

Deliverable D7.3 Standardisation and data policy

Deliverable No.: 7.3 -Standardisation and data policy (WP7)

Prepared by: University of Bremen, Germany

Authors: Robert Huber

Due Date: Month 24

Actual Delivery Date: Month



Grant Agreement number: 312118

Project acronym: COOPEUS

Project title: Strengthening the cooperation between the US and the EU in the field of environmental research infrastructures

Funding Scheme: FP7.INFRASTRUCTURES-2012-1

Project website address: www.coopeus.eu

Report on standardization and data policy within COOPEUS partner research infrastructures

Introduction

International and interdisciplinary collaboration is essential to address the scale and complexity of scientific and societal challenges related to the ecosystem, climate and physical dynamics of our planet. Environmental research infrastructures on both sides of the Atlantic are producing large amounts of research data that can be used to significantly improve our understanding of these processes and issues. Enabling efficient access to and open sharing of these data will facilitate strong collaborations and new scientific understanding globally.

The earth and environmental science community has a long history of sharing and re-using data due to the major societal challenges its sub-disciplines address such as the management of risks associated with geohazards, ecosystems and climate change. Related research needs to be interdisciplinary and builds on data and knowledge of many other disciplines such as biology or physics. Environmental research infrastructures provide very heterogeneous data collected by scientists from a variety of disciplines. The nature of such earth and environmental science data ranges from typical long-tail data such as spreadsheet data collected during laboratory measurements or field experiments up to 'big data' collections acquired by autonomously operating observation systems. This heterogeneity makes effective and open access to and exchange of data essential. Therefore, both, standardization as well as open access friendly data policies are required to ease data exchange among scientists and users involved in Europe's major environmental research infrastructure projects (EISCAT, EPOS, LifeWATCH, EMSO, and ICOS) with their US counterparts (AMISR, EARTHSCOPE, DataONE, OOI and NEON). The COOPEUS mission is to interlink these infrastructures on both sides of the Atlantic to generate new, scientific synergies through cross-discipline infrastructure data access and usage.

This report gives an overview of the current standardization status of COOPEUS related infrastructures, the commonalities with respect to standards identified within the COOPEUS community as well as the role and status of applied data policies in order to define a COOPEUS 'interoperability maturity status' for each infrastructure.

COOPEUS common standards

We have established a matrix of e-infrastructure standards currently used within COOPEUS research infrastructures. The input to the matrix is based on the status quo analysis performed within deliverable D7.1, the 'COOPEUS status quo assessment reports' and subsequent email communication, as well as personal interviews with COOPEUS partners

It became clear during the initial status quo assessment that distinct preferences exist for metadata as well as data exchange. For example the assessment showed that metadata was most often offered in text based formats such as ASCII or XML formats (Dublin Core, DIF, ISO). The file formats for data were less well defined, and no clear preference for data exchange standards could be identified.

This was the status quo for COOPEUS after month 12. Now, after two successful years of COOPEUS integration and harmonization activities, some clear commonalities regarding IT standards have emerged. The COOPEUS commonality matrix below can be used to identify a preliminary set of joint core standards for COOPEUS.

Common metadata standards

	Metadata representation			Metadata format			Metadata interface		
	ASCII	XML	JSON	Dublin Core	ISO 19115	DIF	OAI-PMH	CS-W	Open Search
EISCAT MIT Haystack	x								
ICOS	x								planned
NEON		x			x				
EMSO	x			x	x	x	planned	planned	x
OOI	x			x				planned	
EPOS				x	planned		x		
EarthScope-IRIS		x							
UNAVCO									
EarthScope	x	x	x						x
LIFEWATCH	x	x	x	x	x				x
DataONE		x			x				
Comment									
GEOSS Status				pending	pending	pending	pending	pending	pending

During the first phase of COOPEUS, standardization has progressed. A clear tendency towards XML based standards such as Dublin Core and ISO19115 can be observed. Whereas Dublin Core provides some basic possibilities to describe data, ISO19115 is much more complex and focusses on geographically referenced data. For COOPEUS, an extended Dublin Core format which include the possibilities to metadescribe e.g. geographical, temporal coverage or the measurement types would be highly suitable. Regarding interfaces, OpenSearch seems to be the most commonly protocol to offer access to metadata.

Common data standards

	Data formats			
	ASCII	NetCDF	HDF5	SEED/miniSEED
EISCAT MIT Haystack	x	x	x	
	x	planned	x	
ICOS	x			
NEON	x		x	

EMSO	x	x		planned
OOI		x		
EPOS	x			x
EarthScope-IRIS	x	x		x
UNAVCO				
EarthScope	x	x	x	
LIFEWATCH	x	x	x	
DataONE				

Comment		as HDF-EOS	not listed
GEOSS Status	pending	pending	

After the first phase of COOPEUS, almost all infrastructures offer access to data in ASCII file format. While the way how data is expressed in these ASCII formats may significantly differ between each RI, it at least provides a pragmatic way to access this data for the scientific community. Beneath ASCII, also another format has gained importance within the COOPEUS community. Many RI now offer their data in NetCDF, which has a good potential to become the de facto standard data format for COOPEUS.

Data policies

A profound analysis of available data policies within the COOPEUS community has already been given within deliverable D7.2, the 'Joint core data and IPR policy'. In summary it could be shown that much commonality already exist which could be used to define a joint data policies each partner could agree with:

The COOPEUS joint data policy

1. Research infrastructures under the umbrella of COOPEUS strongly support free and open access to data produced by their facilities and feel committed to work towards the realisation of this principle.
2. Access to data provided by COOPEUS research infrastructures depends on national and international legal and ethical frameworks as well as on their own data policies. COOPEUS infrastructures acknowledge potential reasons to restrict public access to each other data resulting from these obligations.
3. COOPEUS infrastructures aim to identify unnecessary or obsolete barriers to open access to their data and will continuously work towards the stepwise elimination of these obstacles.
4. COOPEUS RIs will provide free and open access, without any restrictions, to the metadata of their data holdings in order to enable data discovery and fitness-for-use evaluation of the data holdings.
5. COOPEUS RIs will make such metadata publicly available without undue delay.

6. Metadata associated with COOPEUS infrastructures data shall meet accepted international standards and shall contain the necessary information to attribute identification, authorship, geographical and temporal coverage, type of measurements and observations, access constraints as well as the responsible organisation and when possible, the principal investigator (PI) for each data set.
7. Data produced by COOPEUS infrastructures will be made available through appropriate e-infrastructures. Published data formats shall meet accepted international standards.
8. Data published by COOPEUS infrastructures shall be made available free of charge. RI specific regulations deviating from this general rule may apply when data is used for e.g. non-scientific or commercial purposes. Separate fees may apply for the reproduction and delivery of data when web-based transfer of data is not possible.
9. In general, data shall be made available by COOPEUS infrastructures as soon as possible and without undue delay. COOPEUS infrastructures may apply exclusive scientific usage rights, which need to be defined within their own data policies. Such exclusivity periods shall not exceed two years.
10. Non-publicized research data shall be made available on demand to researchers representing COOPEUS infrastructures - as possible within the legal, ethical and policy framework of the data holding infrastructure. Each COOPEUS infrastructure shall nominate 'data stewards' to facilitate such data requests.
11. COOPEUS infrastructures acknowledge national and international intellectual property rights regulations. Each infrastructure is responsible for the warranty of copyrights and intellectual property rights which may apply for its data holdings.
12. COOPEUS infrastructures will clearly indicate licenses and terms of use for each dataset in the corresponding metadata.
13. If data or information produced by COOPEUS research infrastructures is used in published or unpublished work, attribution for the used resources is required. Data citations shall exclusively use the information provided within the metadata of each data set. Data which is not publicly accessible but has been disclosed to individual researchers through COOPEUS require explicit permission of the responsible PI prior to the publication of results derived from this data if requested by the data use policy.

The COOPEUS interoperability maturity index and online self-assessment

Criteria for the COOPEUS interoperability maturity index

We have established a set of basic criteria defining the level of interoperability of a research infrastructure in terms of data as well as metadata exchange. These criteria were established based on the COOPEUS status quo assessment, the commonalities matrix shown above, and the gap and commonalities analysis. Similarly, we have used the joint data policy and the gap and commonalities analysis as a template to define a simple set of criteria an infrastructure ideally should fulfill if it aspires to join the COOPEUS community:

After intensive discussions and revisions submitted by several RIs, a final set of criteria for metadata and data interoperability maturity was agreed upon by all partners:

Metadata related criteria:

1. The Research Infrastructure assigns appropriate metadata to each data set
2. The Research Infrastructure provides metadata in a standardised, electronic format(s)
3. The Research Infrastructure provides access to metadata catalogue(s) via standardized internet interface(s)
4. Metadata formats and interfaces are harmonized within my Research Infrastructure
5. Metadata access is harmonized or centralized within my Research Infrastructure (common access point)

Data related criteria

1. The Research Infrastructure provides and shares its data in electronic format via the internet
2. The Research Infrastructure provides its data in (a) standardized electronic format(s)
3. The Research Infrastructure provides its data via standardized internet interface(s)
4. Data formats and interfaces are harmonized within my Research Infrastructure
5. Data access is harmonized or centralized within my Research Infrastructure (common access point)

Data policy criteria for the COOPEUS interoperability maturity index

1. The Research Infrastructure has established and published a data policy which defines clear rules for disclosure or release of data
2. This data policy is harmonized, approved and valid within my Research Infrastructure
3. The Research Infrastructure allows unrestricted, open access to its metadata
4. The Research Infrastructure allows unrestricted, open access to its data
5. The Research Infrastructure allows unrestricted access to its interfaces

Rating and self-assessment

In an initial approach, a Google forms online questionnaire (https://docs.google.com/forms/d/1QmSUMoZyDyU_mluzWo5ZU8Wi1vi0aB2phYGd9WBNbVE/viewform) has been created based on these criteria and was sent to the COOPEUS RI group to test the practical usability of the COOPEUS interoperability maturity index. Several RIs have contributed to this test and based on the results of this initial assessment, the response types for each question (no, yes, partly) have been rated from 1 to 3 (3=yes, 1=no, 2 partly) in order to allow a ranking and numerical analysis of responses.

The screenshot shows the 'STEP 2 OF 3' questionnaire. It contains five questions, each with radio button options for 'no', 'yes', and 'partly'. The 'yes' option is selected for all questions.

COOPEUS INTEROPERABILITY MATURITY INDEX
STEP 2 OF 3

back proceed

Data provision:

My Research Infrastructure provides and shares its data in electronic format via the internet
 no
 yes
 partly

My Research Infrastructure provides its data in (a) standardised electronic format(s)
 no
 yes
 partly

My Research Infrastructure provides its data via standardised internet interface(s)
 no
 yes
 partly

Data formats and interfaces are harmonised within my Research Infrastructure
 no
 yes
 partly

Data access is harmonised or centralized within my Research Infrastructure (common access point)
 no
 yes
 partly

back proceed

Fig. COOPEUS interoperability maturity self-assessment tool

To calculate a numerical value of the COOPEUS interoperability index we simply chose to calculate the average of rated responses, which consequently may lay between 1 and 3. To easily estimate the maturity of each RI we further proposed to rank all average values between 1 and 1.5 as 'poor', between 1.6 and 2.2 as 'fair' and above 2.3 as 'good'.

The screenshot shows the 'RESULT' page of the questionnaire. It displays four rows of results, each with a numerical value and a colored circle indicating the maturity level: 'good' (green) or 'fair' (yellow). A 'back' button is visible at the bottom left.

COOPEUS INTEROPERABILITY MATURITY INDEX
RESULT

Metadata Maturity Index: 2.6 ● good

Data Maturity Index: 2.6 ● good

Data Policy Maturity Index: 2 ● fair

COOPEUS Interoperability Maturity Index: 2.4 ● good

back

Fig. COOPEUS interoperability maturity self-assessment tool result page

To allow COOPEUS and other RIs to perform anonymous self-assessments of their individual interoperability maturity according to the criteria described above we decided to use the COOPEUS index to set up a user-friendly online tool. This tool is available via the COOPEUS website at <http://www.coopeus.eu/documents/>. At this web page, the interested user has to follow the link labelled as 'interoperability index – questionnaire'. Direct access to the tool

is additionally possible at <http://dataportals.pangaea.de/coopeus/maturity.php> . The tool provides three assessment pages (metadata, data and policy) as well as a final rating page. No login is required to access the tool and it allows anonymous self-assessments, which COOPEUS agrees is an initial step in joining the COOPEUS international cooperation framework.