INTERGOVERNMENTAL PANEL ON climate change

Task Group on Data Support for Climate Change Assessments (TG-Data)

The implementation of FAIR data principles in the IPCC AR6 assessment process

Anna Pirani, Andres Alegria, Alaa Al Khourdajie, Wawan Gunawan, José Manuel Gutiérrez, Kirstin Holsman, David Huard, Martin Juckes, Michio Kawamiya, Nana Klutse, Volker Krey, Robin Matthews, Adam Milward, Charlotte Pascoe, Gerard van der Shrier, Alessandro Spinuso, Martina Stockhause, Xiaoshi Xing

Version 1, April 2022

Document history, acknowledgements and citation details

This guidance document is an agreed product of the IPCC Task Group on Data Support for Climate Change Assessments (TG-Data). The mandate of the TG-Data was approved by the IPCC, but this does not imply the Panel's endorsement or approval of documents emanating from the TG-Data or any recommendations contained herein. Furthermore, this document has not been subject to the procedural IPCC review processes.

This guidance document, *The Implementation of FAIR Data Principles in the IPCC AR6 Assessment Process: TG-Data Guidance Document*, constitutes "Supporting Material" of the Intergovernmental Panel on Climate Change (IPCC), according to the Appendix A (Procedures for the Preparation, Review, Acceptance, Adoption, Approval and Publication of IPCC Reports) to the Principles Governing IPCC Work. The document, which was prepared by the FAIR sub-group of the IPCC Task Group on Data Support for Climate Change Assessments' (TG-Data) that was established at the first Face-to-Face TG-Data meeting in October 2019. The document has been reviewd by TG-Data members and ex-officio members. The authors gratefully acknowledge the comments and suggestions received from the review, which have greatly enhanced this document. The information contained herein, including opinions and possible errors are nonetheless solely the responsibility of the authors.

The final version of the document was endorsed by the TG-Data on 5 April 2022.

This document should be referenced as: Pirani, A., A. Alegria, A. Al Khourdajie, W. Gunawan, J. M. Gutiérrez, K. Holsman, D. Huard, M. Juckes, M. Kawamiya, N. Klutse, V. Krey, R. Matthews, A. Milward, C. Pascoe, G. van der Shrier, A. Spinuso, M. Stockhause and X. Xing, 2022: The implementation of FAIR data principles in the IPCC AR6 assessment process: A TG-Data Guidance Document, Supporting Material, Intergovernmental Panel on Climate Change Task Group on Data Support for Climate Change Assessments (TG-Data), 14 pp, DOI:10.5281/zenodo.6504469.

Summary

The IPCC Task Group on Data Support for Climate Change Assessments (TG-Data) recommends the implementation of FAIR (Findable, Accessible, Interoperable, Reusable) data principles and best practices for the documentation and curation of data that is assessed by the IPCC. The motivation to implement FAIR in the IPCC is to increase transparency and accessibility of the assessment, the implementation of the IPCC Error Protocol, and the long-term curation of the assessed digital information.

This document introduces the implementation of FAIR data principles into the IPCC process and reflects the four elements of FAIR to find the data, produce and reproduce figures and, finally, to document the provenance for reusability. It presents standard (basic) measures that are recommended for all digital data that is assessed, intermediate measures that achieve reproducibility of assessed digital information, for example through the use of collaborative platforms for figure development, and also full (advanced) measures to achieve reusability of digital products with complete provenance documentation.

Table of Contents

1. Introduction and IPCC context

The Task Group on Data Support for Climate Change Assessments (TG-Data) advises the IPCC on curation, traceability, stability, availability and transparency of data and scenarios related to the reports of the IPCC. This TG-Data guidance document provides guidance to IPCC authors, Bureau members and Technical Support Unit staff on the implementation of best practices for the documentation and curation of data that is assessed by the IPCC, including analyses used to produce figures and tables that are presented in the reports.

The FAIR data principles aim to facilitate the work of scientists and other users, to ensure that data is widely and reliably reused, and also to support data providers and help achieve the value of their data. It does require some effort to document the data but this ultimately allows users to approach the data from multiple perspectives and needs and leads to a growth of use and further work.

The following are some key motivations for the IPCC implementation of FAIR data principles:

- **Promoting scientific advances based on an open scientific process** by fostering the adoption of best practices of open-source science and increasing the accessibility of the assessment for the scientific community and users more broadly.

- **Implementation of the IPCC Error Protocol** as part of the principles governing IPCC work to document and correct errors that could have been avoided in the context of the information available at the time the report was written. Its use should be reserved for errors of fact or accuracy, not to reflect new knowledge or scientific information that became available only after the literature cut-off date for the report in question.

- **Transparency and reproducibility of the assessment process** through full documentation, including data, metadata and information on the data and software used to produce figures and tables.

- **The long-term curation of the digital information** that is assessed in an IPCC report.

- **Enhance the visibility of IPCC reports** and of the digital information assessed therein. This will increase use by other scientists and users more broadly, and will credit developers with their contribution to the IPCC process.

The implementation of FAIR data principles is prioritised for visual and digital information of the Summaries for Policymakers and Technical Summaries of IPCC reports and the Synthesis Report. It is recommended that they are implemented in the preparation of report chapters. It is also recommended that derivative products that are developed in collaboration with the IPCC or externally also adopt FAIR principles for full transparency and to extend these best practices to products that are developed using assessment products.

1.1 FAIR data principles

The scientific process is best served by implementing FAIR data principles, a set of guiding principles in order to make data findable, accessible, interoperable and reusable (Wilkinson et al., 2016). FAIR is a term coined by the Future of Research Communications and e-Scholarship (FORCE11) community. These principles provide guidance for scientific data management and stewardship and are relevant to all stakeholders of digital information. The complementary

Author Guidelines developed in the 'Enabling FAIR Data Project' provide a widely accepted framework in Earth and Space Sciences for the implementation of the FAIR data principles in scientific publications (Stall et al., 2018).

FAIR stands for: *Findable, Accessible, Interoperable, Reusable*

- **F**indable: Rich metadata describing the resource held in an easily searchable platform;
- **A**ccessible: Adoption of collaborative open-access platforms supporting the export of structured metadata. This requirement is mostly focusing on granting access to the information available through appropriate authorisation and well-defined protocols;
- **I**nteroperable: Adoption of standards (for data, metadata, and software) and defined vocabularies, common workflow and development protocols.
- **R**eusable: All products will have documentation and metadata that integrates different sources of information about authorship, formats, and lineage, for full provenance, traceability, and reproducibility (e.g. input data, metadata, diagnostics, tool version) and implement relevant standards for file formats.

The FAIR Digital Objects Forum is an active new forum that is discussing the application of FAIR principles in practice across different working groups on topics (data management, infrastructure, vocabularies, computational) related to the implementation of FAIR workflows.

1.2 Implementation by IPCC Working Groups

Practical information that is relevant for the implementation of FAIR data principles in the preparation of IPCC reports is provided as Working Group (WG) -specific Supplementary Material to this document. The preparation of this material has been a major achievement that requires considerable collaborative work and detailed planning between the TSUs and DDCs.

- Working Group I Supplementary Material
- Working Group II Supplementary Material
- Working Group III Supplementary Material

2. Data Archival

2.1 IPCC Data Distribution Centre

The IPCC Data Distribution Centre (DDC) has been established to facilitate the distribution of a consistent set of up-to-date scenarios of changes in climate and related environmental and socio-economic factors for use in climate impacts assessments.

The long-term archival and access to input data used in IPCC assessments is supported by the DDC and includes:

- 10+ years storage
- Professional metadata review process
- Long-term data curation, access and user support according to the Core Trust Seal and the emerging TRUST principles for keeping FAIR data FAIR over time (https://bitly.com/trustprinciples)
- Authorized users are allowed to incorporate their own data (provided that certain format and metadata requirements are fulfilled).

2.2 Data categories

There are three categories of data that should be archived: input (source) data, intermediate (post-processed) data, and final (plotted) data:

- **Input data** includes observations, model projections, reanalyses, etc and is the source material for the assessment reports. An example is data from the Coupled Model Intercomparison Project (CMIP) assessed as part of the IPCC assessment that is archived in the DDC Reference Data Archive automatically.
- **Intermediate data** is the outcome of data processing and analysis performed as part of the assessment.
- **Final data** is data directly plotted on a figure, included in a table or is a key numerical value (e.g. climate sensitivity) in the report.

The long-term archival of input data used in IPCC assessments is supported by the IPCC Data Distribution Centre (DDC) which fully implements FAIR data principles and TRUST principles for digital repositories.

2.2.1 Input and intermediate data

Input (source) datasets that are not CMIP and are not already archived in a trusted data repository, and intermediate (post-processed) data can be archived with DDC. A chapter contact is required for the archival of the datasets of a given chapter. The approach adopted during the AR6 has been to either work with chapter scientists or authors, with the TSU facilitating coordination with the DDC.

A guide for archival of input datasets, that is also applicable for intermediate datasets, is available here.

According to the widely accepted Author Guidelines of the 'Enabling FAIR Data Project' the input data should additionally be cited in the report by including data citations in the report reference list. As an example of this being implemented, for practical purposes in how the chapter reference lists were handled, the WGI report resorted to providing the dataset citations as Supplementary Material.

2.2.2 Final (plotted) data

The archival of final data for report figures can be managed in the first instance by the TSUs, uploading the data to the WG document management system. The next step for long term archival involves the data being transferred to the DDC. Documentation checks and tests to recreate the figures are needed before being archived.

2.3 Data formats

All categories of data should be curated using digital formats that ensure accessibility and facilitate reusability. Proprietary data formats should be avoided in favor of open standards that are free to use for everyone.

Further recommendations:

- Simple data formats are preferable to complex data formats.
- Data should be formatted so as to reduce the likelihood of errors by data producers and data users.
- As a whole, the IPCC should strive to use as few different data formats as possible.

2.4 Metadata

Data that is archived by the DDC needs to be accompanied by metadata clearly describing its content. A metadata form needs to be supplied when submitting datasets for archival. For final datasets in WGI, these details have been collected via an online form using the WGI TSU Figure Manager system and then provided to the DDC.

The DDC and WGII-WGIII TSUs have been using the DDC MetadataWorks Onboarding System for final data submission, reviewing, editing, metadata cataloging, and publication. A metadata schema is available here: https://github.com/MetadataWorks/Schemata

2.5 Publishing data

All datasets curated and archived in the DDC are assigned DOIs, enabling data citation by data users. More details are provided by Stockhause et al. (2019). DOI registration with DataCite integrates the data into scholarly publishing with existing services for data usage/credit tracking.

During archival datasets are annotated with metadata providing sufficient information for data reuse. Further information on data provenance and links to the IPCC AR6 as well as to other dataset types stored at a fellow DDC Partner are added where available. The data is disseminated under the CC BY 4.0 license, except for input datasets with already assigned deviating licenses by the data provider. DDC datasets are added to the joint DDC catalog.

2.6 Server-side Workspaces

A new resource that is exclusively provided to IPCC authors for the first time as part of the Sixth Assessment Report is server-side workspace on DDC high performance computers (HPC). DKRZ, CEDA and IFCA have provided IPCC authors access to undertake server-side analysis of global and regional climate datasets.

A host HPC can be selected depending on authors' main requirements. The following is a general guide on which HPC is most suited according to stored datasets:

Virtual Workspace Host	Main Datasets	Additional Information
CEDA JASMIN	CMIP, reanalysis and global observations	Access to Jasmin Analysis Platform tools and services https://github.com/cedadev/jasmin_scivm/wiki/Packages
DKRZ Mistral	CMIP, CORDEX, ERA reanalysis	Supports the implementation of ESMValTool
IFCA IPCC-Hub	Subsets of CMIP, CORDEX, plus several reanalyses and observations	Supports the implementation of Climate4R and the WGI Atlas

3. Code development

Authors are encouraged to implement FAIR in the development of code that can be made available publicly so that figures can be reproduced. Publishing executable notebooks that include an explanation of the analyses and calculations that have been undertaken provides the wider scientific community with an important resource. The use of open-source programming languages and environments is strongly encouraged for code underpinning IPCC figures, tables and data.

Executable notebooks, such as Jupyter notebooks, provide users with a reproducible workflow for chapter figures and to document the analysis in the dataset (e.g. using the global attribute 'history' for NetCDF/CF format). Documentation of the provenance record supports a reusable workflow, as described in Section 5.

3.1 GitHub repositories for chapters

The 'recipe' of each figure can be distributed through GitHub. There are various services offering to host git repositories, such as GitHub. See this brief document for taking your first steps with GitHub.

- The TSUs can host private GitHub repositories where chapters can develop scripts. See for example the AR6 WGI repository: https://github.com/IPCC-WG1
- Repositories should be distinctly named (e.g. AR6-WG1-Ch6), knowing that they will be one among many others.
- Documentation of what is included in the repository should be provided at a minimum with a readme file that includes references to supporting material (datasets, libraries, algorithms, etc), and also a file describing the software environment (libraries and their versions), for example a conda "environment.yml" file.
- Once the report is published, the repositories can be released with the final software versions used to generate the figures and tables.
- The same license type as used for data should be used for code.

3.2 Code version control

Git can be used to version control the code used to generate figures, tables and compute key numerical values. Authors that are unfamiliar with git can consult documentation here. Note that when code is made public, the entire history of the code also becomes public. The repository history can be cleaned, for example using git squash.

3.3 Executable Notebooks

Notebooks are open-source applications that allow you to share reproducible code in an executable environment, accessing datasets, accompanied by explanatory text of analysis steps and visualizations.

Binders are special repositories (Git, Zenodo, etc) that can generate notebooks alongside the required computational environment:

- Mybinder (https://mybinder.readthedocs.io/)
- Pangeo Binder (https://binder.pangeo.io/)

Authors are encouraged to generate executable notebooks to provide users with a reproducible workflow for chapter figures. Useful guidelines on the organisation and dissemination of reproducible notebooks are available in the literature (Rule, A., et al., 2018).

Notebooks are commonly generated using Jupyter. Project Jupyter is an open-source project, born out of the IPython Project supporting interactive data science and scientific computing across all programming languages. See: What are Jupyter Notebooks and why I would want to use them?

Polynote is a new open-source web-based programming notebook tool that was released in 2019, supported by the Netflix Open Source Software Center. It is focused on scientists and machine learning researchers and supports mixing multiple languages in one notebook.

As an illustrative example of the use of these technologies, the Atlas IPCC-WG1 repository allows online reproduction/ reuse of some key results of the Atlas chapter (such as figures with regional future projections) using executable Jupyter notebooks and building on Mybinder.

3.4 Collaborative software environments

Collaborative software environments have been established in the climate science community for the development of scripts and their notebooks to foster open-source science and progress in research and applications. These environments can be used by authors to develop reproducible code that underpins chapter figures.

The following are two examples of collaborative environments that have been used by authors of the AR6 WGI report.

- ESMValTool is a collaborative tool for the evaluation of CMIP Earth System Models and is integrated in the virtual workspace provided by DKRZ, which allows for server-side analysis of CMIP6 and other data. ESMValTool can also be implemented directly on your own computing system, with your own data. ESMValTool outputs some

metadata describing the analyses undertaken. See Righi et al. (2019) and the following for more information: Code, Documentation,Tutorial.

- Climate4R is an R framework that allows for transparent climate data access, post-processing (including bias adjustment and downscaling) and visualization. In addition, a METACLIP-based extension is available for data and post-processing provenance of the analyses performed (see Iturbide et al. 2019; Bedia et al. 2019 for details and worked examples). Climate4R underpins the AR6 WGI Atlas and is available for IPCC authors at the IFCA IPCC-Hub.

Pyam (https://pyam-iamc.readthedocs.io) is a software library related to Jupyter's Binder that provides a range of diagnostic tools and functions for analyzing and visualizing data from integrated assessment models.

3.5 Publishing code

In addition to being archived by the IPCC, the public release of code that underpins figures of the IPCC assessment is encouraged if documentation requirements are met. Published code with an assigned DOI can be cited within the report, in chapter Appendixes and as figure metadata in the IPCC archive, as well as in the broader literature base.

TG-Data can provide advice on best practices to support the publication of code, including its quality control. It is recommended that code is reviewed internally by authors and/or externally.

Once a repository is made public on GitHub, it can be automatically transferred to Zenodo that assigns a DOI to the repository for citation purposes. Note that, each time a release is created on Github, a new DOI is minted at Zenodo. Ideally, each figure/table should incorporate in its caption the code citation including its DOI pointing to the repository in which the data and code used to generate it are located.

Zenodo automatically extracts metadata about the repository from GitHub APIs. The code authors are determined from the repository's contributor statistics. A .zenodo.json file can be added to the root of your GitHub repository to explicitly define the metadata, including specifying the list of authors.

4. Reusability

For true open source science and for data to be meaningfully open, the scientific process needs to be **reusable**. A key element for reusability is to have the full final data provenance. The recommendations and resources described in the preceding sections are key elements to achieving accessible and reproducible scientific results. However, FAIR data principles go beyond sharing information and reproducibility.

Reproducibility refers to the ability of a user to duplicate the results of a prior study, while reusability is the ability to use existing data to address new problems (Bollen, K., et al., 2015). There are multiple uses of the term 'reproducibility' in the literature. Goodman et al., (2016) suggest using "method reproducibility" to refer to the ability to repeat an experiment or simulation and get the same result and "result reproducibility" to refer to the ability to get the same scientific result with an independently designed experiment or simulation.

Different implementations (e.g. software version, model or data version, computation environment) of the same data will not lead to robust (same) results. Reusability requires very careful reporting, or documentation, more than is

currently general practice in scientific publications. Provenance information and good data documentation (metadata) is important to both provide evidence and transparency about where the data has come from and all the assumptions or choices that may have affected the data values.

4.1 Provenance

A key element for reusability is to have the full final data provenance, or 'lineage'. This consists of the data origins, how it was derived, including technical metadata, and the ability to trace versions and errors. This documentation is typically obtained in a semi-automatic way by general-purpose workflow systems or by community tools that embed specific provenance recording mechanisms (e.g. A. Spinuso, et al. 2019). Workflow systems such as Pegasus, dispel4py and CWLTool, support provenance capturing at different levels of detail and have been adopted for computational Earth Science studies. Recently, a few climate software tools and services have been extended with specific provenance capabilities. For instance, this is available in tools such as the Atlas, which implements METACLIP (Bedia, J., et al. 2019) and ESMValtool.

The World Wide Web Consortium (W3C) recommends a conceptual framework for interoperable provenance representation, PROV. This framework supports the provision of machine understandable descriptions of people, institutions, entities, and activities involved in producing data. PROV supports application-specific extensions, for example ProvONE, an extension developed and used by the DataOne project to characterize workflow processes, unifying provenance representation across scientific workflow systems.

Tools exist to produce provenance information (e.g. ProvToolbox), as well as online services (e.g. Open Provenance) that can validate, visualize and store PROV compliant documents. More advanced systems provide support for query services on top of a dedicated database.

How to systematically capture provenance information in a way to be effectively exploited is a challenge, which is still acknowledged in recent literature (Herschel, M, et al., 2017, M. Atkinson, et al., 2017). New software offering provenance template catalogues (e.g. ProvTempCatalogue) has been implemented in order to facilitate discussion and dissemination of reusable provenance scenarios (L. Moreau, et al., 2018). Ultimately, these technical and conceptual efforts are motivated by the importance of the role of provenance data for the advancement of scientific production, at all scales, and its preservation is considered a fundamental issue to be addressed.

Meeting the reusability principle is a fundamental characteristic underpinning the development of interactive IPCC assessment products for full traceability and transparency. The AR6 WGI Interactive Atlas is an IPCC provenance-aware software.

4.2 AR6 WGI Interactive Atlas

An important and novel product of the AR6 is the Interactive Atlas that is part of the Working Group I (WGI) report. The Interactive Atlas allows for flexible spatial and temporal analysis of relevant WGI global and regional climate change information and allows exporting images and underlying data for the maps

The Interactive Atlas builds on and extends the assessments and methodological recommendations made in WGI chapters, providing a more comprehensive analysis of the information beyond the specific scope and limited space of the chapters. In particular, it provides global information in the form of interactive maps, for the climate change signal

of variables and indices (for several future periods, both time slices and warming levels) considering a number of alternative reference baselines. It also provides regional analysis for a number of predefined (reference and typological) regions in the form of plumes, climate stripes, tables and scatter plots (e.g. temperature versus precipitation) of aggregated spatial values. This allows for an in-depth comprehensive analysis (and intercomparison) of global and regional datasets.

The Atlas fully implements the FAIR data principles including reusability (Iturbide et al. 2022). The Atlas uses the METACLIP provenance framework, including a tailored extension for R, as a new component of the climate4R framework for climate data analysis (Bedia, J., et al., 2019). It is a language-independent framework developed for climate product provenance description with a user-friendly web interface.

5. Future progress

TG-Data continues to work in collaboration with the DDC and WGs for progress on data aspects of the assessment and the implementation of FAIR data principles. Recommendations based on the AR6 experience will be submitted to the IPCC at the end of the cycle to inform future assessment cycles.

This document is an outcome of the first steps taken by the IPCC during the AR6 to implement FAIR data principles. Some specific aspects require further development for improved implementation and to extend the adoption of this approach comprehensively as part of the IPCC. These include:

- the inclusion of relevant expertise amongst TSU and author teams suited to contribute to data and code aspects.
- the provision of clear guidance and the development of a streamlined process for authors so that they can contribute to this approach as an integral part of the report preparation process, from the start.
- a more comprehensive review of code as part of the preparation of IPCC reports.
- increasing the efficacy and efficiency in synchronizing the DDC catalogue with the publication of the Report.
- the citation of the datasets and code in the report. Citations can be inserted using the reference management software used for the preparation of the report (e.g. Mendeley). This information can be also be published on the report website alongside figures.
- continuing to strengthen the policy for data archival with the DDC and, for example, the use of Zenodo.
- a rigorous implementation of the IPCC Error Protocol for code and data archived by the DDC and the link to WG GitHub repositories. Errors with influence on final data can originate from the input data as well as the analysis. Thus corrected figures might require storage of corrected input data, scripts and final data.

What is key is the need to initiate work to prepare for the implementation of FAIR immediately at the start of the assessment process. This will require strong coordination, establishing a direct link between dedicated contacts in the TSU, the DDC and TG-Data from the very start of the assessment process.

The major achievements made by the IPCC so far during the AR6 have been referred to as 'culture changing' and 'revolutionary' by TG-Data representatives, given how valuable this work is for the scientific community and

stakeholders of the IPCC at large. These achievements and the lessons learned from the AR6 will support a strengthened and more comprehensive treatment of data aspects by the IPCC in future cycles.

References

Atkinson, A., S. Gesing, J. Montagnat, and I. Taylor, 2017: Scientific workflows: Past, present and future. Future Generation Computer Systems 75:216-227 DOI:10.1016/j.future.2017.05.041

J.Bedia, J., San-Martín, D, Iturbide. M., Herrera. S., Manzana, R., and Gutiérrez, J.M., 2019: The METACLIP semantic provenance framework for climate products. Environmental Modelling & Software Volume 119, 445-457, doi.org/10.1016/j.envsoft.2019.07.005

Bollen, K., Cacioppo, J. T., Kaplan, R. M., Krosnick, J. A., and Olds, J. L., 2015: Social, Behavioral, and Economic Sciences Perspectives on Robust and Reliable Science: Report of the Subcommittee on Replicability in Science Advisory Committee to the National Science Foundation Directorate for Social, Behavior.

Goodman, S. N., Fanelli, D., and Ioannidis, J. P. A.. Science Translational Medicine. 2016: What does research reproducibility mean? Science Translational Medicine, Vol 8, Issue 341, p. 341ps12, DOI: 10.1126/scitranslmed.aaf5027

Herschel, M., Diestelkämper, R., & Ben Lahmar, H. (2017). A survey on provenance: What for? What form? What from? VLDB Journal, 26(6), pp. 881–906. https://doi.org/10.1007/s00778-017-0486-1

Iturbide, M., Bedia, J., Herrera, S., Baño-Medina, J., Fernández, J., Frías, D., Manzanas R., San-Martín, D., Cimadevilla, E., Cofiño, A., S., and M.Gutiérrez, J. M., 2019: The R-based climate4R open framework for reproducible climate data access and post-processing. Environmental Modelling & Software, Vol. 111, pp. 42-54, https://doi.org/10.1016/j.envsoft.2018.09.009

Iturbide M., Fernández, J., Gutiérrez, J. M., Pirani, A., Huard, D., et al. 2022. Implementation of FAIR principles in the IPCC: The WGI AR6 Atlas repository. arXiv, doi: 10.48550/ARXIV.2204.14245.

Moreau, L., Batlajery, B. V., Huynh, T. D., Michaelides, D., & Packer, H. (2018). A Templating System to Generate Provenance. IEEE Transactions on Software Engineering, 44(2), pp. 103–121. https://doi.org/10.1109/TSE.2017.2659745

Righi, M., Andela, B., Eyring, V., Lauer, A., Predoi, V., Schlund, M., Vegas-Regidor, J., Bock, L., Brötz, B., de Mora, L., Diblen, F., Dreyer, L., Drost, N., Earnshaw, P., Hassler, B., Koldunov, N., Little, B., Loosveldt Tomas, S., and Zimmermann, K., 2019: ESMValTool v2.0 – Technical overview, Geosci. Model Dev. Discuss, https://doi.org/10.5194/gmd-13-1179-2020

Rule, A., Birmingham, A., Zuñiga, C., Altintas, I., Huang, S., Knight, R., Moshiri, N., Nguyen, M.H., Rosenthal, S.B., Pérez, F., & Rose, P.W, 2018: Ten Simple Rules for Reproducible Research in Jupyter Notebooks. ArXiv, abs/1810.08055

Spinuso, S., Atkinson, M. and F. Magnoni, F. Active Provenance for Data-Intensive Workflows 2019: Engaging Users and Developers. 2019 15th International Conference on eScience (eScience), pp. 560-569, doi: 10.1109/eScience.2019.00077

Stall, S. et al. (2018), Advancing FAIR data in Earth, space, and environmental science, Eos, 99, https://doi.org/10.1029/2018EO109301. Published on 05 November 2018.

Stockhause, M., Juckes, M., Chen, R., Moufouma Okia, W., Pirani, A., Waterfield, T., Xing, X. and Edmunds, R., 2019. Data Distribution Centre Support for the IPCC Sixth Assessment. Data Science Journal, 18(1), p.20. DOI: http://doi.org/10.5334/dsj-2019-020

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.