IS-ENES3 Deliverable D3.3

Standards Synthesis

Reporting period: 01/01/2022 – 31/03/2023

Authors: Martin Juckes (UKRI), Lars Bärring (SMHI), Antonio S. Cofiño (UNICAN), Martina Stockhause (DKRZ)
Reviewers: Sylvie Joussaume (CNRS-IPSL), Sandro Fiore (University of Trento)
Release date: 15/03/2023

ABSTRACT

This document covers work on building consensus on standards, setting objectives and requirements to guide development work and service delivery priorities. IS-ENES3 is also supporting technical developments enhancing standards in WP10 and delivery of services which extend standards to meet new community requirements in WP7.

Dissemination Level	
PU	Public

Revision table			
Version	**Date**	**Name**	**Comments**
Release for review	25/03/2023	M. Juckes	
Reviewed	14/03/2023	S. Joussaume, S. Fiore	Corrections done by M. Juckes after comments from review
Final version	15/03/2023	S. Morellon	Final formatting

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824084

Table of contents

Executive Summary

Both for the exchange of data within the community and the dissemination beyond, data standards play an essential role. The ENES Climate Data Infrastructure (CDI) provides international leadership in promoting and developing standards which can support the rapidly evolving requirements of the climate modeling research community. This involves work on domain standards such as the Climate and Forecast Convention (CF), Earth-System Documentation (ES-DOC) and the CMIP Data Request, as well as engagement in broader activities such as the Research Data Alliance.
IS-ENES3 has supported the expansion of community standards to support new and emerging requirements. The work reported on here covers existing standards for metadata describing physical parameters and the encoding of data in files and new standards for scientific provenance of model evaluation results.

Standards development work is supported in IS-ENES3 throughout the work programme, including networking, service and development work packages.

The reporting is organised into three streams of activity:

- Vocabularies and Standards for Climate Model Data
- Standards supporting Data Exploitation Services
- Standards for Trust

The standards work is feeding into the planning work for CMIP7, with multiple members of the IS-ENES3 team engaged in key roles in the initial stages of CMIP7 developments.

1. Objectives

Both for the exchange of data within the community and the dissemination beyond, data standards agreed within the community play an essential role. The ENES Climate Data Infrastructure (CDI) provides international leadership in promoting and developing standards which can support the rapidly evolving requirements of the climate modeling research community. This involves work on domain standards for metadata such as the Climate and Forecast Convention (CF), Earth-System Documentation (ES-DOC) and the CMIP Data Request, as well as engagement in broader activities such as the Research Data Alliance.
IS-ENES3 has also supported the expansion of community standards to support new and emerging requirements. The work reported on here therefore covers existing standards for metadata, describing physical parameters and the encoding of data in files, and new standards for scientific provenance of model evaluation results.

2. Description of work: Methodology

This document covers work on building consensus on standards, setting objectives and requirements to guide development work and service delivery priorities. IS-ENES3 is also supporting technical developments enhancing standards in WP10 and delivery of services which extend standards to meet new community requirements in WP7. They are reported in detail in other documents, as listed here:

- **Community Engagement:**
  - **D3.4 (Task 4.2, mo 36): CMIP documentation requirements.** The report focuses on CMIP documentation user requirements for existing and future ES-DOC services and scope. See D3.4 CMIP documentation requirements (Hassel et al. 2021).
  - **M3.5 (Task 1, mo 26) Workshop on climate indices.** The report describes the evaluation of existing indices, requirements for new ones from the various user groups and related technical requirements for the IS-ENES infrastructure. See M3.5 Report: Workshop on climate indices - Eastern Europe perspective (Djurdjevic et al. 2021).
  - **M5.3 (Task 5, mo 20): Requirements for technical standards for diagnostic tools.** Gather all the requirements from the community and produce a list. This list will be the basis to build the technical architecture for the diagnostics plugin (BSC)[1]. [published 18th June, 2021]
  - **D5.2 (Task 5, mo 24): Technical standards for diagnostic tools.** The document emphasizes technical solutions chosen to improve the compatibility between diagnostic tools. See D5.2 Technical standards for diagnostic tools (Vegas et al. 2022).

[1] https://is.enes.org/documents/milestones/m5-3-requirements-for-technical-standards-for-diagnostic-tools/view

- **Research and Development:**
  - **M10.2 (Task 6, mo 18): CMIP data request schema 2.0.** The data request schema used for CMIP6 will be upgraded and released to support CMIP7. The schema proposes an information model to be used for the CMIP7 Data Request (STFC). https://doi.org/10.5281/zenodo.4287148
  - **M10.3 (Task 6, mo 24): Climate indicators/indices and file metadata specifications and tools.** Short report on internationally agreed standards for metadata and DRS specifications for climate indices and indicators. List and description of compliant tools. Description of the integration of a provenance model for processing (M10.3 - Climate indicators/indices and file metadata specifications and tool: Bärring, 2021).
  - **D10.4 (Task 5, mo 46): CMIP6 documentation.** CMIP6 documentation on climate models, simulations and ensembles, experiments and conformance to experimental protocol, delivered via web services and tools developed in IS-ENES3 (CNRS-IPSL).[2]
  - **M10.4 (Task 6, mo 48): Update of the climate indicators/indices and file metadata specifications and tools.** Update of the M10.3 report on internationally agreed standards for metadata and DRS specifications for climate indices and indicators. List and description of compliant tools. Description of the integration of a provenance model for processing (M10.4 – Climate indices and file metadata specifications and tools – Update of M10.3: Bärring, 2023).

- **VA2 Activities**
  - **D7.5 (Task 5, mo 42): Report on operational support for CMIP documentation.** Report on ES-DOC operational support to CMIP (CNRS-IPSL).[3]
  - **D7.3 Second KPI and TA report for ENES CDI data services.** Report on all aspects of the Climate Data Infrastructure, including reporting on CF COnventions support and the CMIP6 Data Request.[4]

The presentation of work has been organised into three streams:

1. Vocabularies and Standards for Climate Model Data
   - There is an extensive ecosystem of domain specific vocabularies, ontologies and protocols anchored in broader cross-domain standards. The domain specific work is critical for communication between different elements of the climate science community.
2. Standards Supporting data Exploitation Services
   - Data standards are often defined in terms of metadata to describe static digital objects. There is a well-defined file or set of files which has been generated by some

[2] https://raw.githubusercontent.com/IS-ENES3/IS-ENES-Website/main/pdf_documents/IS-ENES3_D10.4.pdf
[3] https://raw.githubusercontent.com/IS-ENES3/IS-ENES-Website/main/pdf_documents/IS-ENES3_D7.5.pdf
[4] https://raw.githubusercontent.com/IS-ENES3/IS-ENES-Website/main/pdf_documents/IS-ENES3-D7.3.pdf

completed process and the metadata needs to capture essential information about the process and the file(s) produced. This stream of work, in contrast, deals with data management issues in complex and dynamic situations. The critical complexity here is not in the technical detail but in the distributed decision-making structure, and the dynamic nature of the problems is driven by the efforts of dozens of independent data providers to deliver large data volumes against tough deadlines.

- Developing standards for climate indices commonly used as indicators of climate impacts.

3. Standards for Trust
    - Standards associated with institutional process and transparency are critical for establishing trust. These are cross-domain standards, but applying them in the operational context of climate model data archives, with specific issues around high data volumes and flow rates and heavy dependency on automated processes, creates domain specific challenges.

3. Results

There are a great many different lines of work on standards in the IS-ENES community. Here, we present these in terms of three broad streams of activity:

- Stream 1 dealing with standards for the data generated by climate models and the climate models themselves;
- Stream 2 for standards supporting data exploitation and
- Stream 3 for standards associated with general principles of data curation and trust.

• Stream 1: Vocabularies and Standards for Climate Model Data

A series of meetings have been held and are described in more detail in M3.4 Summary of workshops on standards. These meetings covered standards and protocols used in the creation and management of data files used to store and disseminate CMIP6 data.

These meetings have been important to provide input to following results:

Earth System Model Documentation (ES-DOC)

The ES-DOC[5] software ecosystem facilitates both the provision and the consumption of documentation of the CMIP6 workflow and, where possible, automates the various and often complex stages involved. It offers services for metadata search, comparison and creation, following the CIM standard (Common Information Model) and provides an environment to document the modeling workflow.

[5] Earth System Documentation (https://es-doc.org/)

The ES-DOC eco-system of tools and services are founded upon five pillars:

- a set of controlled vocabularies,
- a common ontology,
- a set of specialisations,
- a public archive,
- an errata registry.

The CIM has evolved over time based upon community feedback for the express purpose of documenting climatology experiments, simulation & models. The underlying schema is governed via a light-weight process that permits the ontology to evolve whilst still ensuring that associated tools & services are kept in sync (D10.4).

In additional work funded by the Copernicus Climate Change Service (C3S), information from ES-DOC flows through the Copernicus Climate Change Service (C3S) platform and into the European Climate Data Explorer (ECDE)[6] portal. During implementation, the reverse process of updating ES-DOC to match the end-user requirements has also shown how a vertically integrated workflow from modeling community to end-users can give feedback to enhancements in the core services. Such enhancements can be lost when down-stream users find it easier to implement local fixes rather than engage in developing upstream solutions. The infrastructure needs to seek out opportunities for good positive feedback.

The CMIP Data Request

The data request of the Coupled Model Intercomparison Project Phase 6 (CMIP6) defined all the quantities from CMIP6 simulations that should be archived. This included both quantities of general interest needed from most of the CMIP6-endorsed model intercomparison projects (MIPs) and quantities that are more specialized and only of interest to a single endorsed MIP. The complexity of the data request has increased from the early days of model intercomparisons, as has the data volume. In contrast with CMIP5, CMIP6 required distinct sets of highly tailored variables to be saved from each of the more than 200 experiments. This placed new demands on the data request information base and leads to a new requirement for development of software that facilitates automated interrogation of the request and retrieval of its technical specifications.

IS-ENES3 supported the finalisation of the CMIP6 data request in version 1.0.33[7] and additional work on developing a revised schema for CMIP7 reported in M10.2 CMIP data request schema 2.0.

M10.2 CMIP data request schema 2.0 sets out the strategic and technical requirements for a revised request designed to support CMIP7.

[6] https://climate-adapt.eea.europa.eu/en/knowledge/european-climate-data-explorer

[7] https://pypi.org/project/dreqPy/1.0.33/, published November 2020. See also Juckes et al. 2020.

Issues and Possibilities

The M10.2 framework specifies a structured approach. This can be used to create a clear definition of the MIP Variables, including complex constraints. The short name of MIP Variables needs to be chosen such that any file names generated are unique. Implementing this will depend on resolution of discussions around the file naming convention (see below). In its current form, the constraint is that the combination "<variable_name>, <frequency>, <realmID/areaID>, <vertID>" is unique. This constraint can be expressed cleanly as an ISO 11179 Data Type provided that the frequency, realmID and vertID vocabularies are defined and citable.

The discussion on realmID versus area type as a file name element will influence the constraint on variable names. In the CMIP6 data request we have variables such as "shrubFrac", "clayFrac", "vegFrac" which all belonged to the "Lmon" table. With the new scheme, it would be possible to call them all "areaFrac".

The concept of "MIP Variable" was introduced in CMIP6 to allow common terms, such as "tas", to be re-used consistently with different data sampling strategies. The somewhat arbitrary nature of the separation of variables into MIP Tables prevented the implementation of a fully systematic approach to managing the relationship between MIP Variables and CMOR Variables. The CMOR variables could have the same title or a modified title, but the concept of inheritance was not well defined in the Data Request schema.

Recent work led by the WGCM Infrastructure Panel is proposing a new approach to CMIP nomenclature which would amount to a revision to the scope_rule to span multiple CMIP eras, and a modification to the uniqueness constraint within the lexical rule. Two versions are under discussion: names should be unique when used in combination with labels of temporal sampling, vertical sampling, horizontal sampling and an area type label[8]. This work is feeding into preparation for CMIP7.

Exploiting CMIP6 Data Request tools for additional MIPs

CMIP is a trendsetting project for the global climate modelling community. Work done in CMIP can also serve other WCRP activities, and in doing so broaden the scope of the standards. Two examples are listed here:

Coupled Chemistry Climate Initiative 2022 -- CCMI 2022

CCMI 2022 is a SPARC[9] initiative involving many groups that participated in CMIP6, but also some new participants. The data management framework is closely based on CMIP6, limiting changes to modifications in vocabularies which do not require changes to software tools and services deployed for CMIP6.

[8] https://docs.google.com/document/d/1FHqbU2qikt92mApcaEgYY-10O2_oJrTWK_pO4omwpo8/

[9] Stratosphere-troposphere Processes And their Role in Climate, a WCRP core project

Checking of validity of file metadata syntax is done with ceda-cc, which has been migrated to python 3 and updated to accept vocabularies in the JSON dictionary format pioneered in CMIP6 by PCMDI. One modification is to combine all the Controlled Vocabulary elements into a single file, "CCMI2022_CV.json". This simplifies the file management aspect of software configuration. The vocabularies and variable tables are held at CEDA.[10]

SNAP (Stratospheric Network for the Assessment of Predictability)

The SNAP[11] project is running a series of seasonal prediction experiments, also following the CMIP6 approach and, following CCMI2022, combining CV dictionaries into a single JSON file. Additional terms have been added to explicitly list the file name template and the file path template in the JSON file:

```
      "filename_template": [
"<variable_id>_<table_id>_<source_id>_<experiment_id>_<subexperiment_id>-
<variant_label>_<grid_label>[_<time_range>].nc"
       ],
       "path_template": [
"<mip_era>/<activity_id>/<institution_id>/<source_id>/<experiment_id>/<subexp
eriment_id>/<table_id>/<variable_id>/"
]
```

Climate and Forecast Conventions (CF)

The Climate and Forecasts Conventions (CF) promote interoperability of climate data by providing a clear and unambiguous standard for representing spatio-temporal information, physical quantities that the data represent, and other ancillary information useful in interpreting the data or comparing it with data from other sources. The CF conventions document and the CF controlled vocabularies such as the standard names (scientific parameter names) are updated periodically in response to proposals put forward by the scientific user community. Proposals to modify the conventions and the standard name table are submitted via GitHub issues in the https://github.com/cf-convention/cf-conventions and https://github.com/cf-convention/discuss repositories, respectively. All proposals are subject to public discussion by the CF community with the aim of achieving consensus on the changes before they are added to the published documents. Members of the CF conventions and standard names committees provide comments and act as moderators in order to move the discussions forward and assist in reaching consensus. This ongoing service is part of WP7.

In addition to the ongoing discussions (part of service support in WP7), a number of organisational issues have been addressed to assist members of the wider scientific community when engaging with the CF processes. Workflows for proposing enhancements and modifications to the CF conventions, including how to use GitHub as part of the process, have been documented on the CF

[10] https://github.com/cedadev/ccmi-2022

[11] Stratospheric Network for the Assessment of Predictability, a SPARC initiative

website. New documentation has also been written to describe the process for proposing, discussing and agreeing changes to the standard name table, http://cfconventions.org/standard_name_rules.html. The workflow for standard names is somewhat different to that for the conventions document, partly to take account of the more rapid timescale from proposal to publication, and also due to differences in the publication process. Synchronized versions of the standard names are published both on the CF website and in the NERC Vocabulary Server (NVS) which is becoming an increasingly important resource for users wishing to consume standard names in both human and machine-readable forms.

A file naming convention for climate model output

IS-ENES3 partners are contributing to two ongoing discussions about evolving metadata standards which are being led within the WGCM Infrastructure Panel:

- Harmonizing Metadata and Filenames Across CMIP Eras and WCRP Activities
    - This work aims to create a framework which enables consistency in file naming patterns and descriptive metadata to be maintained between CMIP eras.
    - The discussion around the paper has clarified many issues involved.
    - The paper is still in draft version, but emerging themes are:
        - The use of MIP table names in file names will be avoided in preference for a system based on objective attributes of the parameter, such as temporal and vertical sampling.
        - File names are likely to be longer and possibly cryptic.
- Climate experimentation beyond CMIP6: CMIP6Plus
    - This discussion attempts to counter the tendency for CMIP data to be a class-apart from other climate simulations because of a range of data services which are used exclusively for CMIP and a small range of additional activities.

Working with the RDA on vocabulary metadata

CEDA contributed to the RDA "InteroperAble Descriptions of Observable Property Terminologies (I-ADOPT) WG Outputs and Recommendations" which set out an ontology for describing parameter vocabulary metadata. That is, a means of describing the terms used to identify parameters in different vocabularies is set out as a means to enhance interoperability between vocabularies, in various communities within and across scientific domains.

Stream 2: Standards supporting data exploitation

In this section we describe standards which support communities making use of climate model data archives, particularly in the areas of climate change assessment and climate services.

Climate Change Assessment: working with the IPCC

IS-ENES3 partners contribute to many areas of standards development for the Intergovernmental Panel on Climate Change (IPCC). With the more flexible working approach that developed during the COVID pandemic this involved multiple ad-hoc meetings, some explicitly sponsored by IS-ENES3 and others sponsored by the IPCC Data Distribution Centre (DDC) or Task Group on Data (TG-DATA). In either case, expertise developed through the IS-ENES3 collaboration contributed to the successful outcomes. This work is supported by DKRZ through WP7.

FAIR Data

IS-ENES3 members from DKRZ, UC, and CEDA (UKRI), worked with TG-DATA and the IPCC WG1 Technical Support Unit to design and implement FAIR Data Guidelines for the IPCC 6th Assessment Report (Pirani 2022a). Implementation of these guidelines will, after publication, result in:

(1) direct links from the online version IPCC 6th Assessment Report to standards compliant catalogue entries in the CEDA catalogue. This will enable greater transparency about the contributions of data from modeling centres to the assessment report.
(2) agreed standards for archiving primary data at DKRZ.
(3) agreed standards for the presentation of data in the WGI Atlas (see MS3.4).

Licensing

IS-ENES3 partners contributed to productive and influential discussions about the CMIP and IPCC licensing frameworks, leading to agreement on licenses which enables a balance between flexible exploitation of CMIP data and protection of the rights of the data creators. The initial work was done through the WGCM Infrastructure Panel (WIP).

Licensing is a crucial aspect of distribution that allows for the wide-spread use of products with legal security. This applies in principle to large software packages such as climate models, small software packages such as evaluation packages, and standards documents. The current state of important products in the community is varying. Many small software packages have adopted Open Source licenses. Large software packages often have to deal with considerable legacy and large consortia, which makes this process inherently more difficult. For community standards this is also an ongoing process, see for example the discussion on-going for the CF Conventions, that have no license at the moment, at https://github.com/cf-convention/cf-convention.github.io/issues/182.

The CMIP6 data is covered, for the first time, by a standard license which gives clarity about the legal status of the data. The license selected is the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 license.

CMIP6 data files contain a "license" attribute with descriptive text, e.g.

```
 license = "CMIP6 model data produced by IPSL is licensed under a Creative
Commons Attribution-NonCommercial-ShareAlike 4.0 International License
(https://creativecommons.org/licenses). Consult
```

```
https://pcmdi.llnl.gov/CMIP6/TermsOfUse for terms of use governing CMIP6
output, including citation requirements and proper acknowledgment. Further
information about this data, including some limitations, can be found via the
further_info_url (recorded as a global attribute in this file) and at
https://cmc.ipsl.fr/. The data producers and data providers make no warranty,
either express or implied, including, but not limited to, warranties of
merchantability and fitness for a particular purpose. All liabilities arising
from the supply of the information (including any liability arising in
negligence) are excluded to the fullest extent permitted by law."
```

The Attribute Convention for Data Discovery recommends that the "license" attribute should «Provide the URL to a standard or specific license, enter "Freely Distributed" or "None", or describe any restrictions to data access and distribution in free text."

The standard licenses are self-contained legal documents which do not allow inclusion of additional text, and use standard legal terms to refer to entities and documents which are relevant for the implementation of the license.

Creative commons refers to attribution, copyright, a notice referring to the license, disclaimer of liabilities and URI for the licensed material which should be included if provided with the "Licensed material". The text provided in the CMIP files fits the idea of "license notice". Attribution information.

Implementing the License

Questions around implementing the license started to arise when the IPCC legal team drew attention to the implications of the license for inclusion of data and figures derived from CMIP data in the IPCC assessment reports. The legal formulation risked introducing unintended obstacles.

IS-ENES3 partners contributed to license guidelines for the IPCC developed by Huard et al. (2022). These guidelines were strongly informed by the discussion of CMIP6 license implementation and its impact on IPCC WGI. The work was carried out through the IPCC Task Group on data (TG-DATA) and the IPCC Data Distribution Centre (DDC) supported in WP7.

Data Citation

The Citation Service delivered through WP7 (a service offered by DKRZ as part of WDCC) is based on existing standards and recommendations of the data publishing community, which includes DataCite, Force11, Earth Science Information Partners (ESIP) and Research Data Alliance (RDA). Digital Object Identifiers (DOIs) conforming to the DataCite metadata standard are registered. The DataCite metadata standard and controlled vocabularies are used for metadata provision and exchange. In addition, the metadata in schema.org standard is implemented in the DOI landing pages, which is used by Google for its Dataset Search and by other social media. The recommendations for the provided metadata content include ORCID for researcher profiles and DOI references to articles documenting the data and to the forcing datasets provided by the project

input4MIPs. The Scholix interface, which exposes standardized data-data and data-literature relations, is regularly accessed and references to papers citing the CMIP6 data are added to the citation metadata.

Citation information is available as HTML (DOI landing page) for humans, as XML and as JSON for machine-access via DOI and persistent URLs, e.g.:

- https://doi.org/10.22033/ESGF/CMIP6.5262 (https://cera-www.dkrz.de/WDCC/meta/CMIP6/CMIP6.ScenarioMIP.IPSL.IPSL-CM6A-LR.ssp126)
- https://cera-www.dkrz.de/WDCC/meta/CMIP6/CMIP6.ScenarioMIP.IPSL.IPSL-CM6A-LR.ssp126.json
- https://cera-www.dkrz.de/WDCC/meta/CMIP6/CMIP6.ScenarioMIP.IPSL.IPSL-CM6A-LR.ssp126.xml

Citation links are provided in the ESGF search results and the CMIP6 citation search (http://bit.ly/CMIP6_Citation_Search with export possibility in csv). They can also be provided in the ES-DOC system, on the furtherinfo landing page which is the target for links embedded in all CMIP6 data files, and

See also the "ENES CDI Strategic Requirements" https://zenodo.org/record/6457603

Data Standards for Climate Indices

This work in WP4 looked at well-established indices which are widely used in the climate impacts community. The work was progressed through a series of workshops and meetings described in MS3.4 Summary of workshops on standards:

- Two online meetings were held with a focus on eastern European users. M3.5: Workshop on climate indices - Eastern Europe perspective.
- Informal discussions with Expert Team on Sector-Specific Climate Indices (ET-SCI).
- A series of online meetings focussed on well-established precipitation indices.

This lead to the development of a community platform for discussing and developing a unified view of metadata elements required to describe climate indices.

Key recommendations from MS3.4 identify the potential gains from working with the CMIP governance structure to embed standards in the next phase of CMIP and working with the Horizon Europe ClimatEurope2 project to develop and support standards relevant to climate services. The need to strengthen and extend standards to respond to advances in climate modeling and societal pressure was also highlighted.

Stream 3: Standards for trust

All of the above contribute to trust by creating transparent and reliable services. Some additional activities looking more directly at trust are described below.

Propagation of documentation and provenance information

IS-ENES3 has worked on the propagation of documentation and provenance information in the model evaluation process.

Documenting provenance across the model evaluation process is an instrumental step for instance in activities where decision making is built on such results, such as for the IPCC Assessment Reports. It allows us to build confidence in evaluation results, by ensuring traceability regarding the processing steps, their inputs and their computing context. It is also a valuable cornerstone for results reproducibility

However, provenance is a very general concept which can be instantiated in extremely various ways, and with numerous tools, even in the restricted context of model evaluation process. Selecting a Web standard such as the PROV family of specifications is of course a necessary step but this alone still allows for semantically non-compatible implementations of provenance handling. The first sensitive item in this respect is the vocabulary, or ontology, to use; this is clearly demonstrated by e.g. the number of entries in the directory of vocabularies and ontologies in Earth, Space and Environmental Sciences maintained by the Research Data Alliance's relevant group. Another key issue is providing tools for instrumenting diagnostic code provided by climate experts with provenance without putting significant additional constraints on such code.

In the context of IS-ENES3 WP3 Task 3.2, besides implementing provenance handling in ESMValTool, handling provenance in model evaluation has been jointly addressed during three events :

- The "Virtual workshop on requirements for a fast and scalable evaluation workflow" (18-19 May 2021, with a dedicated breakout group)
- Two dedicated online meetings, on 14 October and 8 November 2021, with nine (resp. eleven) participants from DLR, KNMI, U.Cantabria, CNRS-IPSL, DKRZ, CERFACS and PREDICTIA

As a result, the proposed framework based on Web standards includes a general setup exemplified by ESMValTool's implementation for provenance handling, an ontology based on METACLIP, the optional use of provenance templates as experimented in Climate4Impact, and possibly the services of the PROV API and the METACLIP viewer. These elements are further described in the "*IS-ENES3 White Paper on provenance handling in the model evaluation process*" (**https://doi.org/10.5281/zenodo.5759571**), together with explanations about how ESMValTool and icclim could take advantage of such tools.

The PROV framework has been implemented in the C4I to produce provenance records for use cases and processing scenarios. This has not been achieved with the ESMValTool because the latter is distributed as code for execution in the users environment, rather than being operated as a web service. This mode of operation creates extra barriers which have not yet been overcome. Specifically, there would be a need for a central portal to curate and host provenance information generated by users.

Although the description of task 3.2 includes a "*work in collaboration with other international efforts and the WGCM/WGNE climate diagnostics and metrics panel*", it appeared that this panel has been relatively inactive the last few years" (see WGCM22 meeting report, Gleckler and Guilyardi, 2019) and that building an alternate "WCRP model evaluation panel or work group" is still in progress.

Long Term Preservation and Trust

Climate data underpins policy decisions like the Paris Agreement. It is important for their traceability to preserve core data as well as core components of the ENES Research Infrastructure (ENES-RI) for the long-term. This is also a prerequisite for the trust in the data and services provided by the ENES-RI and the ENES partners. The data management principles of FAIR data and TRUSTworthy (TRUST: Transparency, Responsibility, User focus, Sustainability and Technology) repositories provide guidance. The CoreTrustSeal criteria, which are part of the Regular Membership application process of the World Data System (WDS), implemented the TRUST principles for research data repositories setting a basic standard for their operations on the long-term preservation of their data and services. The IPCC DDC Partner DKRZ is a Regular Member of the WDS and complies with the CTS standard. The FAIR data principles target on data and metadata quality enhancing its reusability, whereas the TRUST principles and the effort of CTS certified repositories keep the data FAIR on the long-term. IS-ENES3 work in this area is further strengthened by DKRZ represention on the EOSC Long-Term Data Preservation Task Group.

Other guiding principles supporting trust in data and data services are provided by the Open Science/Open Data initiative and Good Scientific Practice, including data citation and open licence aspects. The IPCC FAIR Guidelines combined these principles to enhance the transparency of AR6 results. The IPCC AR6 WGI Atlas included data references in their provenance metadata.

Apart from these guiding principles, long-term preservation needs to be built on international standards to support the maintenance of the service components, such as provided by W3C and the Open Geospatial Consortium (OGC).

3. Difficulties overcome

- **COVID-19**

The COVID-19 pandemic disrupted plans for a series of workshops. The project was able to switch to online meetings, but this was done in a reactive rather than proactive mode to deal with the imposed constraints. With hindsight more could have been done to maximize the value obtained from online meetings. Much of the work was done in the early phase of the pandemic when the community was just starting out on the learning curve for the new ways of working.

- **Engagement**

The value of good community standards is widely recognised, but creation of good standards depends on having broad community involvement in the development of requirements and approaches. This involvement is often difficult to achieve, particularly in the context of international standards with multiple communities. The broad range of stakeholders, with specialties ranging from highly technical computational work associated with high capacity and high velocity data management and analysis through to global policy support work, leads to considerable communication challenges.

- **Requirements**

Achieving clarity on the requirements for standards in sufficient details to enable reliable implementation is an enduring problem. The desired outcome, in the form of frictionless exchange of information, is clear, but few users are able to engage constructively with the level of technical detail required.

5. Conclusions and recommendations

IS-ENES3 has been engaged in a broad range of activities developing and supporting standards. This deliverable provides an overview of work. This work supports the efficient operation of ENES CDI data services, and also has a wider impact through adoption of the standards in the broader climate modeling community.

With the transition to CMIP7 a suite of new "Task Teams" are being established to oversee elements of the process. These task teams will provide a forum for developing and building on standards work started in IS-ENES3. Eight task teams have been launched dealing with:

- Forcings
- Ensemble design
- Scenarios and impacts
- Data request
- Model benchmarking
- Model documentation
- Data access
- Data Citation

The IS-ENES3 standards community is providing three task-team co-chairs (David Hassell (UREAD-NCAS), Martin Juckes (UKRI) and Martina Stockhause (DKRZ)) and multiple task team members. The strong European presence reflects the highly respected role of the IS-ENES community internationally which has been generated through IS-ENES3 and precursor projects.

Work on standards reported in this deliverable also exhibit some further needs with regards to user engagement in standards and how to measure their impact. Indeed:

- The processes around standards development and role of standards in enabling service delivery are rather opaque to users and this inhibits user engagement in standards development. **Pathways for greater user understanding of and engagement in standards development are needed.**
- Usage and impact of standards is not well covered by metrics. The development and exploitation of standards is seen as a high-value activity by data service professionals, but the value is difficult to quantify and convert into the resource streams needed to maintain the activity. **Metrics and indices of the usage and impact of standards are needed.**

References

Hassel et al. 2021: IS-ENES3 Deliverable D3.4: CMIP documentation requirements. https://raw.githubusercontent.com/IS-ENES3/IS-ENES-Website/main/pdf_documents/IS-ENES3_D3.4.pdf

Djurdjevic et al. 2021: IS-ENES3 Milestone MS14/M3.5: Workshop on climate indices - Eastern Europe perspective https://raw.githubusercontent.com/IS-ENES3/IS-ENES-Website/main/pdf_documents/IS-ENES3_M3.5_vf.pdf

Juckes, M., Taylor, K. E., Durack, P. J., Lawrence, B., Mizielinski, M. S., Pamment, A., Peterschmitt, J.-Y., Rixen, M., and Sénési, S.: The CMIP6 Data Request (DREQ, version 01.00.31), Geosci. Model Dev., 13, 201–224, https://doi.org/10.5194/gmd-13-201-2020 , 2020 .

Vegas et al. 2022. IS-ENES3 Deliverable D5.2 Technical standards for diagnostic tools. https://raw.githubusercontent.com/IS-ENES3/IS-ENES-Website/main/pdf_documents/IS-ENES3_D5.2.pdf

Bärring, 2021: IS-ENES3 Milestone M10.3 - Climate indicators/indices(*) and file metadata specifications and tools. https://raw.githubusercontent.com/IS-ENES3/IS-ENES-Website/main/pdf_documents/IS-ENES3_M10.3.pdf

Anna Pirani, Andrés Alegria, Alaa Al Khourdajie, Wawan Gunawan, José Manuel Gutiérrez, Kirstin Holsman, David Huard, Martin Juckes, Michio Kawamiya, Nana Klutse, Volker Krey, Robin Matthews, Adam Milward, Charlotte Pascoe, Gerard van der Shrier, Alessandro Spinuso, Martina Stockhause, & Xiaoshi Xing. (2022). The implementation of FAIR data principles in the IPCC AR6 assessment process. Zenodo. https://doi.org/10.5281/zenodo.6504469

Huard, David, Pirani, Anna, Chen, Robert, Gutiérrez, José Manuel, Juckes, Martin, Krey, Volker, Spinuso, Alessandro, & Stockhause, Martina. (2022). IPCC Data and Code Licensing Policy (Version 1). Zenodo. https://doi.org/10.5281/zenodo.7197274

Bouwe Andela, Joaquin Bedia, Antonio Cofiño, Rémi Kazeroni, Christian Pagé, Daniel San Martín, Stéphane Sénési, Jérôme Servonnat, Alessandro Spinuso, Mats Veldhuizen, Klaus Zimmerman (2021): IS-ENES3 White Paper: Framework for the propagation of documentation and provenance information in the model evaluation process. https://zenodo.org/record/5759571