Improving The Copernicus Climate Data Store Metadata Scheme With The “RDA Metadata Standards Repository”

O2: Organisation type:
Academia/Research
O3: Profile of lead organisation:
Research and Supercomputing Center
Project Team:

First Name:
Iván
Surname:
Cernicharo Ortiz
Organisation:
Barcelona Supercomputing Center - Centro Nacional de Supercomputación
Email:
ivan.cernicharo@bsc.es
Proposed role in project:
Developer
Short bio / relevant expertise:
RDA-Bio-Ivan_Cernicharo_Ortiz.pdf (https://grants.rd-alliance.org/system/files/RDA-Bio-Ivan_Cernicharo_Ortiz.pdf)

First Name:
Francesco

Surname:
Benincasa
Organisation:
Barcelona Supercomputing Center - Centro Nacional de Supercomputación
Email:
francesco.benincasa@bsc.es
Proposed role in project:
Developer
Short bio / relevant expertise:
RDA-Bio-Francesco_Benicasa.pdf (https://grants.rd-alliance.org/system/files/RDA-Bio-Francesco_Benicasa.pdf)

Workflow:
submitted
Related Call:
Call for Europe Adoption Grants
AP1: Who is making this proposal and who else will be involved in the adoption and its testing?:
The Barcelona Supercomputing Center (BSC-CNS) is the organization making the proposal. In addition of hosting the MareNostrum supercomputer, the BSC is a multi-disciplinary research center specialized in Earth, Computer, and Life sciences, dealing on a daily basis with huge amounts of data from very different fields. More precisely, the proposal will be led by Iván Cernicharo Ortiz (research engineer) for his involvement in the Copernicus Climate Change Service (C3S) tender led by BSC on Evaluation of the Climate Data Store (CDS). The exact objectives of C3S512 are described in the sections below. Apart from him who will be involved directly in the testing of the adoption project, another member of the BSC Earth Sciences Department, Francesco Benincasa, who has been involved for several years in RDA activities and was the co-chair of the RDA IG on “Weather, climate and air quality”, will contribute assisting with his previous knowledge of RDA.
O1: Organisation name:
Barcelona Supercomputing Center - Centro Nacional de Supercomputación
AP3: Explain your existing knowledge of the RDA Recommendation(s) and output(s) to be adopted and involvement in RDA overall.:
The BSC has been involved for many years in RDA (RDA2 and 3, RDA Europe and the RDA Spanish node). Our involvement includes between other activities, the organization of the 9th plenary meeting in Barcelona back in 2017, co-chairing an Interest Group (Weather, Climate and air quality), organizing several workshops on provenance in RDA and Earth Sciences as well as leading the creation of the Iberia and Spanish RDA node. About the metadata standards directory output itself, we’ve be dealing with the generation of recommendations on metadata standards for different projects for several years both inside the department in several European projects, mostly based on CF metadata standards, CMOR, but also PROV, which usage was recommended RDA provenance workshops we organized.
AP4: What problem are you addressing in the proposal?:
One of the objectives of the C3S512 tender (URL to C3S512 contract package: https://climate.copernicus.eu/c3s512-quality-assurance-climate-data-store) in which the RDA adoption call will be included is to quality check and document the data stored in the Climate Data Store (URL to CDS: https://cds.climate.copernicus.eu/#!/home). This data repository contains various types of data (satellite and in-situ observations, climate model outputs from many types,...) for which we have to present summaries at the level of variable and dataset. We also have to do recommendations on common data models and metadata standards for data that will be presented in the portal. So far, there is a lack of official and reference metadata repositories to which we can refer to propose the adoption of these standards in the CDS. Adopting and presenting the RDA metadata repository as a trustful reference and using the various standards gathered there will help us make strong recommendations on what should be presented in the CDS.
AP5: Brief description of the planned work. Explain how you will use / implement the Recommendation(s) and output(s).:
The work planned in the adoption call will be closely linked with the timeline of the C3S512 project and its deliverables and milestones. The first stage will be the extensive review of the existing standards presented in the RDA metadata standard repository and how each one can be applied or used for each of the product of the climate data store that we have to check. In a second stage, we will implement in the data checker of the project the different standards found in the first stage and include a flag in the data summaries showing if the each type complies or not with the standard. According to the results of the checker, we will be able to go to the third stage of the proposal

consisting in formulating recommendations about the usage of these standards for the data coming into the CDS.

AP6: Who are the main beneficiaries of your work and how will they be assisted?:

The end beneficiaries of this work will be the users of the climate data store as they will be able to get robust information on which metadata standards are included in the data they are using, making their usage easier. The second beneficiaries will be the CDS managers that, thanks to the work done in this proposal, will get a more trustful and documented data repository. Finally, having recommendations about how to encode their data, the data providers of the CDS will be able to give them more visibility on the CDS. These 3 points about data usability were identified as strong user requirements by the European Commission and Copernicus when the proposal for C3S512 was written.

AP7: What are the expected results and impact?:

As mentioned before, the outcomes will be of two types: - A section in the description of each dataset of the CDS saying which metadata from the RDA metadata standards repository has been used - Written recommendations for the data providers and CDS managers to use the metadata standards in the case they weren’t used so far.

AP8: Proposed timeframe and brief workplan / gannt chart of activities.:

RDA-workplan.pdf (https://grants.rd-alliance.org/system/files/RDA-workplan.pdf)

R1: Describe the resources you will use to undertake the project and how you intend to apply the funding for those resources.:

The detail of costs is described in the Excel file containing the budget in the “Further information” section below. The resources will be mostly spent in staff time (3PM) and in 1 trip for 2 people in the C3S general assembly to present the outcomes of the work developed during the project.

R2: Expectation on dissemination activities for your project:

We plan to present our results and disseminate the RDA recommendations in the C3S general assembly at the end of 2019 (C3S general assembly URL: https://climate.copernicus.eu/c3s-3rd-general-assembly). This event will gather both CDS data providers, CDS managers and users so it will be the perfect opportunity to disseminate our work. The travel costs shown in the previous section correspond to this. We might also need help from RDA Europe to disseminate our work in the RDA 14th general assembly in Finland.

Other information in support of your proposal:

We attach in this section, the detailed budget proposal.

Please upload all relevant supporting documents:

RDA-Budget_for_adoption_proposals.pdf (https://grants.rd-alliance.org/system/files/RDA-Budget_for_adoption_proposals.pdf)

AP2: Recommendations and Outputs to be adopted:

Metadata Standards Directory

RDA Europe 4.0. project has received funding from the European Union's Horizon 2020 (H2020) research and innovation programme under the Grant Agreement no 777388.