A key role will obviously be played by authentication. As each user can have different privileges to access data according to permissions allowed by data providers, all ETCETERA components should be able to manage credentials in a secure way and through secure channels. Cutting-edge approaches like those investigated by e-Infrastructure projects, namely EGI-AARC<sup>26</sup> and the already mentioned EUDAT will be considered. One example can be *OAuth2*<sup>27</sup>. This authentication method is also being implemented in projects like ESGF and Climate4Impact, which supports OAuth2 identity providers from Google and CEDA/BADC, and the consortium is currently assessing the feasibility in the ETCETERA context. The greatest benefit of *OAuth2* is that the site does not have to keep passwords. Credentials (including passwords) are securely stored at the site of the identity provider. Other types of access tokens could be experimented in the form of

<sup>26</sup> <https://www.egi.eu/tag/aarc/>

<sup>27</sup> <https://oauth.net/2/>

universally unique identifiers (UUID<sup>28</sup>). When a user is signed in, he is able to generate and manage his own set of access tokens, which he can use to access all sorts of API's, for example Web Processing Services. For this, the access token will be part of the URL. With *OAuth2* the access token is part of the header. When the portal encounters a valid access token at the right place in the URL, access is granted.

### *The choice of the uses case: air quality*

Air pollution is the first environmental health hazard in Europe, reducing the life expectancy of those affected and contributes to drastically increase of serious illnesses such as heart disease, respiratory disorders and cancers. According to a report published by the European Environment Agency (EEA), air pollution continues to cause more than 524,000 premature deaths in Europe every year and has consequently an important annual economic damage.

The EEA 2015 report on air quality in Europe reviews the European population's exposure to air pollutants and provides an overview of air quality based on data from official monitoring stations in the entire Europe. The report shows that most urban dwellers (90%) remain exposed to levels of air pollution deemed harmful by the World Health Organization (WHO). Starting from those facts, there is an urgent need to improve research and development to provide reliable and continuous related information to stakeholders and policy makers.

A significant existing Copernicus benefit is a 60% of higher accuracy for analysis of the impact of trans-boundaries pollutants on air quality, and this is one of the reason why air quality sector has been identified as an EO downstream service in the Copernicus Market<sup>29</sup> report (issue November 2016). Benefits are likely to increase as more Sentinels satellites become operational and as users progressively discover the Earth Observation's value added services from downstream providers. Facing the current need for a more user friendly and fast access to air quality related data to ensure that society can reduce vulnerability to weather extremes and air quality events, ETCETERA will address this sectorial use case tailored to three cross profiles users' needs, namely **Scientific Community**, **Business sector** and **Public authorities**. They are briefly presented below and will be fully covered under the WP3 Analytics and Diagnostics where the user's needs will be co-designed hand by hand with users representatives, referred as the Advisory Board hereafter. Support letters are provided in Annex and showcase the strong interest of following closely the ETCETERA project along its development, to participate to some meetings and the two workshops and to discuss relevant questions with the consortium. Figure 4 showcases the panel of the potential stakeholders.



Figure 4 ETCETERA Stakeholders

<sup>28</sup> [https://en.wikipedia.org/wiki/Universally\\_unique\\_identifier](https://en.wikipedia.org/wiki/Universally_unique_identifier)

<sup>29</sup> [http://www.copernicus.eu/sites/default/files/library/Copernicus\\_Market\\_Report\\_11\\_2016.pdf](http://www.copernicus.eu/sites/default/files/library/Copernicus_Market_Report_11_2016.pdf)

The different profiles addressed within the use cases are described hereafter:

### Use Case 1 - R&D/Scientific community

All scientific partners namely BSC, ENEA, KNMI and CMCC will be the actors of the use case 1. They will provide inputs and specific requirements for the development of innovative features for data analysis. The benefits from the ETCETERA services on R&D are

1. improve R&D on air quality sector
2. improve forecast on air quality
3. enable an easy access to reliable diagnostics on air quality to foster research
4. outreach on air quality capacities research.

A good application can be the study of Sand and Dust Storms (SDS), which are an important threat to life, health, property, environment and economy in many countries, and play a significant role in different aspects of weather, climate and atmospheric chemistry. There is an increasing need to be prepared to SDS and dust intrusions and to mitigate their impacts. *Reanalysis* (model simulations considering assimilation) can overcome the sparse coverage, low temporal resolution and partial information provided by the measurements, particularly over deserts.

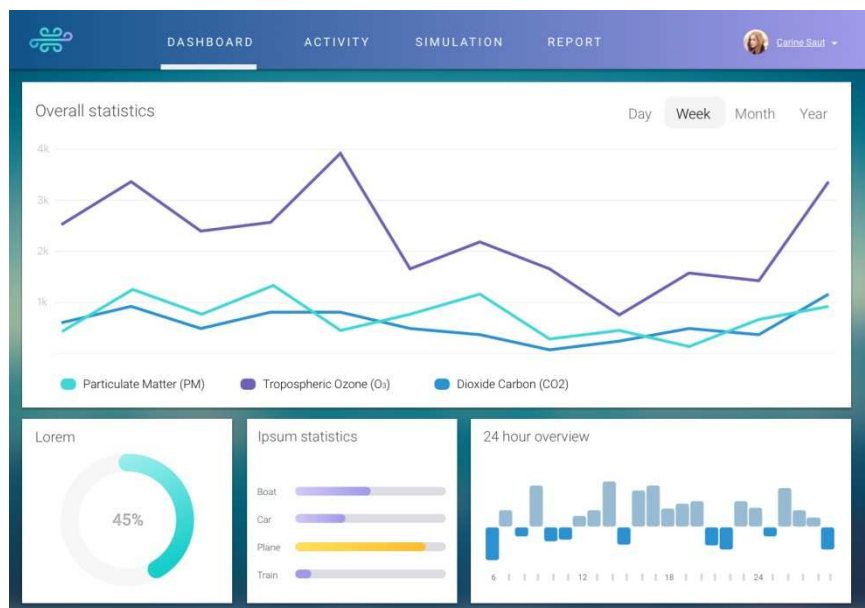


Figure 5 first approach of ETCETERA from R&D Scientific Community

This information is mandatory to develop relevant services in key socio-economic sectors. A framework like ETCETERA becomes mandatory to facilitate a fast access to all needed data from simulations and observations and perform operations over an increasing number of years, which is very difficult and slow nowadays (near-real-time data retrieval and processing). It also slows down significantly the scientific research. Figure 5 introduces a first approach of ETCETERA framework key screen for the R&D Scientific Community. This dash board represents a typical output that can be achieved by ETCETERA with time series and statistics, for instance.

### Uses Case 2 - Business Sector

This use case will mainly be driven by the private partner Green City Solutions which aim is to mitigate climate change and fight air pollution through a highly intelligent and profitable climate infrastructure for liveable cities of tomorrow. For this purpose, Green City Solutions created the CityTree, a 4m high, 3m wide and 60cm deep, freestanding unit, which contains specific moss cultures that eats particulate matter, nitrogen dioxide and ozone – together offsetting 240T CO<sub>2</sub>e/year (CO<sub>2</sub> equivalent).

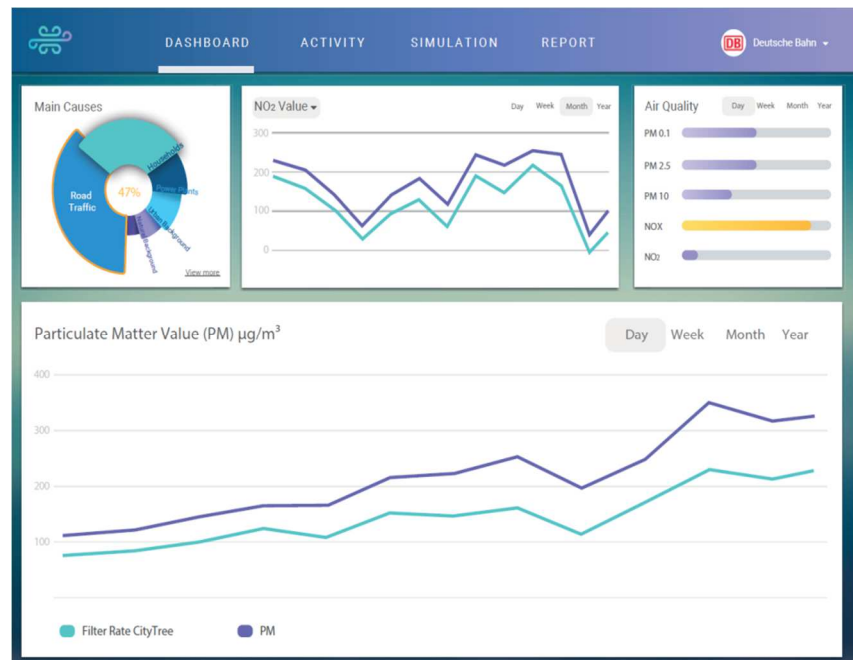


Figure 6 First approach of ETCETERA for Business Sector

Every CityTree has a capacity to clean the air in proximity of up to 50m. To further increase the efficiency of the units, Green City Solutions conducts Computational Fluid Dynamics (CFD) analysis based on data of pollution levels in the street environment of a 3D model. The particle flow is observed and analyzed by using numerical analysis in a simulation and compares them to the actual data gathered by the CityTrees. Integrating ETCETERA's outcomes in their actual product enables the development of added value services or new services which will permit them to grow and expand their business. It is worth noting that Green City Solutions will define a new business model for the industrial and public sector based on a pilot project in the city of Modena Italy with the Proambiente consortium, the EIT and the city of Modena. The Proambiente consortium is a spin-off of the Italian National Research Council and a private non-profit SME who focuses on prototyping and implementing solutions for environmental surveillance and remediation.

The benefits from the ETCETERA framework for those business sectors will create new market opportunities with new players for instance in insurance, tourism, real estate and health sectors. Potential business models could include the combination of Internet of Things (IoT) with air quality to measure the impact of using that data and displaying the composition of the air quality with augmented reality to identify and assess the impact on the above mentioned sectors in an immersive way. A pilot project in the city of Paris which started in 2016 in cooperation with Cisco and the city of Paris, has shown that public private partnerships heavily rely on innovative business models and use cases, bridging the need of public entities for digital information with opportunities for disruptive solutions, which ETCETERA can be considered as to showcase their potential.

Figure 6 introduces a first approach of the ETCETERA framework's key screen for the business sector. As a potential output, this dash board showcases the different air quality indicators (nitrogen dioxide, ozone) and the efficiency of the filter rate CityTree of the GCS versus the Particulate matter value.



### Uses Case 3 - Public Body

This use case 3 will have on board policymakers from the Advisory Board willing to provide requirements for their needs to obtain actionable insights and using ETCETERA service framework as a service for supporting and optimizing their decision. Currently, INERIS<sup>30</sup>, which is the French National competence centre for Industrial Safety and Environmental Protection, has developed expertise in the areas of chronic and hazardous risks. INERIS has expressed its interest and might be consulted for this use case 3. Other public body to be potentially interested in using ETCETERA framework are the local authorities and air quality associations such as AIRPARIF<sup>31</sup> which is an organisation responsible for monitoring the air quality in the Paris agglomeration.

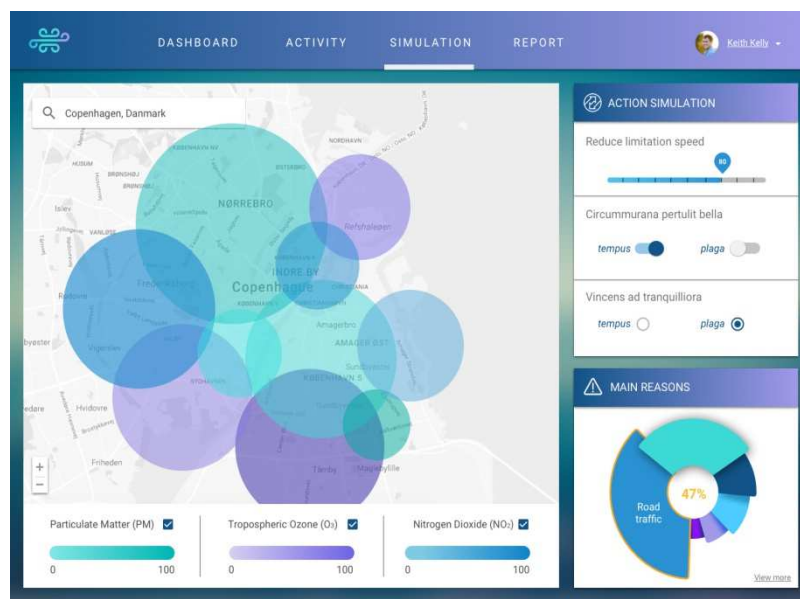


Figure 7 First approach of ETCETERA for public body

ETCETERA will explore ways of exploiting data coming from local networks of air quality sensors. For instance, those deployed by the AiREAS initiative of the city of Eindhoven<sup>32</sup>, local meteorological observations, traffic intensity and velocity will be used.

The benefits from the ETCETERA framework for public authorities are:

1. to inform the population and provide a rapid response
2. to raise awareness of potential epidemics or diseases
3. recommendations to actual European/national/local regulations.

Figure 7 shows a first approach of the ETCETERA framework's key screen for the Public stakeholders. As a potential outcome, the dashboard indicates per district in a European city, the level of the identified air quality indicators, their related main sources and the impact of the actionable insights on their value.

### Gender balanced R&D

ETCETERA will contribute to the EU recommendation to promote the basic principle of equality between women and men. At the start of the project, the number of female experts and researchers represents around 35% of the total involved staff. This leaves some room for improvement. For this reason, ETCETERA project aims to progressively include a larger number of women when recruiting staff at all levels. This means to encourage the recruitment of women at equal scientific or technical merit, especially in the field of Big Data technologies and climate, air quality data management.

<sup>30</sup> <http://www.ineris.fr/>

<sup>31</sup> <http://www.airparif.asso.fr/>

<sup>32</sup> <http://www.aireas.com/welcome-to-aireas/>

All job announcements will encourage women to apply by including a statement that demonstrates an “equal opportunities policy”. This low proportion, however, fits with the lack of women in fields like computational science, climate modelling and impacts. It is noteworthy that three women are in charge of the whole management of the project: Dr. Florence Daverdon and Dr. Alicia Sanchez with support of Dr. Carine Saüt (under WP1) and Ms. Malgorzata Olesiewicz will be in charge of the innovation management in WP5. In addition, 3 more female staff members are already nominated as key staff in the corresponding section.

Gender issues will be considered promoting, in every work package, the following actions:

1. Help the participation of the staff by developing e-conference tools to limit travel, which is more difficult for staff with young children, encouraging a family-friendly organisation of the work; organize child care at meetings, in particular the general assemblies, and conferences as a standard.
2. Create a good working environment by encouraging working-time flexibility.
3. Raise awareness in the Consortium due to the involvement of women in the leadership team, through workshops and training.
4. Communicate within the Consortium and the stakeholder communities the current EU gender legislation.
5. Advertise on gender equality on the project website: links to relevant European web pages, highlighting H2020 initiatives.

## **1.4 Ambition**

With full Copernicus Sentinels satellites measurements as well as Copernicus service data and information rising at the horizon, an increasing amount of EO data is being made available for users with free, full and open access. However, even for expert’s users such as the scientific community, **the access to tens of petabytes of data remains complex and the volume of data is going to increase faster and faster** as still newer generation sensors become operational. There is a crucial need to improve the effectiveness of the access for traditional and non-traditional users’ communities from public stakeholders to private companies. In the current context of Copernicus, data has become a raw material to exploit and to derive new business opportunities.

**Currently, the conventional methods for the processing and exploitation of such volumes are too expensive and too slow for many business services and for scientific exploitation.** The size and the variety of data required has become a part of a crucial problem and it is necessary to innovate creating new solutions to be able to analyse this “Big Data” according to an efficient time and a reasonable operational cost. Also, the diversity and multiplicity of the databases make it difficult to combine data from different sources and compute new variables or statistics accordingly.

In addition, it has been reported that some of the **Sentinel and other Copernicus data were highly underused because of the heterogeneity of the access to the data, their complexity of use or downloading or the level of advertisement** they get. By offering an homogeneous and user-friendly access to several data platforms and some diagnostics tools directly designed thinking of specific types of users, ETCETERA has the ambition to maximise the use of the DIAS data, being the Sentinel satellite data or any other EO data. The use of these data can be maximised targeting a new type of users, who saw this kind of access too difficult from a technical point of view. The number of users from the scientific community that were already using the data can also be increased offering through ETCETERA some tools to validate data across different datasets.

Through the ETCETERA proposal, several main lines of innovation potential are identified regarding the Big Data paradigm:

1. **Providing advanced Big Data functionalities** addressing the needs of final users via the development of a new set of tools. It will be at the nexus of capability in scientific understanding and interpretation for users, challenging IT and robust distributed computational platforms (i.e. ESGF nodes), providing at the same time new methods for processing and analysing huge observational datasets at the Peta- or Exascale.
2. **Developing complementarity features** with Copernicus data & information access services (Core services and Sentinel data) and with other EO data.
3. **Increasing the weather, climate and air quality information** distribution offered by the service framework to traditional and non-traditional communities. ETCETERA will provide a broader range of metrics and co-design new tools that are suitable to drive industry risk models and quantitatively answer these risk-related questions. In the mid- and long- term, these newly developed and improved tools will allow key European business sectors to become more competitive at global scale, providing opportunities for growth and improved resilience.
4. **Facilitating the access and use** of (very) high resolution data and memory intensive diagnostics calculations. All those elements together merge into the ambition to develop new business model and expand new market opportunities to be offered to society.

The development of the technological solutions will be done taking advantage of already existing solutions that have been adopted in several European projects. **The ambition is to use the expertise of the developers in similar contexts and develop innovation based on the tools.** For example, with regard to the Big Data analytics context, the Ophidia framework has been exploited in the pre-Copernicus project EU FP7 CLIP-C for the computation of climate indicators from large scale scientific datasets through OGC-based processing interfaces and in the H2020 INDIGO-DataCloud to run climate model intercomparison analytics experiments in distributed environments with special regard to the CMIP5 context (2PB of data published through ESGF).

EO data on atmospheric composition has been introduced in air quality analysis models only recently; these models are historically based on meteorological data, air composition statistical data and measurements. **This statement shows the innovation potential offered by ETCETERA which will provide Value Added Services through Data Analytics.** In particular, it will establish a collaborative environment where the synergy between software engineers and scientists, will offer devoted implementations of novel data-analysis algorithms and techniques:

1. full exploitation of the environment's capabilities,
2. better software quality,
3. immediate re-usability for other users.

In addition, a strict adherence to revision tracking/control (through git or svn), as well as accurate documentation practice for both data and algorithms, enforced by the data management methodology and carried out by professional information technologists, will ensure the required scientific reproducibility.



## 2. Impact

### 2.1 Expected impacts

Work Program	ETCETERA Contribution	KPIs
<i>Enable value adding services on generic data and information storage and processing facilities which can allow public and commercial users effective production environment to interact with and serve their user base without deploying their own storage and processing facilities.</i>	ETCETERA will create a complete infrastructure of user-friendly accesses to several existing databases of Earth Observation (DIAS, including CAMS, C3S, and CMEMS for the Copernicus related products). The aim is to provide at the same time the extra storage needed to output the diagnostics.	Survey at the end of the project disseminated to the institutes who expressed their interest in the proposal and to external users, to quantify their degree of satisfaction and to evaluate how user-friendly the platform is.  <b>Related work packages: WP2, WP5</b>
<i>Make access to the Copernicus data and information easy and user friendly through scalable dissemination and exploitation software based on international standards.</i>	ETCETERA's main objective is to define complementary and scalable solutions for data access and exploitation. The choice and design of the solution will be made after a review of the state of the art. A key evaluation criteria will be the scalability of the solution and compliance with international standards and software recommendations (INSPIRE, RDA, WMO, OGC Big Data,...). The definition of 3 consistent use cases addressing 3 different communities will help the project to ensure that the platform is in phase with the capacities of the different kinds of users.	Evaluation of the total volume of data accessible through the ETCETERA platform, with the demonstration that the tool can either access to small subsets of the data or to big datasets.  <b>Related work packages: WP2, WP3, WP4</b>
<i>Foster the establishment of interoperable access facilities to all EU Member States.</i>	ETCETERA will support the interoperability of the access through the strict application of guidelines standards issued by OGC Big Data working group <sup>33</sup> .	Even if the data accessed through the ETCETERA platform will be open to anyone, some registration could be useful to keep track of the users. With these statistics, the project will be able to define a KPI monitoring the geographical

<sup>33</sup> <http://www.opengeospatial.org/projects/groups/bigdatadwg>

		<p>repartition of the users, quantifying the access to all the EU member states.</p> <p><b>Related work package: WP2</b></p>
<i>Link with other Big Data initiatives</i>	ETCETERA will federate with FP7 and H2020 initiatives, among others the EarthServer H2020 initiative which is currently already offering several 100 TB of EO and weather data and is working towards offering PB size datacubes. It is worth noting that ETCETERA will be linked with data management projects already ongoing such as the H2020 INDIGO-DataCloud and ESiWACE.	<p>Number of Big Data projects/initiatives (such as RDA IG/WG, workshops, etc) in which the Big Data part of ETCETERA has been evoked.</p> <p><b>Related work package: WP5</b></p>
<i>Provide user community tools including best-practices.</i>	ETCETERA will provide such tools to support community through the re-use of the experience of consortium members in similar Big Data initiatives via the use of the open OGC “Big Earth Data” standard, WCS.	<p>Number and statistics on the use of the user community tools.</p> <p><b>Related work packages: WP2, WP3, WP4</b></p>
<i>Ensure resilience of the overall dissemination and exploitation system</i>	In order to build resilient societies, policy-makers and the public must have access to the right data and information to inform good decision to perform appropriate actions (how to protect economies against future climate impacts...). Under the direct supervision of managers, the incumbent will report to supervisor/Open Data resilience lead/Project Manager.	<p>Documentation of each phase in the collation data and collection of new data, including major documents and data sets. These resources might be available online in open format.</p> <p>Building of relationships between stakeholders through trainings, webinars, and communities of practice.</p> <p>Documentation of lessons learned: Produce reports to documents lessons learned throughout the project to inform future activities, actions.</p>

		<b>Related work package: WP5</b>
<i>Optimise the use of Copernicus data by non-traditional user communities to meet societal challenges</i>	In order to target non-expert users or external communities, the project will ease the use of Copernicus and of the ESGF data through the platform developed in ETCETERA and will provide support to non traditional user communities.	<p>The dissemination part and the inclusion of SMEs in the consortium will be good indicators of the fact that we have met some new users.</p> <p><b>Related work packages: WP3, WP5</b></p>
<i>Create new market opportunities</i>	Many of the elements required for new market opportunities have been laid out in the proposal and all the information and outcomes from ETCETERA will be open source, offering the possibility to the private sector and in particular Green City Solutions (member of the consortium and part of the Advisory Board) to employ such information. For a better market preparation, ETCETERA will enable emerging business opportunities: a use case will be co-defined with and co-designed for a start-up. In addition, the elaboration of an exploitation plan as a task to be performed will showcase that ETCETERA has a clear orientation towards the private sector to push new market opportunities in the sector of green cities. Prediction and monitoring of air pollution levels could be another potential service which would allow private partners to grow its business and expand.	<p>Business model to service based delivered at M30, including expected numbers from the business plan, and expected turnover.</p> <p><b>Related work packages: WP4, WP5</b></p>
<i>Address issues related to air quality</i>	As the use case has been selected and will be defined for a SME specialized in the air quality, ETCETERA outcomes will directly benefit society as they aim at improving cities' air quality. Governmental customers have already shown that evaluation of the measures to prevent air pollution is in high demand.	<p>Number of datasets/variables (or volume) related to air quality (mainly CAMS) made available in a user-friendly way through the ETCETERA platform.</p> <p><b>Related work package: WP4, WP5</b></p>

<i>Important benefits for society related to air quality</i>	Society resilience towards the impacts of the air quality sector is expected given ETCETERA aims at better filling the gap between advanced atmospheric services and related information. The aim is that both business and society can take advantage of synergies between the different climate and air quality research communities and get information from private sector value added services will be key.	Projection of number of users of the value added service proposed by Green City Solutions partner.  <b>Related work package: WP5</b>
<i>Stakeholder engagement</i>	ETCETERA will target the different potential users' communities with ad hoc communication and dissemination activities, also organising workshops and hands-on training activities centered on the advantages in the use of the developed tool. During the project, users' stories about their approach when dealing with EO and Big Data will be collected in order to propose suggestions and to better explain to the community the diversity of the challenges taken on by ETCETERA project. Several stakeholders have already signed official letters expressing their interest in the outcomes of the ETCETERA project, and engaging themselves to use the final products resulting of the work done. Additionally, ETCETERA benefits from the fact that some stakeholders like Green City Solutions, are at the same time ETCETERA partners.	Organisation of two dissemination workshops scheduled at M13 and M30. <b>Related work packages: WP3 WP5.</b>

### ***Barriers to reach impact***

The vision outlined in section 1.1 of facilitating the digital transformation of an entire EO ecosystem using a Big Data platform approach is clearly an ambitious one. In fact, one of the main boundary conditions that ETCETERA will deal with, is the fact that up to now, no implementation of the DIAS system exists. As a consequence, ETCETERA will be built considering existing Copernicus and non-Copernicus data providers, what allows extensibility when needed. Indeed, the possibility to include additional data sources is one of the added values of ETCETERA.

ETCETERA consortium believes sincerely in this vision but is conscious that there will be many barriers and obstacles along the way, both those that are visible today and that are outlined below, but also those that we will discover along the journey towards turning this vision into a reality. This is the reason why the project management will follow the “Lean Startup” approach and methodology for innovation detailed previously. This approach will allow the consortium to remain lucid with regards to the fundamental hypotheses, the solutions proposed and their effectiveness. It also allows the consortium management to adapt efficiently in the case of immovable barriers or obstacles in order to continue moving forward towards the defined vision and maximize impact of the related efforts relative to available budget.

Among the barriers and obstacles anticipated today, we can note the following:

- The ease of data access to EO data. This will be addressed of course as part of the DIAS project, but the solution, interfaces and interaction models remain unknown as of today. In order to cope with this situation, **ETCETERA as a framework will establish and consume existing data source interfaces like for example, those provided via the CAMS Services sources.**
- The impact of the ETCETERA framework and final EO platform will depend greatly on the quality and integrity of data that it stores, manipulates and otherwise makes available. The framework must explicitly treat this both in its delivery of services but also the mechanisms it provides to discover data sets that meet user expectations/requirements. This is a diverse subject and careful prioritisation is required to ensure maximum impact of the platform. **We propose a very well defined data management plan which ensures the establishment of priorities and policies for data quality.**
- Both openness (open source) and confidentiality (privacy as well as protection of intellectual property, whether algorithms, code or documentation) are seen as essential for the adoption of the ETCETERA framework and the ability to be relevant in the digital ecosystem. This sometimes leads to a delicate balancing act that is well understood by the giants of the digital world (Google, Amazon, Netflix...). ETCETERA must also find the right balance to ensure sustainability and allow for all target users of the EO platform to find their own balance. A good example of this is the consortium member Rasdaman, whose product is included in ETCETERA. Rasdaman has both a community open source (OSS) version and a closed source version that is essential to the sustainability of their business model. **ETCETERA modularity will allow to easily combine different components, both open and closed source.**
- Ability to drive rapid adoption is also key. The dissemination activities will obviously be essential to this aspect. However, we have seen many times in the past that frameworks, services or products (henceforth referred to simply as a product) with very rich sets of features have failed to get adoption or long term traction with their target audience despite very hefty dissemination efforts. There are various reasons for this, with the most common being the difficulty of getting started with the product in question. Upfront costs, on-boarding complexity, steep learning curves and many other aspects can slow adoption and ultimately lead to failure. **A continuous monitoring procedure based for example on user-surveys will make it possible to adapt our dissemination approach to reach user-demands accordingly.**

## **2.2 Measures to maximise impact**

The ETCETERA project foresees a continuous monitoring and evaluation of dissemination and exploitation activities, in order to enhance impact of these activities and to make ETCETERA foreground scalable and

replicable. The measure of the impact will be done following three effective early defined **dissemination, exploitation and data management plans**.

1. **The dissemination plan** – pertains to the activities associated with raising awareness regarding the project and its objectives. This phase lasts for the duration of the project, and usually ends with completion of the project. This will also provide a platform for collecting input from the key stakeholders and provide the basis for the exploitation phase described below.

2. **The exploitation plan** – refers to the exploitation of the project outcomes and contains the measures to be implemented during and after the end of the project. Ideally this phase continues long after the end of the project as it promotes the service developed during the execution phase of the project.

3. **The data management plan** – contains all the information related to the data management of the project, from storage hardware to (open) data policy and adoption of international data standards. The Data Management team described in section 3.2 will produce at M6 the initial version of the Data Management Plan under WP3. The plan will be updated annually, an intermediate version will be delivered at M15 and final version at M30.

The project dissemination and communication activities will target a broad audience, covering the value chain from scientific to industrial interests. The results of the project will be promoted by applying direct actions to specific user communities, from researchers to private sector, but also to the interested stakeholders to citizen. Further, the exploitation of the project will be supported by an effective communication and an innovative business model designed by Green City Solutions as described in section 3.1. ETCETERA will also leverage an Advisory Board, a group of professional experts which have recognized knowledge in this field with also interest and/or expertise also on the business aspect and regulation aspects. The User engagement Advisory Board encompasses scientific community, business sector and policy makers, as already shown in section 1.3.

The expected outcomes from ETCETERA should address issues related to climate and air quality designed with and for the final targets. This approach takes on-board our target users from the beginning for the users' needs definition phase, starting from users' knowledge. Passing from information to service framework, ETCETERA will co-design the framework with the users, crossing their knowledge with multiple data sources, competences and actors. The international level will be a priority due to the global scale of Copernicus exploitation within a Big Data context, empowering services and applications demonstrated on a specific use case on air quality focusing in Europe. However, due to the scalable computing environment of the proposed solution, ETCETERA will cover the opportunity to address others sectors deploying other types of applications and at a global scale, meaning out of Europe.

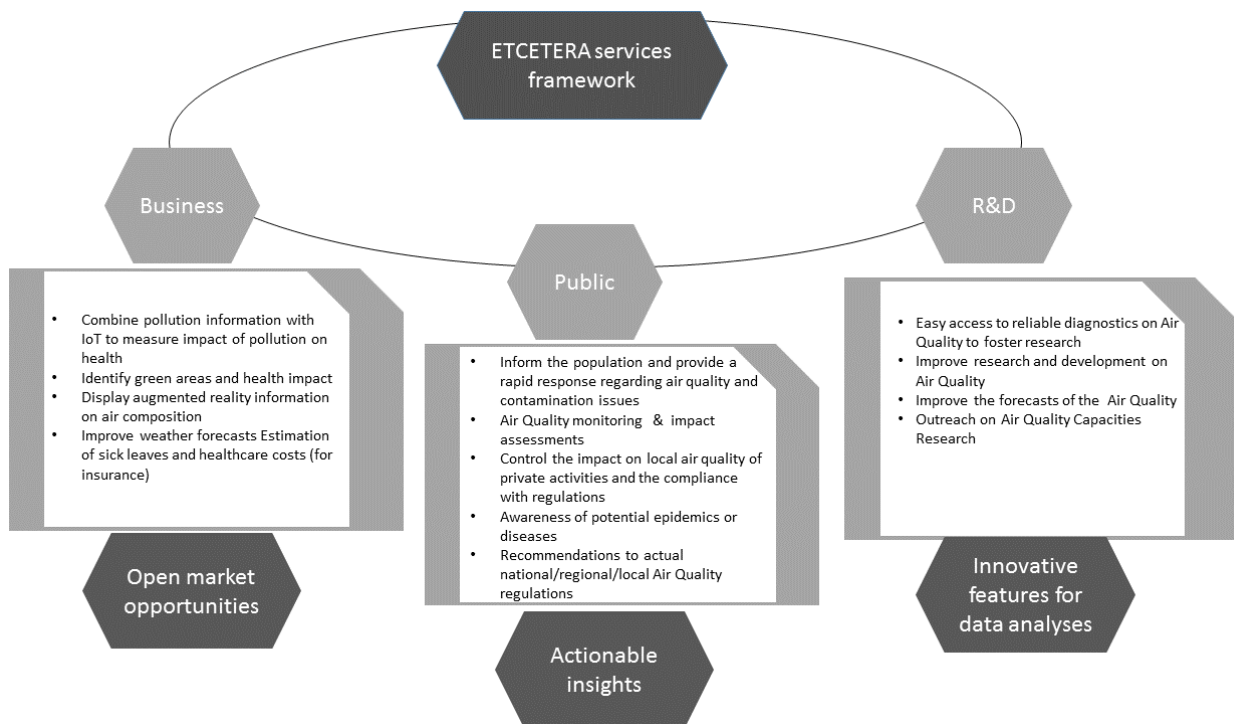


Figure 8 overview of expected impact and ETCETERA opportunities

As shown in Figure 8 above, the expected impacts from the ETCETERA framework might be wide for the different targets:

1. **The Scientific community:** supporting and fostering the excellence of European climate, weather and air quality Research and Development: The European level will be a primary target for ETCETERA due to its high impact through a strong link with Copernicus. ETCETERA will be designed to be compatible with the DIAS which will be running in 2018 on its first operational version. This context will enable researchers from related communities related to access to a set of tools, training, and at the end, improved understanding and knowledge on the domain.
2. **The Business sector:** Bringing directly air quality information to citizens in order to reduce risks on health related to air quality. In order to open new market opportunities through new products or services and eventually innovative business models, the private sector will be addressed as well at different levels:
  - a. the start-up Green City Solutions, member of the consortium, which will in charge of the definition of the private user case
  - b. Rasdaman GmbH, member the consortium, will provide enhanced rasdaman open source license versions to existing and new users (data centers, agencies, industry, etc.)
  - c. through the commitment of the private sector stakeholders, mainly European SMEs, which have expressed their interest and might potentially integrate the ETCETERA's outcomes in their actual service enabling the development of added value services / new services.
3. **The Public Body:** Supporting implementation of improved actionable insights to air quality policies and regulation. The set of international stakeholders that have shown their interests in using the ETCETERA's outcomes and benefit from actionable insights to support their decisions and will contribute to achieve this objective.

### c) Dissemination activities

The primary goal of the ETCETERA dissemination is to federate the different targets enabling to bridge the gap between scientific community, European public stakeholders, national organizations, European business and wider audience. For this, a clear and efficient communication to international research community on the development at technological and scientific levels will be ensured. This will be achieved under the performance of the WP5 *Dissemination & Impact*, defining the dissemination strategy is defined early in the project lifetime. This strategy will be detailed in the *Dissemination Plan (D5.1)* and will coordinate the outreach activities. At national and international level, the ETCETERA dissemination will use the following channels and components:

**Attendance to specific workshops and conferences:** Attendance and participation to external events – including fairs, conferences, exhibitions, workshops and other sector related events - will be ranking top among channels for the dissemination of the project progress and main results. Sectoral conferences will be considered for industrial dissemination: partners usually participate along the year to important international sectoral conferences<sup>34</sup> like the INSPIRE Conferences, International Conference on Web Engineering (ICWE), Data.Space events, the International ACM SIGMOD Workshop on Managing and Mining Enriched Geo-Spatial Data (GeoRich), the International Technical Meeting on Air Pollution Modelling and its Applications (ITM).

Attendance to conferences aims at optimizing the expected dissemination and mainstreaming the results of the project towards users' communities and exploring possibilities for the technology transfer. The partners will take action to maximally disseminate project-related work through scientific conferences, workshops, special events and demonstrations. The participation in suitable industrial events will take place towards the end of the project, in order to communicate the project results to the appropriate communities (e.g. Big Data community, Inside Big Data, Big Data London or BYTE for cross-disciplinary community) and explore the potential for joint ventures with a view to extending the ETCETERA services to other sectors. A calendar of events will be set up in the Dissemination Activities Plan (D5.1 due at M3) and this deliverable will be updated throughout the project lifetime.

### d) Exploitation of the results

In the already mentioned Copernicus Market report, the air quality Data Management is presented among the eight promising downstream domains and user segments ensemble. The actual panel of intermediate users encompass mostly research centres, environmental and meteorological agencies and public institution. Now, the challenge is on reaching non-traditional users such as the private sector and end users and demonstrate them the full benefits of using complementary data and information to improve their business or to support their lifestyles. It is worth noting that demand for commercial EO services still address niche markets with a lot of unexplored potential. There is an increasing number of value added services companies entering the market with air quality forecasting products. In their study in 2015, Eurisy<sup>35</sup> has shown that public authorities are motivated for using satellite data, with the large majority of usage based upon the need to respond to socio-economic or environmental challenges. Monitoring air quality on specific territory, inform the population, support decision making are the possible ETCETERA applications in the view of improved public services

---

<sup>34</sup> See more of international sectoral events in Annex

<sup>35</sup> [http://www.eurisy.org/data\\_files/publications-documents/28/publications\\_document-28.pdf?t=1467808834](http://www.eurisy.org/data_files/publications-documents/28/publications_document-28.pdf?t=1467808834)



and support policy implementation. Cost effectiveness of satellite-based solutions, as compared to other technologies, was also a reported driver for adoption (Eurisy, 2015). With the next Sentinels, the forecasts on the accuracy of the air quality will be improved.

The combined use of EO data versus models evaluation, made available on ETCETERA service platform aims at supporting innovative business development. On top of the air quality domain specifically addressed in this proposal, **Insurance, Renewable energy, Health and Tourism sectors** have been identified as sectors most space-derived data applications. They are expected to be developed as commercial services from ETCETERA platform. Indeed, insurance companies have the willingness to access to air quality forecasts.

The energy sector is the most important source of air pollution. It accounts for 85% of particulate matter and almost all of the sulfur oxides and nitrogen oxides to which 6.5 million deaths are attributed each year<sup>36</sup>. Clean Air Scenario<sup>37</sup> from IEA estimates an investment of 2 billions in advanced pollution control technologies by 2030.

The report proposes a strategy to reconcile the energy sector requirements with a better air quality, opening bridges between the ETCETERA platform and value added service companies. The Health sector and Tourism sector demonstrate the interest on being informed through air quality monitoring tools, to assess impact on environment and public health, to orientate the dissemination of healthy areas, to inform and orientate tourists on pollutants flux, UV or pollen when visiting megalopolis for instance (Shanghai, Beijing, Delhi, Sao Paulo,...).

**New players from the above mentioned sectors may want to use ETCETERA services to develop systems to trace air quality data over the globe.** It is worth noting that improvement of air quality tracking and forecasting will benefit directly to scientific research community. **ETCETERA might be a key framework for R&D on air quality and on climate and weather, especially for air quality in developing countries.** Figure 9 below presents the features of the exploitation plan for the different targets of users, an example of a panel of market opportunities, the actionable insights and impact on R&D.

The Copernicus Market report considers that there is a positive evolution of Copernicus economic impacts depending on the relative size of the domain in the EO downstream market and the penetration rate of Copernicus. With regard to air quality, the expected average annual growth rate of Copernicus benefits up to 2020 is about + 9%. As **ETCETERA can support the emergence of new initiatives in niche markets (tourism, insurance, health) with new players entering, the business sector is expected to present the biggest impact.** While the relative importance and use of Copernicus, the air quality monitoring market is expected to expand at a rate of 8.5% per year according to the *marketandmarkets*<sup>38</sup> report in 2015. Indeed, a rapid scan of the World wide and European air quality companies (among others Numtech, Plume Labs, Iséo, TRC Solutions, Nalco Solutions, GEL engineering, CTP, CSIRO) allow to predict they would take advantage of these new opportunities. Most of the companies have anticipated an increase of their sales based on Copernicus products in the coming years in and outside Europe.

---

<sup>36</sup> According to International Energy Agency, 2016b

<sup>37</sup> <https://www.iea.org/publications/freepublications/publication/WorldEnergyOutlookSpecialReport2016EnergyandAirPollution.pdf>

<sup>38</sup> <http://www.marketsandmarkets.com/top-market-reports.asp>

**An innovative business model:** ETCETERA should enable Green City Solutions, on its initial market, to expand their turnover in a short term on European markets and abroad, where the access to air quality related diagnostics and analytics is missing and access new types of customers (healthcare costs, sick leaves in insurances, for instance). In parallel, as the local policy makers may have legal obligation to monitor and regulate air quality in the future, thus the penetration market by local governments and municipalities is expected to rise. In that context, the use of ETCETERA framework will support such a novel contribution where diagnostics and analytics on pollen, UV and pollutants (nitrogen dioxide, particulate matter, ozone) could be used in applications for health regarding potential epidemics, contamination episode for health. ETCETERA intends to improve precision of air quality composition by 40% and anticipate significant improvement to more local pollutants. Moreover, the exploitation plan will be supported by Consortium partners:

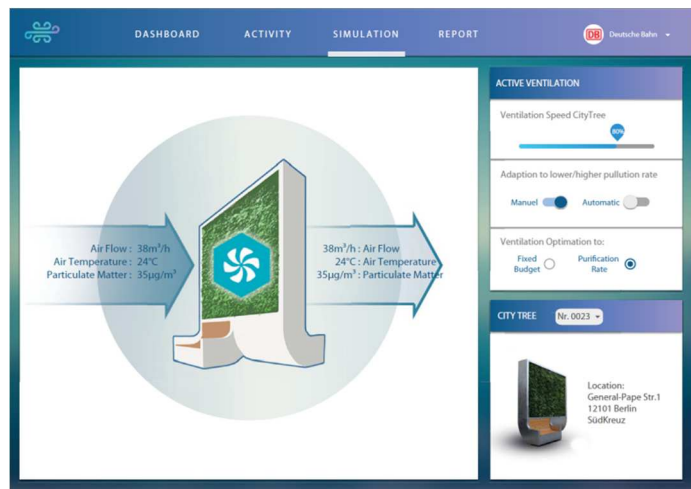


Figure 9 first approach of the ETCETERA outcome tailored to GCS product

- KNMI DataLab will evaluate the framework and collection of services, performing data-driven innovations through the execution of collaborative experiments with public and private sector partners. The DataLab experiments are proof of concepts (prototypes) that combine meteorological data with data from other domains and/or emerging data (e.g. crowd sourced data, IoT). Data wrangling, Big Data analytics, machine learning and the associated reproducibility challenges are playing an important role in the DataLab ongoing activities. These aspects will be driving the evaluation and dissemination activities conducted by the DataLab in WP5, targeting pre-existing partnerships and experiments in the private industry sector of transports.
- Capgemini, as a leading EU provider of technology solution and system integration services has a catalysing role in proposing ETCETERA solutions to its client database in historically strategic sectors for instance finance services (banking and insurances), distribution and transportation, energy and public sectors among others. Furthermore, Capgemini has been running the European Data Portal, on behalf of DG CONNECT, thereby demonstrating its capacity to work cross-border, cross-organisation in the context of platform development and user engagement.
- All scientific partners will benefit from the innovative business model being either by the gain of visibility for future projects or increasing their scientific track record.

Figure 9 presents the first approach of the ETCETERA outcome tailored to GCS current product, showing clearly the added value of having integrated ETCETERA framework.

### e) Data Management

The Data Management Plan (DMP), to be delivered at M6 under work package WP3 Analytics and Diagnostics, will describe the following points, following in particular the recommendations from the “World Climate Research Programme data policy” document

#### 1. Open Data Policy:

- a. Following the EC recommendations and global policy on data sharing, the data generated within ETCETERA (derived variables, diagnostics,...) will be freely accessible, following the same policies as the input data from Copernicus and other sources.

## 2. Data generated and collected:

- a. Full description of the datasets that will be used in the projects (stored in the DIAS and on the other data repositories).
- b. Cache strategy: how much space will be dedicated to the cache of the data retrieved from the data repositories and strategy to define what data to store in the cache.
- c. Once the diagnostics have been physically defined (WP3, T3.2), they will be described in terms of data requirements (volume, format, etc...) in this part of the DMP.

## 3. Data Standards

- a. One guideline during the writing of the proposal and the upcoming work in the project is the compliance to the data standards established in the community. All the WP will be coordinated to follow the same standards (open OGC standards WCS, WCPS, and WMS (including data ingest through WCS-T). As the proposal aims at using data from existing platforms (Copernicus, ESGF,...), the data standards implemented in the input data (CF,...) will be preserved.
- b. Focus will also be put on metadata and provenance to ensure reproducibility and full tracking of the data, always following international standards.
- c. Even if gathered in the DMP, managed by WP3, the standardizations activities will encompass all the work packages and will make sure that the data standards are coherent in all the project and disseminated to the “outside world”.

## 4. Data Curation and preservation

- a. The new derived and experimental data-products will be associated to their lineage and attribution documentation (Provenance). Provenance will be automatically collected during the computation of the new products, with special attention in guaranteeing consistency with the infrastructure’s policies concerning unique identifiers schemas and vocabularies. Further user-driven contextualisation of the provenance will be allowed and explicitly annotated to be differentiated from the established vocabularies. This approach will foster the productive exploitation of the provenance within the experimental phases, supporting data and methods reuse and refinement, towards stable and published results. Provenance will be stored in a dedicated repository via dedicated services. These will also allow its access, discoverability and exploitation. The services will also allow its inclusion within the final data-products, if required and supported by the target data-container (eg. NetCDF), in a machine readable format. For interoperability reasons the exported provenance will be compliant with the W3C-PROV recommendation.

### **f) Knowledge management and protection**

The main short term aim of the knowledge created in the duration of the project is to benefit the large panel related to air quality communities from scientific community, public stakeholders and private sector and citizens. For this reason, fast and open access publication is a key priority of the project, and all of the project publications are expected to follow “Gold Model” open access, i.e. the publications are made with complete immediate open access. Only for very strong reasons can the publications be made with the “Green Model” open access, i.e. the publication is made available in a repository located in ETCETERA website, possibly

after a reasonable embargo period. Requests for publications using the “Green Model” must be made, with detailed explanations for the reason, to ETCETERA project board. Declined requests can only be amended by the General Assembly. As in general requirements of H2020 projects, all ETCETERA publications will be open access. The tasks and related developments performed under the work packages will optimally adopt industrial and open standards, and the software will be managed and released in future as open source. The project can be released as free software on GitHub so that all developers can write their own extensions. For further information on IPR, please consult section f).

### **g) Communication activities**

Maximising the impact of the results is mandatory, as it aims at raising awareness on the outcomes and promoting the ETCETERA project to larger audiences. ETCETERA will rely on an early defined approach on communication activities including a comprehensive communication plan.

The communication measures for promoting the project and their results during the execution of the project will focus on the creation and on maintaining the wider possible community of potentially interested stakeholders. The project addresses a topic of great environmental and societal relevance and the members of consortium will make appropriate efforts to disseminate its results to both the scientific community and towards society at a wider range including citizens, private sector and policy makers.

The planned action for dissemination and in order to raise awareness will allow participants to make use of ETCETERA outcomes through concrete actions in order to manage knowledge and intellectual property and exploit the results in the Big Data community and air quality community. The Advisory Board of experts will be invited to the Annual Meetings (AM at months 12 and 30) to give their feedback on the ETCETERA framework and development. Regarding the ETCETERA identity as a potential funded activity, adequate references to EC funding shall be given in all dissemination and communication materials and channels used. The appropriate and recommended acknowledgement will be used. In addition, a project visual identity, an ETCETERA image will be developed to reflect the project and the key concepts.

This will be achieved through the use of an identifiable ETCETERA logo and graphic design style, which will be used prominently in all dissemination tools (website, e-newsletters, etc.) and printed materials. In the interest of promoting a unified clear image, the logo will be used in all dissemination and communication activities together with the EU emblem given appropriate prominence to the latter when displayed in association with a logo. The proposed communication measures for promoting the project and its outcomes during the activity of the project are listed in the Dissemination & Communication plan (available at M3) and include:

- From M1: Articles and press releases for general and specialised publications, also available on the public website and uploaded to EC websites after approval from contract officer. Opportunities will be sought to publish articles about the project, to issue press releases at kick-off and closing events and in conjunction with major results
- From M6: Brochure and leaflets prepared to illustrate clearly the project and its long-term impact on life quality, research and the environment. The aim is to reach a very large spectrum of users. An electronic version (PDF) of these materials will be available on the public website
- From M12: Newsletters. An annual project newsletter will be emailed to the ETCETERA contact list with a collection of news and achievements from the project. To maximise distribution and to facilitate its ‘readability’, the newsletters will be prepared in e-mail format

In addition to the previously mentioned points, communication will be done through the ETCETERA website and diverse social medias.

**Public website:** The public website will be one of the main promotional tool for publishing, thus diffusing, project results. The public website will be structured as depicted in the figure below

**Social Media:** ETCETERA will make use of social media tools to keep the interest high, a fast-moving newsfeed, extending to various online communities, and similarly to the web platform, having the capacity to reach the widest audience possible and end-users. LinkedIn and Twitter interest groups will be set up for ETCETERA if relevant. This will serve as a discussion forum between

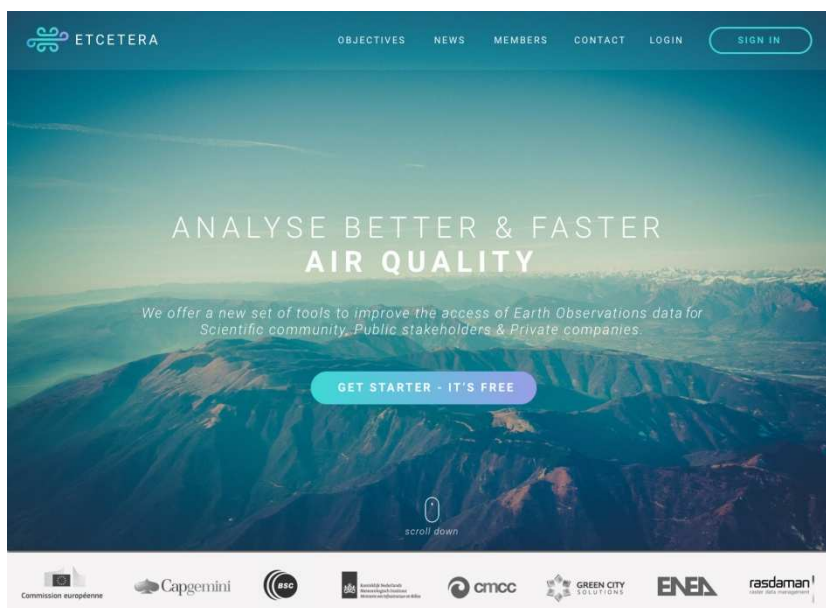


Figure 10: First approach of the ETCETERA Public website

the ETCETERA non-traditional users and public stakeholders. The purpose will be to keep the communications channel open between these two parties throughout the duration of the project, and as such will allow the user community to give the team additional inputs and feedback outside of the user workshops. Twitter feeds will be used to target increasingly mobile stakeholders and interest groups. ENEA, Green City Solutions along with Capgemini will be involved in the management of the social media channels and ensuring that the appropriate team member responds within days to any postings of questions on these fora.

## h) Intellectual property rights (IPR)

The Consortium Agreement (CA) will be set up and agreed between all partners, addressing the internal organization of the consortium in line with the Management structure described within the proposal. Furthermore, the CA will describe rights and responsibilities of all partners with regard to confidentiality, publication and communication of information during the project's lifetime. In addition, the CA will cover aspects as the settlement of internal disputes and Intellectual Property (IP) arrangements. Background – any result generated before the effective entry into force of the CA – shall remain the property of the respective party. Any result generated by a party after said date, during and within the scope of the project (i.e. Result) whether or not it qualifies for Intellectual Property Right (IPR) protection, shall be made available in open access.

A data dissemination protocol will be included in the CA. The ETCETERA data dissemination protocol will aim at promoting access to data and the publication in the scientific literature. This protocol should protect the IPR of the project partners. The protocol will regulate the access to the data by the ETCETERA project, although always promoting the free access to the project data and outcomes by European policy makers, SMEs and the research community. Each party will have the right to exclude specific pre-existing know-how from the other parties, as far as the restrictions are announced before the signature of the funding contract or before the effective joining of a new party. Access rights needed for the project execution according to the agreed work plan will be granted on a non-exclusive basis and shall be made free of any transfer costs. In the event of an invention being the work of a single party of the project and solely the result of these intrinsic skills rather than shared knowledge, this party will be the exclusive owner of the results, subject to granting access rights to the other participants where necessary for their execution of the project. Particular emphasis will be



put on protecting IPR-relevant contributions by the SME partners so as to keep them as their assets. The procedure and the conditions will be fixed in the CA.

#### **i) Contribution to standardization activities**

The standardisation issue is crucial factor to reach a valuable impact through a public funded project. ETCETERA will firstly take into account existing standards in order to guarantee the results respect existing practices, enhance the interoperability and reach a better market application. In addition, promoting the research results to be included into future standards (models, architecture, best practices, additional requirements, or guidelines...) would provide market validation and facilitate the market uptake of the ETCETERA innovations.

Standards allow faster uptake of innovation solutions and enhance the economic value of research and development projects, facilitating the application of the outcomes in practice and thus fostering the innovation uprising from the research results. Standardisations activities related to ETCETERA are considered as valuable tool for the exploitation of the project outcomes, by facilitating future replicability widest use and reducing market acceptance risks. Furthermore, standardizations system constitutes an efficient and fast information and knowledge transfer structure.

Within ETCETERA, all members of the consortium are directly interested in contributing to the standardisation activities in order to foster real applicability of the ETCETERA service platform, while following existing standards or promoting new ones. **They will collaborate in this direction reflecting experiences in OGC and ISO standardizations as well in INSPIRE WCS adoption** where Rasdaman GmbH is actively shaping several such standards as editor. In particular, Rasdaman GmbH makes available for the ETCETERA duration its open-source rasdaman community which “is currently the world leading environment in this domain and the standard working horse for OGC standardisation on these innovative data access inter-faces.” according to Mr. Landgraf, Senior Manager, ESA Ground Segment, 2017. The RASDAMAN product in its rasdaman community edition proposes an open source, coordinated by Jacobs University, fully fledged Array DBMS (which is in active use, eg, with [www.planetserver.eu](http://www.planetserver.eu)); fully supports OGC standards, allows to do data fusion, cross-dimensional queries, etc. In ETCETERA access interfaces will strictly rely on the open OGC standards WCS, WCPS, and WMS (including data ingest through WCS-T). OGC WCS has been adopted by INSPIRE in 2016, and is planned for the process by ISO TC211.

### **3. Implementation**

#### ***3.1 Work plan — Work packages, deliverables***

The present section introduces the work plan of the ETCETERA execution. The overall structure of the work plan is discriminated into five work packages. The timing of the different work packages and their components are presented (Figure 12 Gantt chart) and the ways that the components are interrelated (Figure 13 Pert chart). Then, a detailed work description:

1. The description of each work package (table 3.1a);
2. The list of work packages (table 3.1b);
3. The list of major deliverables (table 3.1c);

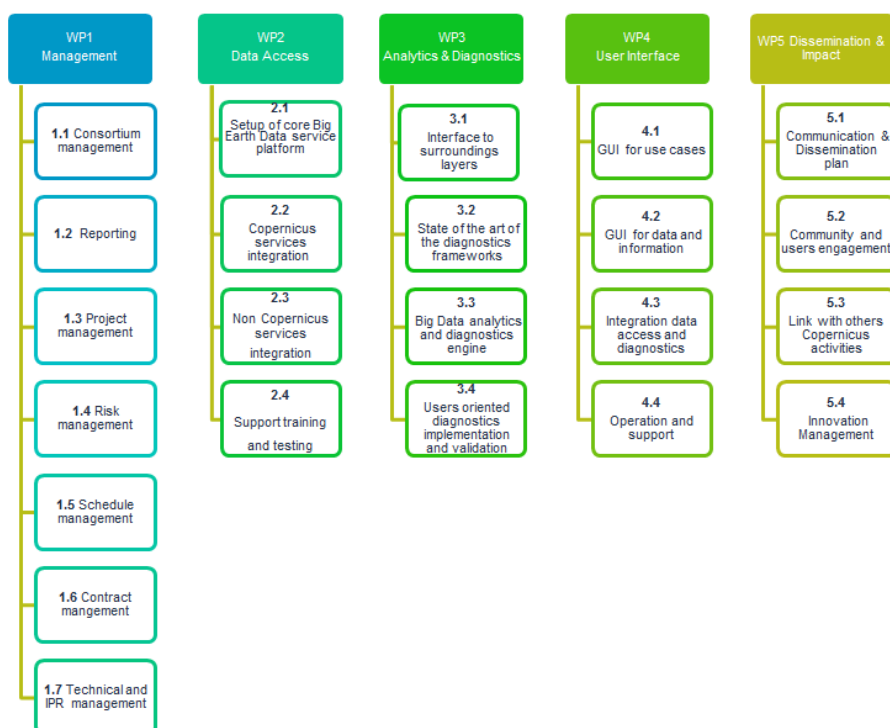
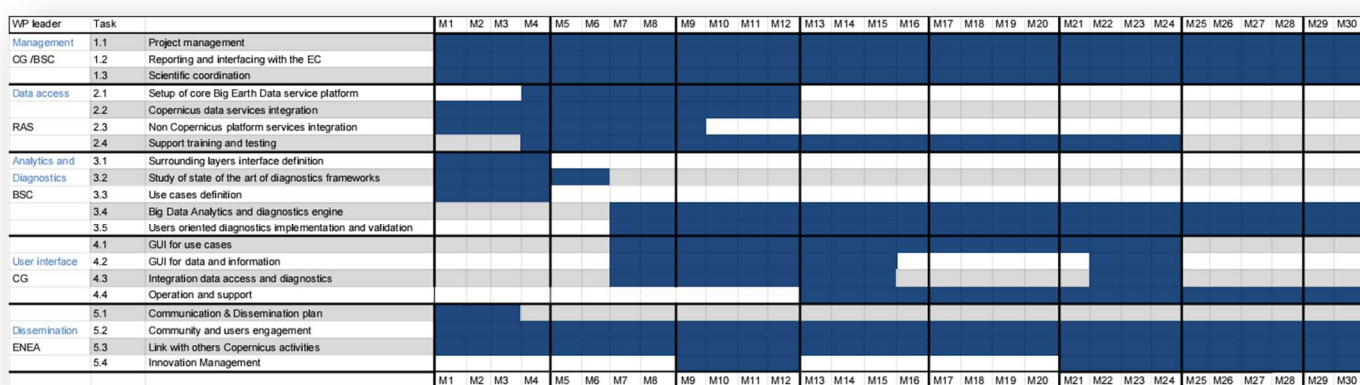


Figure 11: Work packages break down structure

Timing of work packages is shown in figure 12 below. The Gantt chart showcases the complete schedule of ETCETERA and the different elements proposed to achieve the general objectives. ETCETERA proposal has been organised in a 30 months project.

Figure 12: Gantt chart over 30 months



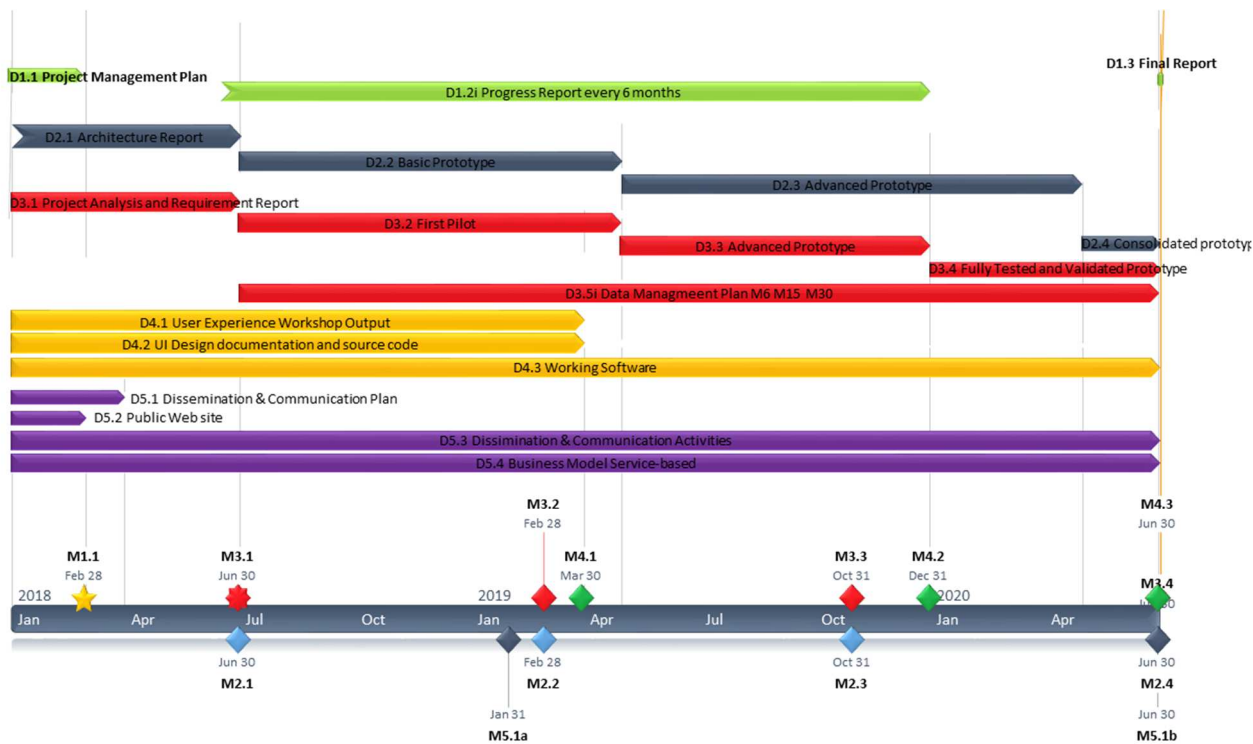


Figure 13: Pert diagram with associated deliverables and milestones

### 3.2 Management structure, milestones and procedures

This section describes the main structures and procedures which will be performed to coordinate, monitor and report the activities of the project. ETCETERA project calls for a strong management structure with strong focus on objectives and milestones, highly knowledgeable technical management skills and strong focus on risk management. Innovation management aspects will be covered as well. The project management activities are implemented along these guidelines in WP1. ETCETERA will be managed by Capgemini in close interaction with all the members of the consortium presented above. An experienced scientific leader from partner BSC will support and report to project manager. All partners have already established a successful working basis since they are cooperating with similar management approaches in several other EU collaborative projects or industrial collaborations.



## Project organization

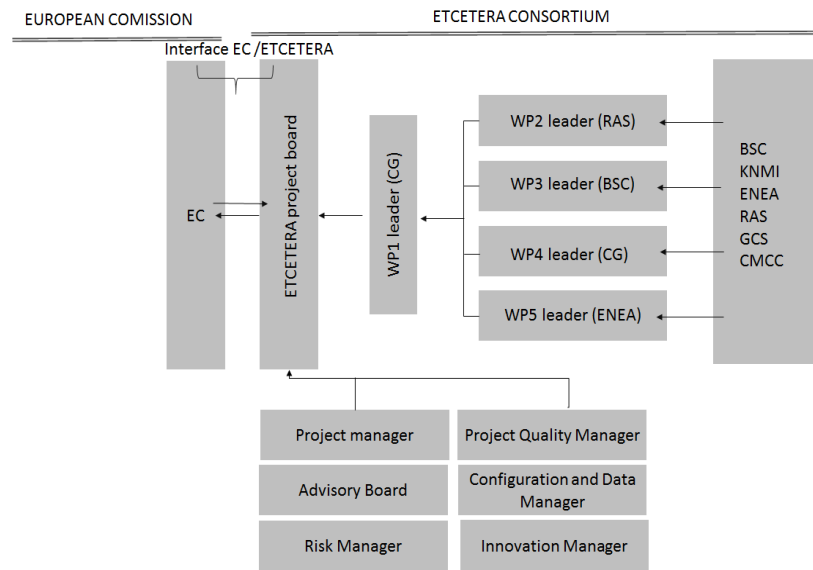


Figure 14: Project organization

### Role and responsibilities

#### Project board

The project board will supervise the project from a strategic level, will be responsible for ensuring that the necessary resources are made available and will oversee the operation of the project management team. They also provide the top escalation level for the project. Key stakeholders of the project, both internal and external, will have a representative on the project board. At a minimum there will be one representative from each partner and one from the customer (European Commission).

#### Project decision making and escalation mechanism

All roles within the project will have the authority to make decisions that are appropriate within the project. Where questions or issues arise requiring a higher view of the project then a defined escalation path will be followed.

#### Project Management team

The project management team will comprise the project manager (who will lead the team) and all the work package managers. They will be responsible for managing the project on a day-to-day basis and collating together all the information to report on the project status. They also provide the second escalation level for the project.

#### The Project Manager

The Project Manager is also assisted by a core team including, Project Assurance Officer (PAO) responsible for ensuring the quality of the delivery, Configuration and Data Manager (CDM), in charge of controlling the configuration within the consortium, and for the deliverables. This role shall be fulfilled by a project manager at Capgemini, Dr Florence Daverdon, who is the coordinating partner for the project and being the EC interface.

### **Work Package Leader**

Each partner will assign a work package manager for each work package they own. The work package manager is responsible for the day-to-day management of the work package related tasks inside a work package and regular reporting to the project manager on the overall status, the progress, the issues, risks and the budget.

### **Advisory Board / External experts**

At users' needs level, we intend to group together high level experts inside an Advisory Board which will be the decision making body of the consortium. This technical board is composed of experts from Partners completed with selected external international experts. The role of this technical Advisory Board will be to support our technical choice, to advice, to recommend and to consider any findings. The board will also consider any findings that may have IPR, and identify an appropriate route for its protection and, if necessary, exploitation.

### **Partner representative**

Each partner in the project shall designate a primary point of contact along with a secondary who cover in case of their absence. This role shall be responsible for coordinating and monitoring all project level communication between their organization and other partner representatives within the consortium. Although peer-to-peer and inter-consortia communication shall be encouraged, it is important to have a single reference point to ensure the project-wide view is considered at all times.

### **Project coordination**

The internal coordination of the ETCETERA project will only be carried out by Capgemini. The project manager from Capgemini will manage internal coordination with support of the Work Package Managers. As coordinator, Capgemini project manager (Dr. Florence Daverdon) will perform the following tasks:

- Organize coordination meeting with work package manager and any required attendees.
- Regularly inform the other consortium participants of the status of the project.
- Ensure the execution of the tasks is performed according to the defined plan.
- Prepare the progress report that integrates the contributions from other participants with proof reading of Capgemini architect (Mr Roger Rutakaza).
- Ensure that the data packs are complete and delivered to the EC on time.
- Be the contractual contact for the EC.
- Follow up the actions of the ETCETERA project.

The Capgemini project manager assigned to ETCETERA will support the Project manager to coordinate the project activities, in the definition of the project's rules (document and development) and in controlling of all deliverables (even from the other participants) provided to EC. The technical aspects will be led by Mr Roger Rutakaza with a team of architects and experts dedicated to the project. The scientific aspects will be led by Dr. Alicia Sanchez coordinating the consortium scientific team. Both technical leader and scientific leaders will be in charge of:

- Organising and coordinating a technical/scientific Advisory Board.
- Coordinating the technical/scientific activities.
- Setting up the common means used by the consortium participants like the project common repository.
- Following up the different development tasks at technical and at scientific levels.

### ***Project Innovation Management***

In the approach to maximise the impact of the project results and its exploitation, ETCETERA will dedicate under the WP5 “*Dissemination & Impact*” a task named ***Innovation Management***. This task will map the suitable business model based service, and it is seen as way of enabling new applications and business opportunities to enable new and innovative business opportunities. The definition of the Business Plan service-based as a foreseen activity during the execution of the project will be delivered at M30. The expression Innovation management is used to share the benefits of European research and development, making related space sector technologies available to the larger industry and as a way of enabling new applications and business opportunities. During ETCETERA, Dr. Matteo De Felice from ENEA, work package leader of the Dissemination and Communication activities, will be in charge of the Innovation management and in particular he will take care of the technology transfer related issues.

### ***Interface with EC***

Capgemini as coordinator will perform the interface with the EC by several means presented in the management procedures:

- Progress and review meetings.
- Progress reports.
- Data packages including the project deliverable items identified for each review.

Immediate communication of any problem or abnormal circumstances detected during the course of the project could be directly carried out with the EC. The potential impacts on the achievement of the project objectives will be evaluated and internally discussed before informing the EC and before proposing an action roadmap for the correction. A justification, an updated schedule and the assessed technical risks will be provided to the EC. An attention will be particularly drawn to minimize the impacts to not delay the time to market. All contractual communications shall be by fax or email as agreed. Telephone and e-mail shall be normally used for day-to-day communications

### ***Project Risks Management***

The formal process to manage and control the project risks is presented here. In the Table 3.2 the Critical risks for implementation are identified as well as the mitigation measures to be applied. The aim of risk management is to identify the project risks and then to keep them continuously under control with dedicated mitigation actions. Risk management is an iterative process, consisting of the following key phases: risk identification, risk assessment, risk handling/control and risk communication. The Capgemini Project Manager and the Project Management Team are in charge of implementing the risk management process for the project. The Project Assurance Manager shall identify risks related to project activity and shall participate to risks Meetings. Project coordinator seeks to establish and maintain a genuine collaboration that shall result in efficient joint risk mitigation strategy and plan. Risk status is addressed in all project committees, whatever their level, from weekly internal progress meeting to the executive steering committee with client sponsors. Risks to be addressed at every level are selected with respect to their respective criticality and manageability. A Capgemini internal escalation procedure is applied in the same way as for problems.

### ***Project Quality Assurance***

Project Assurance is in charge of the control and verification of the compliance of the project rules defined at kick-off meeting. Capgemini project manager will lead this activity and will be responsible for each member

of the consortium in charge of the WP. All project deliverables will be reviewed and assessed in a two steps procedure:

1. Each deliverable will be reviewed by the technical and quality consortium in charge of this deliverable.
2. Consortium reviewing team will perform a final review before the delivery to EC. The consortium maintains the highest standards in delivery quality. One of the basic principles in the implementation of our tasks is the continuing attention to a high quality standard.

As coordinator, Capgemini provides a Quality and Environmental Management System (QEMS) to ensure the quality of process execution and the management and control of products and services. The following table shows a short description of the quality principles which Capgemini uses in the implementation of its projects. Capgemini operates under its own (ISO 9001 certified) quality management system. The Quality Management System (QMS), project management and quality assurance procedures are designed to be compatible with ISO guidelines. This includes abide by the eight quality management principles defined by ISO45 and regularly engaging in internal audits to identify areas where processes could be improved. The table below is a list of the quality management principles and a short overview of our approach to implementing these principles.

<b>Quality management principle</b>	<b>Implementation</b>
Customer EC focus	Regular team company meetings discussing customer needs & requirements and how the consortium can better address these.
Leadership	Clear vision for the consortium in its work within ETCETERA Clear business plan with targets and goals for the business and individual team members. Line management relationships for staff members.
Involvement of consortium members	Line management relationships. Open meetings and discussions.
Process approach	Clear objectives. Fully functional and comprehensive project management software. Systematic management of relationships with partners. Regular internal progress meetings during a project. Maintenance of a risk register.
System approach to Management	Clear budgeting and resource allocation. Fully functional and comprehensive project management software.
Continuous improvement	Regular monitoring and evaluation of QMS (system level). Regular monitoring of feedback received. Dissemination of results, feedback and actions.
Factual approach to decision-making	Fully functional and comprehensive project management software. Culture of open information and access for all consortium members.
Mutually beneficial	Regular partnership working.

supplier relationships	Vetting of partners Open communication channels during contracts and outside specific contracts (e.g. on framework tenders).
------------------------	---

### ***Project Change management***

Formal change management process will be performed throughout the project. All changes requested by internal or external partners will be actively managed by the Project Manager and the project management team to ensure that they are assessed for impact and effort of implementation. Based on the assessed impact and effort cost the project management team will decide if they should be accepted, rejected or escalated to the Project Board. If escalated, then the Project Board will have the final decision.

### ***Legal and ethical considerations***

Legal considerations (IPR): Within the Consortium Agreement, most legal issues pertaining to Ownership, Confidentiality and IPR issues related to resources brought into the project by partners and those developed within the project itself will be dealt with. As stipulated, the Consortium Agreement, based on the Multi Beneficiary General Model Grant Agreement, will be signed by all partners before the beginning of the project. As the issue is complex in this project, given the variety of partners, the background IPR and project developments, identified tasks are dedicated to fully explore the issues of IPR, licensing and eventual access to software/hardware within, after and/or outside the project.

***Ethical considerations:*** No ethical considerations are relevant for the ETCETERA project.

## ***3.3 Consortium as a whole***

As shown in Figure 15, the ETCETERA consortium is made of 7 partners from 5 European countries, representing a wide taste of Europe in terms of population, culture and economic power. The well balanced consortium between business and public sectors includes 4 research institutes and a team of 3 companies with a strong Research and Innovation (R&I) capacities. It includes the necessary and sufficient number of complementary partners covering all the required cross-sectoral and multidisciplinary expertise to successfully carry out the required tasks as well as to assure a manageable project structure and minimize the risk to achieve the ambitious goals of the project.

The project brings together world-class research institutes in Europe with experience in climate and air quality impact modelling, parallel diagnostics and large scale scientific data management together with private companies specialized in air quality sector (Green City Solutions) and Big Data Technologies (Rasdaman GmbH) having leading-edge solutions in the field of EO and climate data. As shown in the participants' description (section 4.1), most of the partners have a robust experience in European projects, having been successfully involved in the coordination of projects or their technical implementation. Most of them are very

familiar with EO and Big Data technologies that are used daily in their work (BSC, CG, ENEA, Rasdaman, KNMI and CMCC) and they all bring to the consortium different technical and scientific knowledge.

**Capgemini** is one of the world's foremost providers of consulting, technology, outsourcing services and local professional services. Present in over 40 countries with almost 180,000 people, the Capgemini Group helps its clients transform in order to improve their performance and competitive positioning. The company offers a panel of integrated services that combine top-of-the-range technology with deep sector expertise and will bring the necessary trustworthiness in terms of project management in complement to their technical knowledge in terms of Big Data solutions. As a leading provider of IT and data services, Capgemini has the expertise and experience in delivering Big Data projects. Open data, along with Big Data and the actionable insights that managed analytics can bring is at the heart of Capgemini's priorities. Capgemini has strong skills about software development. It encompasses all kind of softwares: data processing chains, mission-critical systems and "Digital Customer Experience" websites for collaborative systems, portals, social networks, mobiles applications. This has been reflected in their role of project coordinator and WP4 leader.

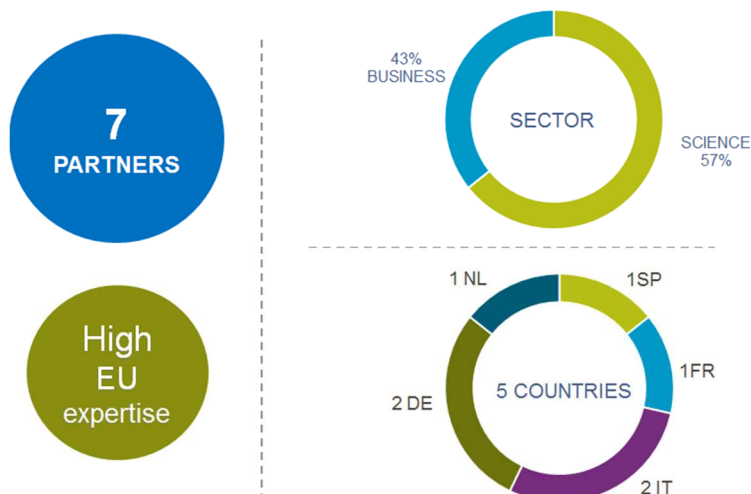


Figure 15: Anatomy of the ETCETERA consortium

**BSC** is participating through their Earth Sciences department will make the bridge between the scientific and technological part of the project, using their long term engagement in air quality modelling. They operates the first Regional Specialized Meteorological Center with activity specialization on Atmospheric Sand and Dust Forecast (RSMC-ASDF) officially recognized by WMO<sup>39</sup> and the WMO Sand and Dust Storms Warning Advisory and Assessment System<sup>40</sup>. Furthermore, they develop an in-house air quality model and developed and operate the CALIOPE<sup>41</sup> air quality forecast system for Europe and Spain. To support and improve research in air quality and climate, they have a deep experience on High Performance Computing (HPC) and Big Data manipulation and analysis with distributed computing tools like COMP Superscalar (COMPSs)<sup>42</sup>, a framework which aims to ease the development and execution of applications for distributed infrastructures, such as Clusters, Grids and Clouds. BSC is leading WP3 on diagnostics and analytics performance

**CMCC** is a well established excellence center, focused on the integrated study of climate change related topics represents a reference for decision makers, institutions, as well as public and private companies seeking technical-scientific support. CMCC Foundation is also member of the European Network for Earth System Modelling (ENES) and partner of the Earth System Grid Federation (ESGF), providing access to 100TB CMIP5 data through its data node deployed at the CMCC Supercomputing Center. Their experience with

<sup>39</sup> [http://dust.aemet.es/about-us/EC65d0431WEATHERISSUESapproved\\_en.pdf](http://dust.aemet.es/about-us/EC65d0431WEATHERISSUESapproved_en.pdf)

<sup>40</sup> <http://sds-was.aemet.es/>

<sup>41</sup> <http://www.bsc.es/caliope>

<sup>42</sup> <https://www.bsc.es/research-and-development/software-and-apps/software-list/comp-superscalar/>

ESGF and the Ophidia<sup>43</sup> Big Data analytics framework, will also allow the consortium to link data that are not in the Copernicus database, as well as to design and implement the WP3 analytics and diagnostics layer in ETCETERA.

**ENEA** is the Italian National Agency for New Technologies, Energy and Sustainable Economic Development aims at the high level dissemination of information in order to increase social awareness on climate change. Therefore, with their recognized experience in this domain, their role of leader of WP5 on dissemination fitted the needs of the proposal

**KNMI** is the national research and information centre for weather, climate and seismology in the Netherlands. With a long tradition in operational and research activities, it has a long standing relationship with Copernicus. It is involved in the development of the IS-ENES backed Climate4Impact portal, C3Ss CLIP-C portal and is a partner in space EO initiatives such as the upcoming TROPOMI satellite. As a key scientific participant in many other earth sciences initiatives, KNMI is the perfect partner to contribute to the design of the diagnostics, dissemination and provenance tooling.

**Green City Solutions**, as SME, brings to the consortium the added value of external users to the Copernicus data, showing real use cases and needs from private companies needing this kind of data. GCS ranked 2nd in the Cisco Innovation Grand Challenge in 2015 and was selected as winner of the EIT Digital Challenge for Digital Cities. With projects spanning from Europe to Asia they have collaborated with corporate customers like CISCO and DB on projects regarding the mitigation and analysis of air pollution in the cities of Paris and Berlin and currently have projects lined up in the UK, Swiss, Italy and Macedonia for 2017.

**Rasdaman GmbH** has long experience in large-scale spatiotemporal EO services, based on its scalable array engine, rasdaman. Databases of spatiotemporal sensor, image, time series, simulation, and statistics data exceed 250TB, single queries have been successfully split across 1,000+ cloud nodes and across continents (Europe and Australia). In January 2017, ESA has characterised rasdaman as the “currently world leading environment in this domain and the standard working horse for OGC standardisation on these innovative data access and interface”.

It is worth noting the complementary between the partners and their coverage of the field and diversity in their organization small and large companies, and research centres. The ETCETERA consortium represents an extraordinary partnership on experts in complex IT integration, researchers experts in climate and air quality impact modelling, parallel diagnostics and large scale scientific data management; data scientists with strong experience in the domains of interests; software architects and computer scientists with recognized expertise in semantic space, climate weather and air quality modelling and interoperability aspects. Project partners, besides offering scientific capacity, experiences and skills to successfully perform the project’s tasks also ensure that remarkable advancement of the state-of-the-art in diagnostics frameworks can be achieved.

All of the partners have been involved in international R&D projects before and most of the partners have worked together in one or more prior projects and will bring their collective knowledge into the ETCETERA project. A core team of research organizations (BSC, CMCC, ENEA, KNMI) provides all the skills to realize the ETCETERA scientific vision. The private partners assure a real professional orientation towards business exploitation of the project results. The private sector team including two start-ups (RAS, GCS) brings the end-user needs and the capacity to deploy data management access and will act as a standardization facilitator, thus helping ETCETERA in maximizing its expected impacts. Capgemini provides from one side a robust

---

<sup>43</sup> <http://ophidia.cmcc.it>

framework and industrial ways of working to ensure a high level of project management and minimizes the potential risks and, for another side, a strong expertise on Big Data technologies. The project management has opted to delegate the scientific project coordinator to BSC which has a long track record in coordination of research projects at EU level while Capgemini will lead the other management aspects and presents a long list of management of projects at the European and national level and in cooperation with industrial partners.

With regard to the multidisciplinary skills provided by partners, the key expertise and relevant research fields of each partner in Europe is presented in the table below

<b>Participant short name</b>	<b>Main functions</b>	<b>Value to the consortium</b>
CG	Project Management and Administration Lead Technical management: user interface leader and ETCETERA technical solution manager	Many years of expertise in coordinating R&D initiatives, including EU projects and large collaborations with industries; strong skills in Big Data technologies and industrial system integrator.
BSC	Scientific management Diagnostics and analytics performance Dissemination activities	Recognized experience in both technical - as HPC center and part of several European data related projects- and scientific -R&D on air quality and climate, development and management of an operational air quality forecast system and a WMO center on mineral dust forecast - domains.
KNMI	Diagnostics and analytics reproducibility. Data integration, dissemination and exploitation	Experience in the exploitation of heterogeneous data for R&D activities in the field of air quality services. Provision of integrated dissemination and reproducibility mechanisms and tooling according to OGC and W3C standards.
ENEA	Diagnostics and analytics performance Dissemination activities	Expertise in dealing with climate data practitioners with a wide range of backgrounds.
RAS	Scalable data access management	Recognized experience in Big Earth Data through the rasdaman array engine as well as editorship of several Big Data standards in OGC and ISO. Link to EarthServer and other international initiatives allowing to integrate ETCETERA and DIAS into federations.
GCS	Business sector user requirements	Expertise in defining and developing



	Business model based service definition	business models relating to air pollution analysis and mitigation for the industrial and public sector.
CMCC	Diagnostics and analytics Dissemination activities	Expertise in large-scale data analytics for climate change, scientific workflows, high performance data management. Experience in dealing with climate and weather data, formats, and tools. CMIP5 contributor and Tier 2 site participating to ESGF. Very active participation over time to dissemination activities (e.g. EGU, AGU, ESGF, RDA, ESA).

### Industrial/ Private involvement

Industrial participation is strong in ETCETERA. The consortium includes a total of 3 industrial/SME partners with multiple and complementary skills and technological expertise, covering the full spectrum of innovation activities foreseen in the project. In particular, two of the partners are specialized SMEs (RAS, GCS) while CG is a large industrial partner from the IT and expert IT applied to Space and Science sector, leader in system integrator market.

The SMEs involved in the consortium are among the leading and innovative companies in their sectors. For instance, both have been recently recognised such Rasdaman has been awarded **Winner of the Big Data Challenge at the Copernicus Masters** in 2014, and at the beginning of 2016 Green City Solutions was selected as the **first German start-up to participate in the Open Data Incubator for Europe (ODINE)**. The project stretched over 6 months and was aiming at using open data datasets to analyse and establish correlations of air pollution to extreme climatic conditions and locations. SME partners share in common their proven ability to apply research results to successful and well established commercial products and services (e.g. CityTree services for public authorities and individuals players provided by Green City Solutions and the advanced array database provided by Rasdaman).

On the other hand, industrial partner CG is actively involved in the management project aspects and in the framework integration aspects (using also User Interface methodology as described in section 1.3), thus this private and industrial team introduces a strong commercial and business perspective to the activities as well as providing a sound base for further commercial development and exploitation within their own production processes and beyond.

Moreover, industrial partner CG, as a leading EU provider of technology solution and system integration services has a key business case in proposing ETCETERA solutions to its client database to other cross-sectoral domains beyond weather, climate and air quality sector. Specific dissemination actions will be carried out, through already established communications channels, networks, and commercial fair trades will be attended by Rasdaman in order to ensure that the results will be conveyed to major stakeholders and therefore further indicating the potential high commercial impact of the project results. The consortium has already initiated relationships and interests from international stakeholders which will be enlarged during the project lifetime.

At this stage, several European SMEs and two European consortiums and cluster of SMEs (Aerospace Valley and Proambiente) have expressed their interests.

### 3.4 Resources to be committed

#### Tables for section 3.1 Work package description

For each work package:

##### WP1 - Management

Work package number	1	Lead beneficiary					CG
Work package title	MANAGEMENT						
Participant number	1						
Short name of participant	CG						
Person months per participant:	25						
Start month	1			End month	30		

#### Objectives

This work package aims at carrying out an effective management of the project in order to:

- Coordinate and organize management activities
- Manage interface with EC and members
- Manage financial activities along the project progress
- Plan and schedule activities for the project
- Organize documentation control (technical and project activity) and to prepare deliveries (doc and components)
- Manage actions & risks

**Description of work** This work package contains the following tasks:

#### Task 1.1 - Consortium management – Period: M1-M30 - Leader: CG

- Establish and provide a Project Management Plan (PMP)
- Organize, monitor and control all project activities and ensure the overall integrity of all Work Packages
- Check all items under configuration
- Coordinate the participation of the end-user organizations to the project
- Coordinate and organize the reviews
- Participate the promotion and dissemination of the results of the project
- Coordinate the international conferences and workshops

**Task 1.2 - Reporting activities to consortium and interface with EC – Period: M1-M30 - Leader: CG; Contributors: All WP leaders**

- Report on the progress of the work on a regular basis to the EC Officer by providing bi-monthly progress reports, starting at the end of the second month
- Organize periodic progress meetings with EC by teleconference, preferably every 6 months
- Provide meeting agendas to the EC, at least 2 weeks prior to each progress meeting (done by teleconference) and 2 weeks prior to each project review (i.e. Kick-Off, Mid Term review, Final Review)
- Write the minutes of all progress meetings and project reviews.

**Task 1.3 - Project management – Period: M1-M30 - Leader: CG**

Documentation control (scientific, technical and project activity), configuration control, deliveries management. Check and review all project deliverables for quality and completeness before delivery.

**Task 1.4 - Risk management – Period: M1-M30 - Leader: CG**

Maintain an up to date Action Items List (AIL) and identify the potential risks. The coordinator should be in charge to mitigate the risks mitigation and propose solution as soon as possible.

**Task 1.5 - Schedule management– Period: M1-M30 - Leader: CG**

Take all the necessary steps to maintain the schedule. In case of deviation from schedule, the coordinator shall identify and propose corrective actions to recover all scheduling overrun.

**Task 1.6 - Contract management– Period: M1-M30 - Leader: CG**

The coordinator should tackle different subjects: authoring of the results, baseline and scope of the project, the commitment and contract visibility. The right and responsibilities of the partners in identifying and managing IPR will be defined in the consortium agreement.

**Task 1.7 - Technical management & IPR management– Period: M1-M30 - Leader: CG**

The coordinator should manage the technical resources needed for the project and shared between the different partners. She ensures the homogeneity of the solution and should consider the technical Advisory Board recommendations. A technical status will be included into the progress reports. The coordinator will draw attention and check the licenses handled into the project. For the project's components to be deployed as open source, it is important to check the non contamination by another license.

**Deliverables** (brief description, **leader**, and month of delivery)

D1.1 Project management Plan, **CG**, M2,

D1.2 Progress report, **CG**, M6, M12, M18, M24

D1.3 Final report, **CG**, M30

**Milestones** (brief description and month of delivery)

### M1.1 Setup of the management structure, M2

#### Dedicated KPIs

K1.1 Relative effort consumption deviation. Measured as the difference between the monthly effort consumption compared with an initial (linear) effort consumption estimation, divided by the initial (linear) effort consumption estimate. This KPI will target a deviation below 20%.

K1.2 Deliverable submission delay. Measured as the total number of days of delay regarding the submission of each deliverable divided by the number of deliverables due at that month. This KPI will target 0.

K1.3 Milestones achievement delay. Measured as the total number of days of delay regarding the submission of each milestone divided by the number of milestones due at that month. This KPI will target 0.

### WP2 - Data Access

Work package number	2		Lead beneficiary			RAS	
Work package title	DATA ACCESS						
Participant number	1	2	3	4	5		
Short name of participant	CG	BSC	KNMI	ENEA	RAS		
Person months per participant:	7	6	6	3	30		
Start month	1			End month	30		

#### Objectives

In the ETCETERA architecture, the Big Data storage and access component will be provided by WP2. Centered on the rasdaman scalable array engine, a distributed platform will be established which enables access to massive spatiotemporal EO and weather data (generally speaking: multidimensional raster data) through the enabling datacube paradigm. **Vertically**, it will form the base layer for WP3 functionality. Access interfaces will strictly rely on the open OGC standards WCS, WCPS, and WMS (including data ingest through WCS-T). **Horizontally**, distributed storage and coupling with project-external archives will be accomplished, such as with the EarthServer federation and ESGF.

#### Description of work

The goal of this task is to set up the Big Data management and query platform, based on the pre-existing rasdaman scalable array analytics engine, plus further tools coupled to rasdaman. The result is the service used by WP3. Therefore, a support task is defined to accommodate all needs of WP3, but also of external users of the ETCETERA data platform. The **WP2 architecture** will be based on the open-source platform

rasdaman community<sup>44</sup> which enables array queries on spatiotemporal datacubes (such as satellite image time series) which are stored using adaptive partitioning, with the possibility of doing data fusion on cubes stored at different disks or data centers. Access interface for WP3 as well as external users will be given by OGC / INSPIRE Web Coverage Service (WCS) standards suite<sup>45</sup> which incorporates the Web Coverage Processing Service (WCPS) geo datacube analytics language. This allows, among others, clients ranging from map browsing (e.g., OpenLayers) over Web GIS (e.g., QGIS) up to analytics (e.g., python).

**Work items** will include setting up, configuring, and tuning rasdaman, establishing project an data specific metadata handling, as well as enabling connection to DIAS and further Copernicus and non-Copernicus services.

**Work strategy** is to have an early ramp-up phase where existing tools – in particular: rasdaman – get installed for a preliminary service; this also serves for a requirements elicitation. After this common task, work will spread out into individual enhancements. Organizationally, three increments will deliver successively advanced services (basic/advanced/consolidated). This ensures that requirements expressed initially, but also requirements emerging can be considered adequately. At every milestone, the status delivered will be discussed with stakeholders, and new steps for the next increment will be prioritized in a consensus process. A dedicated task ensures continued support and quality management.

**Partner work distribution:**

RAS will establish the core Big Earth Data platform based on its open-source rasdaman technology, as well as data integration with DIAS and the emerging H2020 EarthServer federation (in particular, addressing CAMS and C3S through EarthServer partner ECMWF). BSC will provide storage infrastructure for the rasdaman-based services to be established in ETCETERA. Two sets of tape storage will be provided for sharing among the partners as a data cache storage and a first storage to store some data before all the Copernicus data is made available. KNMI will integrate provenance capturing mechanism and services into the platform. CG will perform the Copernicus metadata integration service via DIAS into the platform. At a first stage, the time for DIAS to be operational, the platform will establish and consume the interface with CAMS Services sources. ENEA will work on EO data integration and reconciliation.

**Task 2.1: Setup of core Big Earth Data service platform; period M1-M6; Leader RAS(4PM); Contributors, BSC (2PM), GC (2PM)**

In the initial project period, the core data management and service platform, based on the scalable rasdaman array engine and tools coupled to it, will be established so that in the subsequent WPs application (WP3) and platform (WP2) partners can work in parallel on using and extending it. The common paradigm implemented will be spatiotemporal datacubes linked with metadata and further types of pertinent data. Web service interfaces will rely strictly on open standards, especially OGC WMS<sup>46</sup>, WCS<sup>47</sup>, and WCPS<sup>48</sup>. Work to be performed in Task 2.1 includes:

1. Establishing local installations at BSC, running rasdaman and the further tools needed (in particular, ingest pipelines).
2. Establishing ingest pipelines for growing these installations continuously.

---

<sup>44</sup> [www.rasdaman.org](http://www.rasdaman.org)

<sup>45</sup> [http://external.openeospatial.org/twiki\\_public/CoveragesDWG/WebHome](http://external.openeospatial.org/twiki_public/CoveragesDWG/WebHome)

<sup>46</sup> <http://www.openeospatial.org/standards/wms>

<sup>47</sup> <http://www.openeospatial.org/standards/wcs>

<sup>48</sup> <http://www.openeospatial.org/standards/wcps>

3. Initial training for partners in querying, feeding, and maintaining service platform nodes. The material available at [www.earthserver.eu/webinars](http://www.earthserver.eu/webinars) will be used and extended for this purpose.

At BSC, the dedicated storage for ETCETERA (cache for some of the Copernicus/non-Copernicus data, storage of the diagnostics and intermediate products) will be handled in two systems, hosted by BSC: a FAT node and a tape system. Access to tape will be managed through a workflow manager to handle queries in an asynchronous way. The expected outcome is D2.1.

**Task 2.2: Copernicus service integration, Period: M7-M30, Leader: CG (5PM), RAS (10PM), ENEA (3PM)**

A connection from the ETCETERA service platform to DIAS will be established so that DIAS data and processing capabilities can be integrated with the ETCETERA offering, thereby establishing a common information platform for WP3 and external users. Depending on the capabilities of the forthcoming DIAS service, a reverse integration may be established too, enabling DIAS to utilize the ETCETERA offering. Further, the CAMS Copernicus data source will be integrated (feasibility has been clarified upfront with ECMWF, also their support). This integration is specifically necessary for the air quality use case pursued in WP3. Finally, CMEMS will be investigated for a possible integration. A specific mediator will be developed between Copernicus Data sources (DIAS including CAMS) and ETCETERA WP2 data access layer; it will use WCS-T for setting up ingest pipelines into rasdaman.

Generally, the OGC WCS standards suite - where rasdaman implements the Core and all Extension specifications - will act as integrating abstraction allowing to retrieve data from DIAS and make them available on the ETCETERA platform. This task will be accomplished by RAS. A complementary tool will be a catalogue to be defined enabling to reference a large volume of metadata. The provenance of the metadata might be from different sources, either from Copernicus or non-Copernicus data. CG will be in charge of carrying out this task based on the reuse experience of previous and similar projects (see reference in annex). ENEA/RTU will define methods and techniques for the integration of EO data coming from multiple sources.

To this aim, we will address the problem of data reconciliation in the specific domain of EO data, which is needed since different EO data sources are heterogeneous in content and structure. Following a consolidate practice in the context, EO data reconciliation requires the definition of: a transformation phase, to resolve conflicts between formats and representations of the same attributes (e.g., a date or a geographic coordinate), followed by a data fusion phase (also known as entity resolution), which identifies and merge semantically-related entities. The typical activities involved in the integration of heterogeneous data sources:

1. metadata acquisition,
2. schema alignment, where attributes of different schema are matched, and
3. record linkage (also known as entity resolution or duplicate detection), which identifies and merges semantically-related entities.

**Task 2.3: Non-Copernicus service integration, Period M7-M30, Leader RAS (10PM); Contributor: KNMI (6PM)**

In this task, sample services beyond the Copernicus umbrella will be typically integrated. Integration means that the same ETCETERA clients that access the ETCETERA platform can also access these “external” services, based on the open OGC standards; also fusion between mixed sources will be studied.

One target consists of H2020 **EarthServer**<sup>49</sup>: based on rasdaman, it consists of a federation of international data centers (initially: ECMWF, PML, ESA/MEO, NCI Australia) hosting 3-D x/y/t image time series and 4-D x/y/z/t weather datacubes allowing flexible retrieval and fusion based on WCS/WCPS. As of today, the ESA node hosts already 250+ TB of Sentinel data, growing to a PB. The European Center for Medium range Weather Forecast (ECMWF) is working on unleashing their 120 PB weather archive via rasdaman. This will be accomplished by RAS who is EarthServer partner. In co-development with the air quality use cases definition and realisation of WP3, high resolution data sources will be also evaluated by KNMI as target of the integration to support innovative and accurate services.

These will include **urban air quality local measurements**, for instance such as the AiREAS network for the city of Eindhoven<sup>50</sup>, local meteorological observations<sup>51</sup> and other heterogeneous resources. These may include traffic intensity and velocity Nationale Databank Wegverkeersgegevens<sup>52</sup> (NDW), road classification and population density. With regard to the Traceability and Provenance for Big Data Access, with the end goal of Scientific Reproducibility of ‘data fabric’ workflows for EO systems using Arrays DBs, concretely: rasdaman, the activity undertaken by KNMI will integrate software hooks and tuneable annotations for the run-time extrapolation of lineage information associated with the data extraction and transformation processes performed by the ETCETERA data framework. Consistency and linkage to the dependencies established in the post processing tasks performed in WP3 will be also guaranteed. The task will contribute to the Data Management Plan and it will allow the on-demand production of provenance documentation, in compliance with the W3C-PROV concepts and interoperable representation.

#### **Task 2.4: Support, Training, Testing, Period M7-M3, Leader: RAS (6PM), Contributor: BSC (4PM):**

All WP2 partners will unite in giving support and advanced trainings (see Task 2.1 for initial trainings) to WP3 – which uses the WP2 platform – as well as other, external users wanting to make use of it. This includes production of adequate training material. Finally, RAS will evolve its regression testbed to achieve the necessary software stability, performance, scalability, and standards coherence. Further, training and material will be provided for the metadata and provenance components. BSC will undertake performance benchmarking.

#### **Deliverables** (brief description and month of delivery)

- D2.1: Architecture and interfaces report & Installation report (Task 2.1 description of installation, data holdings, data management plan, access guidance), **RAS**, M6
- D2.2 : Basic prototype (software; Tasks 2.2-2.5), **RAS**, M14
- D2.3 : Advanced prototype (software; Tasks 2.2-2.5), **RAS**, M22
- D2.4 : Consolidated prototype (software; Tasks 2.2-2.5), **RAS**, M30

#### **Milestones** (brief description and month of delivery)

- M2.1: Initial services established, associated to D2.1, M6
- M2.2: Basic prototype associated to D2.2, M14
- M2.3: Advanced prototype associated to D2.3, M22
- M2.4: Basic prototype associated to D2.4, M30

<sup>49</sup> [www.earthserver.eu](http://www.earthserver.eu)

<sup>50</sup> <http://data.aireas.com/csv/>

<sup>51</sup> <http://projects.knmi.nl/klimatologie/uurgegevens/selectie.cgi>

<sup>52</sup> [http://www.ndw.nu/pagina/nl/103/datalevering/120/open\\_data/](http://www.ndw.nu/pagina/nl/103/datalevering/120/open_data/)

**Dedicated KPIs:**

- K2.1: Target at M30 consolidated prototype installed and working at 100%
- K2.2: Target at M30 the provenance coverage for attribution, data resources and lineage metadata will be evaluated against the specification indicated by the Data Management Plan. Aiming at 100% coverage.



### WP3 - Diagnostics & Analytics

Work package number	3	Lead beneficiary				BSC
Work package title	ANALYTICS AND DIAGNOSTICS					
Participant number	1	2	3	4	5	6
Short name of participant	CG	BSC	KNMI	ENEA	GCS	CMCC
Person months per participant:	6	39	10	10	9	21.5
Start month	1			End month	30	

#### Objectives

The main objective of this work package is to create a diagnostics and analytics layer between the *data access* developed in WP2 and the *user interface* developed in WP4. This layer will be designed taking into account:

1. needs from user community (through a set of use cases),
2. the state of the art in the diagnostics and analytics fields as well as
3. interoperability, extensibility and performance principles.

#### Description of work

This work package defines and implements the architectural component (engine) responsible for providing the diagnostics and analytics functionalities in ETCETERA, a set of diagnostics driven by three different use cases jointly with the ETCETERA data management plan.

This work will be linked to WP2 regarding optimal access to the data storage, WP4 in terms of functionalities exposed to the interface and WP5 for the non-expert users' requirements. It will provide the proper intermediate layer between EO and model data from Copernicus and other platforms (input) and diagnostics (output). Starting from existing tools/frameworks in the Big Data landscape, this WP will identify the most suitable solutions with respect to the project objectives, jointly with the proper adaptation activities to deliver a scalable, extensible, modular and robust analytics and diagnostics layer. A close link between this WP and the CDS (Copernicus Climate Data Store) will be established to follow the advancement of the CDS Toolbox<sup>53</sup> to avoid duplication of work and ensure complementarity regarding the developments of the diagnostics. Even if, at this stage, the other Copernicus services (CAMs, CMEMS, CLEMS, EMS) don't have any similar "toolbox", close attention will be paid to the different ongoing efforts regarding these components. The exact type of satellite of data (level 0, 1 or 2 as defined in Copernicus<sup>54</sup>), that the diagnostics will access will be defined by the use case themselves.

This WP will be also in charge of defining the Data Management Plan.

<sup>53</sup> <https://climate.copernicus.eu/tenders/c3s25-software-development-climate-data-store-cds-toolbox>

<sup>54</sup> <https://sentinels.copernicus.eu/web/sentinel/user-guides/sentinel-1-sar/product-types-processing-levels>

**Task 3.1: Surrounding layers interface definition - Period:M1-M4, leader: CMCC (2PM) contributors: BSC (2PM), CG (1PM)**

This task led by CMCC will define the proper interface between the analytics & diagnostics layer and the user interfaces provided by WP4. Taking into account interoperability, this interface will properly decouple the end-users applications from the analytics and diagnostics engine.

**Task 3.2: Study of state of the art of diagnostics frameworks - Period:M1 - M6, leader: KNMI (4PM), contributors: BSC (4PM), CMCC (2PM), CG(2PM)**

Review of existing frameworks enabling diagnostics based on their compatibility with the project objectives. The list of some commonly used frameworks includes (but it is not limited to): Ophidia, CAFE, CMDA, ESMValTool, s2dverification, KNMI Prov, BSC COMPs, Climate4Impact portal. The compatibility of each of the platforms with the project objectives will be evaluated according to the following criterias: portability, scalability, modularity, as well as adaptability with the user requirements coming from the use cases defined in T3.3.

**Task 3.3: Use cases definition - Period:M1 - M4, leader : GCS (4PM), contributors: BSC (4PM), KNMI (2PM), CG(2PM)**

Study and definition of end users use cases. The possibility to access at the same time and in the same way to a huge variety of data give us the possibility to imagine different scenarios for very diverse kind of customers which could be able to exploit services of the ETCETERA platform or to create their own ones on top of it. In particular, we will define three use cases:

1. **Scientific Community (BSC).** Due to the increasing complexity of atmospheric models, their evaluation is an increasingly issue. In the last few years many studies demonstrated the role of mineral dust in a huge variety of human activities and in general in the whole society (health, agriculture, transports, industry, ...). For these reasons, this use case will be defined as a (near) real-time performance analysis of a dust models ensemble through calculation of various statistics with the aim to compare models' outputs against Earth Observations (in-situ and satellite products). This model evaluation will be extended to other types of pollutants (ozone, NO<sub>x</sub>, ...) to cover a wider spectrum and to enable researchers to better understand and evaluate model predictions and improve the science within the models themselves.
2. **Business Sector (GCS).** Development of specific use cases for the industrial sector, where the analysis of air quality can reduce costs or increase value of existing processes and infrastructure. E.g. analysis and assessment of real estate properties according to their outdoor pollution levels and increase of value with pollution reduction measurements. The creation of predictive computational pollution level models for particles and gases based on satellite images would have benefits for customers from the health sector and could be expanded to insurance providers and even further. Particulate Matter of the class PM<sub>2.5</sub> and UFP (Ultrafine Particles), as well as NO<sub>x</sub> and their concentration in µg/m<sup>3</sup> have the highest research priority, causing an annual economic damages of over 1.48 billion € in Europe.
3. **Public Body (KNMI)** Policy makers (city halls, local governments as well as at national, European and international level) have already officially expressed interest on having the evaluation of the measures to prevent air pollution and access to tools offering analysing and comparison of different measurements to reduce air pollution, especially regarding PM (Particulate Matter) and nitrous

oxides (NO<sub>x</sub>). A use case providing for instance the hourly concentration fields of NO<sub>2</sub> for an urban area might be defined and validated through the integration and exploitation of operational alternative air quality networks as part of the ETCETERA data access platform provided by WP2 to evaluate the heterogeneity and high-resolution capabilities of the data. By establishing the scientific use case, implemented measurements in the past can be analysed and evaluated on their efficiency, to establish a comparison on efficiency.

**Task 3.4: Big Data analytics and diagnostics engine - Period: M7 - M30 leader : BSC (11PM)**

**contributors: KNMI (4PM), CMCC (6PM), ENEA (4PM) , CG(1PM)**

This task relates to the implementation of a server-side Big Data analytics and diagnostics engine. Starting from selected framework in T3.2, this task will include all the needed implementation and adaptation activities to deliver the proper diagnostics component to WP4, exposing the standard interfaces defined in T3.1. The engine will be based on two separate modules: the former addressing basic analytics functionalities, the latter, built on top, able to orchestrate multiple analytics operators into complex dataflows. Optimizations exploiting equivalence rules regarding the analytics layer will be implemented to address efficient dataflow execution. Based on the requirements coming from T3.3, the analytics module will be properly extended to include all the core building blocks for the diagnostics.

The analytics engine (which will perform the computations as close as possible to the data to avoid unnecessary data transfer overhead), will be connected to the WP2 access layer, through a connector (developed in this task) to retrieve the input data needed for the diagnostics. This task will also cover the delivery of the diagnostics output in different formats. Finally, enabling Traceability and Provenance tools for Earth Sciences and EO Diagnostics Tools will be investigated, features for traceability within the software model will be evaluated and integration of the W3C-PROV concepts and compliant solutions around the diagnostic package will be performed.

**Task 3.5: Users oriented diagnostics implementation and validation:Period:M7 - M30 leader: BSC (15PM) contributors: CMCC (11.5PM), ENEA (6PM), GCS (5PM) , CG(1PM)**

According to the results of tasks 3.3, a second block of diagnostics and analytics definition, configuration and querying will be implemented through an iterative joint collaboration process between the scientific use case responsables and software engineers. A test-driven methodology will be adopted during the development phase of the diagnostics implementation. The diagnostics type of each task will be defined according to the use cases and the needs expressed by the partners. These diagnostics and analytics will include at least time-series, spatial subsettings, and statistical verification of models outputs against Earth Observations. Once implemented, the diagnostics will be validated against the use case they were built for by the users themselves. This task will be discriminated into three sub tasks described below.

**T3.5.1:** Science oriented diagnostics: implementation and testing - leader: BSC - contributors: CMCC, ENEA M7 - M30

**T3.5.2:** Industry oriented diagnostics: implementation and testing - leader: GCS- contributors: BSC CMCC, ENEA M7 - M30

**T3.5.3:** Policy makers oriented diagnostics: implementation and testing - leader: BSC contributors: KNMI, CMCC, ENEA M7 - M30

**Task 3.6: Data Management Plan: Period M1-M6, leader BSC (2PM), Contributors: BSC**

As described in section 2.2.c, a detailed data management plan will be delivered at M6, in which precise description of data storage, formats and standards used in the project will be described. The plan will be updated annually and an intermediate version will be delivered at M15 and final version at M30.

**Deliverables** (brief description and month of delivery)

- D3.1: Project analysis and requirements report (Tasks 3.2-3.3 outcomes description, will include architecture and interfaces report), **BSC**, M6,
- D3.2: First pilot (Tasks 3.1-3.4-3.5), **BSC**, M14
- D3.3: Advanced prototype (Tasks 3.1-3.4-3.5), **BSC**, M22
- D3.4: Fully tested and validated prototype (Tasks 3.1-3.4-3.5), **BSC**, M30
- D3.5a: Data management plan, (Tasks 3.6), **BSC**, M6
- D3.5b : intermediate Data management plan, **BSC**, M15
- D3.5c : final Data management plan, **BSC**, M30

**Milestones** (brief description and month of delivery)

- M3.1: System design finalized (associated to D3.1), M6
- M3.2: First pilot (basic functionalities) (associated to D3.2), M14
- M3.3: Advanced prototype (extended functionalities) (associated to D3.3), M22
- M3.4: Fully tested and validated prototype (First production version), M30

**Dedicated KPIs:**

- K3.1: Number of core diagnostics implemented in T3.5. Target for M15 is 3, for M30 is 10.
- K3.2: Number of implementation and adaptation activities to deliver the proper diagnostics engine. Target for M15 is 2, for M30 is 5.
- K3.3: Standardized questionnaires will be used to validate the acceptance of the implementation in terms of usability, effectiveness and relevance. Target by M30 is acceptance  $\geq 70\%$ .

**WP4 - User Interface**

Work package number	4	Lead beneficiary				CG	
Work package title	User Interface						
Participant number	1	2	3	4	5	6	7
Short name of participant	CG	BSC	KNMI	ENEA	RAS	GCS	CMC C
Person months per participant:	30	2	5			9.5	4
Start month	7			End month	30		

**Objectives**

The objectives of work package WP4 is to offer users of the system an interface to interact with the solution. The primary focus is on the graphical user interface for the end users of the use cases that will demonstrate the platform. Interfaces will also be hosted to help users that are using the platform to develop new uses

cases. The work package also includes the hosting and operation of the user interface layer and its components, as well as the integration with the lower layers provided by WP2 and WP3.

### **Description of work**

This work package will define, develop, deploy, host and operate the user interface layer. It will be linked with WP2 for optimal access to the data and information, with WP3 for execution of analytics and display of results and WP5 for users' requirements. This work package requires inputs from WP2.4 and WP 3.1

**Task 4.1: GUI for use cases, Period M7-M15, leader: CG (8PM), contributors: BSC (1PM), GCS (2PM):** This task will ensure the creation of the user interfaces for the use cases to be developed as part of the demonstrator of value of ETCETERA, as outlined in §1.1. Specifically an application will be developed in order to allow end-users to access air quality information and derive direct engagement of with those users through a well-designed application user interface. It will be developed according to the methodologies of user-centered design and user experience, as outlined in §1.3. Specific activities in this task include:

1. Three User Experience (UX) workshops one per profile users and design for 10 use cases
2. Web or Mobile user interface adapted to uses cases, based on UX design, with a maximum of 20 user interface screens
3. Link with T3.5.1 and T3.5.2 from WP3 in which the diagnostics for the individual use cases will be implemented (via task T4.3)

**Task 4.2: GUI for data and information, Periods M7-15; M22-M24, leader: KNMI (5PM) contributors : CG (4PM), CMCC (4PM)**

The task will integrate common and dedicated User Interface (UI) tools to support developers, analysts and curators in controlling, validating and managing the computations and the results obtained through ETCETERA. More specifically, the interfaces will enable:

1. Browsing and visualising data adopting when possible OGC compliant viewers, according with the access services provided by WP2.
2. Observing and supervising diagnostics processes
3. Accessing Provenance information of EO transformation methods and derived products at various level of detail, for validation and reproducibility.
4. Implementation of a diagnostics editor user interface to help the generation and validation of diagnostics dataflows by the end users.

The interfaces will leverage from the interactive functionalities offered by the components in WP2 and WP3. The task will take care of performing adaptations to new requirements and graphical harmonisation.

**Task 4.3: Integration to Data Access and Diagnostics, Period M7-M24, leader: CG (10PM) contributors: BSC (1PM), GCS (7.5PM)**

This task will ensure the link with the underlying components. It will ensure that the required information and processes are properly exposed to the application user interface so that the end users of the solution can interact efficiently and naturally with the solution. It includes the following activities.

1. Communication between user interface components and data access (WP2) and diagnostics layers (WP3) via their respective API
2. Projection of views adapted to use cases
3. Logic and workflows adapted to use cases

**Task 4.4 Operations & Support M13-M30, leader: CG (8PM)**

This task will ensure that the user interface components can be hosted with an acceptable service level from the time of first minimal viable product release (M6) until the end of the project (M30).

4. Hosting of the user interface components (web application & logic for use case)
5. Web or Mobile user interfaces adapted to uses cases, based on UX design

**Deliverables** (brief description and month of delivery)

D4.1: User Experience workshop outputs (minutes, storyboards, wireframes, design elements), (Task 4.1) CG, M15

D4.2: UI design documentation and source code, (Task 4.2), CG, M15

D4.3: Working software in production with 99% availability from date of first release to M30, (Task 4.3) CG, M30

**Milestones** (brief description and month of delivery)

M4.1: wireframes of the display per user profile, M15

M4.2: UI design document acceptance, M24

M4.3: Acceptance of the ETCETERA framework, M30

**Dedicated KPIs**

K4.1: Absolute number and rate of increase of organizations that sign up to the air quality use case user interface and increase of active users of the air quality use case user interface.

K4.2: Stability, performance and availability of the air quality use case user interface after initial release.

K4.3: Number of displays available per month per user profile, M15 Target: to complete 3 displays per users profiles by M30

**WP5 - Dissemination & Impact**

Work package number	5	Lead beneficiary				ENEA	
Work package title	Dissemination & Impact						
Participant number	1	2	3	4	5	6	7
Short name of participant	CG	BSC	KNMI	ENEA	RAS	GCS	CMC
Person months per participant:	5	5	1	10	4,5	15	4
Start month	1			End month	30		

## Objectives

This work package ensures an effective and high quality project dissemination and technology transfer through the design of an exploitation plan for private sector. All project partners have been assigned work within this WP to ensure the largest dissemination and exploitation of the results of the project.

## Description of work

The activities described in this section aim at supporting a wide-range dissemination of the framework/service and the developed demonstrator, and to implement a program of training activities focused on the application of Big Data technologies on EO. One of the big challenges of this project is to strengthen the link between the EO/Climate and Big Data communities, to this end communication and dissemination activities will take part also during events like scientific conferences and industrial workshops for both communities.

### **Task 5.1: Communication & Dissemination plan, Period M1-M3 Lead ENEA (1PM), Contributors: CG(1PM)**

A dissemination and communication plan provides all the information about the planned activities along the entire duration of the project. Those activities have as objective:

1. Assess the impact of the improved workflow enabled by the developed framework/service on the user community.
2. Wide-range dissemination of the framework/service and the developed demonstrator
3. Implement a program of training activities focused on the application of Big Data technologies on Copernicus EO data & information
4. Establish a link with policy makers and data providers
5. Maximise the cross-fertilisation with the Big Data community

### **Task 5.2: Community & User Engagement, Period M1-M30, Lead ENEA (6PM), Contributors: CG (2PM), BSC (1PM), RAS (2.5PM), CMCC (2PM)**

Dissemination and Communication activities will be carried out, according to the plan developed in Task 5.1, during the entire duration of the project. The main target of those activities are the potential users of the developed framework, i.e. SMEs which businesses strongly relies on EO data, scientific centres and EO data providers, etc.

The activities are the following:

1. Dissemination of promotional materials (brochure, leaflets, roll-ups...)
2. Preparation of dissemination material targeted to specialised users (project brief, factsheets, newsletter)
3. Creation of a project strong visual identity
4. Creation of a project website
5. Attendance to events in both the EO and Big Data domains, including commercial fairs
6. Organization of workshops for a wide range of potential users

Two workshops held in Rome and Toulouse will provide a description of the developed framework/service with hand-on sessions for scientific users. Furthermore, a set of topics on BD and data-driven technologies will be presented. The workshop in Rome will be organised by ENEA at M13 and will be connected with the Data Driven Innovation Open Summit that is held every year in February to have the opportunity to engage also the SMEs and the start-up accelerators involved in the event. A high visibility workshop will

be organised by CG at the end of the project in Toulouse at M30 to present the final stage of the framework/service and the demonstrator. This workshop will involve audience from different domains and communities. The list of events and conferences related to the project is presented in section 2.2a.

**Task 5.3: Coordination with other initiatives, Period M1-M30, Lead ENEA (3PM) Contributors: BSC(3PM), RAS (2PM), GCS (1PM), CMCC(2PM)**

This task will aim at the creation of synergies with other European and global initiatives (a list is provided in the table at Section 2.2.a) with the aim to:

1. Establish a two-way communication link with the Copernicus initiatives on the implementation of sustainable industrial solutions and the Big Data challenges
2. Taking into account any emerging need from the Copernicus user community
3. Promote the project activities through the Research Data Alliance (RDA) working and interest groups
4. Coordinate the project activities with the Copernicus CDS, the Earth Server Federation, national Big Data Networks and the Earth System Grid Federation

**Task 5.4: Technology Transfer: Periods M9-M12, M21-M30- Lead: Green City Solutions (14PM), Contributors: KNMI (1PM), CG (1PM)**

Green City Solutions will conduct an analysis on the accuracy of the tool utilizing and comparing data and results gathered from measuring equipment from pilot projects with data from the ETCETERA framework. Data sets of a research project in Modena, Italy can be used for this purpose. A pilot project with 6 sensor implementations and scientific measuring equipment will be installed in April 2017 and gather data for a period of up to two years. Upon confirmation of the correlation and accuracy of both data sets different customer groups from the private and public sector will be approached and another survey will be conducted to establish a business model for services revolving around air quality.

The results from those surveys will be used to determine feasible services derived from the ETCETERA framework and to adapt the framework to the newly identified business models and the user interface, linking it with WP4 of the proposal. The intended outcome are the establishment of pilot projects with partners and customers from the private and public sector to apply the technology in real use cases and transfer them to service based business models to customers from those segments. KNMI DataLab will contribute to the exploitation plan for private sector by performing experiments that combine meteorological data, offered by DIAS, with Safety and/or Health data from public/private partners hence demonstrating the use and exploitation of DIAS. It is worth noting GCS effort for this task is greater than the work package leader effort since the main effort of GCS will be concentrate in this crucial task of the definition of the technology transfer.

**Deliverables** (brief description and month of delivery)

D5.1: Dissemination & Communication plan, **ENEA**, Task 5.1,M3

D5.2: Public Website, **CG**, Task 5.2,M2

D5.3: Report on the impact of Dissemination & Communication Activities, **ENEA**, Task 5.2,M30

D5.4: Business model based service, **GCS**, Task 5.4, M30



<b>Milestones</b> (brief description and month of delivery) M5.1: Workshop 1 in Rome, M13 M5.2: Workshop 2 in Toulouse, M30 <b>Dedicated KPI:</b> K5.1: Number of events attended per month, included in every Progress report K5.2: Number of organised workshops, included in every Progress report
--

**Table 3.1b:**     *List of work packages*

WP No	Work Package Title	Lead Participant No	Lead Participant Short Name	Total Person-Months	Start Month	End month
1	Management	1	CG	35	1	30
2	Data access	5	RAS	49	1	24
3	Analytics and Diagnostics	2	BSC	91,5	1	30
4	User Interface	1	CG	50,5	7	30
5	Dissemination	4	ENEA	44,5	1	30
				Total person-months	270,5	

**Table 3.1c:**     *List of Deliverables*

Deliverable	Deliverable name	Work package number	Leader	Type	Dissemination level	Delivery date (in months)
D1.1	Project management Plan	1	CG	R	CO	M2
D1.2i	Progress report	1	CG	R	PU	M6, M12, M18, M24
D1.3	Final Report	1	CG	R	PU	M30
D2.1	Architecture report	2	RAS	R	CO	M6
D2.2	Basic prototype	2	RAS	DEM	PU	M14

D2.3	Advanced prototype	2	RAS	DEM	PU	M22
D2.4	Consolidated prototype	2	RAS	DEM	PU	M30
D3.1	Project analysis and requirements report	3	BSC	R	PU	M6
D3.2	First pilot	3	BSC	DEM	PU	M14
D3.3	Advanced prototype	3	BSC	DEM	PU	M22
D3.4	Fully tested and validated prototype	3	BSC	DEM	PU	M30
D3.5i	Data management plan	3	BSC	R	PU	M6, M15, M30
D4.1	User Experience workshop outputs	4	CG	OTHER	PU	M15
D4.2	UI design documentation and source code	4	CG	R	PU	M15
D4.3	Working software	4	CG	OTHER	PU	M30
D5.1	Dissemination & Communication plan	5	ENEA	R	PU	M3
D5.2	Public Website	5	CG	DEC	PU	M2
D5.3	Dissemination & Communication Activities,	5	ENEA	R	PU	M30
D5.4	Business model service-based	5	GCS	R	CO	M30

In the table above, the abbreviations correspond to the following definitions:

Type:

R: Document, report (excluding the periodic and final reports)

DEM: Demonstrator, pilot, prototype, plan designs

DEC: Websites, patents filing, press & media actions, videos, etc.

OTHER: Software, technical diagram, etc.

Dissemination level:

PU = Public, fully open, e.g. web

CO = Confidential, restricted under conditions set out in Model Grant Agreement

CI = Classified, information as referred to in Commission Decision 2001/844/EC.

**Table 3.2a:**     *List of milestones*

Milestone number	Milestone name	Related work package	Due date	Means of verification
M1.1	Setup of the management structure	WP1	M2	Progress report
M2.1	Initial services	WP2	M6	Training, and planning accomplished
M2.2	Basic prototype	WP2	M14	New functionality at level “proof of concept”
M2.3	Advanced prototype	WP2	M22	New functionality at level “basically usable”
M2.4	Basic prototype	WP2	M30	New functionality at level “fully usable & documented”
M3.1	System design finalized	WP3	M6	Progress report
M3.2	First pilot	WP3	M14	Basic functionalities
M3.3	Advanced prototype	WP3	M22	Extended functionalities
M3.4	Fully tested and validated prototype	WP3	M30	First production version
M4.1	Wireframes of the display per user profile	WP4	M15	Number of wireframes
M4.2	UI design document	WP4	M24	Progress report
M4.3	Acceptance of the ETCETERA framework	WP4	M30	Framework released include in final report validated by EC
M5.1i	Workshops held in Rome, and Toulouse	WP5	M13, M30	Progress report

**Table 3.2b:**     *Critical risks for implementation*

### ***Risk Management***

This section presents the formal procedure to manage and control the project risks and the list of the identified risks. Risks to be addressed at every level are selected with respect to their respective criticality and manageability. An anticipated identification of the risks and the definition of mitigating actions will be part of every progress meeting or review.

*Note: L: Low, M: Medium, H: High*

	Description of risk	Level of likelihood	WP(s) involved	Proposed risk-mitigation measures
IT	Difficulties to deliver on time an operational framework & tools	L	WP4, WP5	Capgemini is a highly experienced industrial partner in distributed framework implementation using cloud and Big Data technologies. Besides, Capgemini proposes to mitigate this risk applying a set of procedures and processes that have a proven to be efficient on numerous of scientific or non-scientific ambitious and successful projects
	Technical performances not reached and not enough relevant in comparison with “classic” means	L	WP2, WP4	Iterative approach with Agile development method should raise earlier the potential technical performances. Many prototypes will be developed to identify the bottlenecks. Besides, a Technical Advisory Board (internal technical experts) will validate and control the technical choices and will give some recommendations
	Delay in DIAS provision	M	WP2, WP4	Access to non Copernicus (and already existing) platforms will be provided, allowing to performs diagnostics and create the user interface with non DIAS data.
	Delay in the implementation of the WP2 solutions, impacting WP3 and 4	L	WP2, WP3, WP4	Academic partners are recognized as experts and have a great experience in different domains and have proven to be groundbreaking actors. Moreover, the Technical Advisory Board will give recommendations to go beyond the state-of-the-art and improve the European Big Data skills.
Use cases	Availability of data sets to experiment the use case	L	WP2	A data availability schedule will be established to identify the due date and data needed along the project
	Disruption in data procurement (Sentinel, Air quality weather datasets) Less diagnostics made available (or less scientifically robust) in the WP3 and 4 because of lack of data	L		
	Not reaching end user requirements from scientific and tools usability for exploration and visualization	L	WP3, WP4	The organization and the Advisory Board representing end users are involved in workshops during the modules definition. The iterative development mode enables to collect feedback along the project life and to adjust and then mitigates the risk of failing to meet the expected requirements

<b>Other</b>	IPR and legal issues	L	WP1, WP5	During the project, the IPRs will be managed in WP1. The WP5 technology transfer will tackle this topic for the future exploitation. The legal issues will be managed by the Consortium agreement.
	General management issues	M	WP1	During the project, any conflict will be managed by the Consortium agreement

#### Definition critical risk:

*A critical risk is a plausible event or issue that could have a high adverse impact on the ability of the project to achieve its objectives.*

#### Tables for section 3.4

**Table 3.4a: Summary of staff effort**

The resources required to perform this project have been budgeted using a bottom up approach. The budget shows a balanced budget in terms of effort and costs. The use of resources is efficient to complete the required tasks successfully. The project's work plan has been broken down in work packages, tasks and individual work elements of each task. Each task has been time scheduled in details and each partner's effort has been budgeted in person hours per activity type. The result of these efforts is a very accurate planning of the partner's efforts needed to complete the project and achieve the objectives. With total effort of 270.5 person months spread over 30 months, and with 7 strong partners in the consortium, ETCETERA has the critical mass required to accomplish its goals. The following table summarizes the use of resources, indicating the number of person/months over the whole duration of the planned work, for each work package, and for each participant. The work-package leader for each WP is indicated by showing the relevant person-month figure in bold.

	WP1	WP2	WP3	WP4	WP5	Total Person-Months per Participant
<b>CG</b>	<b>25</b>	7	6	<b>30</b>	5	<b>73</b>
<b>BSC</b>		6	<b>39</b>	5	5	<b>55</b>
<b>KNMI</b>		6	10	5	1	<b>22</b>
<b>ENEA</b>		3	10		<b>10</b>	<b>23</b>
<b>RAS</b>		<b>30</b>			4,5	<b>34,5</b>
<b>GCS</b>			9	9,5	15	<b>33,5</b>
<b>CMCC</b>			21,5	4	4	<b>29,5</b>
<b>Total PM</b>	<b>25</b>	<b>52</b>	<b>95,5</b>	<b>53,5</b>	<b>44,5</b>	<b>270,5</b>

The following charts (figure 16) represent the use of resources per partner over the different work packages and showcases the balanced share of work between the partners.



Figure 16: Use of resources per partner and per work packages

More specifically, the resources foreseen are used as the following distribution: the project management (WP1 9%) led by Capgemini covers all managerial and coordination activities. The effort is kept limited thanks to the experience of the project coordinator and of the consortium as a whole, in particular the existing cohesion between the scientific partners in previous and current collaborations under H2020 and Copernicus projects. WP2 representing 19% ensures the data access in others terms, the setup of the core service platform, the integration of Copernicus and non Copernicus service as well as the support, training and testing. WP3 concentrating the main effort with 35%, deals with the effective diagnostics and analytics of atmospheric observations interface to perform analyses and interpretations calculations, including the state of the art of the existing frameworks, the setup of the engine for analytics and diagnostics and the users' needs definition. WP4 (20%) ensures integration and validations of the users cases in the final framework, including the operational and support aspects. WP5 (16%) properly support the dissemination and communication activities, the technology transfer via the exploitation plan for business sector via the definition of a business model service based.

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

The table below describes for each participant the sum of the costs for 'travel', 'equipment', and 'goods and services'

CG	Cost (€)	Justification
<b>Travel</b>	<b>6000</b>	2 persons per 3 project meeting (500€ per meeting)for a total of 3000 € and 2 persons in 3 other project/ UX workshops/dissemination meetings for 3000€
<b>Audits</b>	<b>7500</b>	Audits costs when EC funding is greater than 325 000€
<b>Total</b>	<b>13500</b>	

BSC	Cost (€)	Justification
<b>Travel</b>	<b>6000</b>	2 persons per 3 project meeting (500€ per meeting)for a total of 3000 € and 2 persons in 3 other project/ UX workshops/dissemination meetings for 3000€
<b>Equipment</b>	<b>14000</b>	BSC will provide storage infrastructure for the rasdaman-based services to be established in ETCETERA. Two sets of tapes (Data Cartridge 20 Pack

		HPE LTO7, 6TB per tape) are requested to provide cache storage for the data and to have space to work before the DIAS is available. BSC also asked for a fat node (SUPERSERVER 1U SYS-6018Ultra INTEL BROADWELL-EP 10C E5-2630V4, 256GB DDR4, T.RED 10GB 1P PCI-Eto, SATA 4TB) to compute the diagnostics in WP3 (more details in WP3 description online). Both the fat nodes and the tapes will be of course made available to the all consortium.
<b>Publications</b>	<b>2000</b>	Publication in journals such as Computational Geosciences, Environmental Modelling and software, Computer and Geosciences, Big Data Research.
<b>Audits</b>	<b>1500</b>	Audits costs when EC funding is greater than 325 000€.
<b>Total</b>	<b>23500</b>	

<b>KNMI</b>	<b>Cost (€)</b>	<b>Justification</b>
<b>Travel</b>	<b>6000</b>	2 persons per 3 project meeting (500€ per meeting)for a total of 3000 € and 2 persons in 3 other project/ UX workshops/dissemination meetings for 3000€
<b>Equipment</b>	<b>2000</b>	Contribution for the temporary storage and computational resources allocated on local VMs, for air quality sensor networks data and meteorological information for the realisation of the ETCETERA use case. Test-bed VMs for the evaluation and outreach tasks (WP5) of the DataLab. Initial storage resource hosting the provenance repository and services.
<b>Total</b>	<b>8000</b>	

<b>ENEA</b>	<b>Cost (€)</b>	<b>Justification</b>
<b>Travel</b>	<b>6000</b>	2 persons per 3 project meeting (500€ per meeting)for a total of 3000 € and 2 persons in 3 other project/ UX workshops/dissemination meetings for 3000€
<b>Equipment</b>	<b>6000</b>	Disks for storage Silicon Graphics IS5160 currently connected via double fibre channel with a server Silicon Graphics UV2000 (256 cores, 1TB dRAM). The equipment will to be used to store the data and as a cache for WP2 and WP3 activities.
<b>Other goods and services</b>	<b>11000</b>	The costs covering the organization of two dissemination workshops to be held in Rome and in Toulouse.
<b>Total</b>	<b>23000</b>	

<b>RAS</b>	<b>Cost (€)</b>	<b>Justification</b>
------------	-----------------	----------------------

<b>Travel</b>	<b>6000</b>	2 persons per 3 project meeting (500€ per meeting)for a total of 3000 € and 2 persons in 3 other project/ UX workshops/dissemination meetings for 3000€
<b>Other goods and services</b>	<b>2000</b>	Trade fair material (Brochures, leaflets, ...)
<b>Total</b>	<b>8000</b>	

<b>GCS</b>	<b>Cost (€)</b>	<b>Justification</b>
<b>Travel</b>	<b>6000</b>	2 persons per 3 project meeting (500€ per meeting)for a total of 3000 € and 2 persons in 3 other project/ UX workshops/dissemination meetings for 3000€
<b>Total</b>	<b>6000</b>	

<b>CMCC</b>	<b>Cost (€)</b>	<b>Justification</b>
<b>Travel</b>	<b>6000</b>	2 persons per 3 project meeting (500€ per meeting)for a total of 3000 € and 2 persons in 3 other project/ UX workshops/dissemination meetings for 3000€
<b>Total</b>	<b>6000</b>	



## Section 4: Members of the consortium

*This section is not covered by the page limit.*

### 4.1. Participants

This section provides with a description of each partners and its main tasks, with an explanation of how its profile matches the tasks in the proposal.

#### **Partner 1 Capgemini**

##### Description & Tasks assigned

Capgemini Technology Services France, in particular, the sector Capgemini Aerospace & Defense, with the two practices Custom Software Development and Insights & Data dedicated to the Space domain, are involved in the execution of the project. Capgemini has developed Aerospace & Defense Centres of Excellence (CoE) applied to the four sectors of Defense, Air Traffic Management and Space & Navigation with more than 4,000 engineers working on engineering services across main locations in France, Germany, UK, Italy, USA, Canada, Sweden, Netherlands, India and China. The Capgemini Aerospace & Defense CoE can rely on a highly talented group of specialists with varied skill sets from different disciplines in the AeroSpace domain and an average of seven years of experience. Capgemini invested many proofs of concept to demonstrate the benefit of the big data technologies and the added value in Space domain for the data processing in term of performance, scalability and flexibility of the solution. An internal initiative is being industrializing (ESA GEORICE project) and will bring to the market soon: Insights and Big Data as a service. This platform establish strong links with experts of analytics and with the innovative big data players such EMC, Teradata, SAS, Cloudera or Pivotal. As a leading provider of IT and data services, Capgemini has the expertise and experience in delivering big data projects. Open data, along with big data and the actionable insights that managed analytics can bring is at the heart of Capgemini's priorities. Capgemini has strong skills about software development. It encompasses all kind of softwares: data processing chains, mission-critical systems and "Digital Customer Experience" websites for collaborative systems, portals, social networks, mobiles applications.

**Regarding Big data and Cloud technologies**, Capgemini assets could be grouped according to the following categories:

- 1) Data hub and processing engines: several project and benchmarks have been performed, especially around the Hadoop ecosystem. Capgemini has been experimenting these solutions for spatial and more generally science applications since 2011 for 1 Sentinel-1 image for instance.
- 2) Cloud administration to manage central and local(VMs) environments for operational services or collaborative development environment sharing resources.
- 3) Common services: Capgemini team has implemented some proofs of concept by using streamed events or Data Hub Software (DHuS) Sentinel ingestion.

Capgemini brings together leading data technology experts with business and industry sector expertise to help get better value from both corporate and external data. The company knows the opportunities presented by new big data sets and help its customer to channel these to ensure him stay one step ahead. Capgemini can also help to take control of Big Data and transform the information landscape in a way that delivers continual value and where the information is managed based on the value it delivers. Working directly with the engineering teams of some of the software industry's most innovative startups to drive a more consistent managed approach to the big data ecosystem, the company has completed many big data projects successfully, including:

- Helping a global employment agency to improve their insight into localized job markets, matching job seekers with openings. Users have been amazed by the accuracy of our proof of concept.
- Improving a global Telco insight into network usage and identified new commercial opportunities through Big Data transformation.

- Fostering a Big Data and Predictive Analytics Help HMRC Process Debt Payments More Quickly  
More references on Big Data here<sup>1</sup>.

**Regarding the scientific aspect**, Capgemini has a good knowledge of the science issues through the Scientific Office (SO), a pool of 20 PhD researchers on physics, mathematics and engineers. The SO ensures the operational implementation of scientific results in close collaborations with research centres. The innovation goal relies on the improvement of Big Data technologies and the framework to integrate scientific algorithms, in a Cloud infrastructure. The Scientific Office develops data processing algorithms for Earth Observation (radar, optics, with Sentinel-1, Landsat, Radarsat, Envisat, SPOT, Pléiades, MERIS, SWOT, etc.) and carry out data analysis from other domain (industry, aeronautic, etc.), define models to discover and extract added value from huge datasets; Regarding data science, Capgemini has accumulated an expertise both with “basic” machine learning algorithms (classification, clustering, regression, recommender systems, etc.) and with advanced algorithms (eg: neural networks, advanced optimization, descriptive and predictive analysis) that are designed by the mathematicians from the Scientific Office. Since 2015, Capgemini has been invested in its own **Big Data platform dedicated to Earth Observation** named **Tech4Earth**. This platform is the first step of the platform Insights & Data as a service for the applications handled space data. Capgemini demonstrates its high commitment to address this challenge in collaboration with our clients and partners through the industrialized developed solutions. Over a 30 years period, Capgemini has been consolidated its expertise based on a consistent high quality delivery services from our Center of Excellence in the area of Earth Observation and Navigation. Our Space & Navigation Center of Excellence has built over the past 30 years a truly impressive track record with many hundreds of successful projects in collaboration with our partners across Europe. Capgemini’s financial strength, company size and its presence in all ESA member states, our 300 experienced engineers, project & program managers with 20-years experience of managing complex space related projects, our Scientific Office and our industrialized ways of working, consolidate Capgemini’s pole position to manage complex, cross-border projects more easily, effectively and more efficient. Furthermore, Capgemini has been running the European Data Portal, on behalf of DG CONNECT, thereby demonstrating its capacity to work cross-border, cross-organisation in the context of platform development and user engagement.

List of previous projects and activities connected to the subject of this proposal:

**Fully Automated Aqua Processing Service (FAAPS)** funded by ESA is a service in near real time for the flooding monitoring using space data and crowdsourcing information. An additional flood risk prediction service crossing some soil moisture measured by the satellite SMOS and precipitation forecasts has been added in collaboration with a French research laboratory (CESBIO),

**GEORICE (ESA, 2014-2016)** is a dedicated service providing to the rice monitoring stakeholders’ information and products developed using Copernicus Sentinel-1 4 data. In parallel, Capgemini has developed various demonstrators including for the marine and coastal pollution monitoring using optical satellite images using Copernicus Sentinel-2 data

**EU open data core platform (H2020, 2014-2017):** In addition, in his strategy for the big data technology, Capgemini has been awarded a contract for the deployment of an **EU open data core platform**. The goal of this project is to develop, launch and operated a single large-scale pan-European Open Data Portal which bring together public data resources from all over Europe in one portal covering all European countries, improving the discoverability and the ability to re-use this data.

**MOONSON (H2020, 2016-2019)-** MOdel-based coNtrol framework for Site-wide OptimizatiON of data-intensive processes - aims to establish data-driven methodology to support identification and exploitation of

---

<sup>1</sup> <https://www.emc.com/collateral/partner-solution-overview/idg-brief-capgemini-emc-turning-big-data-into-big-opportunities.pdf>

optimization potentials by applying model-based predictive controls so as to perform plant and site-wide optimization of production process. The ambition of MONSOON project is shared by 2 significant process industries from the sectors of aluminium and plastic.

In the present proposal, Capgemini will be take care of the general management and coordination with EC as leader of the consortium under WP1. Capgemini is also involved in tasks related for the data access protocol under work package WP2, to ensure the whole consistency with the activities defining the user interface under the WP4, where acting as leader. Capgemini is involved as well in the dissemination and impact activities facilitating the link with other activities and communication actions.

#### Qualification of key personnel

##### **Dr. Florence Daverdon (female)**

Holding a PhD in Applied Mathematics in the Bordeaux University (1992), Florence has been working for the past 20 years on various Space & Science related projects. Since she has a vast experience in complex and international projects, Dr. Florence Daverdon will endorse the role of project following closely the project activities and will ensure the highest quality of the final delivery.

**Ms. Lee Cerchiari** (female) Lee is web designer (or UI designer) with a more than 20 years' experience. Lee has worked on many ambitious projects, designing and implementing the design of digitals interfaces . Lee contributions cover from graphic design to the formalization of interactions. Lee is Creative and innovative and as an UI designer Lee takes into account constraints of *ergonomics, usability and accessibility*.

##### **Mr. Roger Rutakaza Maneno (male)**

Roger is a senior IT Architect accumulated 20 years of experience on Industries, Space and Aeronautic domains. He masters JEE (Java), RDBMS, SOA, and distributed systems and Micro services technologies. His experience comes from projects implemented for CNES, ESA, Airbus, Snecma and other major actors of industries, space and aeronautic domains. Roger will be in charge of leading the design and implementation of WP4.

**Dr. Carine Saüt (female)**: After receiving her PhD in Atmospheric Chemistry in the University of Toulouse (France) in 2003, Carine then worked as a post-doctoral fellowship researcher at Harvard University (USA). Then for the past 10 years, she has fully orientated her career towards the European R&D project management. Now, she is currently business developer for science & space applications in the European context with a special focus on H2020 projects and Copernicus frameworks. She will be backup Dr Florence Daverdon, Carine will be available to take over the project management tasks at any time should a problem occur or additional human resources will be needed at some specific point to manage the ETCETERA project.

## Partner 2 BSC

### Description & Tasks assigned

**The Barcelona Supercomputing Center (BSC)** was established in 2005 and 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-CNS 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. The Center houses MareNostrum, one of the most powerful supercomputers in Europe with 48,128 cores and 1.1 Pflops capacity. The mission of BSC is to research, develop and manage information technologies in order to facilitate scientific progress. BSC combines HPC service provision, and R&D into both computer and computational science (life, earth and engineering sciences) under one roof and currently has over 450 staff from 44 countries. BSC has collaborated with industry since its creation, and participates in various bilateral joint research centers with companies such as IBM, Microsoft, Intel, NVIDIA and Spanish oil company Repsol. The centre has been extremely active in the EC Framework Programmes and has participated in over 100 projects funded by it. BSC is a founding member of HiPEAC, the ETP4HPC and other international fora. The ES-BSC activities with the focus on global climate modelling and prediction are based on research, development and predictions with the EC-Earth climate forecast system. EC-Earth is the state-of-the art coupled climate model that is being developed and used for climate predictions and projections by the European consortium of more than 20 research and operational institutions from European Centre for Mid-range weather Forecasts (ECMWF is provider of the atmospheric and land components) to ES-BSC. Beside contributing to the 5th phase of the Coupled Model Intercomparison Project (CMIP5) critical for the UN IPCC Fifth Assessment Report (AR5), global climate research activities at ES-BSC enable provision of various historical reconstructions and initial conditions to the EC-Earth community for analysis of climate dynamics and for seasonal to decadal climate predictions. The ES-BSC is a contributor to the IS- ENES FP7 European project fostering the integration of the European climate modelling community and the development of Earth System Models (ESM) for advancing the understanding and predictions of climate variability and change. The ES-BSC is already active in the planning and design of the future coupled climate model intercomparison project, CMIP6, and is preparing to make key contributions including the groundbreaking high-resolution climate simulations with EC-Earth.

### List of previous projects and activities connected to the subject of this proposal

**COPERNICUS project QA4Seas - *Quality Assessment Strategies for Multi-model Seasonal Forecasts*:** BSC is the leader of this project which is aimed at developing a strategy for the evaluation and quality control (EQC) of the multi-model seasonal forecasts provided by the Copernicus Climate Change Service (C3S) to respond to the needs identified among a wide range of stakeholders. The quality assessment will be user driven and will put at work the best expertise available on the evaluation of the multi-faceted quality aspects of state-of-the-art seasonal forecast systems.

**COPERNICUS project MAGIC - *Metrics and Access to Global Indices for Climate Projections*:** this solution consists of extending an existing web portal designed for analysing climate data (climate4impact.eu) with the capability to manipulate and combine the data according to the specifications set forth in the ITT. The necessary software for these manipulations will mainly be taken from existing tools that have been developed by the partners. The climate4impact.eu web site will be equipped with an interface that accepts commands from the user and displays the results. The system design allows interaction with different data stores and (web) interfaces, and will therefore act as a hub both for the SIS project portals to collect tailored information from the CDS, but also for expert users that do not need the intermediate step of the SIS.

**COPERNICUS project CLIM4ENERGY** will bring together the complementary expertise of 7 climate research and service centers and 11 energy practitioners to demonstrate, from case studies, the value chain from ECVs to actionable information in the energy sector. It will deliver 9 energy-relevant pan-European indicators of climate trends and variability with a cross sectoral consistency, appropriate documentation and guidance, estimation of uncertainties, and a demonstration of use.

**COPERNICUS project CAMS 84-Global and regional a posteriori validation, including focus on the Arctic and Mediterranean areas:** The main outputs of CAMS-84 are a series of 3-monthly validation reports for the near-real-time global products and for those from experimental chains running in parallel. Validation activities also provide a set of plots and quality monitoring statistics, which are integrated in the CAMS website and are easily accessible together with the products from the CAMS catalogue.

**H2020 project ESIWACE Excellence in Simulation of Weather and Climate in Europe (GA 675191):** will substantially improve efficiency and productivity of numerical weather and climate simulation on high performance computing platforms by supporting the end-to-end workflow of global Earth system modelling in HPC environment. This will be obtained by improving and supporting (1) scalability of models, tools and data management on state-of-the-art supercomputer systems (2) Usability of models and tools throughout the European HPC ecosystem, and (3) the Exploitability of the huge amount of resulting data.

#### List of products connected to the subject of this proposal

**CALIOPE<sup>2 3</sup>**(Air Quality Forecast System): Caliope, an air quality forecast system, is also a mobile application which has been developed thanks to support received through the MYGEOSS project, as part of which it received an award as one of the best environmental applications using open data.

**WMO Dust Centers:** Sand and Dust Storm Warning Advisory and Assessment System (SDS-WAS), R+D project operated by AEMET and BSC under the umbrella of WMO to study R+SDS modelling and Barcelona Dust Forecast Center<sup>4</sup> (BDFC)

**EUDAT** (EUropean DATa infrastructure), BSC presents as service provider<sup>5</sup> and as use case pilot<sup>6</sup>

**RDA** (Research Data Alliance): BSC is leader of the Interest Group on Weather, Climate and Air Quality.

#### Qualification of key personnel:

**Francesco Benincasa** (male) holds a Master's Degree in "Software Engineering" from the "Alma Mater Studiorum - Bologna University" (Italy) in 2004. 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, 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. He is in charge of data processing and management and web development and maintenance of the WMO SDS-WAS NA-ME-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

---

<sup>2</sup><http://digitalearthlab.jrc.ec.europa.eu/activities/mygeoss-applications-your-environment/57752>

<sup>3</sup> <https://www.bsc.es/es/news/bsc-news/caliope-eu-application-displays-air-pollution-forecasts-throughout-europe>

<sup>4</sup> the first SDS forecast operational system officially supported by WMO: <http://dust.aemet.es/>

<sup>5</sup> <https://www.eudat.eu/eudat-cdi/partners>

<sup>6</sup> <https://www.eudat.eu/communities/support-to-scientific-research-on-seasonal-to-decadal-climate-and-air-quality-modelling>

mediterranean area (BDFC). He is also chairman of the "weather, climate and air quality Interest Group" inside the Research Data Alliance (RDA).

**Pierre-Antoine Bretonnière** (male), holds a Masters Degree in "Mathematical and Mechanical Modelling" from the Matmeca engineer school in Bordeaux (France). Graduated in 2010, he has worked in several climate research institutes (CERFACS - Toulouse - France, Catalan Institute of Climate Sciences - Barcelona - Spain and the Earth Sciences Department of the Barcelona Supercomputing Center). His work focuses on climate models outputs, data management and model coupling. He was the person in charge of the data management plan and data conventions definitions in the SPECS FP7 project and has participated in several other European projects. He is also involved in the Research Data Alliance (RDA) framework as chairman of the "Weather, climate and air quality" interest group.

**Alicia Sánchez Lorente** (female), holds a PhD in "High-Precision Gamma and X-Ray Spectroscopy" from the Johannes Gutenberg University, Mainz (Germany). After working at the Helmholtz Institute Mainz (Germany) in the Nuclear Physics Division as a research manager, she joined the Barcelona Supercomputing Center Earth-Sciences co-leading the computational Earth Sciences group. She has a long experience in developing computational techniques as well as data analysis algorithms to deal with huge amount of data (Big Data) and improve Signal-to-Noise Ratio. She has participated on several international Scientific Collaborations worldwide and has helped with her interdisciplinary knowledge to consolidate important european research projects within the framework of High Precision Energy technologies. Currently, her interest focuses on developing Big Data and Machine Learning techniques in the field of earth sciences.

#### **Infrastructure:**

BSC-CNS is the National Supercomputing Facility of Spain and hosts a range of high-performance computing (HPC) systems including MareNostrum IV the new supercomputer, will be 12.4 times more powerful than the current MareNostrum 3 that will have a performance capacity of 13, 7 Petaflop/s. The general purpose element, will have 48 racks with more than 3,400 nodes with next generation Intel Xeon processors and a central memory of 390 Terabytes. The second element of MareNostrum 4 will be 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 US and Japan to accelerate the arrival of the new generation of pre-exascale supercomputers.

The BSC-CNS 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-CNS 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.

## Partner 3 KNMI

### Description & Tasks assigned

The Royal Netherlands Meteorological Institute ( KNMI ) is the national research and information centre for weather, seismology, climate and climate change in the Netherlands. KNMI has a long tradition in operational and scientific activities. As an integral part of the Ministry of Infrastructure and Environment, KNMI provides on a day-to-day (24/7) basis advice on weather, seismological events and climate to national, regional and local authorities. KNMI is participating in many European projects on climate, weather, seismological and space research, and keeps close ties with many of its stakeholders. KNMI produces climate scenarios to support stakeholders for developing adaptation and mitigation strategies. KNMI has initiated the development of the global climate model EC-Earth and hosts one of the nodes of the ESGF consortium, contributing to the collection of simulations and observational data for climate change research. The R&D activities conducted within the institute in the field of Data-Technologies have produced tools and services which are adopted across different disciplines in Earth Science. The Data-Technology team will contribute to the project with their expertise in data visualisation, adoption and implementation of OGC and metadata standards, integration of workflows for systems orchestration and data-intensive applications and data provenance. Moreover, KNMI will provide access to the ESGF data hosted by the local node and support training and outreach activities towards related Copernicus projects and external stakeholders.

### Relevant previous projects or activities:

**CLIPC** ([www.clipc.eu](http://www.clipc.eu)) : Climate Information Platform for Copernicus (CLIPC). The Climate Information Platform for Copernicus (CLIPC) project has developed an integrated web- platform of Climate Data Services to provide a single point of access for authoritative scientific data and information on climate variability and change, and the impacts of these. CLIPC supports the Copernicus Climate Change Services (C3S), which will deliver the next generation of climate and climate impacts data for Europe's citizens.

**CAMS-84** (<https://atmosphere.copernicus.eu>): Copernicus project CAMS84, Global and regional a posteriori validation, including focus on the Arctic and Mediterranean areas. CAMS-84 is a sub-project of CAMS, dealing with the validation of the services. CAMS-84 provides 3-monthly updates of validation reports for the global and regional services.

**C3S-MAGIC** (ongoing): Copernicus project on climate: Metrics and Access to Global Indices for Climate Projections. The requested output of this ITT is software to calculate and present metrics, statistics, time series, and some tailored applications from climate model data.

**EPOS<sup>7</sup>** Research Infrastructure and E-Science for Data and Observatories on Earthquakes, Volcanoes, Surface Dynamics and Tectonics. EPOS aims at creating a pan-European infrastructure for solid Earth science to support a safe and sustainable society. EPOS will enable innovative multidisciplinary research for a better understanding of the Earth's physical and chemical processes that control earthquakes, volcanic eruptions, ground instability, tsunami, and all those processes driving tectonics and Earth's surface dynamics.

**VERCE<sup>8</sup>**: Virtual Earthquake and seismology Research Community e-science environment in Europe (). VERCE aims at supporting the computational seismology community by developing a data-intensive e-science environment to enable innovative data analysis and data modeling methods that fully exploit the increasing wealth of open data generated by the observational and monitoring systems of the global seismological community.

---

<sup>7</sup> <http://www.epos-eu.org>

<sup>8</sup> <http://verce.eu>



#### List of products connected to the subject of this proposal

##### **Climate4Impact<sup>9</sup>:**

Climate4impact portal is the result of a collaborative effort within the European FP-7 project IS-ENES2 an effort directed to Bridging Climate Research Data and the Needs of the Impact Community. The aim of Climate4impact is to enhance the use of research data and to support other climate portals. It has been developed within the European projects IS-ENES, IS-ENES2 and CLIPC. Climate4impact is connected to the Earth System Grid Federation, using certificate based authentication, ESGF search, openid, opendap and downloading datasets.

##### **ADAGUC<sup>10</sup>:**

ADAGUC is a geographical information system to visualize netCDF files via the web. The software consists of a server side C++ application and a client side JavaScript application. The software provides several features to access and visualize data over the web, it uses OGC standards for data dissemination.

#### Qualification of key personnel:

**Alessandro Spinuso** is a researcher at the R&D Observations and Data Technology division of the KNMI (Royal Netherlands Meteorological Institute). He is involved in a number of international initiatives focusing on the realisation of e-science infrastructures for Climate and Solid Earth Science Research (CLIPC, VERCE, EPOS ). At the moment, he is responsible for the Computational Earth Science activities conducted by the EPOS (European Plate Observing System) project and collaborates in the investigation of scalable and provenance aware Web Processing Services (WPS) for climate studies.

**Andrej Mihajlovski (MSc)** is a skilled robotics software engineer with years of experience in Java development. After his work in commercial trading, developing trading algorithms at several companies, he decided to go back to his research roots. He is now working at KNMI, developing OGC services for Copernicus and EU projects.

**Bas Mijling (male)** is a specialist in algorithm development and data analysis of air pollutant measurements. He received his Ph.D in 2012 for developing an advanced algorithm that combines satellite data with state-of-the-art atmospheric models to calculate emissions from known and unknown sources. His research now focuses on new data-assimilation techniques of these alternative networks in high resolution urban air quality models.

---

<sup>9</sup> [www.climate4impact.eu](http://www.climate4impact.eu)

<sup>10</sup> [www.adaguc.knmi.nl](http://www.adaguc.knmi.nl)



## Partner 4 ENEA

### Description & Tasks assigned

ENEA is the Italian government agency responsible for the areas of new technology, energy and the sustainable economy. Its two fundamental tasks are to conduct research in these areas and to diffuse the results nationally. ENEA's activities in the Environmental sector involve: Environmental surveying and monitoring; Climate modelling and analysis at global and regional scale; Global change assessment; Research and assessment of the impact of productive activities on the human and natural environments; Development of advanced technologies and new products with low environmental impact. The Agency has around 3000 staff throughout Italy, operating at nine major Research Centres and a number of smaller facilities. ENEA has a special laboratory devoted to Climate Modelling and Impacts (SSPT-MET-CLIM) with a staff of 15 employees. This laboratory combines long standing expertise both in the area of oceanic and atmospheric modelling (regional and global) and impact evaluation in relevant sectors (energy, infrastructure, agriculture, ecosystems). Its major fields of research are Mediterranean and African climate, the analysis of the regional hydrological cycle and related teleconnections, energy and environmental modelling, forecasting applications for renewable energy, electric grid, infrastructure maintenance and planning. ENEA participates in the many large international networks/ initiatives, e.g. EERA (European Energy Research Alliance), ECRA (European Climate Research Alliance), European Energy Network, MEDENER (Mediterranean Association of the National Agencies for Energy Conservation), TAFTIE (The Association For Technology Implementation in Europe), Enterprise Europe Network. In the last five years, the SSPT-MET-CLIM laboratory coordinated the EU-FP7 CLIM-RUN project on climate services, and it has contributed to IMPACT2C, SPECS and EUPORIAS, SINGULAR, PERSEUS EU-FP7 projects and the COPERNICUS SIS ECEM. It is now participating to CRESCENDO H2020 initiative and to the national program Next-Data to build a national system for the retrieval, storage, access and diffusion of environmental and climate data from mountain and marine areas This laboratory co-leads and participates to Med-CORDEX initiative, hosting the Med-CORDEX database on ENEA dedicated servers

### Qualification of key personnel:

**Dr. Matteo De Felice (male)** he is currently staff scientist at the Climate Impacts and Modelling laboratory in the ENEA Research Centre in Bologna. He has a Ph.D. in Informatics and Automation, which revolved around the application of nature-inspired optimisation and modelling methods (namely evolutionary algorithms or ensemble of neural networks) to energy-related problems. In 2011, he moved to his current laboratory, and since then he has been taking advantage of my background in data-driven modelling and machine learning to deal with several real-world applications where weather and climate play an important role. He was the supervisor of the ENEA service for the Dispatching Office of Terna, the Italian TSO, and he is the ENEA supervisor for the C3S ECEM project. He was also involved in CLIM-RUN, SPECS and EUPORIAS FP7 projects.

**Dr. Massimo D'Isidoro (male)** obtained a degree in physics from University of Bologna in 1998. In 2003 he received his PhD in physics from the University of Ferrara. His research interests focus on mesoscale atmospheric modelling, including transport and dispersion of atmospheric pollutants, and data assimilation techniques in atmospheric models. He has been involved in various international projects (WMO-COMPARE-III, EU-STOWASUS2100, EU-FP6-GEMS, EU-FP7-MACC; EU-FP7-CITYZEN) as well as in national projects in Italy (ROSA, CLIMAGRI, MINNI). Since 2010 he is a researcher at ENEA, Air Pollution Laboratory.

### List of previous projects and activities connected to the subject of this proposal

**COPERNICUS project ECEM – European Climatic Energy Mixes** ENEA is partner of this project which objective is to enable the energy industry and policy makers to assess how well energy supply will meet demand in Europe over different time horizons, focusing on the role climate has on energy supply and demand. It will develop, in close collaboration with the energy sectors, a demonstrator that will offer a coherent approach for the climate variables/indicators used in power demand/supply balance, an added value with respect to current practice in the sector, where climate data and derived ESCIIs are not always physically homogeneous and/or in balance.

**FP7 SPECS – Seasonal to decadal climate prediction for the improvement of European Climate Services:** ENEA was partner of this project and coordinator of the pilot activities. The project analysed the impact of improved climate information, provided by climate forecasts, on the energy and agriculture sectors.

**FP7 MACC - Monitoring Atmospheric Composition and Climate:** Paving the way for the future GMES Atmospheric Service, MACC combines computer model simulations with world-wide observations to monitor the composition of the Earth's atmosphere and predict regional air quality.

**FP7 CityZen:** CityZen aims at determining the air pollution distribution and change in and around hotspots over the last decade from extensive satellite and in-situ observations and employs a series of different scale models to analyze the impacts of air pollution hot spots on regional and global air quality, including potential future changes for various climate scenarios.

## Partner 5 RASDAMAN GmbH

### Description

Research spinoff rasdaman GmbH has been established for commercializing the rasdaman technology which effectively has pioneered Array Databases and is world technology leader in this domain. In 2008, rasdaman GmbH has disclosed large parts of its product to establish the open-source rasdaman community version in parallel to the commercial rasdaman enterprise version, adopting a dual-licensing model. Downward compatible rasdaman enterprise comprises highly effective performance boosters, extra functionality, and convenience utilities. For both versions the company offers support, development, and maintenance. The rasdaman technology has received numerous innovation awards; since 2016, it is incorporated in the top Big Data technologies list maintained by US magazine CIO Review.

Tasks assigned WP2 leader and WP2 technical contributor (scalable array services) WP5: standardization and further outreach (RDA), exploitation planning

### List of previous projects and activities connected to the subject of this proposal

EarthServer-2 (EU H2020 INFRA)

BigPicture (German Ministry for Food and Agriculture)

EVO-ODAS (ESA)

EarthServer (EU FP7 INFRA, finished)

Viseo-Fusion (ESA, finished)

PublicaMundi (EU FP7 ICT, finished)

### List of publications or services

- Sample services running rasdaman: <http://eodataservice.org>, <http://earthserver.ecmwf.int/>, <http://www.planetserver.eu>
- Software: rasdaman, [www.rasdaman.org](http://www.rasdaman.org), [www.rasdaman.com](http://www.rasdaman.com); selected awards and appreciations:
  - ESA Senior Manager Guenther Landgraf: "The RASDAMAN product is currently the world leading environment in this domain and the standard working horse for OGC standardisation on these innovative data access interfaces."
  - US Magazine CIO Review: rasdaman Top Big Data Solution
  - Winner, Big Data Challenge, Copernicus Masters 2014
  - OGC Kenneth Gardels Award 2014 for "extraordinary contribution in significant and enduring advances in technical standards"
  - NASA WorldWind Europe Challenge winner, 2104
  - Geospatial World Forum Innovation Award 2013

### Qualification of key personnel:

**Peter Baumann (male)** is founder and CEO of rasdaman GmbH. He holds a PhD degree in Computer Science from TU Darmstadt, Germany and is Professor of Computer Science at Jacobs University, Germany. He is Principal Architect of the rasdaman technology and has ample experience in international project research and management. Peter Baumann is active, often leading contributor to standardization in the Open Geospatial Consortium (OGC) and ISO bodies, being editor of a series of adopted international standards. In OGC he is chairing "Big Geo Data" working groups and is editor of 12 adopted standards, among them the WCS suite. In ISO, he is working on the forthcoming SQL / MDA standard which extends SQL with n-D arrays, and on modernizing the ISO Big Geo Data standards. In the Research Data Alliance (RDA), he co-chairs the Big Data and Geospatial Interest Groups as well as the Array Database Assessment Working Group. In 2014, OGC has honored his contribution to Big Data standardization with the prestigious Kenneth Gardels Award.

**Heike Hoenig** (female) is Director, Marketing at rasdaman GmbH. She holds a Magister in Germanistics from Oldenburg University and has 20+ years of professional experience in journalism, PR and marketing, e.g., in doing market studies. In EU EarthServer, she has performed all outreach work for rasdaman GmbH, such as trade fairs, newsletters, and social media. Further, she has played a key role in the concept of the 1h TV documentary “Big Earth Data” done in EarthServer. Based on this success she is leading social media outreach and exploitation tasks in EarthServer-2.

**Dimitar Misev** (male) is Head of Product Development and a substantial code contributor to both rasdaman community and enterprise. He holds an MSc degree in CS from Jacobs University. In parallel to his employment with rasdaman GmbH he is pursuing a PhD in CS.

## Partner 6 Green City Solutions

### Description & Tasks assigned

Green City Solutions was founded in April 2014 in Dresden Germany, out of a research project in collaboration with the Technical University of Dresden and the University of Applied Sciences Dresden. The vision of the company is to mitigate climate change and fight air pollution through a highly intelligent and profitable climate infrastructure for liveable cities of tomorrow. For this purpose Green City Solutions created the CityTree; a 4 meter high 3m wide and 60cm deep, freestanding unit, which contains specific moss cultures that eats particulate matter (PM), nitrogen dioxide and ozone – together offsetting 240 t CO<sub>2</sub>e/ year. The construction contains smart sensors collecting environmental as well as climatic data, to regulate and control the unit and ensure that moss culture survives. Thus, the CityTree has the same effect as 275 normal urban trees, but requires 99% less space and 90% less investment. Every CityTree has a capacity to clean the air in proximity of up to 50m. To further increase the efficiency of the units, Green City Solutions conducts CFD (computational fluid dynamics) analysis based on data of pollution levels in street environment in a 3D model. The particle flow is observed in a simulation and gets compared to the actual data gathered by the CityTrees. In the beginning of 2016 Green City Solutions was selected as the first German start-up to participate in the Open Data Incubator for Europe (ODINE). The project stretched over 6 months and was aiming to use open data datasets to analyse and establish correlations of air pollution to extreme climatic conditions and locations. With this information a particle flow simulation could be conducted to calculate the filter and adaptation potential of the CityTree, and in a next step visualize the effect of every single CityTree on its surroundings with a web application. Green City Solutions' role in the project will primarily focus on the use case definitions in combination with diagnostics of the data from a scientific view and an end-user view (WP3). They will focus on the user impact of the use cases stretching from municipalities to companies and even citizens as existing customers of their services (WP5).

### Qualification of key personnel:

**Zhengliang Wu** (male) (CIO)- Business Administration and IT, Work experience in Germany & China, Focus: Controlling & IT.

**Malgorzata Olesiewicz** (female) (Senior Manager Business Development) : First class MA in Economics at the University of Aberdeen (UK) , two-year work experience in low carbon transition consultancy, specialised in market research and economic modelling. Work experience: Pale Blue Dot Energy, Aberdeen Centre for Research in Energy Economics and Finance.

## Partner 7 CMCC

### Description & Tasks assigned

The Euro-Mediterranean Center on Climate Change Foundation (CMCC Foundation - <http://www.cmcc.it/>) is a non-profit research institution funded in 2005 with the financial support of the Italian Ministry of Education, University and Research, that aims at establishing a Center of excellence focused on integrated study of climate change related topics in Italy. It involves and links private and public institutions jointly investigating multidisciplinary topics related to climate science research in order to provide full analyses of climate impacts on various systems such as agriculture, ecosystems, coasts, water resources, health, and economics. Thanks to its networked structure, CMCC Foundation brings together highly qualified experts/professionals from different climate research areas in a single unique institution to develop a global and extensive approach to climate research including: mathematics, physics, economics and computer science. The following eight research Divisions work together in an interdisciplinary manner: ASC (Advanced Scientific Computing), CSP (Climate Simulation and Prediction), ECIP (Economic Analysis of Impact and Policy), IAFES (Impacts on Agriculture, Forests and Ecosystems Services), ODA (Ocean Modelling and Data Assimilation), OPA (Ocean Predictions and Applications), RAS (Risk Assessment and Adaptation Strategies), REMHI (Regional Models and Hydrogeological Impacts). The Advanced Scientific Computing (ASC) Division of CMCC Foundation carries out Research & Development activities on Computational Sciences applied to Climate Change. In particular, the Division works both on the optimization of numerical models on HPC architectures and the management of big volumes of scientific data looking forward at exascale scenarios. CMCC Foundation is also member of the European Network for Earth System Modelling (ENES) and partner of the Earth System Grid Federation (ESGF), providing access to 100TB CMIP5 data through its data node deployed at the CMCC Supercomputing Center. Finally, CMCC Foundation is partner of several European and other national and international projects, working on scientific data management, Science gateways, server-side data analysis and PaaS for data analytics. The HPC infrastructure managed at the CMCC Supercomputing Centre is composed of a 960 cores IBM Power6 cluster (peak performance 18 TFlops) and a 8000 cores Intel Xeon Sandy Bridge (peak performance 160 TFlops). The total system provides 180 Tflops of computing power, 1.2 Petabyte on-line storage and 3 PetaBytes Archiving capacity. Part of this infrastructure (1 cluster running 60 cores, 30TB storage, 1 TB RAM) will be committed to this project for testing/validation purposes regarding the big data solution to be implemented in the project.

### Qualification of key personnel:

**Sandro Fiore** (male) Ph.D., is the Director of the Advanced Scientific Computing (ASC) Division of the Euro-Mediterranean Centre on Climate Change. His research activities focus on parallel, distributed, grid and cloud computing, in particular on distributed data management, data analytics/mining and high performance database management. He is Visiting Scientist at Lawrence Livermore National Laboratory (LLNL) working at PCMDI in the context of the Earth System Grid Federation (ESGF). Since 2004, he has been involved into several national and international projects like: EGEE (the 3 cycles), EGI-InSPIRE, IS-ENES1 and IS-ENES2, EUBRAZILCC, ExArch, ORIENTGATE, TESSA, OPHIDIA, CLIP-C, INDIGO-DataCloud, ESIWACE working on data management topics. Since 2010, he is the Principal Investigator of the Ophidia project, a research project on high performance data analytics and mining for eScience. He is author and co-author of more than 50 papers in refereed books/journals/proceedings on distributed and grid computing and holds a patent on data management topics. He is editor of the book “Grid and Cloud Database Management” (Springer, 2011). He is ACM Member.

**Donatello Elia** (male), is a post-degree at the Euro-Mediterranean Center for Climate Change. He is involved in scientific data management research activities with special regard to big data analytics, parallel and cloud computing. Currently he is involved in the EU FP7 EUBrazilCC project, working on the development of a

PaaS framework for parallel data analytics (Ophidia) as well as on the design and implementation of a Science Gateway for climate change and biodiversity. Furthermore, he has working on tool for infrastructure monitoring (Cacti). He has a good knowledge on IaaS middleware (OpenNebula, OpenStack) and interfaces (OCCI, CDMI).

List of previous projects and activities connected to the subject of this proposal

**EU H2020 Project ESIWACE (Centre of Excellence in Simulation of Weather and Climate in Europe; 2015-2019)** ESIWACE is a user-driven Centre of Excellence in Simulation of Climate and Weather in Europe. In particular, CMCC Foundation participates to several activities including: enhancing community capacity in HPC; scheduling and workflow capabilities; implementation of new storage layouts for Earth System Data to be evaluated on fast (in-memory) data analytics frameworks for scientific data management.

**EU H2020 Project EUBra-BIGSEA (Europe – Brazil Collaboration of BIG Data Scientific Research through Cloud-Centric Applications; 2016-2017)** EUBra-BIGSEA aims at providing an abstract framework for the development of distributed Big Data applications. CMCC Foundation leads WP4 on the provisioning of a fast and Big Data platform integrating multiple classes of Big Data systems to address multifaceted issues (e.g. data storage, querying, transformation, analysis, and mining).

**EU H2020 Project INDIGO-DataCloud (INtegrating Distributed data Infrastructures for Global ExpLOitation; 2015-2017)** INDIGO-DataCloud aims at developing a data/computing platform targeted at scientific communities, deployable on multiple hardware, and provisioned over hybrid (private or public) e-infrastructures. CMCC Foundation contributes to WP2 tasks on requirements gathering from the user communities and dissemination as well as to WP6 activities on libraries and toolkits for user interfaces and big data workflows support (WaaS) for eScience. □

**EU H2020 Project PRIMAVERA (PRocess-based climate sIMulation: AdVances in high resolution modelling and European climate Risk Assessment)** The goal of PRIMAVERA is to deliver novel, advanced and well-evaluated high-resolution global climate models (GCMs), capable of simulating and predicting regional climate with unprecedented fidelity, out to 2050. The ASC division of the CMCC Foundation will work in WP9 addressing HPC and Data management challenges.

## 4.2. Third parties involved in the project (including use of third party resources)

### ENEA

<b>Does the participant plan to subcontract certain tasks (please note that core tasks of the project should not be subcontracted)</b>	<b><i>X</i></b>
During the execution phase of project, ENEA will subcontract tasks related to the issues related to data reconciliation and transformation to the development of new principles, methods and tools for big data management. The subcontractor should have expertise on Big Data & Databases, with a strong theoretical background. The project should benefit from this expertise, which will complete the panel of Consortium current skills.	
<b>Does the participant envisage that part of its work is performed by linked third parties[1]</b>	<b><i>No third parties involved</i></b>
<i>If yes, please describe the third party, the link of the participant to the third party, and describe and justify the foreseen tasks to be performed by the third party</i>	
<b>Does the participant envisage the use of contributions in kind provided by third parties (Articles 11 and 12 of the General Model Grant Agreement)</b>	<b><i>No third parties involved</i></b>
<i>If yes, please describe the third party and their contributions</i>	



## **Section 5: Ethics and Security**

*This section is not covered by the page limit.*

### **5.1 Ethics**

No ethical issues have been entered in the ethical issue table in the administrative proposal forms.

### **5.2 Security[2]**

The ETCETERA project contains

- activities or results raising security issues: **NO**
- 'EU-classified information' as background or results: **NO**

### **5.3 Support letters**

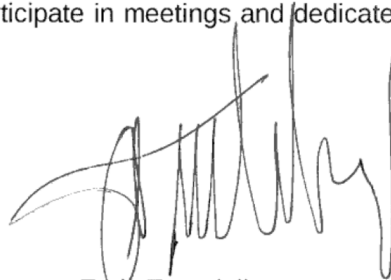
## Letter of support

To Whom It May Concern,

This letter is designed to supporting to the project entitled *ETCETERA* to be submitted to the European Commission under the EO-2-2017 Call

ETCETERA is a proposal lead by Capgemini Technology Services dedicated to provide an extensible and scalable framework which should be able to orchestrate existing EO/non EO dissemination platforms providing to different profiles of users with a homogeneous interface to retrieve air quality related products such as observational data, numerical models outputs, analytics and diagnostics calculated. The project brings together world-class research institutes in Europe with experience in climate and air quality impact modeling such as the Barcelona Supercomputing Center (BSC), the Italian National Agency for New Technologies, Energy, and Sustainable Economic Development (ENEA), the Netherlands Meteorological Institute (KNMI), the Euro-Mediterranean Center on Climate Change (CMCC) together with private companies specialized in air quality sector (Green City Solutions) and Big Data Technologies (Rasdaman) having a diversity of climate-related issues. From the beginning of the project, the aim is to engage with potential users to shape various aspects of the work to be done

The **Sand and dust Storm – Warning Advisory and Assessment System (SDS-WAS)** was launched by **World Meteorological Organization (WMO)** with the mission to enhance the ability of countries to deliver to end-users dust observations, forecasts, information and knowledge to reduce the adverse impacts of airborne dust. Its **Steering Committee** expects that SDS-WAS will benefit from the project results. So, it strongly supports the funding of ETCETERA and is available and willing to participate in meetings and dedicated discussions during the three-year period



Enric Terradellas  
State Meteorological Agency of Spain  
Chair of WMO SDS-WAS Steering Committee



**Pôle de compétitivité Aéronautique, Espace, Systèmes embarqués**

CAPGEMINI Technology Services  
Application Services France  
Carine Saüt  
109, avenue Eisenhower  
31036 Toulouse Cedex  
FRANCE

Toulouse, Feb. 9<sup>th</sup>, 2017

Réf. : 17-ADMIN-CS-029

Subject : Support to the proposal entitled "ETCETERA – ExTensible and sCalable models EvaluaTion vs Earth obseRvations frAMework" submitted to the EC EO-2-2017 call.

Dear Madam,

ETCETERA is a proposal lead by Capgemini Technology Services dedicated to provide an extensible and scalable framework which will be able to orchestrate existing Earth Observation/non EO dissemination platforms providing to different profiles of users with a homogeneous interface to retrieve air quality related products such as observational data, numerical models outputs, analytics and diagnostics calculated. The project brings together world-class research institutes in Europe with experience in climate and air quality impact modeling together with private companies specialized in air quality sector and Big Data technology. From the beginning of the project, the aim is to engage with potential users to shape various aspects of the work to be done.

We consider that this project is directly in line with the development strategy of Aerospace Valley competitiveness cluster (France) in the field of digital technology and space applications.

CapGemini Technology Services, member of Aerospace Valley cluster, is the leader of the ETCETERA proposal.

Therefore Aerospace Valley fully supports the ETCETERA proposal.

Faithfully yours.

Patrick Desiré  
Chief Executive Officer

A handwritten signature in blue ink, appearing to read "Patrick Desiré", with a blue arrow pointing from the text above to the signature.

Copy : Grégory Pradels / Philippe Lattes – "TOP – Telecom./Observation/Position." technical domain

**AEROSPACE VALLEY**  
118 route de Narbonne – CS 94244 – 31432 TOULOUSE CEDEX 4  
Tél. 05 61 14 80 37 – Fax : 05 62 26 46 25 - e-mail : [contact@aerospace-valley.com](mailto:contact@aerospace-valley.com)  
<http://www.aerospace-valley.com>  
N° SIRET : 484 284 526 000 56 – Code APE : 9499Z

February 2017

### Letter of support

To Whom It May Concern,

This letter is designed to support the proposal entitled: *ExTensible and sCalable models EvaluaTion vs Earth obseRvations framework (ETCETERA)* to be submitted to the European Commission under the EO-2-2017 Call.

ETCETERA is a proposal lead by Capgemini Technology Services, providing an extensible and scalable framework which will be able to orchestrate existing EO/non EO dissemination platforms providing to different profiles of users with a homogeneous interface to retrieve air quality related products such as observational data, numerical model outputs, analytics and diagnostics calculated. The project brings together world-class research institutes in Europe with experience in climate and air quality impact modelling, parallel diagnostics and large scale scientific data management such as the Barcelona Supercomputing Center (BSC), the Italian National Agency for New Technologies, Energy, and Sustainable Economic Development (ENEA), the Netherlands Meteorological Institute (KNMI), the Euro-Mediterranean Center on Climate Change Foundation (Fondazione CMCC) together with private companies specialized in air quality sector (Green City Solutions) and Big Data Technologies (Rasdaman) having a diversity of climate-related issues. From the beginning of the project, the aim is to engage with potential users to shape various aspects of the work to be done

Climate-KIC is a Europe's largest public-private partnership working to identify climate innovation opportunity. We support the project, because if successful it will provide stakeholders with new access mechanism to data allowing adaptable analysis. More importantly from a Climate-KIC perspective, it will provide entrepreneurial SMEs the possibility to expand their business as users of the platform and provide policy makers a way to exploit scientific outcomes as an effective decision support and engage actionable insights.

Kindest regards



Ebrahim Mohamed

Director Education, Climate-KIC



Climate-KIC is supported by the  
EIT, a body of the European Union





IEA

406 Philip Lyle Building  
University of Reading, Whiteknights  
Reading RG6 6BX

phone +44 (0)118 6820

email [info@the-iea.org](mailto:info@the-iea.org)

9<sup>th</sup> February 2017

### Letter of support

To Whom It May Concern,

This letter is designed to supporting to the proposal entitled: *ExTensible and sCalable models EvaluaTion vs Earth obseRvations framework (ETCETERA)* to be submitted to the European Commission under the EO-2-2017 Call

ETCETERA is a proposal led by Capgemini Technology Services dedicated to provide an extensible and scalable framework which will be able to orchestrate existing EO/non EO dissemination platforms providing to different profiles of users with a homogeneous interface to retrieve air quality related products such as observational data, numerical models outputs, analytics and diagnostics calculated. The project brings together world-class research institutes in Europe with experience in climate and air quality impact modeling, parallel diagnostics and large scale scientific data management such as the Barcelona Supercomputing Center (BSC), the Italian National Agency for New Technologies, Energy, and Sustainable Economic Development (ENEA), the Netherlands Meteorological Institute (KNMI), the Euro-Mediterranean Center on Climate Change Foundation (Fondazione CMCC) together with private companies specialized in air quality sector (Green City Solutions) and Big Data Technologies (Rasdaman) having a diversity of climate-related issues. From the beginning of the project, the aim is to engage with potential users to shape various aspects of the work to be done.

The Institute for Environmental Analytics (IEA) is a cutting edge R&D organization focused on applying big data analytics in the environmental sector. The IEA supports the funding of the project, because of the benefits that it will bring to our organization. The IEA works with other organizations to help them turn data into competitive advantage, hence having easier access to relevant data is highly beneficial to us. We hope that ETCETERA will provide, for air quality, a reliable platform where to retrieve the necessary data for us to analyse and use in targeted applications.

Date: 6 February 2017

Signature: 

Jon Blower, IEA Chief Technical Officer

---

DELIVERING VALUE FROM BIG DATA

**Letter of support**

To Whom It May Concern,

This letter is designed to supporting to the proposal entitled: *ExTensible and sCalable models EvaluaTion vs Earth obseRvations framework (ETCETERA)* to be submitted to the European Commission under the EO-2-2017 Call

ETCETERA is a proposal lead by Capgemini Technology Services dedicated to provide an extensible and scalable framework which will be able to orchestrate existing EO/non EO dissemination platforms providing to different profiles of users with a homogeneous interface to retrieve air quality related products such as observational data, numerical models outputs, analytics and diagnostics calculated. The project brings together world-class research institutes in Europe with experience in climate and air quality impact modeling such as the Barcelona Supercomputing Center (BSC), the Italian National Agency for New Technologies, Energy, and Sustainable Economic Development (ENEA), the Netherlands Meteorological Institute (KNMI) together with private companies specialized in air quality sector (Green City Solutions) and Big Data Technologies (Rasdaman) having a diversity of climate-related issues. From the beginning of the project, the aim is to engage with potential users to shape various aspects of the work to be done

Eurac research is a private research center with specific research activity focused on the Earth Observation and the monitoring of environment. Eurac research strongly supports the funding of the project. Eurac research is working with BigData Analytics and more specific with the Rasdaman DataCube Technology. The platform the project will create could be an important tool to be integrated within the infrastructure running at eurac research and could also be used in several running or upcoming projects where eurac research will be involved.

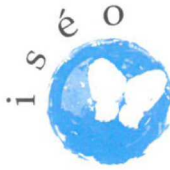
In order to better shape and co-design the tools, eurac research is willing to follow the project along its development by participating to a few meetings along the three-year periods, and possibly dedicated discussions with individual teams.

Date: 01/02/2017

Signature:   
**EURAC**  
research  
(Legal representative)

Drususallee/Viale Druso 1 • 39100 Bozen/Bolzano  
Tel. +39 0471 055 055 • Fax +39 0471 055 059





## Letter of support

To Whom It May Concern,

This letter is designed to supporting to the proposal entitled: *ExTensible and sCalable models EvaluaTion vs Earth obseRvations framework (ETCETERA)* to be submitted to the European Commission under the EO-2-2017 Call

ETCETERA is a proposal lead by Capgemini Technology Services dedicated to provide an extensible and scalable framework which will be able to orchestrate existing EO/non EO dissemination platforms providing to different profiles of users with a homogeneous interface to retrieve air quality related products such as observational data, numerical models outputs, analytics and diagnostics calculated. The project brings together world-class research institutes in Europe with experience in climate and air quality impact modeling such as the Barcelona Supercomputing Center (BSC), the Italian National Agency for New Technologies, Energy, and Sustainable Economic Development (ENEA), the Netherlands Meteorological Institute (KNMI) together with private companies specialized in air quality sector (Green City Solutions) and Big Data Technologies (Rasdaman) having a diversity of climate-related issues. From the beginning of the project, the aim is to engage with potential users to shape various aspects of the work to be done

iséo is a company of the Environnement SA group with more than 25 years of experience and an internationally recognized expertise in the development of specialized software and data acquisition systems in the field of the air quality monitoring. The customers of iséo are mainly the Air Quality Monitoring Networks and organizations that are in charge of the air quality monitoring in large European cities and regions. The possibility of accessing the new services and the data provided inside this project may benefit to iséo and its customers. As a consequence, iséo Environnement SA strongly supports the funding of the project.

In order to better shape and co-design the tools, iséo Environnement SA is willing to follow the project along its development by participating to a few meetings along the three-year periods, and possibly dedicated discussions with individual teams.

Date: ...10 February 2017

Signature: ..... **Jacques MEDIAVILLA**

(Legal representative) **Directeur**





JPI - Connecting Climate Knowledge for Europe

Petra Manderscheid  
Executive Director Central Secretariat  
Avenue Louise 231  
B – 1050 Brussels/Belgium

Date 27/02/2017

**Letter of Support**

To Whom It May Concern,

With this letter the Joint Programming Initiative Connecting Climate Knowledge for Europe (JPI Climate) expresses its interest in and support for the research project proposal titled *ExTensible and sCalable models EvaluaTion vs Earth obseRvations framework (ETCETERA)* to be submitted to the European Commission under the HORIZON 2020 call number SC5-31-2017.

ETCETERA is a proposal lead by Capgemini Technology Services dedicated to provide an extensible and scalable framework which will be able to orchestrate existing EO/non EO dissemination platforms providing to different profiles of users with a homogeneous interface to retrieve air quality related products such as observational data, numerical models outputs, analytics and diagnostics calculated. The project brings together world-class research institutes in Europe with experience in climate and air quality impact modeling, parallel diagnostics and large scale scientific data management such as the Barcelona Supercomputing Center (BSC), the Italian National Agency for New Technologies, Energy, and Sustainable Economic Development (ENEA), the Netherlands Meteorological Institute (KNMI), the Euro-Mediterranean Center on Climate Change Foundation (Fondazione CMCC) together with private companies specialized in air quality sector (Green City Solutions) and Big Data Technologies (Rasdaman) having a diversity of climate-related issues. From the beginning of the project, the aim is to engage with potential users to shape various aspects of the work to be done.

JPI Climate is an initiative of European member states and associated countries to align national programmes by jointly coordinating their climate research and funding new transnational research activities.

Considering the potential relevance of ETCETERA for the overall goals of JPI Climate due to its role in enabling cross-border research and science-practice interaction and also in fostering interdisciplinarity and societal relevance, JPI Climate hereby confirms its support for the ETCETERA proposal.

Yours sincerely

A handwritten signature in blue ink, appearing to read 'Petra Manderscheid'.

Petra Manderscheid  
Executive Director of the Central Secretariat

Central Secretariat  
Joint Programming Initiative JPI - Connecting Climate Knowledge for Europe  
Avenue Louise 231 – 1050 Brussels/Belgium – T. +32 (0)2 238 37 02 - 32 (0)4 78 784 262  
secretariat@jpi-climate.belspo.be  
www.jpi-climate.eu





UMR 3589 CNRM  
Météo-France - CNRS  
42 av. G Coriolis  
31057 Toulouse cedex

Toulouse, 14 February 2017.

### Letter of support

To Whom It May Concern,

This letter is designed to supporting to the proposal entitled: *ExTensible and sCalable models EvaluaTion vs Earth obseRvations framework (ETCETERA)* to be submitted to the European Commission under the EO-2-2017 Call.

ETCETERA is a proposal lead by Capgemini Technology Services dedicated to provide an extensible and scalable framework which will be able to orchestrate existing EO/non EO dissemination platforms providing to different profiles of users with a homogeneous interface to retrieve air quality related products such as observational data, numerical models outputs, analytics and diagnostics calculated. The project brings together world-class research institutes in Europe with experience in climate and air quality impact modeling such as the Barcelona Supercomputing Center (BSC), the Italian National Agency for New Technologies, Energy, and Sustainable Economic Development (ENEA), the Netherlands Meteorological Institute (KNMI) together with private companies specialized in air quality sector (Green City Solutions) and Big Data Technologies (Rasdaman) having a diversity of climate-related issues. From the beginning of the project, the aim is to engage with potential users to shape various aspects of the work to be done.

CNRM is the research branch of Météo France which priorities amongst several are notably the multi-space and time scale weather forecast, the physics and dynamics of the atmosphere, the knowledge and evolution of the climate, with particular emphasis on air quality. Since the aim of the project is to enhance crucial environmental information through a smart combination, CNRM strongly supports the funding of the project because it will offer new interesting perspectives of application to the meteorological outcomes.

In order to better shape and co-design the tools, CNRM is willing to follow the project along its development by participating to a few meetings along the three-year periods, and possibly dedicated discussions with individual teams.

Le Directeur de l'UMR 3589-CNRM



Marc PONTAUD



**Prof. Dr. Guy Brasseur**

Senior Scientist  
Former Director

Max-Planck-Institut für Meteorologie  
Bundesstr. 53  
20146 Hamburg  
Deutschland

Tel.: +49 - (0)40 - 41173 - 209  
Fax: +49 - (0)40 - 41173 - 390  
guy.brasseur@mpimet.mpg.de

Hamburg, 31.1.2017

### Letter of support

To Whom It May Concern,

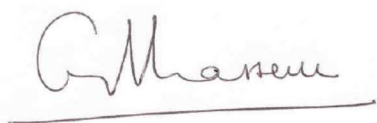
This letter is designed to supporting to the proposal entitled: *ExTensible and sCalable models EvaluaTion vs Earth obseRvations framework (ETCETERA)* to be submitted to the European Commission under the EO-2-2017 Call

ETCETERA is a proposal lead by Capgemini Technology Services dedicated to provide an extensible and scalable framework which will be able to orchestrate existing EO/non EO dissemination platforms providing to different profiles of users with a homogeneous interface to retrieve air quality related products such as observational data, numerical models outputs, analytics and diagnostics calculated. The project brings together world-class research institutes in Europe with experience in climate and air quality impact modeling such as the Barcelona Supercomputing Center (BSC), the Italian National Agency for New Technologies, Energy, and Sustainable Economic Development (ENEA), the Netherlands Meteorological Institute (KNMI) together with private companies specialized in air quality sector (Green City Solutions) and Big Data Technologies (Rasdaman) having a diversity of climate-related issues. From the beginning of the project, the aim is to engage with potential users to shape various aspects of the work to be done

The Max Planck Institute for Meteorology is a research institution of the Max Planck Society. The Environmental Group of the Institute strongly supports the funding of the project, because the Group will directly benefit from its results. By creating a Big Data platform that allows the scientists to access Copernicus and other data extract data, it new and effective diagnostics and analytics of atmospheric observations will be greatly facilitated. This tool will be very useful in particular for the Air Quality and Climate communities to interface and retrieve data from models and observations, and to perform analyses and interpretation calculations.

In order to better shape and co-design the tools, scientists from our institution are willing to closely follow the project along its development and to participate to a few meetings along the three-year periods, and to discuss relevant questions with individual teams involved in the proposed project.

Sincerely,



Guy Brasseur

## LETTER OF SUPPORT

### To Whom It May Concern

*This letter is designed to supporting to the proposal entitled: ExTensible and sCalable models EvaluaTion vs Earth obseRvations framework (ETCETERA) to be submitted to the European Commission under the EO-2-2017 Call*

ETCETERA is a proposal lead by Capgemini Technology Services dedicated to provide an extensible and scalable framework which will be able to orchestrate existing EO/non EO dissemination platforms providing to different profiles of users with a homogeneous interface to retrieve air quality related products such as observational data, numerical models outputs, analytics and diagnostics calculated. The project brings together world-class research institutes in Europe with experience in climate and air quality impact modeling, parallel diagnostics and large scale scientific data management such as the Barcelona Supercomputing Center (BSC), the Italian National Agency for New Technologies, Energy, and Sustainable Economic Development (ENEA), the Netherlands Meteorological Institute (KNMI), the Euro-Mediterranean Center on Climate Change Foundation (Fondazione CMCC) together with private companies specialized in air quality sector (Green City Solutions) and Big Data Technologies (Rasdaman) having a diversity of climate-related issues. From the beginning of the project, the aim is to engage with potential users to shape various aspects of the work to be done

Proambiente s.c.r.l. is a spin-off of the Italian National Research Council, a R&D private no profit SME, creating prototypes and solutions for environmental surveillance and remediation. We are involved very interested in new collaborations for innovative solutions/services finalized to the improvement of Air Quality and mitigation of Climate Change.

Indeed, the proposed actions are performed by means of advanced and innovative services offered by the consortium, or through the fabrication of sensors, instrumentation and systems suited for environmental monitoring and analysis, so fitting very well with the mission of Proambiente. For these reasons Proambiente is willing to follow the project along its development by participating to a few meetings along the three-year periods, and possibly discuss new projects and future collaboration with individual teams.

Bologna, 13 February 2017

**Proambiente S.c.r.l.**  
Via Gobetti, 101  
40129 Bologna  
C.F./P.IVA 03305011201



Sandro Fuzzi  
President and Legal representative



## Letter of support

To Whom it May Concern,

This letter is designed to supporting to the proposal entitled: *ExTensible and sCalable models EvaluaTion vs Earth obseRvations framework (ETCETERA)* to be submitted to the European Commission under the EO-2-2017 Call

ETCETERA is a proposal lead by Capgemini Technology Services dedicated to provide an extensible and scalable framework which will be able to orchestrate existing EO/non EO dissemination platforms providing to different profiles of users with a homogeneous interface to retrieve air quality related products such as observational data, numerical models outputs, analytics and diagnostics calculated. The project brings together world-class research institutes in Europe with experience in climate and air quality impact modeling, parallel diagnostics and large scale scientific data management such as the Barcelona Supercomputing Center (BSC), the Italian National Agency for New Technologies, Energy, and Sustainable Economic Development (ENEA), the Netherlands Meteorological Institute (KNMI), the Euro-Mediterranean Center on Climate Change Foundation (Fondazione CMCC) together with private companies specialized in air quality sector (Green City Solutions) and Big Data Technologies (Rasdaman) having a diversity of climate-related issues. From the beginning of the project, the aim is to engage with potential users to shape various aspects of the work to be done

Predictia is a Spanish SME that develops data management software solutions for the meteorology sector. It strongly supports the funding of the project, because if funded, Predictia will benefit from its results. The technology and services proposed in this project will incredibly ease the usage of air quality data as well as the development of derived products.

Date: 3rd Febraury 2017

Signature: Daniel San Martín

  
General Manager

Avda. de los Castros s/n · CDTUC Fase A, Módulo 203 · 39005 Santander, España  
Tfno: +34 942 764410 · [www.predictia.es](http://www.predictia.es) · [predictia@predictia.es](mailto:predictia@predictia.es)

Paris, February, 9<sup>th</sup> 2017

Bertrand BESSAGNET (HDR, PhD., Ing.,)  
Chief Scientist  
INERIS, Parc technologique ALATA  
60550 Verneuil en Halatte  
Tél : +33 (0)344556533  
e-mail : [bertrand.bessaqnet@ineris.fr](mailto:bertrand.bessaqnet@ineris.fr)

To Whom It May Concern,


This letter is designed to supporting the proposal entitled: ExTensible and sCalable models EvaluaTion vs Earth obseRvations framework (ETCETERA) to be submitted to the European Commission under the EO-2-2017 Call.

ETCETERA is a proposal lead by Capgemini Technology Services dedicated to provide an extensible and scalable framework which will be able to orchestrate existing EO/non EO dissemination platforms providing to different profiles of users with a homogeneous interface to retrieve air quality related products such as observational data, numerical models outputs, analytics and diagnostics calculated. The project brings together world-class research institutes in Europe with experience in climate and air quality impact modeling such as the Barcelona Supercomputing Center (BSC), the Italian National Agency for New Technologies, Energy, and Sustainable Economic Development (ENEA), the Netherlands Meteorological Institute (KNMI) together with private companies specialized in air quality sector (Green City Solutions) and Big Data Technologies (Rasdaman) having a diversity of climate-related issues. From the beginning of the project, the aim is to engage with potential users to shape various aspects of the work to be done.

INERIS is a company working in the field of Air Quality Management services for public and private bodies. INERIS strongly supports the funding of the project, because if funded, INERIS and the community of Air Quality in Europe will benefit from its results. There is a urgent need to develop big data procedures to analyse the huge amount of data generated by models and instrument (ground and satellite data) to better assess the air quality mitigation strategies.

In order to better shape and co-design the tools, INERIS is willing to follow the project along its development by participating to a few meetings along the three-year periods, and possibly dedicated discussions with individual teams.

Paris - February, 9th 2017



Bertrand BESSAGNET

---

Parc Technologique Alata BP 2 F-60550 Verneuil-en-Halatte  
Tél +33(0)3 44 55 66 77 fax +33(0)3 44 55 66 99 internet [www.ineris.fr](http://www.ineris.fr)  
Institut national de l'environnement industriel et des risques

Etablissement public à caractère industriel et commercial – RCS Compiègne B 381 984 921 – Siret 381 984 921 00019 – APE 7120B – TVA Intracom FR 73 381 984 921

Page 1/1

## 5.4 Glossary

### Common vocabulary, glossary and acronym references to websites

Analytics: discovery, interpretation, and communication of meaningful patterns in data. It includes both calculation of statistics from comparison between model outputs variables and Earth Observations (in situ, satellite, ...) and statistics over time or space.

API: Application Program Interface

B2B: Business to Business

CA: Consortium Agreement

CAMS: Copernicus Atmosphere Monitoring Service

C3S: Copernicus Climate Change Service

CLEMS: Copernicus Land Monitoring Service

CMEMS: Copernicus Marine Environment Monitoring Service

Copernicus core services: service information related to land, marine (CMEMS), atmosphere (CAMS), emergency security and climate (C3S) data

CSW: Catalog Service for the Web

DIAS: Data and Information Access Services

Diagnostics: calculation of new variables from combination of existing ones

DMP: Data Management Plan

EMS: Emergency Monitoring Service

EC: European Commission

EEA: European Environment Agency

EO: Earth Observation

ESA: European Space Agency

ESGF: Earth System Grid Federation

EUMETSAT: European Organization for the Exploitation of Meteorological Satellites

ECMWF: European Centre for Medium-range Weather Forecasts

GUI: Graphical User Interface

[ETCETERA]

IG: Interest Group

IGS: Integrated Ground Segment

IoT: Internet of Things

IP: Intellectual Property

IPR: Intellectual Property Rights

OGC standards: OGC(R) standards are technical documents that detail interfaces or encodings. Software developers use these documents to build open interfaces and encodings into their products and services. These standards are the main "products" of the Open Geospatial Consortium and have been developed by the membership to address specific interoperability challenges

MVP: Minimum Viable Product

QEMS: Quality and Environmental Management System

QMS: Quality Management System

RDA: Research Data Alliance

RDV: Rapid Design Visualization

SME: Small and Medium Enterprises

SaaS: Software as a Service

TRL: Technology Readiness Level

UI: Users Interfaces

UUID: Universally Unique Identifiers

WCPS: Web Coverage Processing Service

WCS: Web Coverage Service

WG: Working Group

WMO: World Meteorological Organization

WMO GAW<sup>11</sup>: WMO Global Atmosphere Watch - WMO program aiming at understanding and control the increasing influence of human activity on the global atmosphere

WMS: Web Map Service

---

<sup>11</sup> <http://www.wmo.int/pages/prog/arep/gaw/history.html>



### Lists of publications relevant to this proposal:

1. Author: Spinuso, A., Cheney, J., & Atkinson, M. (2013). Provenance for seismological processing pipelines in a distributed streaming workflow. In EDBT/ICDT Workshops (pp. 307–312). doi:10.1145/2457317.2457369
2. Co-Author, Filgueira, R., Krause, A., Atkinson, M., Klampanos, I., Spinuso, A., & Sanchez-Exposito, S. (2015). dispel4py: An Agile Framework for Data-Intensive eScience. 2015 IEEE 11th International Conference on E-Science, 454–464. doi:10.1109/eScience.2015.40
3. Co-Author Atkinson, M., Carpené, M., Casarotti, E., Claus, S., Filgueira, R., Frank, A., ... Vilotte, J.-P. (2015). VERCE delivers a productive e-Science environment for seismology research. E-Science (e-Science), 2015 IEEE 11th International Conference on, 224–236. doi:10.1109/eScience.2015.38
4. Co-Author, Gesing, Sandra and Atkinson, Malcolm and Filgueira, Rosa and Taylor, Ian and Jones, Andrew and Stankovski, Vlado and Liew, Chee Sun and Spinuso, Alessandro and Terstyanszky, Gabor and Kacsuk, Peter (2014, November). Workflows in a dashboard: a new generation of usability. In Workflows in Support of Large-Scale Science (WORKS), 2014 9th Workshop on (pp. 82-93). IEEE.
5. W. J. Som de Cerff, Monique Petitdidier, A. Gemünd, L. Horstink, H. Schwichtenberg (july 2009). Earth science test suites to evaluate grid tools and middleware - Examples for grid data access tools, Earth Science Informatics, DOI: 10.1007/s12145-009-0022-y
6. L. Cinquini, D. Crichton, C. Mattmann, J. Harney, G. Shipman, F. Wang, R. Ananthakrishnan, N. Miller, S. Denvil, M. Morgan, Z. Pobre, G. M. Bell, C. Doutriaux, R. Drach, D. Williams, P. Kershaw, S. Pascoe, E. Gonzalez, **S. Fiore**, R. Schweitzer, *The Earth System Grid Federation: An open infrastructure for access to distributed geospatial data*, Future Generation Computer Systems, Volume 36, July 2014, pp. 400-417, ISSN 0167-739X, <http://dx.doi.org/10.1016/j.future.2013.07.002>.  
□
7. **S. Fiore**, A. D'Anca, C. Palazzo, Ian T. Foster, Dean N. Williams, **G. Aloisio**: *Ophidia: Toward Big Data Analytics for eScience*. ICCS 2013, June 5-7, 2013 Barcelona, Spain, ICCS, volume 18 of Procedia Computer Science, pp. 2376-2385. Elsevier, (2013).
8. **D. Elia**, **S. Fiore**, A. D'Anca, C. Palazzo, I. T. Foster, D. N. Williams, **G. Aloisio**: *An in-memory based framework for scientific data analytics*. ACM Conference Computing Frontiers 2016, pp. 424-429.
9. M. Plóciennik, **S. Fiore**, G. Donvito, M. Owsiak, M. Fargetta, R. Barbera, R. Bruno, E. Giorgio, D. N. Williams, **G. Aloisio**: *Two-level Dynamic Workflow Orchestration in the INDIGO DataCloud for Large-scale, Climate Change Data Analytics Experiments*. International Conference on Computational Science, ICCS 2016, pp. 722-733.
10. **S. Fiore** et al, “*Distributed and cloud-based multi-model analytics experiments on large volumes of climate change data in the Earth System Grid Federation eco-system*”, IEEE Big Data Conference 2016, December 5-8, 2016, Washington [to appear on IEEEExplore].
11. **[Software]** CMCC provides the Ophidia software (<http://www.ophidia.cmcc.it>), a cross-domain big data analytics framework for the analysis of scientific, multi-dimensional datasets. The framework exploits a server-side, declarative, parallel approach for data analysis and mining.
12. P. Baumann, I. Manolescu-Goujot, L. Trani, Y.E. Ioannidis, G.G. Barnaföldi, L. Dobos, E. Bányai: *Proceedings of the 28th International Conference on Scientific and Statistical Database Management, SSDBM 2016*, Budapest, Hungary, July 18-20, 2016. ACM 2016
13. P. Baumann, V. Merticariu, A. Dumitru, D. Misev: *Standards-Based Services for Big Spatio-Temporal Data*. Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci., XLI-B4, 691-699, doi:10.5194/isprs-archives-XLI-B4-691-2016, 2016.
14. G. Merticariu, D. Misev, P. Baumann: *Measuring Storage Access Performance in Array Databases*. Proc. 7th Workshop on Big Data Benchmarking (WBDB), December 14-15, 2015, New Delhi, India

15. D. Misev, P. Baumann: *Enhancing Science Support in SQL*. Proc. Workshop Data and Computational Science Technologies for Earth Science Research (co-located with IEEE Big Data), Santa Clara, US, October 29, 2015
16. A. Dumitru, V. Merticariu, P. Baumann: *Exploring Cloud Opportunities from an Array Database Perspective*. Proc ACM SIGMOD Workshop on Data analytics in the Cloud (DanaC'2014), June 22 - 27, 2014, Snowbird, USA
17. P. Baumann: *On the Management of Multidimensional Discrete Data*. VLDB Journal 4(3)1994, Special Issue on Spatial Database Systems, pp. 401 - 444 (note: paper introducing the concept of Array Databases)

#### **List of identified international sectoral events**

Data Driven Innovation Open Summit: <http://2017.datadriveninnovation.org/>

International Conference on Web Engineering: <http://icwe2017.webengineering.org/>

Int. ACM SIGMOD Workshop on Managing and Mining Enriched Geo-Spatial Data (GeoRich): <http://www.public.asu.edu/~jiayu2/georich17/>

Copernicus Symposium on Climate Services: <http://www.the-iaea.org/events/copernicus-symposium-climate-services/>

Data.Space: <http://www.dataspace.xyz/>

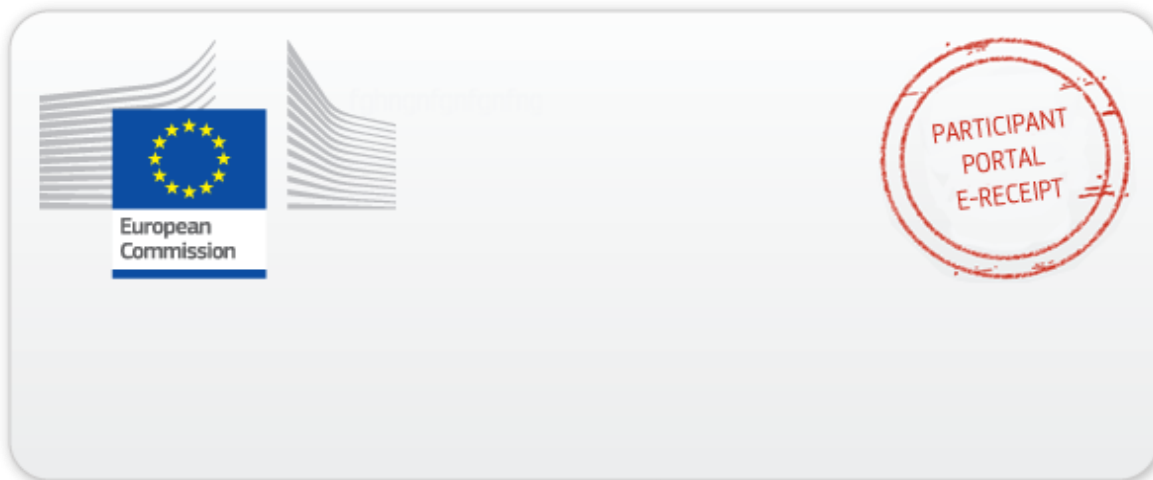
EUMETSAT Meteorological Satellite Conference:

[http://www.eumetsat.int/website/home/News/ConferencesandEvents/DAT\\_3212307.html](http://www.eumetsat.int/website/home/News/ConferencesandEvents/DAT_3212307.html)

---

**[1] A third party that is an affiliated entity or has a legal link to a participant implying a collaboration not limited to the action. (Article 14 of the Model Grant Agreement).**

**[2] See article 37 of the Model Grant Agreement**



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>)