Climateurope2

Guidelines of good practices for vocabularies, formats, metadata and technical standards, including FAIR principles for data and software

Deliverable 2.6

Authors: Anna-Lena Flügel, Fanny Adloff, Lars Bärring, Xavier Domingo Soldado, Aleksandra Krzic, Julien Lenhardt, Christian Pagé, Alison Pamment, Charlotte Pascoe

Contributors: Nakiete K. Msemo, Amir H. Delju, Biljana Milic-Petrovic, Martina Stockhause

Funded by the European Union

Climateurope2

This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101056933.

Document Information

GRANT AGREEMENT	101056933
PROJECT TITLE	Supporting and standardising climate services in Europe and beyond
PROJECT ACRONYM	Climateurope2
PROJECT START DATE	01/09/2022
RELATED WORK PACKAGE	WP2
RELATED TASK(S)	T2.1
LEAD ORGANISATION	DKRZ
AUTHORS	Anna-Lena Flügel (DKRZ), Fanny Adloff (DKRZ), Lars Bärring (SMHI), Xavier Domingo Soldado (BSC), Aleksandra Krzic (RHMZ), Julien Lenhardt (SMHI), Christian Pagé (CERFACS), Alison Pamment (STFC), and Charlotte Pascoe (STFC)
SUBMISSION DATE	31/03/2026
DISSEMINATION LEVEL	public

History

DATE	SUBMITTED BY	REVIEWED BY	VISION (NOTES)
02/03/2026	Anna-Lena Flügel	Harilaos Loukos	CE2 internal review
24/03/2026	Harilaos Loukos		Review received
31/03/2026	Anna-Lena Flügel		Deliverable submitted

Please cite this report as: Flügel, A., Adloff, F., Bärring, L., Domingo Soldado, X., Krzic, A., Lenhardt, J., Pagé, C., Pamment, A., Pascoe, C. (2026), Guidelines of good practices for vocabularies, formats, metadata and technical standards, including FAIR principles for data and software, D2.6 of the Climateurope2 project.

Disclaimer: *Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Climate, Infrastructure and*

Environment Executive Agency (CINEA). Neither the European Union nor the granting authority can be held responsible for them

Table of Contents

About Climateurope2

Timely delivery and effective use of climate information is fundamental for a green recovery and a resilient, climate neutral Europe, in response to climate change and variability. Climate services address this through the provision of climate information for use in decision-making to manage risks and realise opportunities.

The market and needs for climate information has seen impressive progress in recent years and is expected to grow in the foreseeable future. However, the communities involved in the development and provision of climate services are often unaware of each other and lack interdisciplinary and transdisciplinary knowledge. In addition, quality assurance, relevant standards, and other forms of assurance (such as guidelines, and good practices) for climate services are lagging behind. These are needed to ensure the saliency, credibility, legitimacy, and authoritativeness of climate services, and build two-way trust between supply and demand.

Climateurope2 aims to develop future equitable and quality-assured climate services to all sectors of society by:

- Developing standardisation procedures for climate services
- Supporting an equitable European climate services community
- Enhancing the uptake of quality-assured climate services to support adaptation and mitigation to climate change and variability

The project will identify the support and standardisation needs of climate services, including criteria for certification and labelling, as well as the user-driven criteria needed to support climate action. This information will be used to propose a taxonomy of climate services, suggest community-based good practices and guidelines, and propose standards where possible. A large variety of activities to support the communities involved in European climate services will also be organised.

Executive Summary

Deliverable 2.6 of the Climateurope2 (CE2) project provides a coherent set of guidelines for good practices in vocabularies, formats, metadata, licensing, software, and data management, aligned with the FAIR principles (Findable, Accessible, Interoperable, Reusable). Its objective is to support CE2 to formulate recommendations for increased consistency, interoperability, transparency, and long-term sustainability across the climate services value chain in Europe and beyond.

Rather than introducing new standards, the report consolidates and structures widely adopted community standards and practices — including WMO frameworks, CF conventions, metadata standards, and FAIR4RS principles — into actionable recommendations tailored to climate services. The report formulates recommendations and identifies research gaps to support the standardisation work of Climateurope2 regarding good practices for FAIR vocabularies, formats, metadata and technical standards, for data as well as for software. Key guidance includes the use of established metadata standards, CF-compliant NetCDF or Zarr formats, FAIR and resolvable controlled vocabularies, open and clear licensing, DOI assignment for datasets and products, and lifecycle-oriented software management practices. The report also highlights the importance of regulatory alignment (e.g. GDPR, EU AI Act, Accessibility Act), Data Management Plans as living governance tools, and recognised quality and security standards.

The recommendations are informed by a review of existing standards and literature, as well as insights gathered through interviews and a targeted survey of climate service providers. Community feedback confirms the need for harmonised terminology, improved interoperability (especially across APIs and datasets), stronger software and documentation practices, and clearer incentives for certification. It builds on, updates and consolidates the preliminary recommendations presented in Deliverable 2.1 (Stockhause et al., 2024) of this project.

Several research gaps are identified, including the need for harmonised FAIR software metrics, verifiable implementation frameworks, improved interoperability standards, and clearer benchmarking approaches.

Overall, this deliverable establishes a structured, community-informed reference framework to strengthen the quality, credibility, and sustainability of climate services and to support future standardisation suggestions within Climateurope2.

Keywords

Climate service, FAIR, guidelines, good practices, data, metadata, software, standards, interoperability, sustainability, climate change, vocabulary, format, interviews, survey.

List of Abbreviations

CE2	Climateurope2 project
CF	Climate and Forecast metadata conventions
CMIP	Coupled Model Intercomparison Project
CORDEX	Coordinated Regional Climate Downscaling Experiment
CS	Climate Service(s)
C3S	Copernicus Climate Change Service
DOI	Digital Object Identifier
FAIR	**F**indable, **A**ccessible, **I**nteroperable, **R**e-usable
NMHS	National Meteorological and Hydrological Services
WCRP	World Climate Research Programme

1. Introduction

The growing number and increasing diversity of climate services in Europe—and worldwide—require a consistent and authoritative set of guidelines on good practices following the FAIR principles[1]. These guidelines are essential to support the quality, functionality, compatibility and interoperability of data, tools and processes across the climate services value chain. In the absence of shared guidance, the reuse, comparability and the long-term sustainability of climate services is hampered by fragmentation in vocabularies, formats, metadata and technical standards.

In response to this need, the objective of this deliverable is to recommend a coherent set of guidelines on good practices for vocabularies, formats, metadata and technical standards, including the application of FAIR principles to data and software. The guidelines are intended to provide a common reference framework for climate service providers, developers and users, facilitating alignment across sectors and promoting consistency, interoperability and sustainability. Rather than introducing new standards, the deliverable synthesises and structures existing standards and widely adopted practices into actionable guidance.

Within the Climateurope2 project, the the authors of this deliverable have undertaken a systematic review of current practices in the field of climate services related to data and software processing. Recognising the heterogeneity of practices across climate service communities, this document prioritises those data management approaches that demonstrate a relatively high level of uptake and maturity, and that can therefore be recommended as good practice. This deliverable intends to inform Climateurope2 in its task to generate recommendations and identify research gaps that assist in the standardisation process of climate services in Europe. The scope of D2.6 includes guidelines on FAIR vocabularies, formats, metadata and technical standards for data and software. It thus complements the other upcoming deliverables of Climateurope2, which will cover recommendations on provenance of processing methodologies (D2.7), uncertainty quantification and communication (D2.8), procedures for data verification, certification and quality management (D2.9) and the impact assessment that underpins climate risk analysis (D2.10).

The guidelines presented in this report consolidate the preliminary framework established in D2.1. (Stockhause et al., 2024). They were largely drawn from existing literature and reports, but also from the experience of the co-authors. The results have been further refined and expanded through additional input from the climate services community through qualitative interviews and a complementary survey. The report is structured into four main sections: (i) guidelines for good practices supporting FAIR climate data and software, (ii) key messages derived from interviews and surveys, (iii) recommendations to Work Package 1 for synthesis of the project results, and (iv) identified research gaps. The recommendations and the research gaps can be found in boxes within the respective chapters as well as in a compiled list in chapters 4 and 5.

2. Guidelines for FAIR Climate Services

As demand for high-quality data to support climate services—such as disaster risk reduction and climate change adaptation and mitigation—increases, a robust regulatory framework that defines interoperable standards, recommended practices, and procedures for data management in accordance with the FAIR principles is essential. The following guidelines are based on recommendations from some of the most influential organizations of the climate science community. However, they have also

[1] https://doi.org/10.1038/sdata.2016.18

been selected from the perspective and to the best knowledge of the authors and may not fully capture all discipline- or sector specific demand or needs.

The *WMO Manual on the High-quality Global Data Management Framework for Climate* (WMO-1238, 2023)[2] can be regarded as a good example of high-level standards and recommended practices for data stewardship to ensure that climate data on regional, national and global scale are reliable, accessible, sustainable, and as complete as possible. These practices address historic and real-time observational data from a wide range of platforms, including remote sensing, in situ, marine, and aircraft observations as well as climate model data for projections of future climate scenarios.

The provisions in the Manual are based on universally agreed data management goals and relate essentially to the following fundamental aspects, many of which can be mapped to various FAIR principles[3] (see *respective FAIR principle* in round brackets):

1. **Accessibility**: In the context of data management, "accessibility" refers to ensuring that the dataset is easily and conveniently downloadable or, in case of large datasets, can be processed within a cloud computing environment by users (*A1*);
2. **Data integrity**: Ensuring that data are recorded, preserved and are free from corruption or loss when transferred between systems or in storage throughout the data life cycle;
3. **Data quality control**: Using community best practices, including ensuring that the results of quality control are also documented (*F2*);
4. **Discoverability**: Ensuring that data and relevant information about the dataset may readily be found, including visibility in online catalogues (*F4*);
5. **Documentation**: Including all elements necessary to access, provide guidance for users, understand, and process the data, and including documentation on replication, reprocessing and updates(*F2, R1*);
6. **Governance**: Ensuring that accountability, responsibility and compliance mechanisms are well-defined and transparent (*R1.2*);
7. **Metadata**: Ensuring that information about data is publicly available, including full details of the origin and processing history of raw observations and derived products ("provenance metadata"), to ensure full traceability of the processing chain (*R1, R1.2*);
8. **Portability**: Ensuring that data are easily incorporated into users' working environment based on community standards (*R1.3*);
9. **Preservation**: Ensuring that data and metadata are protected from loss and preserved for future use in line with well-specified retention policies;
10. **Quality assessment of datasets**: Ensuring that datasets and data products are routinely reviewed, and the results of the review are transparent;
11. **Uncertainty**: Ensuring that uncertainty estimates are documented and made available. The quantification and communication of climate uncertainty are discussed in Climateurope2 deliverable D2.8..
12. **Usability and usage**: "Usability" describes how easily the data product may be understood and used by users and incorporated into a user's own working environment. It incorporates aspects of compatibility of the publication medium with community standards and supporting documentation. "Usage" refers to the degree of scientific credibility of the data among users, including through citations in peer-reviewed literature (*R1*).

[2] https://library.wmo.int/idurl/4/56975
[3] c.f https://www.go-fair.org/fair-principles/

The following subsections offer additional guidelines for ways in which these high-level goals may be met in practice regarding metadata, formats, vocabularies, data licensing and citation as well as software and data management. Key recommendations as well as research gaps have been highlighted in coloured boxes within each chapter. In addition, they can be found as a compiled list in chapter 4 “Recommendations” for convenient readability.

2.1 Metadata

High-quality metadata is crucial for achieving FAIR data management goals regarding information processing and dissemination by climate services to ensure data integrity, findability, accessibility, and (re-)usability. Based on documentation from the World Meteorological Organization (WMO), the Global Framework for Climate Services (GFCS), the Open Geospatial Consortium (OGC), the Group on Earth Observations (GEO), and other actors in the metadata sphere for climate services, best practices for climate services metadata in connection with data can be summarised as follows[4]:

1. **Standardised Metadata Formats:** Use established metadata standards such as WMO Core Metadata Profile (WCMP) (latest published version: v2.2.0, 2026[5]) or WIGOS Metadata Standard (WMDS)[6] for weather and climate observation data, CMIP metadata standards[7] for model output data, and ISO 19115-1[8] or DataCite Metadata Schema[9] for the findability of datasets in interdisciplinary contexts. All of these standards provide a structured way to document information.

2. **Clear and Comprehensive Descriptions:** Provide detailed information about the dataset. Include titles, abstracts, keywords, and descriptions that explain the purpose, content, and context of the data. This helps users understand what the data is about.

3. **Data Provenance:** Document the data source, including the instruments used, data collection methods, and organisations responsible for data collection. This helps establish data credibility and traceability (see CE2 Deliverable 2.7 on *Provenance* for more details).

4. **Temporal and Spatial Information:** Clearly specify the temporal and spatial coverage of the data. This should include start and end times, as well as geographical coordinates, resolution, and projection information.

5. **Data Quality and Uncertainty:** Include information on data quality, accuracy, precision, and any known uncertainties. This helps users assess the reliability of the data (see CE2 Deliverable D2.8 on Uncertainty and Climate Risk Assessment for more details).

6. **File Naming Conventions:** Use consistent and informative file naming conventions for data files. Include relevant timestamps, variables, and other metadata in the file names.

7. **Version Control:** Maintain version control for datasets. Document changes, updates, and revisions to ensure users are aware of the dataset's history.

[4] Stockhause et al., 2024
[5] https://wmo-im.github.io/wcmp2/standard/wcmp2-STABLE.html
[6] https://library.wmo.int/idurl/4/55626
[7] https://bit.ly/CMIP7-DReq-v1_2_2_3
[8] https://www.iso.org/standard/80275.html
[9] https://schema.datacite.org/

8. **Units and Parameters:** Clearly define the units of measurement for variables and parameters in the dataset. Include standard abbreviations and conversion factors where applicable.

9. **Data Access and Distribution:** Specify how users can access the data, including download links, APIs, or other methods. Ensure that data are available in open and machine-readable formats.

10. **License and Usage Terms:** Clearly state the terms of use, licensing, and any restrictions on data usage. Make it explicit if the data are open access or require permissions (see section 2.4 on Licensing and Citation).

11. **Metadata Cross-Referencing:** Link related datasets, publications, and documentation. This helps users find additional information and resources.

12. **Data Citation:** Encourage data citation by providing a standardised citation format. This helps acknowledge the data source and promotes proper attribution (see section 2.4 on Licensing and Citation).

13. **Data Visualization:** Include visual representations of data coverage, quality, and relevant graphs, maps, or images that help users quickly understand the dataset's context.

14. **Metadata Updates:** Ensure that metadata is regularly updated to reflect changes in the dataset, project, or data source.

15. **Metadata Accessibility:** Make metadata discoverable and accessible through metadata catalogues, data repositories, or data portals. Use standardised metadata search and discovery mechanisms.

16. **Interoperability:** Ensure that metadata conform to international standards and can be easily integrated with other climate data sources and tools, including machine interpretable or machine-to-machine exchange.

17. **User Support:** Provide contact information for users who have questions or need assistance with the data. Offer user support and FAQs when possible.

18. **Training and Documentation:** Offer documentation, tutorials, and training materials to help users understand and use the data effectively.

19. **Engage Stakeholders:** Involve stakeholders, including the scientific community and end-users, in the development of metadata standards and requirements to ensure they meet user needs.

20. **Review and Quality Assurance:** Implement a review and quality assurance process for metadata to ensure consistency and accuracy.

Metadata standards for observational as well as climate model data have been developed by the dedicated authorities, which are introduced in the following subsections.

Metadata for observational data

A comprehensive metadata standard covering all types of weather and climate observations should enable the (future) user to identify the conditions under which the observation or measurement was made as well as any aspects that may affect its use to determine whether the observations are fit for purpose. As an example, National Meteorological and Hydrological Services (NMHSs) affiliated with the WMO establish, maintain and update explanatory metadata and thus ensure data homogeneity. Data homogeneity is very important for the calculation of climate normals and averages.

The WMO Integrated Global Observing System (WIGOS) observations consist of an exceedingly wide range of data, from manual observations to complex combinations of satellite hyper-spectral frequency bands, measured in situ or remotely, from single dimension to multiple dimensions, and those involving processing. The WIGOS metadata describe the observed variable, the conditions under which it was observed, how it was measured or classified, and how the data have been processed, in order to provide users with confidence that the data are appropriate for their application. The **WIGOS Metadata Standard (WMDS)**[10] is a best practice example for metadata for observational data. It defines a comprehensive framework for documenting observational data to ensure interoperability and utility, covering 10 categories (see Table A1 in Appendix) and over 90 metadata elements. All of the categories are considered to be important for the documentation and interpretation of observations made, and even for their use in the distant future. Hence, the standard currently declares many elements that are not needed for applications intended for immediate use of observations, such as numerical weather prediction, aeronautical, transport sector applications and advisories, or other climate services. For these applications, profiles of the standard may be developed. The categories reflect the need to specify the observed variable; to answer why, where and how the observation was made; how the raw data were processed; and what the quality of the observation is. For a more in-depth look at WMDS, see CE2 Deliverable D2.1 (Stockhause et al., 2024, p. 8-10).

Metadata for climate model output

Climate and earth system models are the essential tools for producing projections of future climate pathways, which is an important information source for climate services. The overarching requirements on metadata for climate model data largely follow those of climate data from observational and remote sensing sources. However, data homogeneity is not so much an issue for climate model data. Descriptions of instrument and measurement technique and environment are replaced by description of the modelling system, its configuration and performance, which is a complex and far-reaching requirement. Basically, it falls into two categories: (i) technical description of the modelling system and its output, and (ii) the model's performance and related uncertainties.

For the technical description of the modeling system, several well established community standards exist (see Table 2). Together these form a comprehensive set of community standards for documenting climate model data and in particular global and regional climate projections. While the different standards are managed independently, there is close collaboration because of substantial overlap and interaction between the communities. As these standards are essential for implementing both **CMIP and CORDEX**, which are core projects of the World Climate Research Programme (WCRP), these standards fall under the authority of WCRP.

Table 1: Key community standards for documenting climate model data and experiments

Standard name	Focus area	Key Reference
Climate and Forecasting (CF) Conventions	NetCDF metadata conventions and vocabularies for describing climate data, including climate model data.	https://cfconventions.org Eaton et al. (2023)

[10] WIGOS Metadata Standard: https://library.wmo.int/idurl/4/55626

Attribute Convention for Data Discovery (ACDD)	Discoverability of climate datasets.	https://wiki.esipfed.org/Attribute_Convention_for_Data_Discovery_1-3
CMIP Specification	Specification of global climate and earth system model experiments, model output data formats, dissemination and archiving.	CMIP6 Homepage: https://pcmdi.llnl.gov/CMIP6/ CMIP6 Data Request (v1.00.31): https://doi.org/10.5194/gmd-13-201-2020 CMIP 7 Homepage: https://wcrp-cmip.org/cmip-phases/cmip7/ CMIP 7 Data Request (v1.2.2.3): https://bit.ly/CMIP7-DReq-v1_2_2_3
CORDEX Specification	Specification of downscaling experiments by regional climate models, model output format, dissemination and archiving.	https://cordex.org/ https://github.com/WCRP-CORDEX
ES-DOC	System and database for documentation of CMIP experiment and participating modelling systems and configurations.	https://github.com/es-doc

Regarding the model performance and related uncertainties, the situation is more complex as there is no direct way to assess performance and uncertainties in future projections. This is an area of intense research where definitive standards yet have to be defined. However, as there is a pressing need from users and stakeholders to provide such information, various best practices have emerged. For more information on Uncertainty and Climate Risk Assessment see CE2 Deliverable 2.8.

Metadata for interdisciplinary contexts

To increase the findability of climate data and CS data products in interdisciplinary research environments, using less discipline-specific metadata standards like DataCite Metadata Schema[11] or ISO 19115-1[12] is required to ingest the metadata into large-scale interdisciplinary metadata aggregators and search portals like OpenAIRE[13].

DataCite provides a standardized, machine-readable schema that is widely adopted in the research community. By assigning DOIs (Digital Object Identifiers) and including structured fields for creators, titles, subjects, descriptions, and funding information, DataCite ensures that climate datasets can be persistently identified, indexed, and linked to related publications, projects, and researchers. This strong integration into the scholarly communication ecosystem makes datasets more visible and interoperable across repositories and aggregators.

International Organization for Standardization's ISO 19115-1 standard complements this by offering detailed, domain-specific metadata for geospatial information. ISO 19115-1 supports precise

[11] https://schema.datacite.org/
[12] https://www.iso.org/standard/80275.html
[13] https://www.openaire.eu/

descriptions of geographic coverage, temporal extent, resolution, and data lineage. This enables more accurate filtering, map-based search, and expert-level discovery within aggregators.

Recommendation 1 (R1): Adopt established metadata standards
Use established metadata standards such as WMO Core Metadata Profile (WCMP) or WIGOS Metadata Standard (WMDS) for weather and climate **observation** data, CMIP metadata standards for **model data**, and the more general metadata schemas ISO 19115-1 or DataCite Metadata Schema for metadata in **interdisciplinary** contexts. These standards provide a structured way to document information and enhance findability, accessibility, and thus reusability of the data.

Recommendation 2 (R2): Provide high-quality metadata to enable data reuse
Provide high quality, standardised, and complete metadata including information on observation methods or model design, uncertainty and variable descriptions. This facilitates data homogeneity, understanding and future (re-)use of the data.

2.2 Formats

A comprehensive collection of data and file formats used for meteorological and climatological data are maintained by WMO[14]. Once the data has been processed and enriched to produce more user-oriented climate data and information other data formats may become more relevant either to interface with geographical information systems (GIS) and/or graphical data formats.

In climate services, there are several formats used in this context. The reader is referred to the section 2.1.3 of D2.1 to have a full description of all the formats. Here the focus is on guidelines and recommendations to the formats for climate services given the current landscape, standards and practices. From targeted interviews and landscape, it can be shown that the formats used highly depend on which part of the processing workflow they belong to. Three main parts of the workflows, that are generally used in most of the climate services, can be identified. In those, specific formats can be bound.

13. INPUT DATA: Input data from original data sources. Those data sources can have data that is close to raw or in other cases data has already been processed to distribute it in proper usable form.
14. MIDDLEWARE DATA: In-house data formats: the formats used within the climate services for custom data processing. This step is not present in some cases, like in the C3S that distributes data in a form that is closer to the input data sources.
15. OUTPUT DATA: Output formats: suitable for the end users, so it depends on which specific users the climate services are targeting.

The formats used in those generic three layers do not vary much from one climate service to another.

[14] https://codes.wmo.int/wmdr/DataFormat

For input data, the **NetCDF** format is very common and is by far the most widely used (**NetCDF** is using **HDF** as a backend). **Zarr** is sometimes used as well and it is highly recommended when using cloud platforms. It is understandable that we have NetCDF and Zarr because those formats are very adapted for climate data and a lot of tools and interfaces support those formats. Starting with **NetCDF** 4.8.0, the *NCZarr* implements a limited set of extensions to the Zarr data model.

- **NetCDF** (Network Common Data Form) is a format that essentially is defined by the reference software library developed and maintained by University Corporation for Atmospheric Research (UCAR). It is a collection of several closely related machine-independent binary data formats. The formats target multi-dimensional array-oriented datasets. The recent version builds on the HDF5 data format. The NetCDF formats are self-describing, which means that the data files contain essential metadata rather than relying on external sources. In combination with the Climate and Forecasting CF conventions the NetCDF format has become the *de facto* standard format for exchange and dissemination of climate model data. The documentation is hosted by the UCAR Unidata program[15]. NetCDF is adopted by OGC through a suite of standards[16].
- **HDF** (Hierarchical Data Format) is a self-describing machine-independent general purpose binary data format for storing numeric data. It is actively maintained by the producer[17]. As a binary data format it includes a range of data compression mechanisms that allows for efficient storage of data, which allows for handling huge data volumes. Except for what is needed to describe the internal data structures as such there are no specific requirements on metadata, which instead have to be organised by user communities. Such communities exist for various sectors, for example the climate data community make use of HDF via the NetCDF format, and the earth observation and remote sensing community make use of HDF-EOS. HDF5 is included as OGC Core standard OGC 18-043r3[18].
- **Zarr**[19] is a recent data format specifically targeted at storing huge multidimensional numeric datasets on cloud platforms. It is a community effort, where the format specification is now at version 3, which is in development. Similar to HDF it is a generic format where metadata standards have to be added on top of the basic Zarr specification. Zarr version 2 is adopted as OGC standard 21-050r1[20] [N.B. file download]), which specifically mentions links to NetCDF and CF Conventions for metadata standards. Specific functionality and conventions to handle geolocated data is under development through the recently formed OGC GeoZarr Standards Working Group.

For MIDDLEWARE DATA, **NetCDF** is used in the vast majority, because it is very convenient and has good performance. It is not a surprise since this format is also the native format used to export, distribute and exchange climate modelling data.

For OUTPUT DATA, formats used are more diversified. **NetCDF** is also used in majority, but for less experienced users, **CSV** is generally preferred for data, parquet/geoparquet are a very good alternative as well. Markup is another option that is also recommended. For end-user products, plots and images are provided in **PNG**. For reports, **PDF** (*PDF/A standard*) is good but **HTML5** may be a better solution depending on the context, because it can be more accessible and reusable (more FAIR) than PDF.

---

[15] https://www.unidata.ucar.edu/software/netcdf/
[16] https://www.ogc.org/standard/netcdf/
[17] https://www.hdfgroup.org/
[18] https://www.ogc.org/standard/hdf5/
[19] https://zarr.dev/
[20] http://www.opengis.net/doc/CS/zarr/2.0

Recommendation 3 (R3): Use CF-compliant NetCDF or Zarr as primary formats for climate data workflows
Given the large number of community tools available to process NetCDF or Zarr file formats, it is recommended to use these formats for the maximum number of data and processing steps within workflows. NetCDF is preferred for non-cloud environments, GRIB could also be used but is more complex but more efficient. Zarr is suitable for cloud environments. Both of these formats must use Metadata Standards and Vocabularies (such as CF-Conventions: see section 2.3 on Vocabularies).

Recommendation 4 (R4): Use community-specific formats for output products
Output of CS should use community-specific file formats. NetCDF/Zarr for most of the research communities, CSV can be used for smaller datasets to make data format more accessible for non-experienced users. BADC-CSV is recommended for CSV data; it is a structured format with metadata that is interoperable with netCDF metadata conventions. For reports, PDF (PDF/A standard) is good but HTML5 or Markup may be a better solution depending on the context, because it can be more accessible and reusable (more FAIR) than PDF, and machine-readable.

2.3 Vocabularies

Controlled vocabularies are related to three of the four FAIR principles: Findability, Interoperability and Reusability. They are used to provide detailed semantic information about many aspects of data, including:

- model, observing platform or instrument used to create or collect the data, i.e. provenance metadata
- experiment protocols and methodologies used in data collection and processing
- scientific parameter naming
- units of measure

Findability: Individual terms from within a controlled vocabulary can be used as search terms in data catalogues, and are key to finding data relevant to a specific query.

Interoperability: Terms within a controlled vocabulary should have clear and unambiguous descriptions to help ensure correct scientific interpretation of the data. Descriptions, together with units of measure, facilitate the construction of mappings between similar terms in different controlled vocabularies. Controlled vocabularies have often originated from within user communities and identifying synonyms and other relationships between different vocabularies increases the interoperability of data that adhere to different metadata standards. Mappings allow searches to be expanded to include similar terms from multiple vocabularies.

Reusability: Clearly and unambiguously defined vocabulary terms support data processing for immediate use cases, and enable repurposing to future use cases which may be very different from the reasons the data was originally collected or produced.

FAIR principle I2 requires that '(meta)data use vocabularies that follow FAIR principles' (Wilkinson, M., Dumontier, M., Aalbersberg, I. et al., 2016). This means that the vocabulary itself should follow the

FAIR principles[21]. A FAIR controlled vocabulary used to describe datasets needs to be documented and resolvable using globally unique and persistent identifiers. The vocabulary documentation needs to be easily findable and accessible by anyone who uses the dataset.

In practice, vocabularies may take a number of forms such as a list of terms, a glossary, a taxonomy or an ontology. The collective name for the information stored in these different forms of vocabulary is 'semantic artefacts'.

Semantic services such as the NERC Vocabulary Server (NVS)[22], the Research Vocabularies Australia vocabulary server[23] and the ICES vocabulary server[24] provide access to many vocabularies covering a range of environmental disciplines. Users may also submit vocabularies for publication in these services, subject to quality checks. Vocabularies published in these services satisfy the FAIR principles by being assigned unique persistent identifiers for individual terms as well as providing access to vocabulary documentation, such as purpose and governance authority. Ontology lookup services[25] present commonly used terms and their relationship to one another, again covering a wide variety of disciplines, such as environmental science. For example, http://purl.obolibrary.org/obo/ENVO_01001242 is a unique identifier for the term 'canopy' which is a narrower term than 'vegetation layer' but broader than specific instances such as 'crop canopy' or 'forest canopy'.

Selecting input data

Wherever possible, input data that is described using FAIR vocabularies should be chosen in preference to using data labelled with proprietary identifiers whose documentation may be difficult to access, or undocumented identifiers whose meaning may be open to misinterpretation. This is important in helping to establish data provenance, e.g. which instrument or model produced the data, and in documenting which input data any subsequent products were derived from.

Processing

Detailed description of processing algorithms is usually provided by software documentation, user manuals or scientific publications and is beyond the scope of controlled vocabularies. However, controlled vocabularies should be used to describe the variables and units of measure of inputs, intermediate products and final products resulting from a processing chain.

Sharing products

In many cases, the same controlled vocabularies that are used for input data can also be applied to output products, for example, units of measure and CF standard names. Where possible, a term from an existing controlled vocabulary should be chosen to label the end product of a climate service. If no suitable existing term can be found, it is good practice to select the most appropriate vocabulary and engage with the governance process to propose a new term. Controlled vocabularies often expand and evolve in response to user needs, for example, as climate models increase in sophistication or when new products are defined. The benefit of using a FAIR vocabulary is that its governance processes are well documented and discoverable so that members of the user community can readily engage with the vocabulary maintainers.

[21] https://www.go-fair.org/fair-principles/i2-metadata-use-vocabularies-follow-fair-principles/
[22] https://vocab.nerc.ac.uk/
[23] https://vocabs.ardc.edu.au/
[24] https://vocab.ices.dk/
[25] e.g. OLS, here the ENVO ontology: https://www.ebi.ac.uk/ols4/ontologies/envo

In addition to using a controlled vocabulary to describe a data product, it may be necessary to provide an alternative label which describes the product using terminology that is familiar to the intended audience. This may take the form of a single term or a longer piece of descriptive text. This does not necessarily break established metadata conventions, for example, CF metadata allows for a 'long name' in addition to the 'standard name'. The long name is a non-standardised text string and could be used to store the data description most familiar and meaningful to the end user.

General aspects of quality management and user engagement

Engagement with the user community is vital for maintainers of FAIR vocabularies because this is the only way to ensure that a vocabulary includes all the necessary terms. It is important that the description of a newly proposed term and its units of measure are recorded at the same time the term is added to a vocabulary and the data producer is usually best placed to provide this information.

Examples of FAIR vocabularies

i. Observing platforms and instruments

The World Meteorological Organisation (WMO) maintains a series of manuals, the Guide to Instruments and Methods of Observation[26], in particular Volume III – Observing Systems[27], but currently does not publish controlled vocabularies listing the instruments and methods as individual semantic artefacts.

For instrumental data, the NASA GCMD (National Aeronautics and Space Administration Global Change Master Directory) is a taxonomy covering many categories related to earth science, including platforms and instruments. Platforms are divided into categories such as "air-based" (e.g. radiosondes, Uncrewed Aerial Vehicles (UAV)), "water-based" (e.g., buoys, ships), "space-based" (a list of satellite missions), "ground-based" (includes meteorological stations) and "other" (e.g., computers, climate models). Instruments are divided into categories for "Earth remote sensing instruments", "in-situ/laboratory instruments" and "solar/space observing instruments". The NASA GCMD keywords can be browsed using the GCMD Keyword Viewer[28], downloaded in a variety of formats (RDF, JSON, XML, CSV) or accessed via a RESTful API service[29].

A comprehensive list of Earth Observation satellites, the instruments on board and other details of their mission such as start and end dates is available at https://www.eoportal.org/. The SeaVoX Device Catalogue of marine observing instruments is available from the NERC Vocabulary Server (NVS)[30].

ii. Experiment protocols and methodologies used in data collection and processing

The WMO manuals Guide to Instruments and Methods of Observation contain very detailed descriptions of observing instruments and the methods of using them. These manuals should be regarded as the reference text when making meteorological observations of any kind, but the various methods of data collection are not currently published as semantic artefacts.

For observing methods or experimental protocols, some examples of vocabularies that contain terms to describe methods used in observations, field studies and sample preparation can be found at

[26] https://library.wmo.int/records/item/41650-guide-to-instruments-and-methods-of-observation?language_id=13&back=&offset=
[27] https://library.wmo.int/records/item/68661-guide-to-instruments-and-methods-of-observation?language_id=13&back=&offset=
[28] https://gcmd.earthdata.nasa.gov/KeywordViewer/
[29] https://gcmd.earthdata.nasa.gov/kms/capabilities?format=html
[30] https://vocab.nerc.ac.uk/collection/L22/current/

https://linkeddata.tern.org.au/prez/tern-cv/v/super, https://vocabs.lter-europe.net/en/ (the hierarchical view of which is the quickest way to see the terms relating to methods) and https://vocab.nerc.ac.uk/collection/S03/current/.

iii. Scientific parameter naming

Climate model projections, for example, CMIP6 data are widely available in NetCDF format and conform to the CF (Climate-Forecast) Conventions, (Eaton, B., Gregory, J., Drach, B., Taylor, K., Hankin, S., et al, 2024)[31]. CF Standard Names[32] are a controlled vocabulary for scientific parameter naming which sits within the wider CF conventions. New Standard Names can be proposed as needed and undergo a well defined process of community discussion prior to acceptance and publication. The increasing sophistication of climate models and earth system models has driven much of the vocabulary development, for example, many ocean biogeochemistry and sea ice dynamics terms were introduced in 2018 in response to the data requirements of the CMIP6 project. Over time, CF Standard Names have also been adopted beyond the climate modelling community and now include terms applicable to satellite products, radar remote sensing, ocean gliders and other in situ observations.

Another widely used vocabulary in environmental science is the BODC (British Oceanographic Data Centre) Parameter Usage Vocabulary[33]. The CF Standard Names (P07) and P01 vocabularies have some overlapping content and are sometimes used alongside one another, for example, both are required in the metadata for data shared through SeaDataNet (Pan-European infrastructure for ocean and marine data management)[34].

NASA GCMD keywords include parameter names for many areas of earth science, including the atmosphere. The full list of available terms can be viewed in a browser using the GCMD Keyword Viewer[35] and opening the Earth Science section of the hierarchy. The vocabulary terms can also be accessed in RDF[36] or JSON[37] formats.

Specialised vocabularies can be used in combination with the more general environmental ones, for example, to ensure standard nomenclature is used when referring to chemical and biological species. As an example of this, CF standard names refer to two biological taxonomies, WoRMS (World Register of Marine Species)[38] for ocean species and ITIS (Integrated Taxonomic Information System)[39] for terrestrial and freshwater species. IUPAC chemical species names are used either in the CF standard name itself or referred to in the standard name description.

iv. Units of measure

Units vocabularies are used alongside parameter names and are essential for correct interpretation of the quantities in the data. For example, a quantity that is labelled as a “concentration” could be a mass concentration, molar concentration or volume concentration. Some parameter vocabularies, such as CF Standard Names, distinguish between such quantities while others use the more generic label of “concentration” and the scientific variable can only be understood by also referring to the units.

---

31 https://doi.org/10.5281/zenodo.14275599
32 https://vocab.nerc.ac.uk/collection/P07/current/
33 http://vocab.nerc.ac.uk/collection/P01/current/
34 https://www.seadatanet.org/Publications/Technical-documentation/Metadata
35 https://gcmd.earthdata.nasa.gov/KeywordViewer/
36 https://gcmd.earthdata.nasa.gov/kms/concept/1f0765e3-4ea3-42be-8ed5-3e26bdebb219?format=rdf
37 https://gcmd.earthdata.nasa.gov/kms/concept/1f0765e3-4ea3-42be-8ed5-3e26bdebb219?format=json
38 https://www.marinespecies.org/
39 https://www.usgs.gov/tools/integrated-taxonomic-information-system-itis

Within the CF Conventions, units are required to be compatible with the UDUNITS software package[40] which contains its own units database, mostly consisting of SI units and derived units. The NVS serves a unit collection, the BODC-approved (British Oceanographic Data Centre) data storage units[41], to describe the measurement units for data held in its repositories. The NVS representation of the CF Standard Names uses mappings between individual terms in the P07 and P06 collections to express the association between each standard name and its appropriate unit.

UCUM (Unified Code for Units of Measure)[42] is ‘a code system intended to include all units of measures being contemporarily used in international science, engineering, and business.’ It imposes a strict syntax on all units, combinations of units and symbols. It is primarily intended for unambiguous electronic communication of quantities and units.

QUDT (Quantities, Quantity kinds, Units, Dimensions, and data Types)[43] provides a collection of ontologies and vocabularies based on SI units of measure and also includes CGS units for mechanics, CGS EMU (electromagnetic) units, CGS-ESU (electrostatic) units, Gaussian units for electrodynamics, and the Planck system of natural units. Many US Customary and British Imperial units for length, weight, and heat are also included. The aim of the QUDT ontologies is to improve interoperability of data by providing unambiguous and machine readable units specifications and supporting conversion between equivalent units and quantities. The QUDT units vocabulary[44] provides multilingual support for unit names and is available in a variety of formats (CSV, RDF/XML, TURTLE, JSON, JSON-LD).

Recommendation 5 (R5): Use FAIR vocabularies to annotate data throughout the workflow
A FAIR vocabulary is one that is documented and where each term is resolvable using a globally unique and persistent identifier. The vocabulary documentation needs to be easily findable and accessible by anyone who uses the dataset. Vocabularies provided via semantic services such as the NERC Vocabulary Server https://vocab.nerc.ac.uk/ satisfy these requirements.

Recommendation 6 (R6): Request new terms in an existing vocabulary when necessary
If there is an existing FAIR vocabulary that covers the domain of interest, but it does not contain a suitable term for your data, e.g. a parameter name, it is best practice to engage with the governance processes for the vocabulary to request the addition of a new term. This is highly preferable to creating a new, separate, undocumented vocabulary.

Recommendation 7 (R7): Use CF vocabularies for NetCDF or Zarr data
The CF metadata conventions are designed to work alongside the NetCDF format and are also compatible with Zarr. They contain controlled vocabularies for parameter naming (CF standard names), surface characteristics (CF area type) and geographical areas with complex boundaries. The CF conventions stipulate the use of UDunits for providing information about units of measure.

[40] https://doi.org/10.5065/D6KD1WN0
[41] http://vocab.nerc.ac.uk/collection/P06/current/
[42] https://ucum.org/
[43] https://doi.org/10.25504/FAIRsharing.d3pqw7
[44] https://doi.org/10.25504/FAIRsharing.d3pqw7

2.4 Data licensing and citation

Licensing

Data licensing establishes the terms and conditions under which data can be accessed, utilised and shared by granting legal permission. Licences are assigned by the data owner or copyright holder. An overview of existing licences with their legal conditions is available at https://spdx.org/licenses/.

Open licences allow the reuse of the data by all stakeholders. Open data licensing is encouraged by the Open Data and FAIR data principles aims. Imposed restrictions limit the reuse of the data and the creation of derivative products. As climate services often use data from multiple copyright holders, the handling of these imposed restrictions, which might be incompatible, can curtail the publication of derivative products.

Examples of data use restrictions that limit the use of data in climate services are: non-commercial (not allowing any commercial use), and nonderivative (not allowing to share data derivatives). Even the share-alike attribution of data disseminated under an open licence might restrict commercial reuse, as data derivatives have to be distributed under the same open licence.

Climate service providers have to comply with the data licences and all the restrictions imposed by the copyright holder of the data it uses when distributing its products. In
in order to ensure maximal reusability of the data, IPCC TG-Data recommends to
licence data derivatives under the Creative Commons Attribution 4.0 International (CC BY 4.0) licence, where this does not infringe the interests of relevant licence holders (Huard et al., 2022). This open licence permits the reuse of data under the condition that appropriate credit is given, which includes credit to the data creators (see https://creativecommons.org/licenses/by/4.0).

Citation

In science credit is commonly given by citation. Data citation principles were formulated by FORCE11 in the Joint Declaration of Data Citation Principles (Data Citation Synthesis Group, 2014). The international membership organisation DataCite for example provides the infrastructure and services for the citation of research data and further scholarly records and the registration of standardised metadata. For a fee, DataCite mints DOIs for their members as a registration agency of the International DOI Foundation.

Most data providers have implemented DOIs on data and have included a "cite as" statement on their data pages, which climate service provider should comply with when citing the used data for credit assignment. In addition, the modification of the data should be indicated, which also supports the transparency and traceability of the climate service product. Products of public climate services should be assigned a DOI to make them referenceable and citable for climate service customers. Organisations providing best practices for data citation include apart from DataCite and FORCE11, the Data Curation Centre (DCC), the Earth Science Information Partners (ESIP) and the Coalition for Publishing Data in the Earth and Space Sciences (COPDESS).

Recommendation 8 (R8): Include an explicit licence for data
Licence data and data derivatives under the Creative Commons Attribution 4.0 International (CC

BY 4.0) where this does not infringe the interests of relevant licence holders. The open licence CC BY 4.0 permits the reuse of data under the condition that appropriate credit is given. https://creativecommons.org/licenses/by/4.0

Recommendation 9 (R9): Issue DOIs for data products
Data products of public CS should be issued with a DOI and a "cite as" statement should also be included. This makes the products referenceable and citable for climate service customers.

Recommendation 10 (R10): Document all data modifications
Any modification of data should be indicated in order to support the transparency and traceability of climate service products.

2.5 Software

Software underpins climate services across the whole value chain, from research and modelling of climate-relevant processes and variables to operational production and dissemination. It shapes how data is processed, interpreted, and transformed into actionable information. Software may be developed in-house, co-developed with partners, or adopted from external sources.

Consistent guidelines and good practices are crucial across the whole software lifecycle (e.g. EOSC Research Software Lifecycle[45]), and need to focus on both internal and external software, with consideration of their respective requirements, risks, and assessment criteria.

Ensuring quality, interoperability, and long-term usability is essential for reproducibility, transparency, and reusability. To this extent, the FAIR (Wilkinson et al., 2016) principles, and in particular the FAIR principles for Research Software (FAIR4RS; Barker et al., 2022) are adopted here as a guiding framework, and interpreted in ways relevant to both research-oriented and operational climate service software.

Research Gap 1 (G1): Harmonizing FAIR software tools and metrics
Existing tools for supporting and assessing FAIR software practices rely on heterogeneous frameworks and metrics limiting consistent evaluation, comparison, and automation. This fragmentation hinders the integration of FAIR software assessments into development workflows and increases the burden on developers and users. Harmonised FAIR software metrics and interoperable assessment tools are needed to enable consistent, transparent evaluation, and to support informed software adoption and reuses.

Some of the points raised in FAIR4RS principles can be encapsulated by establishing and updating a software management plan (see for example ISO/IEC/IEEE 16326:2019[46] or Anderson et al., 2007). Such a plan helps to formalise ways and means to ensure that the software is and stays accessible and

[45] https://zenodo.org/records/8324828
[46] https://www.iso.org/standard/75276.html

reusable. A software management plan has the benefit to clarify responsibilities and processes, to alleviate potential risks, and overall to foster the use and reuse of the software.

Recommendation 11 (R11): Adopt FAIR-aligned software management practices across the full software lifecycle.
Software developed or adopted for climate services should align with FAIR and FAIR4RS principles across its whole lifecycle. This should be operationalised through a software management plan covering development, release, and maintenance. Key elements include adherence to community standards for coding, testing, versioning, and review, and defined responsibilities for maintenance and long-term sustainability. Applying these practices proportionally improves transparency, reproducibility, and long-term reusability of software.

Planning, FAIR, provenance, and community engagement

Before developing or using software, the decisions made at the planning stage define if a software will be accessible, usable and trustworthy over time.

For internally developed software, defining objectives, the intended operational context, a roadmap, and stakeholder engagement early allows to plan for the findability, accessibility, interoperability, and reusability of the future software. Requirements on provenance and traceability (e.g. links between software, data, and outputs) should be incorporated from the outset (see D2.7 on these topics). Properly identifying stakeholders and expected user communities allow for a better link to community engagement during the whole software's lifecycle.

Several steps strengthen trust before adopting external software: namely, assessing if the software aligns with the FAIR4RS principles, provides clear documentation and metadata, and is supported by an active and open community.

Recommendation 12 (R12): Address governance, provenance, and user requirements during software planning
Decisions made at the planning stage should explicitly address intended use, provenance, governance, and user communities, in order to ensure that FAIR4RS principles, traceability requirements, and stakeholder engagement are embedded before development or adoption begins.

Standards and best practices in software development

Adherence to shared vocabulary (e.g. ISO/IEC/IEEE 24765:2017[47]) and development standards enable software to be understood, maintained, and integrated consistently by its developers and users.

Established and recognised coding standards should be used from the first written line of code, following language- and community-specific standards. Relevant standards include PEP8[48] for Python, Tidyverse for R, and SEI CERT for C/C++. Automatic code quality checks such as linters, formatters, or pre-commit hooks can be used to ensure written code is consistent. In order to design stable

[47] https://www.iso.org/standard/71952.html
[48] https://peps.python.org/pep-0008/

interfaces or APIs, configuration should be favored in place of hard-coded parameters. Providing execution environments, and using version control and semantic versioning standards (e.g. Semantic Versioning 2.0.0[49]) ensures the reproducibility of results and use. Support of standard file formats and protocols further helps the interoperability and reusability of the developed software.

When integrating external software into the climate service's pipeline, specific attention needs to be put on the adherence to renowned standards, proper versioning, the availability of reproducible environments and documentation, and mainly the compatibility with existing systems and workflows.

Recommendation 13 (R13): Follow recognised software development standards
Software development should follow recognised language- and community-specific standards for code structure, configuration, interfaces, and versioning. This is supported by automated quality checks and reproducible execution environments to ensure maintainability and interoperability.

Quality assurance, testing, and scientific validation

Systematic testing and validation are essential aspects to ensure reliable results and maintain scientific integrity as the software evolves. These efforts can be structured around two complementary pillars: technical quality assurance and scientific validation.

On the technical side, the implementation of unit, integration, and system tests assures the stable functioning of the software. Additionally, this can be automatically and consistently assessed through continuous integration and deployment (CI/CD) pipelines to automate testing and quality checks. Regression testing further ensures consistency across versions, and peer review of code provides an additional layer of transparency and robustness. External tools can help both report and assess the quality of a software's code, and a large panel of software quality attributes and metrics exist and should be considered when producing or using software (see for example EOSC's task force report on software quality, David et al., 2024). Established standards like ISO/IEC 25002:2024[50], 25010:2023[51], or 25019:2023[52] provide useful elements towards building and evaluating software quality.

On the scientific side, validation specifically verifies that the underlying methods, parametrisations, and outputs are physically consistent and aligned with domain knowledge and accepted standards in the climate sciences (e.g. units or definition of climate indicators). This includes numerical validation against reference datasets or benchmarks, verification that the software reproduces known results for pre-defined cases, and assessment of physical consistency across a range of use cases. Regression testing against previous scientific results further ensures that software evolution does not introduce unintended changes in scientific behavior. Peer review by domain experts, distinct from technical code review, is a fundamental step in ensuring the scientific integrity and transparency of the software.

Together, these two types of quality and validation procedures pave the way towards a proper software release process.

Recommendation 14 (R14): Ensure quality and scientific validation of software

[49] https://semver.org/
[50] https://www.iso.org/standard/78175.html
[51] https://www.iso.org/standard/78176.html
[52] https://www.iso.org/standard/78177.html

Software used in climate services should be subject to both rigorous technical quality assurance and thorough scientific validation. Technical quality assurance should include automated testing pipelines (unit, integration, and system tests), continuous integration and deployment, and code peer review against recognised quality standards. Scientific validation should verify that outputs are physically consistent, reproducible against reference datasets, and reviewed by domain experts. Together, these practices ensure the reliability, scientific integrity, and consistency of software across versions and as it evolves.

Release, licensing, identification, and metadata

Clear licensing, persistent identification, and comprehensive metadata are the cornerstones for enabling legal reuse, proper attribution, and long-term discoverability of software, directly contributing towards the FAIR principles for software.

For internally developed software, choose clear, permissive or copyleft open-source licenses to foster accessibility. Providing machine-readable metadata tremendously helps make software easily findable in registries: from standardised software's metadata formats like Citation File Format[53] files or CodeMeta[54], to more elaborate interoperable representations of software and their internal workflows with RO-CRATE[55] or ARCO[56] (Mourkoussis et al., 2003). Assigning persistent identifiers (PIDs, such as digital object identifiers) to each released version clarifies access guidelines and supports software reuse. In addition, archiving previous releases ensures long-term access and usage of the software, while supporting reproducibility. The EU Licensing Assistant[57] for software licenses can also assist in choosing the appropriate license

For external software, licensing terms should be compatible with the intended use, versions should be identifiable through PIDs, and archived versions should be available in trusted repositories.

Recommendation 15 (R15): Ensure findability and proper citation of software
Released software should be accompanied by clear and compatible licenses, persistent identifiers, archived versions, and machine-readable metadata. These aspects enable the discoverability, legal reuse, citation, and long-term access of software assets.

Maintenance, sustainability, and long-term usability

Sustainable maintenance and clear support mechanisms are essential for ensuring the long-term usability, reliability, and relevance of software throughout its lifecycle.

User support channels and issue tracking (e.g. GitHub issues or mailing lists) are great ways to facilitate and encourage community engagement and problem resolution. Maintenance responsibilities and update policies should be defined to ensure continuous improvement of the software. It is crucial to document deprecation and end-of-life procedures to inform users of upcoming changes and migration paths.

[53] https://citation-file-format.github.io/
[54] https://codemeta.github.io/
[55] https://www.researchobject.org/ro-crate/
[56] https://www.arco-web.org/
[57] https://interoperable-europe.ec.europa.eu/collection/eupl/solution/licensing-assistant/find-and-compare-software-licenses

The maintenance status of external software needs to be assessed to evaluate their reliability: from actively developed, to passively maintained or archived. The size and activity of both the user and developer community as well as governance structures are key indicators of potential risks of disruptions for critical components related to dependency longevity and support.

Recommendation 16 (R16): Establish maintenance and risk-management practices for software
Software should have defined maintenance responsibilities, support mechanisms, and deprecation policies. The sustainability of external dependencies should be regularly assessed to manage operational risks and ensure long-term usability.

2.6 Data Management

While the data lifecycle is often complex and non-linear, its successful development relies on a core sequence of actions: planning, collecting, processing, analyzing, preserving, sharing, and reusing (see Fig. 1 below).

Although the implementation of these stages varies depending on the specific context, each is a necessary component of a high-functioning service. This complexity raises a critical question: how can these diverse actions be managed efficiently?

We propose that **a robust Data Management Plan (DMP) is the only way to effectively meet this objective**. However, before exploring the development of DMPs, it is essential to understand the legal framework surrounding information services and why it serves as the foundation for these practices.

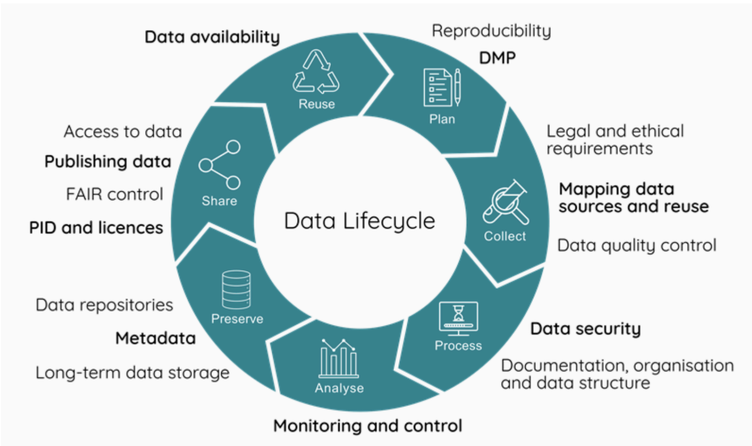


Fig. 1: Representation of the data lifecycle developed by the European Open Science Cloud (EOSC).

Legal framework

It's essential to understand that under the eyes of the law not all data can be collected, stored, processed, and presented equally. Some examples of regulations applied in Europe are:

- **The General Data Protection Regulation — Regulation (EU) 2016/679**
  While personal or proprietary data is distinct from climate information, its storage and processing are often essential. For instance, a climate service may need to handle sensitive location data for a bank's critical assets, requiring rigorous security protocols to ensure its protection.

- **The EU AI Act — Regulation (EU) 2024/1689**
  As climate services increasingly integrate AI for downscaling, model emulation information provision, and others, they have entered the purview of a new risk-based regulatory framework. Under the EU AI Act, systems supporting critical infrastructure or emergency response are classified as 'high-risk,' triggering stringent requirements for accuracy, robustness, and transparency. Central to this mandate is the necessity for human oversight and a fundamental shift in data governance. Because the Act requires training and validation datasets to be as representative and error-free as possible, uncertainty quantification could become essential as a means to comply with the mandate.

- **The EU Accessibility Act — Directive (EU) 2019/882**
  From June 2025, digital services provided in the EU must ensure that critical information is "perceivable, operable, understandable, and robust" for all users, including those with disabilities. Under this framework, climate platforms are now required to move beyond standard data visualization toward inclusive design, incorporating features like screen-reader compatibility, high-contrast visuals, and alternative text for complex climate maps. By mandating these accessibility standards, the Act transforms inclusive communication from a voluntary "best practice" into a fundamental legal requirement, ensuring that life-saving climate information and risk assessments are accessible to every citizen on an equal footing.

In addition, there's the aforementioned topic of data licensing (see Section 2.4), which is essential to guarantee a legal and appropriate reuse of resources, particularly when they are shared in public repositories.

Recommendation 17 (R17): Consider regulatory alignment in climate service design
The design of climate services necessitates a robust understanding of the EU legal framework. Because these services often process sensitive data, utilize complex modeling for high-stakes decision-making, and serve diverse stakeholders, they fall directly within the scope of existing EU regulations. Addressing these legal requirements during the initial design phase is essential to ensure both compliance and operational integrity.

Conventions

As we have already discussed, there are certain conventions which guide the development of information services. One of them is of course the FAIR principles convention (Wilkinson et al., 2016)[58], which focuses on the technical aspects of research data. There's also its adaptation to research software, namely FAIR4RS (Barker et al., 2022)[59]. Another popular framework in data management is the Data Management Body of Knowledge (DMBOK) provided by DAMA[60]. However, an information service is much more than a data file or a piece of code. It's also about the ethical

[58] https://doi.org/10.1038/sdata.2016.18
[59] https://doi.org/10.1038/s41597-022-01710-x
[60] https://dama.org/learning-resources/dama-data-management-body-of-knowledge-dmbok/

context surrounding it, the reliability of the repositories where data is stored and much more. Examples of principles taking such concepts into account are:

- The **CARE principles** (Carroll et al., 2021)[61], which stand for Collective Benefit, Authority to Control, Responsibility and Ethics, deal with equity, power, and ethical use of indigenous data.

- The **TRUST principles** (Lin et al., 2020)[62], which stand for Transparency, Responsibility, User focus, Sustainability, and Technology, and their aim is to ensure that repositories are reliable custodians of data.

Since the implementation of these conventions can be ambiguous, the community tries to agree on methodologies that can be applied operationally and under similar contexts. For example, for the FAIR principles there are a couple of implementation proposals: RDMkit[63], FAIR cookbook[64], and FAIRsharing.org[65].

Research Gap 2 (G2) : Concrete operational frameworks are needed to translate high-level principles (e.g. FAIR, CARE, TRUST) into verifiable implementation practices
The lack of standardised interpretations for high-level principles remains a significant hurdle. Without clear, reproducible implementations, stakeholders cannot move beyond subjective assessments toward rigorous verification. Establishing these concrete benchmarks is a necessary step to facilitate transparent compliance audits and meaningful cross-service comparisons.

Standards

Even though the list of available standards is overwhelming here we want to highlight two of them, which are currently used and requested both in the public and private industry:

- **ISO 9001:2015 Quality management systems**: It is used to guarantee that a service is managed properly, which is different from stating that a service provides a trustable product, and is discussed more deeply in Deliverable D2.9, which is to be published in April 2026.

- **ISO/IEC 27001:2022 Information security, cybersecurity and privacy protection — Information security management systems — Requirements**: It is used by services to provide a secure environment in which confidential data can be stored and processed.

Both of these standards are certifiable, which means that certified institutions get a unique label that can be shared openly or directly with clients. The certification process is again something contextual, which means that factors like the auditor and the type and size of the company that is being certified can affect the quality of the outcome and the effort necessary to obtain the certification label.

For WMO member states and affiliated institutions, WMO's High-quality Global Data Management Framework for Climate (WMO-No. 1238) establishes standards and recommended practices for sourcing, securing, managing, assessing, and cataloguing climate data, and for sharing infrastructure and responsibilities for, for example, data exchange, analysis and data service provision.

[61] https://doi.org/10.1038/s41597-021-00892-0
[62] https://doi.org/10.1038/s41597-020-0486-7
[63] https://rdmkit.elixir-europe.org/
[64] https://faircookbook.elixir-europe.org/content/home.html
[65] http://FAIRsharing.org

Recommendation 18 (R18): Adopt standards to strengthen operational quality and security
Attaining ISO 9001 and ISO 27001 certifications is a critical milestone for climate service providers seeking to formalize their operational excellence. These standards provide a dual-layered assurance: ISO 9001 validates the robustness of the quality management system, while ISO 27001 ensures that sensitive environmental and proprietary data are managed in strict accordance with international security protocols. Together, they instill the stakeholder confidence necessary for high-stakes decision support.

Research Gap 3 (G3): Bridge the recognition gap for non-certifiable standards
A significant gap exists in the market recognition of non-certifiable standards within the climate services community. While frameworks such as the ISO/IEC 25000 series provide essential guidelines for software quality, they often remain underutilized due to the absence of a formal validation mechanism. To bridge this gap and incentivize adoption, a dedicated certification framework should be developed, providing a structured pathway for providers to demonstrate compliance and for users to verify service excellence.

Data management plan

Conventions, standards, regulations and many other governance factors influence the strategy that a company or institution might take towards managing data. All this information jointly with the description of all the steps the data goes through, like collection, processing, archiving, preservation, publication, etc., should be documented in what is called a Data Management Plan (DMP).

A DMP should be a living-document, which means that it should be reviewed and edited not only at the start and end of a project, but also at every significant milestone. In other words, every change of plans that occurs during the development or maintenance of a service should be documented.

The shape and format of a DMP can vary depending on the context, but many templates are available. See for example the templates from Science Europe[66], Horizon Europe[67], the Data Curation Center[68], and the European Research Council Open Research[69]. Moreover, there are interactive tools that guide users through the development of a DMP. See for example, Data Stewardship Wizard[70], DMPTool[71], EasyDMP[72] and DMP Online[73].

A service developing team might not have the capacity to write a comprehensive DMP at the beginning of a project. However, at least a draft should be produced, it should evolve according to the needs of the project and never be left obsolete, as it represents an essential tool to make sure that all actions in the lifecycle of data (Fig. 1) are aligned and the governance context is well-understood.

[66] https://scienceeurope.org/our-priorities/open-science/research-data-management/
[67] https://cohortcoordinationboard.eu/toolkit/horizon-europe-data-management-plan-template/
[68] https://dcc.ac.uk/resources/data-management-plans
[69] https://erc.europa.eu/sites/default/files/document/file/ERC_info_document-Open_Research_Data_and_Data_Management_Plans.pdf
[70] https://ds-wizard.org/
[71] https://dmptool.org/
[72] https://easydmp.no/
[73] https://dmponline.dcc.ac.uk/

Research Gap 4 (G4): Standardised and machine-actionable frameworks for Data Management Plans could improve transparency, interoperability, and automation of data governance processes
Standardizing Data Management Plans (DMPs) offers a significant opportunity to enhance operational transparency and ensure the structural completeness of data management strategies. However, widespread and consistent adoption depends on the availability of clear, actionable implementation frameworks. Establishing such clarity would facilitate the systematic comparison of services and the identification of strategic gaps. Furthermore, the integration of machine-actionable formats would allow for automated updates and real-time synchronization, shifting DMPs from static documents to dynamic management tools.

3. Key Messages from Interviews and Survey

To complement the desk research and gather qualitative insights into the experiences, services and applied standards of CS providers, interviews and a survey were conducted. All collected information is presented in the following subchapters, without personal or company information to protect the privacy of the participants. **The findings are based on a relatively small sample of CS organizations (30) and individuals (8), and may not fully represent the diversity of CS providers.**

3.1 Interviews

To qualitatively explore the services, experiences, challenges, and needs of CS providers related to data, software, service provision and communication, including aspects concerning sector-specific issues and opportunities for improvement, partners involved in preparing this deliverable conducted interviews with CS providers.

While these interviews were carried out specifically for this deliverable, similar interviews were conducted for preparing deliverables of other WP2 tasks in Climateurope2. Given that some task focuses may overlap, and that interviewees may elaborate beyond the initial scope of questions, all interviews conducted within WP2 by the end of January were reviewed. Each task applied its own tailored interview protocol. The protocol for the D2.6 interviews is provided in Appendix A2, while protocols for other tasks are available in their respective deliverables. Additionally, interviews conducted within WP4 were also revised.

Overall, interviews with 20 private companies and consultants, as well as 8 public CS providers, were analysed. These organisations include companies and startups developing data, predictions, risk models, IT platforms, as well as insurance and reinsurance companies. Their services cover climate risk assessment (floods, droughts, storms, fires), support to the energy, finance, infrastructure, and public sectors, and the transformation of complex climate data into decision-ready information for non-expert users.

The analysis of the interviews was conducted using a qualitative approach with descriptive language. Responses were often broad and unspecific, so prioritisation, extra analysis, or grouping of responses was not feasible.

3.1.1 Data sources

There are a large number of climate data sources and providers, and some respondents see a challenge in selecting the right one. During the creation of climate services, some private data providers were abandoned because clients prefer the national public data providers; others decided against buying in expertise because the added value is estimated to be too low; and for third, acquisition/guidance on datasource selection rely partly on personal networks.

Current data providers vary in how well they meet expectations. The external datasets come from different providers, who might use different formats, and even for the same data product changes can occur from version to version. The level of detail regarding method description, temporal coverage, metadata, and processing steps is uneven among data providers and standardisation was suggested.

There are also differences in data accessibility. Data access and data connectivity is still considered limited and could be improved. Interviewed CS providers think that data from the public sector should be open source and standardised. Standardisation of open source data would improve usability, especially in cross-sector applications. However, for model outputs this is difficult to apply consistently due to dataset diversity. There is an opinion that data sharing, exchange, and mapping should be standardised as well.

Large data providers, like Copernicus and climate centres, and their major datasets are trusted as stable and reliable, but users would value (in no particular order):

- advance notice of dataset updates,
- access to version information via APIs,
- delivery in cloud-optimized formats and CSV,
- harmonization of regional and global models,
- higher spatial resolution of data, and
- clearer information on data skill/quality.

Challenges include:

- accessing and downloading data depending on the source,
- availability of local data for downscaling and bias correction,
- managing large data volumes,
- handling long time series that often require manual activities.

A complete list of mentioned data providers with their datasets, used formats, compliance with standards and user comments on gaps and needs is given in the Appendix A3.

Research Gap 5 (G5): Guidelines to navigate data are needed
Since there are many data providers as well as numerous data types and datasets with varying availability and different formats, users often struggle to select the most appropriate one. Clear guidelines are needed to help users navigate the vast amount of available data and obtain the data that best suits their purpose.

Research Gap 6 (G6): Keep working on overcoming barriers in data access and interoperability
Differences in data access protocols, metadata conventions, and data formats across climate data providers continue to limit interoperability and efficient integration of datasets.

3.1.2 Data processing

Interviewed CS providers use either solely in-house developed software or a combination of internally and externally developed software (open-sourced or proprietary). For internal software development, they follow standardised and best practices for software and data management:

- test and production servers,
- CI/CD process,
- specific utility code,
- orchestration,
- unit testing,
- automated deployment,
- DAMA-DMBOK data management principles and implementation of standard formats such as ARCO.

For software development itself, some CS providers implement open-source tools with publicly available code repositories. GitHub and GitLab are used to version-control the software and track changes, and metadata is added to specify the code that generated a dataset. In some cases, national guiding principles and standards (e.g. VDI - Verein Deutscher Ingenieure/The Association of German Engineers) require application of specific software.

Some CS providers use computer resources from external providers or outsource hosting due to high energy costs or for easier collaboration within the teams and with clients. Software that is a CS product and sold commercially is licensed.

Overall, interviewed CS providers reported following established software engineering and data management practices.

3.1.3 Data and information product sharing and communication

Data and information product sharing

CS providers are required to be able to provide services in multiple ways adapted to the client (Table 2). Users of the interviewed CS providers, that are small companies or technical users, are mostly fine with the interface and platforms, as these are designed to be generic and useful for a broad audience. Platforms can be customised only to a certain extent, which may not always suit client requirements. In case of larger companies, which have more time and resources to explore and leverage the tool or to use it as an input data for their own processes, an API connection is often requested. The API allows them to evaluate the performance of the tool using their own methodologies in a more comfortable way.

In general, CS providers offer a variety of products and services through multiple delivery channels. These channels range from general-purpose solutions, like websites and platforms designed for broad audiences, to specialised outputs, such as expert analyses and prototype model services, to support diverse climate-related decision-making needs.

Table 2: List of CS products, methods of their sharing and targeted user community

Products	Product sharing	Target user community
Datasets	• (Web or Data) Platform • APIs • Downloadable CSV packages with metadata • Website • Interface • Interactive dashboards • Climate Data Store	• Private and Public sector • Infrastructure sector incl Architecture • Financial sector • Renewable energy (wind)
Tools	• Filters • Chatbot • Other (not stated which tool)	• Private and Public sector • Banking sector • Insurance sector
Software	Software plugins	
Decision-Support Services and Indicators	• APIs • White-label websites • Platforms • Applications	• Private and Public sector • Banking sector • Insurance sector
Data Catalogue (Definitions, metadata, units, last update, notes about the limits of the data)		• Private and Public sector • Banking sector • Insurance sector
Analyses	• Expert reports (containing all the necessary information, presented as text, tables and graphics, including methodology) • Interactive maps • Applications • APIs • Web platforms • Web sockets	• Private and Public sector • Insurance and reinsurance sector • Renewable energy (wind) • Industrial sectors like energy, finance, and infrastructure • Forestry • Consulting companies • Climate action planning
Consulting	• Sustainability reports (include methodology, maps and diagrams) • Model service as prototype • Web-based tools • e-Mails • Meetings	• Private and public sector • Insurance and reinsurance sector • Banking sector • Responsible Investment sector • Architecture
Map-verified outputs	-	-

Training modules	-	-

Community Insight 1 (C1): Engage with users to co-design the delivery system
Climate services can be provided in multiple ways, and the choice should be made according to the needs of the client or targeted user community.

Standards in use for sharing data and information products

The data products are stored in databases such as SQL and accompanied by metadata compliant with ISO and WMO metadata standards.

Datasets provided through Copernicus, more specifically C3S Climate Data Store, are in NetCDF and GRIB formats while following CF metadata conventions and ISO metadata standards (specifically ISO 19115 and ISO 19139). C3S Datasets possess a DOI or PID and a specified license. Data access methods include OpenDAP data adapters, OGC standards, and controlled vocabularies for standardizing variable names and units. C3S provides both programmatic access via RESTful API and user interfaces, ensuring consistency across all access methods.

Some other data providers also follow Copernicus CDM-Obs-Core (Common Data Model - Observation Core) and OGC API EDR (Environmental Data Retrieval) standards. The CDM-Obs-Core standard is considered to be a metadata format that is easier to handle than other metadata standards currently used.

Interviewed public data providers adhere to the INSPIRE directive and EU High-Value Datasets (HVD) Directive as well. The EU HVD Act requires certain public-sector datasets to be made available in harmonised, open, machine-readable ways, but not in a specified format. According to this directive an API should be available with climate data, where one can use bulk download or filter data in it. This can be done in many different ways and without proper alignment, each European country could implement an API whose output is incompatible with the implementation from other countries.

The Creative Commons framework of standardised licenses, especially the CC-BY license suitable for Open Access data and publications, is also found helpful when reaching data sharing agreements between multiple countries.

Lastly, products as well as expert and sustainability reports are standardised according to EU taxonomy and EU Corporate Sustainability Reporting Directive (CSRD).

Overall, while a wide range of established standards and frameworks already support data storage, metadata structuring, data access and licensing, their implementation remains heterogeneous. This is particularly evident in the development of APIs and data access solutions, where the lack of harmonised approaches can result in incompatible outputs between providers.

Research Gap 7 (G7): Lack of harmonized API implementations
Providing a service most of the time requires having an API. API creation can be done in multiple ways, which can lead to incompatible outputs among different providers.

Transparency

CS providers aim to be as transparent as possible without compromising the company's intellectual property and competitive advantages. They are transparent about how their data and products are generated regarding input data and sources, and which methodologies were used. However, most interviewed private companies do not share the input data itself or detailed processing methodology (formulas, algorithms, code or workflows) as these constitute the company's intellectual property (industrial secrecy protection system). In some cases, access may be granted upon signing a confidentiality agreement.

A company's detailed knowledge of its methods is seen as the basis of trust and quality. Providers that are transparent about their methodology, consider this a competitive advantage, as few competitors share their methods, which reassures clients about the rigor of the approach. Also, a challenge in sharing methodologies "*is that you can find articles with different assessment results from the same dataset, which causes confusion and hesitation towards its applicability*".

Transparency is maintained through:

- documentation,
- data catalogues (definitions, metadata, units, version-tagged files, deprecation policies, notes on data limitations, disclaimers on data quality),
- public forums,
- catalogue banners, and
- known issues tables.

Community Insight 2 (C2): Build trust through transparency
As transparency builds trust and provides evidence of quality, it is recommended that CS providers be as transparent as possible about their products without compromising their intellectual property or competitiveness in the market. In cases involving proprietary information, it is preferable to grant some level of transparency through confidentiality agreements.

User engagement

The type and depth of engagement and feedback largely depend on the client. Services and their priorities are developed primarily through direct and regular consultations with clients. In some cases, long-term customers have entered partnerships for joint development. Specifically for public providers, user engagement also includes a user requirements database, interviews, workshops, conferences and general assemblies. All these activities provide direct feedback that feeds back into service improvements, demonstrating a commitment to user-driven development.

Interviewed CS providers often take additional time for training the clients. Users are generally more interested in obtaining "good results" than in understanding the underlying algorithms, methodologies, or scientific details, and many lack the expertise or interest to define technical requirements.

Customer satisfaction and trust are considered essential to consolidate the position in the market and the company's competitive advantage.

Community Insight 3 (C3): Commit to user-driven development
User engagement is seen as essential for developing the service, regardless of the type. Direct and regular feedback can contribute to service improvements and user satisfaction. Climate Service products should be developed in collaboration with users or based on dedicated studies.

User support

There are multiple ways in which interviewed CS providers support their users, including:

- data catalogues (definitions, metadata, units, last update, notes on data limits),
- technical documentation and user manuals for APIs and tools,
- technical and scientific documentation for products, guidelines and good practices,
- notifications about updates and versioning system,
- targeted communication, including consultations with experts for explanations and usage advice,
- educational communication, tutorials, and training.

In some cases, the level of customer support offered by private providers depends on the subscription type.

C3S provides the most comprehensive user support system. It covers the complete user journey from discovery to application. Services include step-by-step video tutorials, example notebooks, e-learning resources, webinars, training materials and in-person training events. A tiered support system features a help desk with JIRA ticketing, multiple support levels including data provider consultation, and a user forum for community discussions and self-help.

Community Insight 4 (C4): Establish user support to maximize product uptake
It is advisable that CS provide an appropriate level of user support depending on the user type and expertise, such as data catalogues, technical documentation, notifications, communication channels or educational resources.

3.1.4 Organisational standards

This section outlines the experiences, opinions, and needs of interviewed CS providers related to organisational standards, which organisations establish to ensure consistency, quality, and efficiency in their operations. While only few providers have implemented such standards, many shared their views, mainly as “needs” and “should”, regarding the standardisation and certification of CS. There is a wide range of opinions that rarely align. Yet, this does not imply a lack of consensus, but rather agreement was not articulated in this study, possible due to diversity of interviewers and interviewees. This diversity is also valuable because it showcases genuine, unfiltered opinions, free from any influence or guidance.

Which standards are in use?

Organizational and sectoral standards and accreditations are considered important by most of the interviewed CS providers, however, only a limited number have implemented ISO9001 for quality management, and ISO27001 for processing and storing of confidential data. Instead of formal ISO standards, some interviewed CS providers mostly rely on recognized best practices, which they find more useful.

Indirect certification or implementation of standards are mentioned in the case of a financial deal or application for funds, when providers undergo audits of the technology, the data and software quality, etc.

Insights on standardisation implementation

Standards are often not implemented because their adoption is considered too costly in terms of time and resources (e.g. ISO and AFNOR) or there is no awareness or perceived need within their current commercial context. Additionally, there is an opinion that relevant standards like ISO, DIN or VDI (Association of German Engineers) are often outdated before they are published or updated for a highly innovative field like climate adaptation planning, and that some regulations are biased towards a specific product.

Opinions on standardisation

Quality standards are considered to provide a certain level of quality assurance of CS and a principle-based standardisation approach, with e.g. an audit, is suggested. Providers also expressed interest in certification which could be a DIN norm on climate adaptation concept development or similar. Some providers believe ISO certification of CS would not increase competitiveness, while others consider certification would help distinguish high-quality service from the growing number of competitors on the CS market. There is also a suggestion that certification focuses on usefulness for customers, but client trust in the CS provider and its methods is considered equally important.

Interviewees acknowledge the importance of product provision standards (harmonized methods for providing products across regions) and product quality standards. Standards in provision of CS are desirable as they can help companies reduce costs and deliver requested products more quickly, but their applicability is limited because customer requests remain diverse and specific. Therefore, it was recommended that standardisation of CS be practical (e.g. following the methodology such as from the Horizon Europe project CLIMAAX), flexible, and focused on defining principles of best practice, rather than standardizing every aspect. This approach could help avoid commoditisation of the market, stifling innovation and disadvantaging small providers. In addition, there is an opinion that standardisation facilitates prioritization by explaining to customers what they need so it is a market focused on investors but rather it needs to focus on clients.

Research Gap 8 (G8): Find the balance between different levels of obligation
While standards are seen as important for ensuring quality and credibility, most CS providers prefer relying on recognized best practices rather than strictly following formal ISO standards, as they find them more practical and flexible. Therefore, different degrees of obligation within standardisation frameworks could be explored, allowing for context-sensitive application.

Some providers believe that the guidelines should come from a public institution so that there's no conflict of interest from private companies, and so that big companies are not heard more than small ones. Finally, the qualification and education of CS employees is highlighted as an essential factor for ensuring service quality and implementation of standards and guidelines.

Needs

Setting benchmarks to position providers was suggested, i.e. reaching a consensus on which benchmarks to use, either sector specific or one that will cover general needs.

Some interviewees, in both the private and public sector, also expressed a need for clear and harmonized terminology. In terms of how standardisation can support CS providers, the development of a common vocabulary could facilitate communication and establish a shared language between practitioners and end users. However, many still consider creating the common vocabulary, and standardisation of CS in general, secondary to improving availability, quality and resolution/skill of data, tailoring services to user needs, and other. In addition, some providers think that the EU taxonomy is still relatively limited in terms of regulation, and standardisation of sustainability reports was suggested.

Community Insight 5 (C5): Qualifications and continuous education of CS employees are essential for service quality
Ensuring all technical and quality requirements for development and delivery of climate services is impossible without qualified employees. Qualifications, continuous training, and professional development of service providers are recognised as essential for guaranteeing service quality and implementation of standards and best practices.

Research Gap 9 (G9): Will standardisation kill innovations and "What's in it for me?"
There is a concern that standardisation may stifle innovation. Additionally, many CS providers often lack a clear understanding of the benefits that certification, such as for QMS standards, would bring.

Research Gap 10 (G10): Reach agreement on benchmarks to be used
In addition to standards, some CS providers think that the climate services market also needs benchmarks to help position providers, and needs to find an agreement on which benchmarks to adopt, whether sector-specific or broader ones that address general needs.

Research Gap 11 (G11): "What are we talking about?" i.e. ensuring consistent terminology
Some CS providers ask for clear and harmonized terminology. Establishing a common vocabulary would improve communication and create a shared language between practitioners and end users. Furthermore, some find that the EU taxonomy remains relatively limited, suggesting more standardised reporting according to EU legislation.

3.1.5 Key messages and take aways

The interviews highlighted several common practices and challenges faced by climate service providers:

- Data source selection remains challenging, due to the diversity of available datasets and inconsistent documentation, despite the widespread use of both public and private data sources.
- Public data providers are generally preferred, mainly because of their perceived credibility, stability, and accessibility.
- Data interoperability and access remain significant challenges, particularly due to differences in formats, metadata conventions, and access mechanisms across providers.
- Modern software engineering practices are widely adopted to manage complex data processing workflows, including version control, CI/CD pipelines, and automated deployment.
- Climate services are delivered through diverse channels, such as platforms, APIs, dashboards, reports, and consulting services, reflecting the varied needs of users.
- Transparency regarding data sources and methodologies is recognised as essential for building trust, although some processing methods remain protected as intellectual property.
- User engagement and co-design with clients play a central role in the development and improvement of climate services.
- Adoption of formal organisational standards remains limited, although certifications such as ISO standards are recognised as valuable.

3.2 Survey

In order to inform the formulation of useful and consistent guidelines for data, software, and services used and developed by CS, the WP2 team also created a survey to gather information from climate service providers. The survey is composed of sections along the stages of a climate service provision as well as one section on the organization's standards and certificates. The thematic sections include, among others, questions about good and bad examples to be shared as well as questions on requirements for more standardisation.[74]

The survey was disseminated and actively promoted through Climateurope2 newsletters and social media, professional networks and personal contacts. By the end of January 2026, 10 fully completed questionnaires were received. The majority of responders were scientists and data scientists (70%, see Appendix A4, left) with more than 5 years of experience working as CS providers (Appendix A4, right). Surveyed organizations primarily originate from the governmental sector, as well as research and academia (Appendix A5), and provide services for a wide range of users including public authorities, private companies, the general public and scientists (Appendix A6). These services include:

[74] https://ec.europa.eu/eusurvey/runner/Climateurope2_WP2_1_data_standards

- data, analyses and assessments of past, present and future climate (including early warnings and derived products) for public support, policy-relevant applications, as well as applications in adaptation and impact studies;
- consultations and development of prototypes;
- assisting national meteorological organizations and their climate centres, as well as providing them with guidelines and standards.

Unfortunately, none of the surveyed CS providers represent the private sector, which introduces a bias toward the public sector in presented results.

3.2.1 Data sources

Data sources

Surveyed CS providers predominantly use both internally generated and externally sourced datasets to create CS products (80%, see Appendix A7).

A wide range of data and sources are utilized, including:

- National Meteorological and Hydrological Services (NMHSs),
- Global climate centers (JMA, NOAA, KNMI);
- ECMWF: ERA5,
- Copernicus C3S: ERA5,
- ESGF nodes (mostly German Climate Computing Center DKRZ): CMIP, CORDEX and EUROCORDEX data,
- observations obtained via various vendors,
- Zenodo,
- own digital data archive,
- in-house regional climate simulations,
- satellite products,
- Universities,
- other governmental agencies.

Experience with obtaining data varies depending on the source. Survey participants mentioned unsatisfactory experiences with certain National Meteorological and Hydrological Services (NMHS) data being non-digitized, inconsistent, and inadequately quality controlled. Government-related limitations on data sharing as well as the absence of clear policies for pricing and sharing the data were negatively perceived as well. However, there are also good examples where digitized and quality controlled data are available for sharing in accordance with WMO data-sharing policy. Users consider JMA (Japan Meteorological Agency) to provide good climate data for the entire world, while NOAA (National Oceanic and Atmospheric Administration) data are not always available due to occasional government shutdowns. Additionally, there is good ingestion of climate data via EDR (OGC Environmental Data Retrieval API), but there are also some disadvantages such as inefficient data structures and sudden interruptions in uploads.

Centralized access through the ESGF portal or C3S website was positively evaluated in most cases. ECMWF and Copernicus datasets are well-documented and quality-controlled, with consistent formats and reliable long-term availability.

Surveyed CS providers find that access to and download of CMIP data have significantly improved over the last 10 years; however, some shortcomings still persist:

- data need to be downloaded from various nodes;
- there are different formats depending on the type of climate projection used;
- there is a lack of documentation (no catalogue available);
- the process for obtaining new datasets is lengthy;
- when issues with dataset are identified, it is difficult to know how to report them to the data provider;
- extraction of local regions of interest is not possible.

In general, and depending on the data source, issues are related with inconsistent metadata, format changes, limited documentation, or delays in data availability.

Data access

Not all datasets needed by the surveyed CS providers are openly accessible or freely re-usable (Fig. 2). In some cases, there are restrictions such as on commercial use and data sharing with third-parties. These restrictions mainly concern specialised datasets, whereas some datasets, e.g. core meteorological observations, are openly accessible under defined conditions. Public data may also be provided under an open licence (e.g. the Etalab licence[75] was mentioned as a positive example) as well as attestations and certificates.

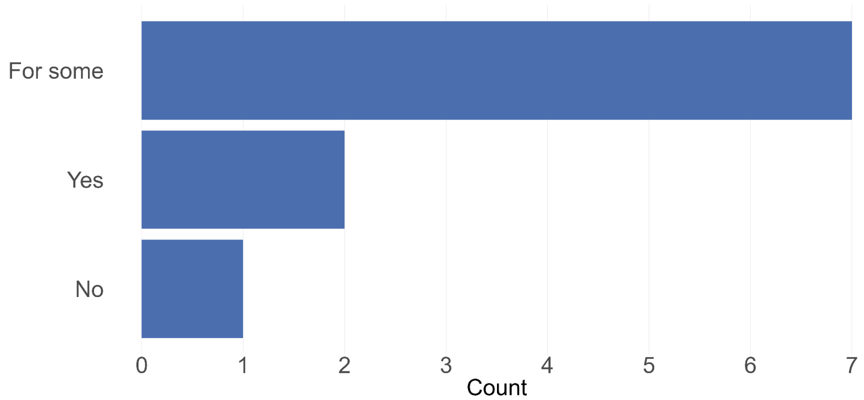


Fig. 2: Answers on the question "Are there license or access/usage restrictions for the source data?"

Data can be accessed in multiple ways: through direct download from server or cloud storage (the most common method among surveyed providers), via in-house storage or using API access. Portals, web applications and webpages are less used (Fig. 3).

[75] https://www.etalab.gouv.fr/wp-content/uploads/2014/05/Open_Licence.pdf . Etalab is a French government department responsible for digital transformation, opening up administrative data, and promoting open source software in the public sector.

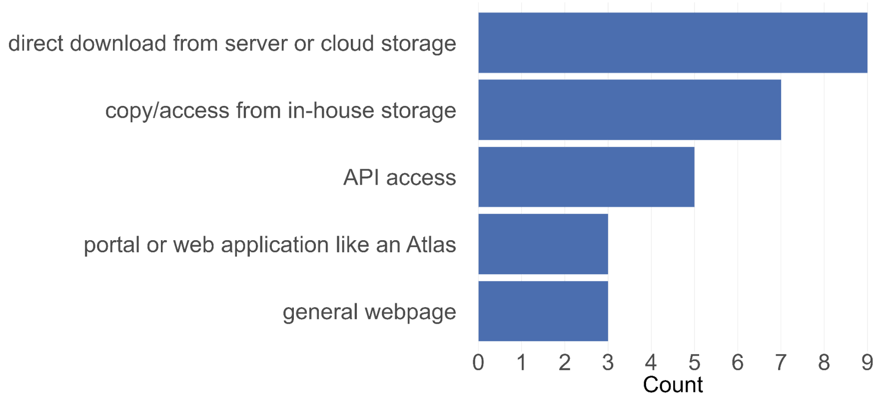


Fig. 3: Ways of accessing data. Multiple answers were allowed.

Data formats

Data formats differ among data owners. The most common format is NetCDF, followed by CSV (Fig. 4). Other formats that are also in use, but to a lesser extent, include Shapefile, JSON, GRIB, xlsx, geoTiff, PDF and plain text.

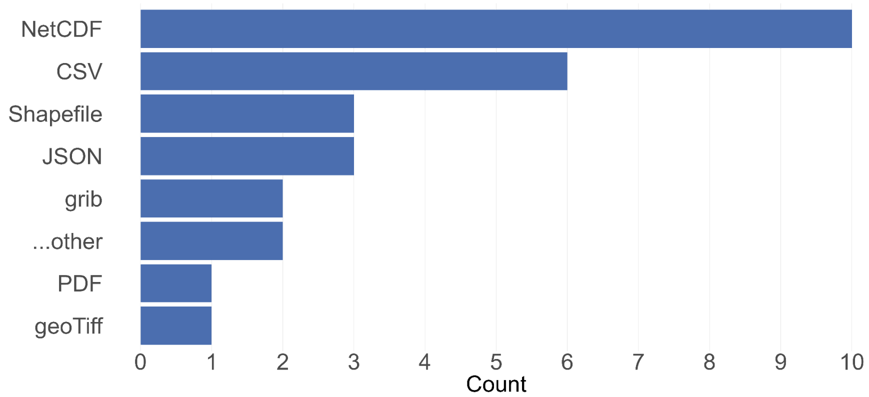


Fig. 4: Formats of source data. Multiple answers allowed.

Most of the used data comply with community standards, such as CF for the NetCDF format; however, further progress in the development of data standards is still expected. Exceptions in complying to standards are some observations, although in this case the metadata are considered sufficient, as well as legacy datasets that use non-standard formats.

What is still missing are temporal outputs and calendars (it is unclear which datasets this refers to) and a standard for versioning datasets. Additionally, it was suggested that NetCDF chunking be standardised, or at least recommended and properly documented.

FAIR & Standards

Among the FAIR data principles and data quality, on average, interoperability has the lowest importance rating (3.2) compared to other principles which are rated >=3.7 (Fig. 5). In case of data provided by NMHSs, there is a significant gap in their capacities, which leads to differences in data quality and interoperability. To address this, application of WMO Technical Regulations, Volume I – General Meteorological Standards and Recommended Practices No. 49, GCOS standards and the

WMO Stewardship Maturity Matrix for Climate Data (SMM-CD) is suggested. Additionally, interoperability between climate models remains challenging despite adherence to FAIR principles.

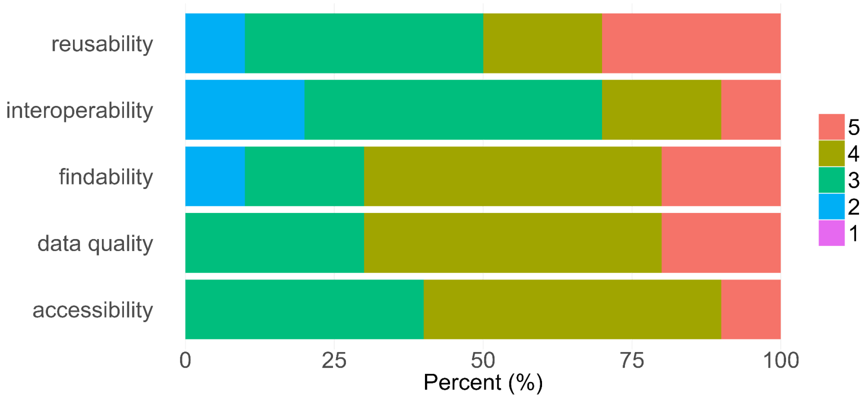


Fig. 5: Ratings of the source data in terms of the FAIR data principles and data quality (1 = low/bad to 5 = high/good).

In general, standards are considered important for data formats, machine accessibility of data, and naming conventions and controlled vocabularies, whereas data licences, and machine-accessible metadata and documentation are considered somewhat less important (Fig. 6).

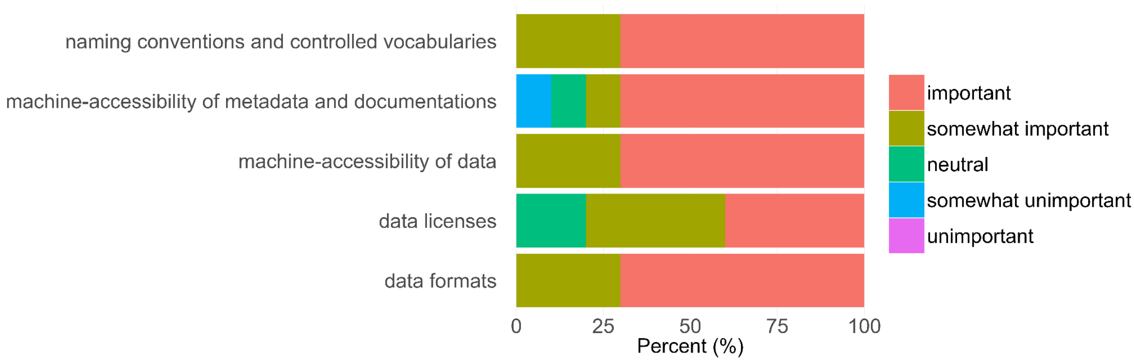


Fig. 6: Importance of standards for different aspects of data.

Community Insight 6 (C6): Input and output datasets should be quality-controlled and FAIR

It is recommended that datasets used as input for product development be quality-controlled and well-documented (including metadata), with consistent formats and reliable long-term availability. To enhance trust and usability, it is advisable that data products be provided in the formats and accompanied by metadata according to community recognised and accepted standards (e.g. NetCDF and CF convention), and each product to be assigned a PID to ensure traceability.

Research Gap 12 (G12): Continue advancing community standards

The majority of the datasets currently in use follow established community standards, but additional progress in developing and harmonizing these standards is still expected.

Research Gap 13 (G13): Let's improve interoperability!
Interoperability of source data is considered the most critical FAIR principle. To mitigate this challenge, application of WMO Technical Regulations No. 49, GCOS standards and WMO Stewardship Maturity Matrix for Climate Data (SMM-CD) could be considered.

3.2.2 Data processing

Surveyed CS providers mostly apply both in-house and external **software** (70%, see Appendix A8). Software is applied at each step of product and service preparation and provision: data storage and retrieval, data processing, analysis, development of customised products, visualisation and dissemination.

Experiences in applying software differs and depend on the software used. There is a generally positive attitude towards open-source and non-commercial software, that has a helpdesk and regular updates. Nevertheless, some software is discontinued, with no helpdesk, no updates, and with limited visualization capability. Ideally, and where applicable, it is suggested that software be adaptive to different climatic regimes of various regions rather than being a purely global application.

Common python packages for climate indices (e.g. Xclim and ICClim) have seen strong uptake, but not everyone shares the code for the customized indices, which would be useful for open science. One provider has better experience with CDO and NCO than with Python.

Advantages of in-house developed software and well-established workflows include agility, ability to process data reliably and in a timely manner, and the capacity to provide operational services with customer assessment. Yet, different challenges may arise such as input data issues (e.g. heterogeneous data formats), occasional gaps or inconsistencies in external datasets that increase processing time, stacking of a lot of product steps, and different use of vocabularies.

Python is the most commonly used **programming language** among surveyed providers, followed by R and Fortran (see Fig. 7 below). Other languages in use include Java/Javascript, SQL and C/C++. In addition to these, CS providers employ NCL, PHP, Perl, as well as Shell scripts.

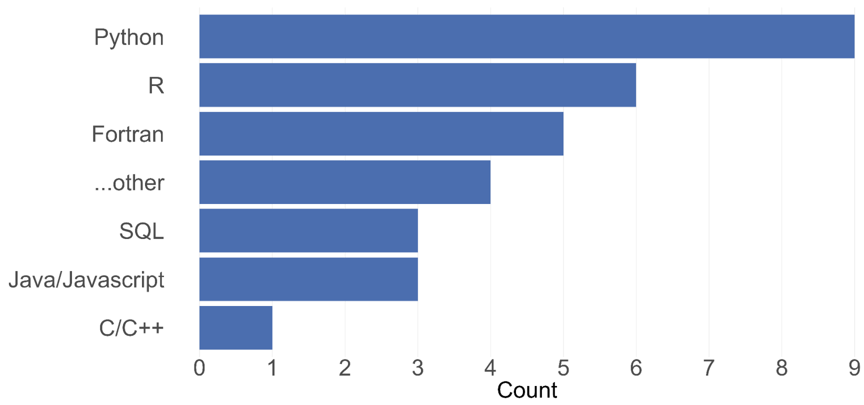


Fig. 7: Programming languages used in CS for data processing, visualization and/or digital service provision (multiple answers allowed)

There are also numerous **software packages** used for production and visualisation of products. In addition to programming languages for direct data manipulation, external software like Excel, Clisys, Clidata and ClimPACT is applied. For visualisation, GIS software like QGIS and ArcGIS, as well as Bokeh are commonly used. **AI/ML** is still not widely implemented in CS processes. Only two providers stated that they apply AI/ML, and the other three are planning to do so (see Appendix A9). An overview of **digital services** provided by CS is given in the table in Appendix A10.

With regard to **criteria, guidelines and best practices** for software and service development, survey responders mainly apply open-source software usage and sharing as well as software versioning (see Fig. 8 below). Equally, some providers are not informed about these standards or do not apply them. CI/CD as well as standards for software and service security and code review, are applied by three and two providers respectively. Additionally, ISO 9001/2015 is also used for software and service development.

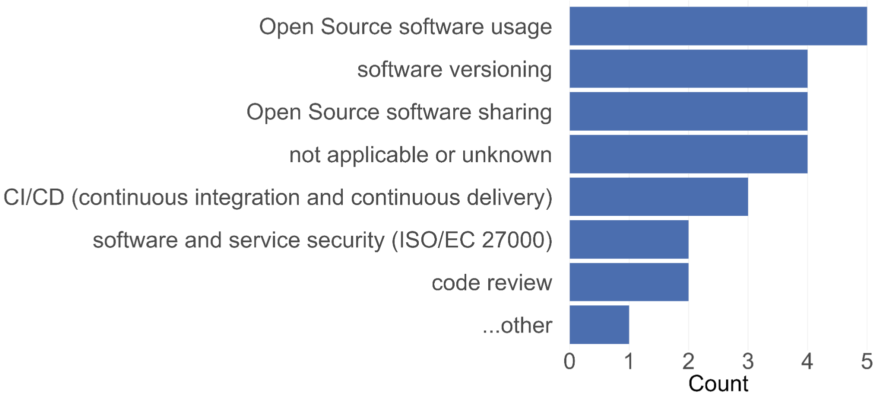


Fig. 8: Criteria or guidelines/best practices applied for software and service development. Multiple answers were allowed.

On average, surveyed CS providers rate their service maturity (3.6) and the transparency of applied methods (3.4) the highest, while the weakest aspect of CS is software and service documentation (3.0; Fig. 9). As many services are developed during project funding, the consistency of service delivery and maturity are affected, and documentation is rarely prioritized at the end of projects. Some of these issues are addressed through software tracking, but it is not as effective as having people with experience.

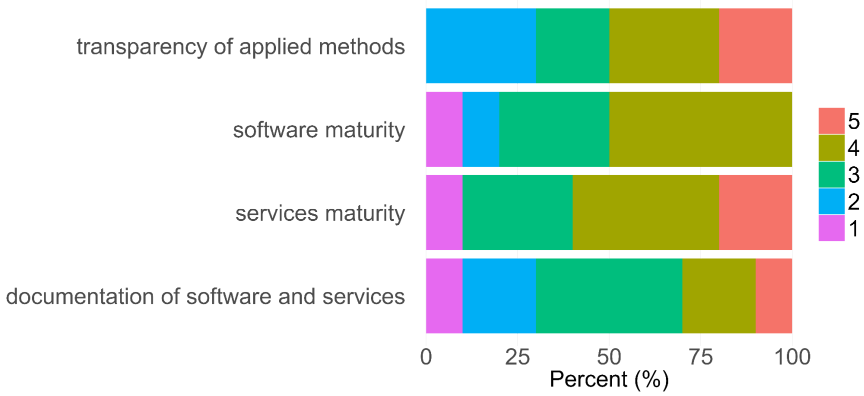


Fig. 9: Rating of different aspects of software and digital service (1 = low/bad to 5 = high/good).

In general, guidelines for software licensing and security are considered the most important, compared to software versioning and ethics (Fig. 10). Open software publications, like GitHub, are considered as important as journal publications for open science, although they are not valued as academic metrics. Survey respondents also stated that data security guidelines exist but did not specify them, as well as suggested exploring WMO e-library.

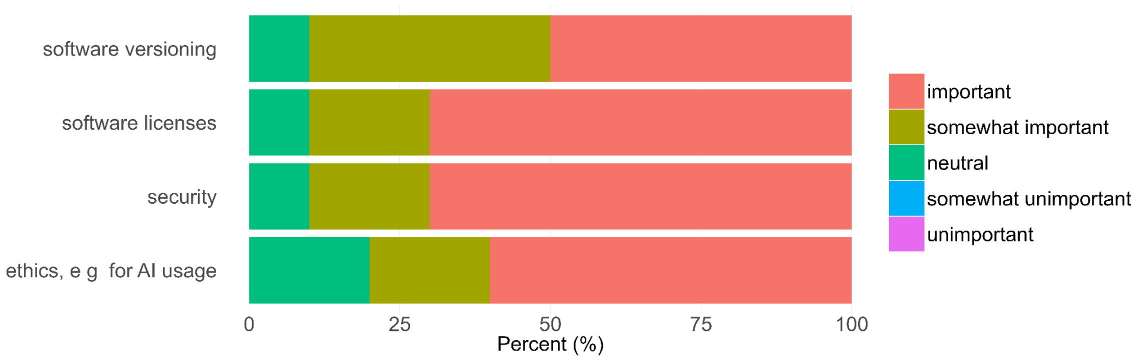


Fig. 10: Rating of different aspects for software and digital services

Community Insight 7 (C7): If possible establish user support for software
It is desirable that open-source and non-commercial software have a helpdesk and regular updates.

Community Insight 8 (C8): Always include rigorous software documentation
One of the weakest aspects of climate services seems to be software and service documentation. Since many services are developed within the framework of project funding, the consistency and maturity of service delivery are often affected by the end of the project, and documentation is rarely prioritized. Some of these challenges are partially addressed through software tracking tools.

Research Gap 14 (G14): Raise awareness about software and service security standards and best practices
Software and service security standards, as well as CI/CD processes, are rarely implemented or are not known among surveyed CS providers. Active promotion of these standards and best practices might be needed.

Research Gap 15 (G15): Value open software publications more
It is suggested that open software publications be regarded as equally important as peer-reviewed publications.

3.2.3 Data and information product sharing and communication

With regard to products that CS providers offer, all of survey responders produce data products and expert reports, and the majority of them provide service applications as well (Fig. 11, left). Products are available online on CS websites or can be obtained upon request, while some products, like reports, are sent directly to the customer (Fig. 11, right). Depending on the product type, several CS providers offer online machine-access or GUI as well. Products can also be customised based on contracts with clients or specific requests.

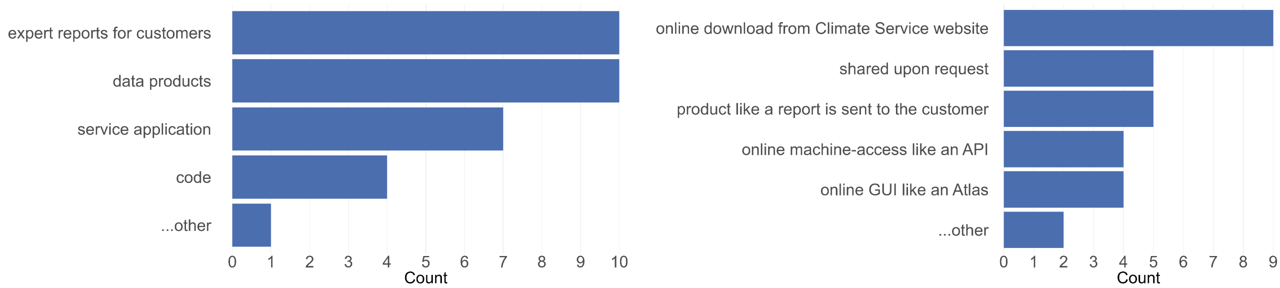


Fig. 11: Types of CS products (left) and their access (right). Multiple answers were allowed.

Both good and bad **experiences** arise from developing products or services. Good outcomes often result from cooperation with users and dedicated studies, for example, providing and documenting bias-adjusted data, initiating new monthly heating and cooling analyses for the energy sector, or creating a dashboard with custom climate indices and time scales covering the past, near and distant future. For the services developed within timebound projects, challenges are changes in the scope and focus of the project during its implementation, or a lack of responsibility for software maintenance after project completion. Additional issues may arise from input data limitations, including limited data availability, the need to reiterate the whole development process when discovering the problem with the input, not having input information on the needed resolution, or downtime caused by third party providers.

Data and information products are provided in various **formats**, with NetCDF, CSV and PDF being the most common (see Fig. 12 below), while XLSX and plain text are the least used. Products comply with community standards like the CF Convention and the Attribute Convention for Data Discovery (ACDD) for NetCDF, OGC standard for geospatial products, as well as WMO and ISO standards.

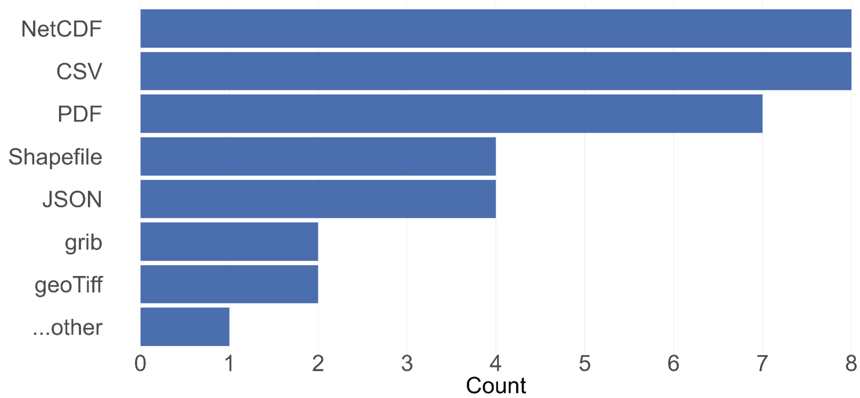


Fig. 12: Formats of data and information products (multiple answers allowed)

Customer support is provided through documentation and education, and to a lesser extent, through listening to clients' needs by collecting their feedback on a regular basis, providing notification, as well as communicating via email and providing a ticketing system (see Fig. 13 below).

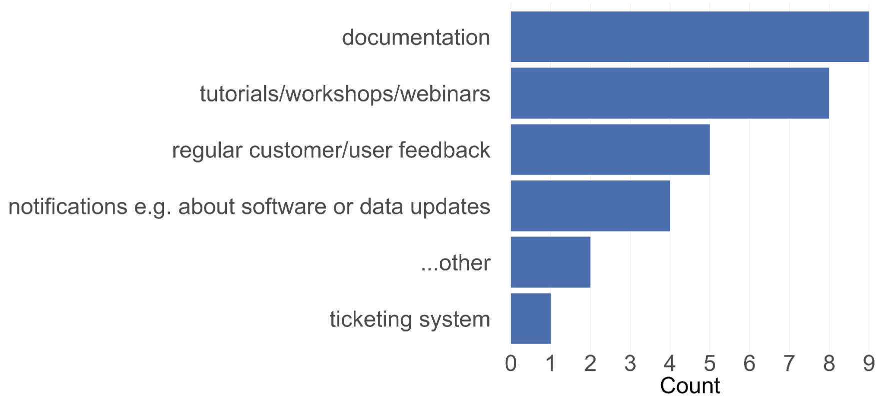


Fig. 13: Methods of customer support (multiple answers allowed)

Surveyed CS providers are satisfied with the simplicity of their products, user engagement, and their compliance to standards (average rating 3.7/5; Fig. 14). There is room for improvement, particularly in product development and versioning (average rating 3.1/5).

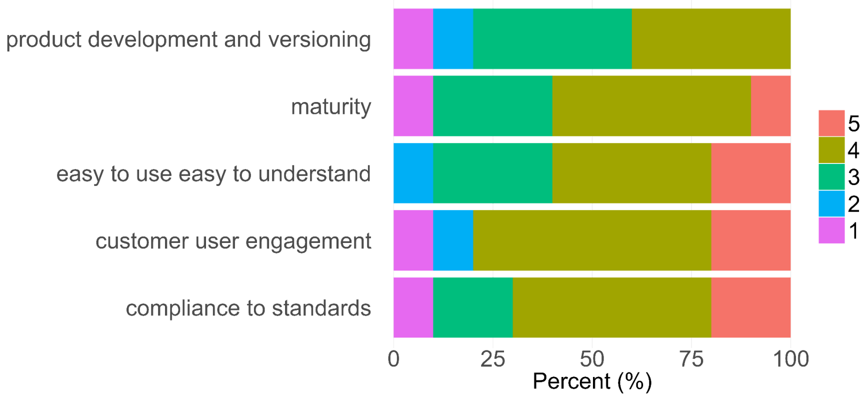


Fig. 14: Rating for different aspects of products (1 = low/bad to 5 = high/good).

Standards and guidelines related to products are considered very important for data formats and product sharing, while naming conventions and vocabularies are considered somewhat important (Fig. 15). The least importance is assigned to licenses and machine accessibility and readability. One respondent noted that attention should be paid to WMO, especially the GFCS, guidelines.

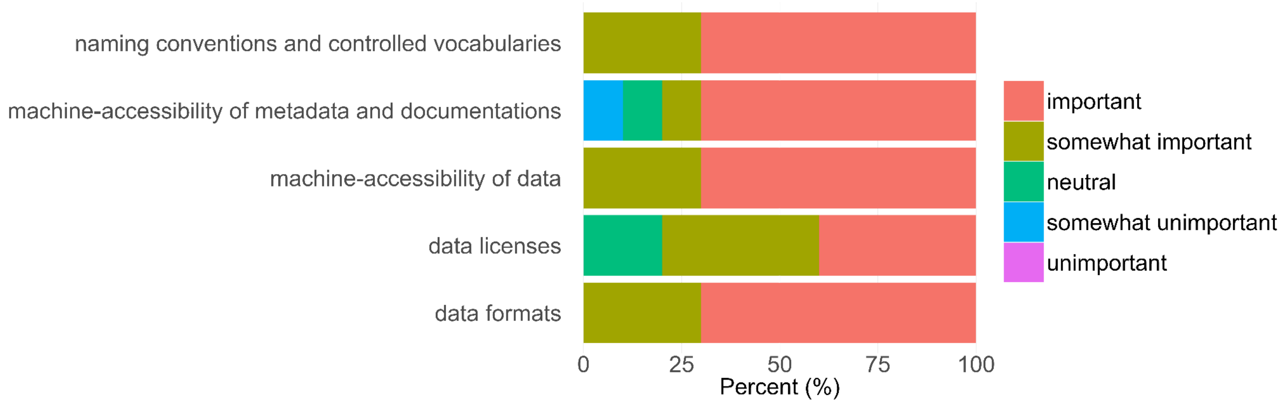


Fig. 15: Importance of standards for different aspects of products.

Community Insight 9 (C9): Raise awareness about licensing, machine accessibility and readability
Standards and guidelines regarding licensing, machine accessibility and readability are not recognized as the most important aspects of products. It might be necessary to actively promote these standards and best practices.

3.2.4 Organisational standards

The majority of surveyed CS providers have implemented either ISO 9001, in-house developed or published guidelines (Fig. 16). None of the respondents have applied for the CoreTrustSeal certificate for data preservation self-assessment (one of the response options).

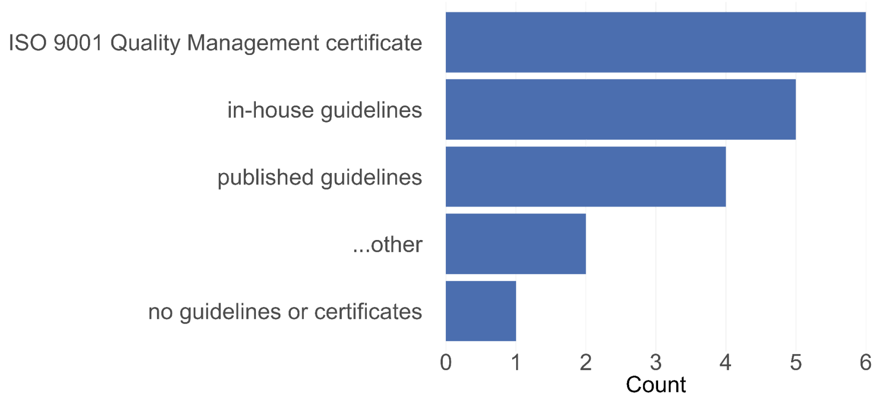


Fig. 16: Number of responses on the question "Does your organization have a guideline or certificate on its general operations?" (multiple answers allowed)

A stated benefit of implementing ISO 9001 is the requirement for feedback from external stakeholder surveys. Additionally, stakeholders often value the certificates, although in some cases having a certificate may lead to conflicts of interest. Positive experience in implementing Quality Management Systems in climate services have contributed to the issuance of WMO guidelines: WMO No. 1221

(Guidelines on Quality Management in Climate Services (2018))[76], WMO No. 1100 (Guide to the Implementation of Quality Management Systems for National Meteorological and Hydrological Services and Other Relevant Service Providers (2017))[77], and WMO No. 1285 (Guidelines for the Assessment of Competencies for Provision of Climate Services (2022))[78].

Internal guidelines ensure consistent workflows, reliable data handling, and clear responsibilities across staff. However, unexpected updates to these guidelines can sometimes prevent timely deployment.

Standardisation of data, products, and services is considered to facilitate the work of producers and users, making products comparable and reliable. It also helps filter out companies that do not follow good practices and "oversell" their products. Through operational consistency, data quality, service reliability, and end user support, standards strengthen trust and transparency with users, provide convenience to decision makers and help organizations remain competitive in the market. Additionally, standards or guidelines may be legally required.

The surveyed CS providers highly value the reliability of services and certification, as well as sustainability of services and transparency of operations (Fig. 17). These aspects are considered as essential, as there is recognised need for more sustainable and operationalized services. However, developing formal guidelines or applying for certification can also be time-consuming and resource-intensive, especially when external standards or audits are involved.

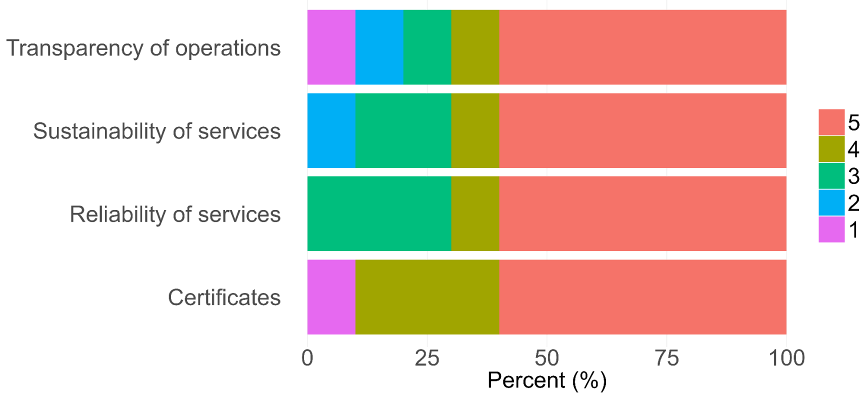


Fig. 17: Rating of different aspects of CS (1 = unimportant to 5 = very important).

3.2.5 Key messages and take aways

The survey results broadly confirm the findings from the interviews and highlight several recurring challenges and practices among climate service providers:

- Access to reliable and well-documented climate datasets remains a key concern.
- Interoperability issues continue to affect the usability of climate data, particularly due to differences in formats, metadata conventions, and access mechanisms across providers.

[76] https://library.wmo.int/idurl/4/56255
[77] https://library.wmo.int/idurl/4/50552
[78] https://library.wmo.int/idurl/4/58175

- Standardised documentation and metadata are considered essential to improve transparency and facilitate the reuse of climate data products.
- Clear communication of data processing methods and assumptions is recognised as important for building trust in climate services.
- User-oriented delivery formats are widely used, including data platforms, dashboards, APIs, and tailored reports, reflecting the diversity of user needs.
- Operational constraints, including data volume and processing complexity, remain important considerations in the development and delivery of climate services.
- Respondents emphasise the importance of collaboration and community standards to improve interoperability and reduce fragmentation in the climate services ecosystem.

4. Recommendations

The following section presents recommendations on good practices for metadata, vocabularies, data licensing and citation, software development, as well as governance and compliance, including FAIR principles for both data and software. The literature-derived recommendations in subsection 4.1 are also consistent with the authors' expertise in the field of data and software management in Climate Services. However, due to diversity of data handled by CS, not all sector-specific data needs can be covered in this deliverable.
The community insights in subsection 4.2 present additional insights from the qualitative interviews and survey conducted with CS providers. Due to the relatively small sample size of answers, however, these should be read as personal insights from single CS practitioners in their daily work rather than representative results across the CS landscape.

4.1 Literature-derived recommendations

The following recommendations are organised into five thematic groups covering metadata and standards, vocabularies and semantics, licensing and citation, software practices, and governance. They are based on thorough desk-based research of community standards and good practices as well as on the professional expertise of the authors.

Metadata

R1: Adopt established metadata standards
Use established metadata standards such as WMO Core Metadata Profile (WCMP) or WIGOS Metadata Standard (WMDS) for weather and climate **observation** data, CMIP metadata standards for **model data**, and the more general metadata schemas ISO 19115-1 or DataCite Metadata Schema for metadata in **interdisciplinary** contexts. These standards provide a structured way to document information and enhance findability, accessibility, and thus reusability of the data.

R2: Provide high-quality metadata to enable data reuse

Provide high quality, standardised, and complete metadata including information on observation methods or model design, uncertainty and variable descriptions. This facilitates data homogeneity, understanding and future (re-)use of the data.

R3: Use CF-compliant NetCDF or Zarr as primary formats for climate data workflows
Given the large number of community tools available to process NetCDF or Zarr file formats, it is recommended to use these formats for the maximum number of data and processing steps within workflows. NetCDF is preferred for non-cloud environments, GRIB could also be used but is more complex but more efficient. Zarr is suitable for cloud environments. Both of these formats must use Metadata Standards and Vocabularies (such as CF-Conventions: see section 2.3 on Vocabularies).

R4: Use community-specific formats for output products
Output of CS should use community-specific file formats. NetCDF/Zarr for most of the research communities, CSV can be used for smaller datasets to make data format more accessible for non-experienced users. BADC-CSV[79] is recommended for CSV data; it is a structured format with metadata that is interoperable with NetCDF metadata conventions. For reports, PDF (PDF/A standard) is good but HTML5 or Markup may be a better solution depending on the context, because it can be more accessible and reusable (more FAIR) than PDF, and machine-readable.

Vocabularies

R5: Use FAIR vocabularies to annotate data throughout the workflow
A FAIR vocabulary is one that is documented and where each term is resolvable using a globally unique and persistent identifier. The vocabulary documentation needs to be easily findable and accessible by anyone who uses the dataset. Vocabularies provided via semantic services such as the NERC Vocabulary Server (https://vocab.nerc.ac.uk/) satisfy these requirements.

R6: Request a new term in an existing vocabulary when needed
If there is an existing FAIR vocabulary that covers the domain of interest, but it does not contain a suitable term for your data, e.g. a parameter name, it is best practice to engage with the governance processes for the vocabulary to request the addition of a new term. This is highly preferable to creating a new, separate, undocumented vocabulary.

R7: Use CF vocabularies for NetCDF or Zarr data
The CF metadata conventions are designed to work alongside the NetCDF format and are also compatible with Zarr. They contain controlled vocabularies for parameter naming (CF standard names), surface characteristics (CF area type) and geographical areas with complex boundaries. The CF conventions stipulate the use of UDunits for providing information about units of measure.

Data Licensing and Citation

R8: Include an explicit licence for data
Licence data and data derivatives under the Creative Commons Attribution 4.0 International (CC BY 4.0) where this does not infringe the interests of relevant licence holders. The open licence CC BY 4.0 permits the reuse of data under the condition that appropriate credit is given. https://creativecommons.org/licenses/by/4.0

[79] https://doi.org/10.5281/zenodo.7355840 (pp1-6)

R9: Issue DOIs for data products
Data products of public CS should be issued with a DOI and a "cite as" statement should also be included. This makes the products referenceable and citable for climate service customers.

R10: Document all data modifications
Any modification of data should be indicated in order to support the transparency and traceability of climate service products.

Software Development

R11: Adopt FAIR-aligned software management practices across the full software lifecycle
Software developed or adopted for climate services should align with FAIR and FAIR4RS principles across its whole lifecycle. This should be operationalised through a software management plan covering development, release, and maintenance. Key elements include adherence to community standards for coding, testing, versioning, and review, and defined responsibilities for maintenance and long-term sustainability. Applying these practices proportionally improves transparency, reproducibility, and long-term reusability of software.

R12: Address governance, provenance, and user requirements during software planning
Decisions made at the planning stage should explicitly address intended use, provenance, governance, and user communities, in order to ensure that FAIR4RS principles, traceability requirements, and stakeholder engagement are embedded before development or adoption begins.

R13: Follow recognised software development standards
Software development should follow recognised language- and community-specific standards for code structure, configuration, interfaces, and versioning. This is supported by automated quality checks and reproducible execution environments to ensure maintainability and interoperability.

R14: Ensure quality and scientific validation of software
Software used in climate services should be subject to both rigorous technical quality assurance and thorough scientific validation. Technical quality assurance should include automated testing pipelines (unit, integration, and system tests), continuous integration and deployment, and code peer review against recognised quality standards. Scientific validation should verify that outputs are physically consistent, reproducible against reference datasets, and reviewed by domain experts. Together, these practices ensure the reliability, scientific integrity, and consistency of software across versions and as it evolves.

R15: Ensure findability and proper citation of software
Released software should be accompanied by clear and compatible licenses, persistent identifiers, archived versions, and machine-readable metadata. These aspects enable the discoverability, legal reuse, citation, and long-term access of software assets.

R16: Establish maintenance and risk-management practices for software
Software should have defined maintenance responsibilities, support mechanisms, and deprecation policies. The sustainability of external dependencies should be regularly assessed to manage operational risks and ensure long-term usability.

Governance and Compliance

R17: Consider regulatory alignment in climate service design
The design of climate services necessitates a robust understanding of the EU legal framework. Because these services often process sensitive data, utilize complex modeling for high-stakes decision-making, and serve diverse stakeholders, they fall directly within the scope of existing EU regulations. Addressing these legal requirements during the initial design phase is essential to ensure both compliance and operational integrity.

R18: Adopt standards to strengthen operational quality and security
Attaining ISO 9001 and ISO 27001 certifications is a critical milestone for climate service providers seeking to formalize their operational excellence. These standards provide a dual-layered assurance: ISO 9001 validates the robustness of the quality management system, while ISO 27001 ensures that sensitive environmental and proprietary data are managed in strict accordance with international security protocols. Together, they instill the stakeholder confidence necessary for high-stakes decision support.

4.2 Additional community insights from interviews and survey

The additional community insights offer further observations from the CS community. However, they are set apart from the recommendations due to the relatively small number of participants. They may reflect individual opinions and might not be representative of the CS community as a whole. Nevertheless, these insights offer interesting practical perspectives from CS providers on top of the literature-based recommendations.

C1: Engage with users to co-design the delivery system
Climate services can be provided in multiple ways, and the choice should be made according to the needs of the client or targeted user community.

C2: Build trust through transparency
As transparency builds trust and provides evidence of quality, it is recommended that CS providers be as transparent as possible about their products without compromising their intellectual property or competitiveness in the market. In cases involving proprietary information, it is preferable to grant some level of transparency through confidentiality agreements.

C3: Commit to user-driven development
User engagement is seen as essential for developing the service, regardless of the type. Direct and regular feedback can contribute to service improvements and user satisfaction. Climate Service products should be developed in collaboration with users or based on dedicated studies.

C4: Establish user support to maximize product uptake

It is advisable that CS provide an appropriate level of user support depending on the user type and expertise, such as data catalogues, technical documentation, notifications, communication channels or educational resources.

C5: Qualifications and continuous education of CS employees are essential for service quality
Ensuring all technical and quality requirements for development and delivery of climate services is impossible without qualified employees. Qualifications, continuous training, and professional development of service providers are recognised as essential for guaranteeing service quality and implementation of standards and best practices.

C6: Input and output datasets should be quality-controlled and FAIR
It is recommended that datasets used as input for product development be quality-controlled and well-documented (including metadata), with consistent formats and reliable long-term availability. To enhance trust and usability, it is advisable that data products be provided in the formats and accompanied by metadata according to community recognised and accepted standards (e.g. NetCDF and CF convention), and each product to be assigned a PID to ensure traceability.

C7: If possible establish user support for software
It is desirable that open-source and non-commercial software have a helpdesk and regular updates.

C8: Always include rigorous software documentation
One of the weakest aspects of climate services seems to be software and service documentation. Since many services are developed within the framework of project funding, the consistency and maturity of service delivery are often affected by the end of the project, and documentation is rarely prioritized. Some of these challenges are partially addressed through software tracking tools.

C9: Raise awareness about licensing, machine accessibility and readability
Standards and guidelines regarding licensing, machine accessibility and readability are not recognized as the most important aspects of products. It might be necessary to actively promote these standards and best practices.

4.3 Feedback from the Deltares Use Case

Our recommendations were reviewed by Deltares from their use case perspective. They acknowledged that these recommendations are interesting and useful to reflect on the standardisation of climate services. As a provider of software tools that translate climate data into actionable insights, Deltares would benefit from the alignment of these recommendations with the FAIR principles, as it would help identify potential gaps and areas for improvement in their existing approaches. We anticipate that this alignment will be further consolidated in the overarching high-level guidelines to be synthesized collaboratively within the Climateurope2 project and along the existing standardisation components.

5. Key Research Gaps

The following research gaps have been identified from literature and report research, but also out of expertise and experience of the co-authors of this deliverable as well as from interviews and survey answers by members of the Climate Services Community.

5.1 Literature-derived research gaps

G1: Harmonizing FAIR software tools and metrics
Existing tools for supporting and assessing FAIR software practices rely on heterogeneous frameworks and metrics limiting consistent evaluation, comparison, and automation. This fragmentation hinders the integration of FAIR software assessments into development workflows and increases the burden on developers and users. Harmonised FAIR software metrics and interoperable assessment tools are needed to enable consistent, transparent evaluation, and to support informed software adoption and reuses.

G2: Concrete operational frameworks are needed to translate high-level principles (e.g. FAIR, CARE, TRUST) into verifiable implementation practices
The lack of standardised interpretations for high-level principles remains a significant hurdle. Without clear, reproducible implementations, stakeholders cannot move beyond subjective assessments toward rigorous verification. Establishing these concrete benchmarks is a necessary step to facilitate transparent compliance audits and meaningful cross-service comparisons.

G3: Bridge the recognition gap for non-certifiable standards
A significant gap exists in the market recognition of non-certifiable standards within the climate services community. While frameworks such as the ISO/IEC 25000 series provide essential guidelines for software quality, they often remain underutilized due to the absence of a formal validation mechanism. To bridge this gap and incentivize adoption, a dedicated certification framework should be developed, providing a structured pathway for providers to demonstrate compliance and for users to verify service excellence.

G4: Standardised and machine-actionable frameworks for Data Management Plans could improve transparency, interoperability, and automation of data governance processes
Standardizing Data Management Plans (DMPs) offers a significant opportunity to enhance operational transparency and ensure the structural completeness of data management strategies. However, widespread and consistent adoption depends on the availability of clear, actionable implementation frameworks. Establishing such clarity would facilitate the systematic comparison of services and the identification of strategic gaps. Furthermore, the integration of machine-actionable formats would allow for automated updates and real-time synchronization, shifting DMPs from static documents to dynamic management tools.

5.2 Community-perceived research gaps from interviews and survey

G5: Guidelines to navigate data are needed
Since there are many data providers as well as numerous data types and datasets with varying availability and different formats, users often struggle to select the most appropriate one. Clear

guidelines are needed to help users navigate the vast amount of available data and obtain the data that best suits their purpose.

G6: Keep working on overcoming barriers in data access and interoperability
Differences in data access protocols, metadata conventions, and data formats across climate data providers continue to limit interoperability and efficient integration of datasets.

G7: Lack of harmonized API implementations
Providing a service most of the time requires having an API. API creation can be done in multiple ways, which can lead to incompatible outputs among different providers.

G8: Find the balance between different levels of obligation
While standards are seen as important for ensuring quality and credibility, most CS providers prefer relying on recognized best practices rather than strictly following formal ISO standards, as they find them more practical and flexible. Therefore, different degrees of obligation within standardisation frameworks could be explored, allowing for context-sensitive application.

G9: Will standardisation kill innovations and "What's in it for me?"
There is a concern that standardisation may stifle innovation. Additionally, many CS providers often lack a clear understanding of the benefits that certification, such as for QMS standards, would bring.

G10: Reach agreement on benchmarks to be used
In addition to standards, some CS providers think that the climate services market also needs benchmarks to help position providers, and needs to find an agreement on which benchmarks to adopt, whether sector-specific or broader ones that address general needs.

G11: "What are we talking about?" i.e. ensuring consistent terminology
Some CS providers ask for clear and harmonized terminology. Establishing a common vocabulary would improve communication and create a shared language between practitioners and end users. Furthermore, some find that the EU taxonomy remains relatively limited, suggesting more standardised sustainability reporting.

G12: Continue advancing community standards
The majority of the datasets currently in use follow established community standards, but additional progress in developing and harmonizing these standards is still expected.

G13: Let's improve interoperability!
Interoperability of source data is considered the most critical FAIR principle. To mitigate this challenge, application of WMO Technical Regulations No. 49, GCOS standards and WMO Stewardship Maturity Matrix for Climate Data (SMM-CD) could be considered.

G14: Raise awareness about software and service security standards and best practices
Software and service security standards, as well as CI/CD processes, are rarely implemented or are not known among surveyed CS providers. Active promotion of these standards and best practices might be needed.

G15: Value open software publications more
It is suggested that open software publications be regarded as equally important as peer-reviewed publications.

6. Conclusion

Deliverable 2.6 establishes a structured, community-informed framework for implementing good practices in using FAIR metadata, formats, vocabularies, licensing, software management, and governance within climate services. By aligning with established international standards and the FAIR principles, and by integrating community experience, it contributes recommendations to strengthen the saliency, credibility, legitimacy, and sustainability of climate services in Europe and beyond.

Ultimately, consistent application of these guidelines will enhance interoperability across sectors, support innovation without compromising quality, and foster long-term trust between data providers, service developers, and users in the climate services ecosystem.

References

Anderson, W. B., Morris, E. J., Smith, D. B., & Ward, M. C. (2018). COTS and Reusable Software Management Planning: A Template for Life-Cycle Management (Version 1). Carnegie Mellon University. https://doi.org/10.1184/R1/6572816.v1

Barker, M., Chue Hong, N.P., Katz, D.S. et al. Introducing the FAIR Principles for research software. Sci Data 9, 622 (2022). https://doi.org/10.1038/s41597-022-01710-x

Carroll, S.R., Herczog, E., Hudson, M. et al. Operationalizing the CARE and FAIR Principles for Indigenous data futures. Sci Data 8, 108 (2021). https://doi.org/10.1038/s41597-021-00892-0

Data Citation Synthesis Group (2014). Joint Declaration of Data Citation Principles. Martone M. (ed.) San Diego CA: FORCE11. https://doi.org/10.25490/a97f-egyk

David, M., Colom, M., Garijo, D., Castro, L. J., Louvet, V., Ronchieri, E., Torquati, M., del Caño, L., Cerlane, L., Van den Bossche, M., Campos, I., & Di Cosmo, R. (2024). Ensure Software Quality (1.0.0). Zenodo. https://doi.org/10.5281/zenodo.10723608.

Eaton, B., Gregory, J., Drach, B., Taylor, K., Hankin, S., et al. NetCDF Climate and Forecast (CF) Metadata Conventions (2024). Zenodo. https://doi.org/10.5281/zenodo.14275599

Huard, D., Pirani, A., Chen, R., Gutiérrez, J. M., Juckes, M., Krey, V., Spinuso, A., & Stockhause, M. (2022). IPCC Data and Code Licensing Guidelines (1.1). Zenodo. https://doi.org/10.5281/zenodo.7431834

Lin, D., Crabtree, J., Dillo, I. et al. The TRUST Principles for digital repositories. Sci Data 7, 144 (2020). https://doi.org/10.1038/s41597-020-0486-7

Mourkoussis, N., White, M., Patel, M., Chmielewski, J., & Walczak, K. (2003). AMS--Metadata for Cultural Exhibitions using Virtual Reality. Proceedings of the International Conference on Dublin Core and Metadata Applications, 2003. https://doi.org/10.23106/dcmi.952107475

Parton, G. A., & Pepler, S. (2009). The BADC-CSV Format: Meeting user and metadata requirements (pp. 1–6). Zenodo. https://doi.org/10.5281/zenodo.7355840

Stockhause, M., Bärring, L., Bretonnière, P.-A., Cripe, D., Delju, A. H., Krijger, T., Krzic, A., Mihic, D., Muñoz, Á. G., Pagé, C., Pamment, A., Pascoe, C. L., Spinuso, A., & Thijsse, P. (2024). Preliminary consistent guidelines of good practices for vocabularies, formats, metadata and technical standards, including FAIR principles for data and software - Deliverable D2.1 of the Climateurope2 project (1.0). Zenodo. https://doi.org/10.5281/zenodo.10912719

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

World Meteorological Organization (WMO). Guide to the Implementation of Quality Management Systems for National Meteorological and Hydrological Services and Other Relevant Service Providers (WMO-No. 1100). Geneva, 2017. https://library.wmo.int/idurl/4/50552

World Meteorological Organization (WMO). Guidelines on Quality Management in Climate Services (WMO-No. 1221). Geneva, 2018. https://library.wmo.int/idurl/4/56255

World Meteorological Organization (WMO). WIGOS metadata standard (WMO-No 1192). Geneva, 2019. https://library.wmo.int/idurl/4/55626

World Meteorological Organization (WMO). Guide to Instruments and Methods of Observation (WMO-No. 8). Geneva, 2021–2025. https://library.wmo.int/idurl/4/41650

World Meteorological Organization (WMO). Guidelines for the Assessment of Competencies for Provision of Climate Services (WMO-No.1285). Geneva, 2022. https://library.wmo.int/idurl/4/58175

World Meteorological Organization (WMO). *Manual on the WMO Information System* (WMO-No. 1060), Volume I. Geneva, 2023. Appendix C. https://library.wmo.int/idurl/4/35315

World Meteorological Organization (WMO). Manual on the High-quality Global Data Management Framework for Climate (WMO-No. 1238). Geneva, 2023. https://library.wmo.int/idurl/4/56975

World Meteorological Organization (WMO). Guide to Instruments and Methods of Observation (WMO-No. 8), Volume III. Geneva, 2024. https://doi.org/10.59327/WMO/CIMO/3

Appendix

A1 Table of WIGOS metadata categories

Category	Description
1. Observed variable	Specifies the basic characteristics of the observed variable and the resulting datasets. It includes an element describing the spatial representativeness of the observation as well as the biogeophysical compartment the observation describes.
2. Purpose of observation	Specifies the main application area(s) of the observation and the observing programme(s) and networks the observation is affiliated to.

3. Station/platform	Specifies the observing facility, including fixed station, moving equipment or remote-sensing platform, at which the observation is made.
4. Environment	Describes the geographical environment within which the observation is made. It also provides an unstructured element for additional meta-information that is considered relevant for adequate use of the observations and that is not captured anywhere else in this standard.
5. Instruments and methods of observation	Specifies the method of observation and describes characteristics of the instrument(s) used to make the observation. If multiple instruments are used to generate the observation, then this category should be repeated.
6. Sampling	Specifies how sampling and/or analysis are used to derive the reported observation or how a specimen is collected.
7. Data processing and reporting	Specifies how raw data are transferred into the observed variables and reported to the users.
8. Data quality	Specifies the data quality and traceability of the observation.
9. Ownership and data policy	Specifies who is responsible for the observation and owns it.
10. Contact	Specifies where information about the observation or dataset can be obtained.

A2 Protocol for interviews conducted in CE2 WP2, T2.1 for Deliverable D2.6

Topics to ask about:

- ask about the interviewee
- ask about user experience/knowledge/CS example
- ask about challenges/limitations/obstacles
- ask about gaps/needs/suggestions
- ask about sector-specific issues
- ask about further aspects

Stages of a CS:

1. Data / Software:
    a. Input data

    b. Data distributor
    c. Processing/calculating indices
2. Service / Communication:
    a. Visualisation
    b. Communication
    c. Feedback
3. Target audience: CS providers

Proposed Questions (ask along the workflow steps):

1. Information about the interviewee and CS example
    a. Interviewee: experience/knowledge/role
    b. CS example: pilot/operational/research project, general/sector-specific
2. Data Source (for Usage/Production):
    a. What data sources do you use? Are these sources inhouse or external?
    b. Are there specific data requirements from your sector/CS?
    c. What standards are used for the data sources (if known)? E.g. standards for observations etc.
    d. (For external data sources) What experience do you have in using data with respect to finding, selecting, downloading and integrating them into your CS? Think of an example for what works well and one which does not work well.
    e. (For external data sources) Are there any gaps, obstacles or limitations in accessing and using the data such as formats, license or access restrictions, terminology?
    f. What are the most important data needs, which are not fully met? Do you have suggestions on how obstacles and difficulties you encounter can be overcome?
3. Data processing
    a. What data processing do you apply?
    b. Do you develop your software inhouse or do you use software, which is externally developed?
    c. Do you use standards and best practices in software and software development? E.g. an open and common programming language, software versioning repositories (github), code review, unit and system testing, building, testing, deploying, ISO/IEC 25000 family.
    d. Do you share and document your software incl. methods applied? How transparent are your CS?
    e. How mature do you regard your processing? Please explain your answer.
4. Data and Information Product Sharing and Communication:
    a. What kind of data or information do you provide?
    b. Who is your target user community? Does your climate service support diverse user communities? What knowledge do they require to use your service?
    c. How do you share your data and information products? Do you support humans with GUIs as well as machines with APIs?
    d. Which data/information (standard) formats do you use?

    e. How do you support users of data and information products with tutorials, documentation, webinars and workshops, ticketing system, tool advice for further processing the data?
    f. How do you deal with data and information product errors and product improvements in terms of versioning concept, user notification etc.?
    g. Do you engage with users to improve your CS?
5. Organization/Transparency:
    a. How would you describe your corporate culture or working principles?
    b. Is your organization certified, e.g. ISO9001, CoreTrustSeal or other? How important is this/are these certificate(s) for you?
    c. How transparent are your CS products? Which details on methods, provenance and way of working do you publicly share?
    d. Do you have a commitment for a stable and sustainable service and products?

A3 List of data providers incl. types of datasets used, format, compliance with standards and user comments on gaps and needs

	Data Provider	**Datasets**	**Format**	**Standard**	**Experiences and needs**
Public	COPERNICUS & C3S	climate data seasonal forecast ERA5-Land CMIP6			There are technical challenges for future datasets like CMIP7 and ERA6 – there is a need for innovative serving methods due to increasing data volumes and user bases. At the moment these challenges are accommodated by expansion of documentation, training materials, and support systems. Needs: Better data management. Data on damages are missing.
Public	-	Reanalyses ERA5 ERA5-Land	NetCDF		Shortcomings: ERA5 regarding wind Needs: Challenges of potential future integration of weather station data into ERA6

	WCRP and others	CMIP (5 and 6)	NetCDF	CMIP is practically an ISO, it standardises, and imposes some formats like NetCDF CMIP-related vocabularies	There is not enough information about the models, so that informed decisions can be made downstream, e.g. What are the parameters they are using? What about the assumptions? How does this affect my work?
	ESGF & ESGF/DRIAS	climate data	NetCDF	CF conventions	ESGF is considered "outdated" and CDS better designed thanks to the Python package. Challenges posed by large data volumes, the necessity of verifying file integrity using hashes, and the development of tools to rigorously automate these processes. Needs: Clearer versioning and change logs, and machine-readable metadata.
	IPCC	future data CMIP assessment findings			The IPCC provides information/data on storm intensity etc. and how it affects current nature and infrastructure but not the consequences.

	National Meteorological and Hydrological Services	observational climate & hydrological data (incl. In situ observations)			Shortcomings: Metadata: Historical metadata is often incomplete, inconsistent, or lost. Observational metadata and QC processes exist but are not maintained at publishable quality. Even basic metadata like units, observation times, update intervals differ across countries. Availability: Some national services retain data for unknown or unstated reasons. Sharing: Many services find having an API challenging, while others make everything available, but not everything is valuable, and navigating through it is complicated and time-consuming. Some dataset descriptions are in the native languages. Data format is not consistent in time. Other: They can't fully adhere to WMO guidelines because of a lack of resources and capacity. Suggestion: It would be ideal if there were a standard format or protocol through which data are collected and delivered. Legislation can support data sharing by recognising its importance and enabling requests for resources.
	French data portals		CSV		
	GeoRisque (France)	climate impact and risk data	shapefile		

	Severe storms database (US by NOAA)	data on extreme events (e.g. tornadoes) including casualties and damages	excel tables	Data provider has its own standard for data quality.	The change in the intensity scale of tornadoes in the past (the Enhanced Fujita scale replaced the Fujita scale) causes inhomogeneity of time series. A good knowledge of the differences in scales, as well as vulnerability of assets, is needed in order to make the best decision in the user's interest when selecting input data.
	Public service providers	climate hazards publications from local authorities fact sheets on projections			The climate data and scenarios need to be scientifically sound and the data needs to be as up-to-date as possible and of a high resolution. Scenarios need to be consistent with IPCC's approaches. The climate data specifically for a location is so diverse. There is a large variation across states.
Private	Data provided by the client	Wind measurements on multiple heights	excel tables	Measurements are done by anemometers with calibration certificates. Calibration is done annually or biannually. Anemometers are required to meet a certain IEC (International Electrotechnical Commission) standard.	The instruments used to measure wind parameters are of extremely high quality, and measurements are taken 10 times per second. However, only 10-minute averages, which are needed to assess the wind potential, are recorded and stored. The raw data would be extremely valuable for research work, especially in the field of turbulence.
		existing data for specific projects, including geodata and concepts			Many different standards across regions are in use, requiring its harmonization. Data requires harmonization e.g. harmonizing resolution of input data across regions.
		vulnerability data			

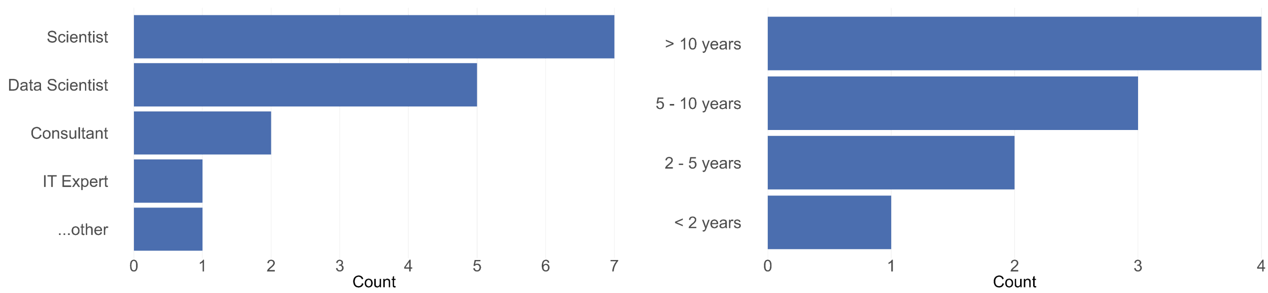
Additional data providers and datasets, that were discussed in interviews, are:

- public data providers and datasets:
    - *CORDEX & Euro-CORDEX* as well as in-house obtained RCM-outputs,
    - *World Bank* for historical data,
    - *NASA*,
    - *IGN* for high-resolution aerial imagery;

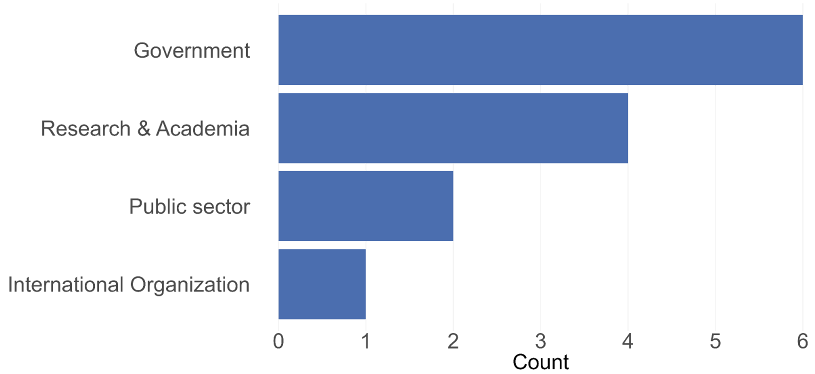
- private data providers and datasets:
    - *mapbox* for cartography, geometrical data of buildings, addresses,
    - *Vortex*,
    - private providers for climate projections for different scenarios,
    - reinsurance companies for hazard risks,
    - in-house collected/produced data like: city measurements and priority datasets,
    - sectoral data like morbidity and mortality;

- other data sources and datasets include:
    - satellite data and remote-sensing products,
    - numerical model outputs,
    - national forest inventories,
    - land-use/land-cover datasets,
    - geocoded, high-resolution asset maps,
    - crowdsourced data – very dense, but their use is limited to certain projects,
    - *ROM*,
    - *Safran*,
    - *Explore2*.

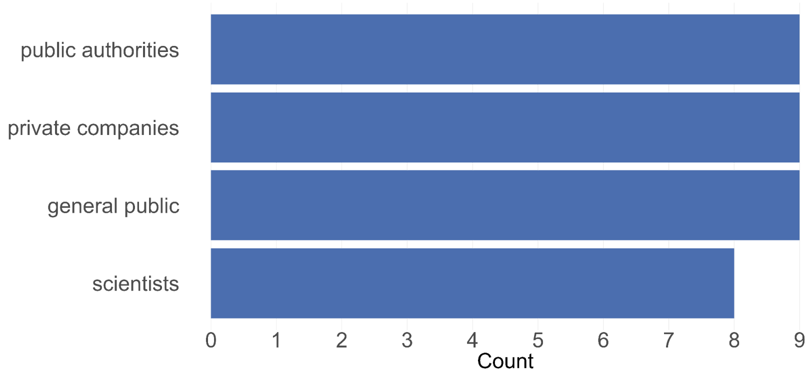
A4 Professional roles of survey participants in their CS (left) and the number of years working as CS provider (right)



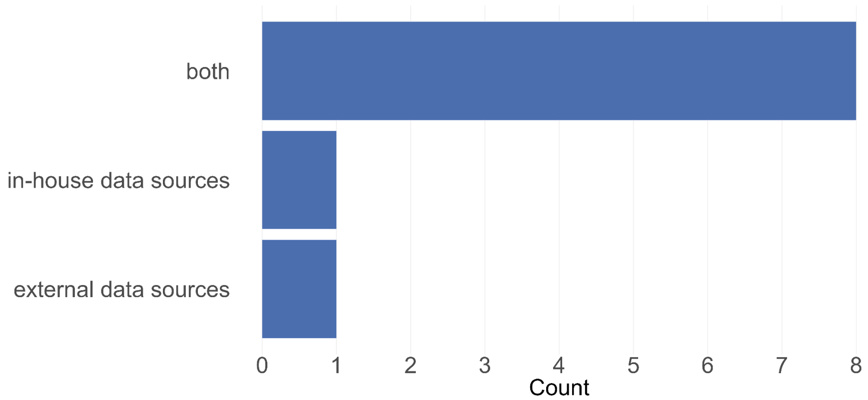
A5 Surveyed CS providers by sector (multiple answers allowed)



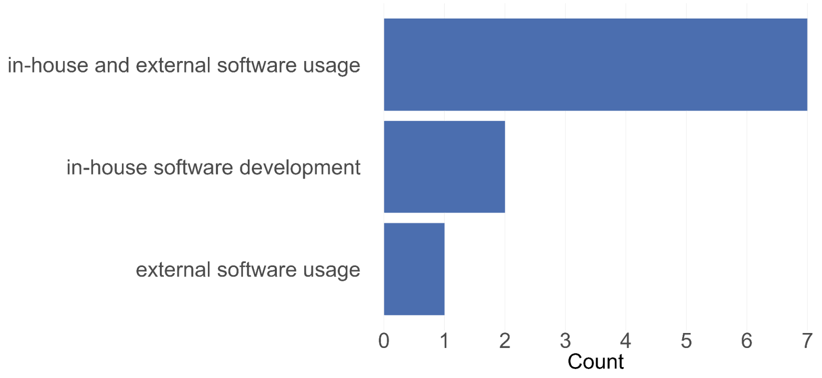
A6 Target users of surveyed CS providers. Multiple answers were allowed.



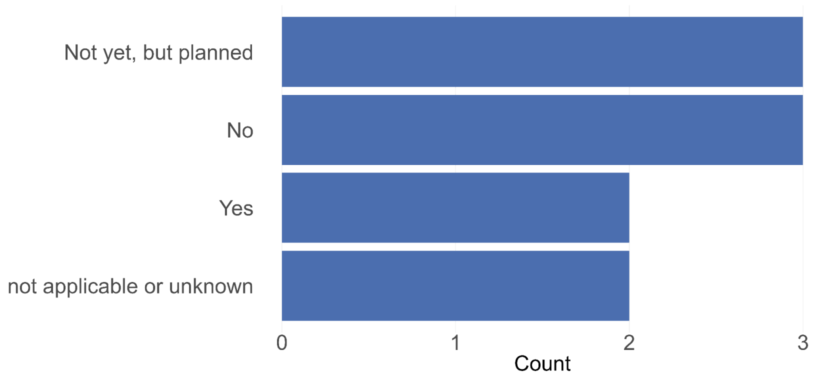
A7 Data sources used by CS



A8 In-house software development vs. external software usage



A9 Application of AI/ML



A10 Digital services provided by CS

Software and Packages	• Xclim • C3S API for ERA5-HEAT
Servers and Data Access Protocols	• THREDDS server • WMS (Web Map Service) • JSON API
Platforms and Visualisation	• Web-based data platforms • Climate Data Web Application (KLIMADAT) / Climate Atlas • Graphs
Services and User Access	• Data access (including production of specific climate indicators based on multi-model approach) • Map access • On-demand services • Technical and scientific hotline

Operational Services	• Climate data provision • Operational forecasts • Data services via web platforms and direct requests • Warnings • Reports
Types of Data	• Observations • Weather statistics • Projection data