

REVIEW

Implementing FAIR data principles in the IPCC seventh assessment cycle: Lessons learned and future prospects

Martina Stockhause¹*, David Huard², Alaa Al Khourdajie^{3,4}, José M. Gutiérrez⁵, Michio Kawamiya⁶, Nana Ama Browne Klutse⁷, Volker Krey⁴, David Milward⁸, Andrew E. Okem⁹, Anna Pirani¹⁰, Lina E. Sitz⁵, Silvina A. Solman¹¹, Alessandro Spinuso¹², Xiaoshi Xing¹³



1 Deutsches Klimarechenzentrum (DKRZ), Hamburg, Germany, **2** Ouranos, Montréal, Canada, **3** Department of Chemical Engineering, Imperial College London and International Institute for Applied System Analysis, London, United Kingdom, **4** International Institute for Applied System Analysis (IIASA), Laxenburg, Austria, **5** Instituto de Física de Cantabria (CSIC-UC), Santander, Spain, **6** Japan Agency for Marine-Earth Science and Technology (JAMSTEC) and WPI-AIMEC, Tohoku University, Yokohama, Japan, **7** Department of Physics, University of Ghana, Accra, Ghana, **8** MetadataWorks Limited, Warwick, United Kingdom, **9** Stichting Deltares, Delft, The Netherlands and University of KwaZulu-Natal, Durban, South Africa, **10** Centro Euro-Mediterraneo sui Cambiamenti Climatici (CMCC), Venice, Italy, **11** Department of Atmospheric Sciences, School of Sciences, Instituto Franco-Argentino para el Estudio del Clima y sus Impactos (IRL 3352, IFAECI), University of Buenos Aires, Research Center for the Sea and the Atmosphere (CIMA), CONICET-UBA, Buenos Aires, Argentina, **12** Royal Netherlands Meteorological Institute (KNMI), De Bilt, The Netherlands, **13** Center for International Earth Science Information Network (CIESIN), Climate School, Columbia University in New York, New York, New York, United States of America

* These authors contributed equally to this work.

* stockhause@dkrz.de

OPEN ACCESS

Citation: Stockhause M, Huard D, Al Khourdajie A, Gutiérrez JM, Kawamiya M, Klutse NAB, et al. (2024) Implementing FAIR data principles in the IPCC seventh assessment cycle: Lessons learned and future prospects. PLOS Clim 3(12): e0000533. <https://doi.org/10.1371/journal.pclm.0000533>

Editor: Juan A. Añel, Universidade de Vigo, SPAIN

Published: December 9, 2024

Copyright: © 2024 Stockhause et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: AAK was supported by the European Union’s Horizon Europe research and innovation programme under grant agreement no. 101056306 (IAM COMPACT). The funder had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. The other authors received no specific funding for this work.

Competing interests: The authors have declared that no competing interests exist.

Abbreviations: AR, Assessment Report; AR6, Sixth Assessment Report; AR7, Seventh Assessment Report; CMIP, Coupled Model Intercomparison Project; CMIP6, Coupled Model Intercomparison Project Phase 6; CMIP7, Coupled Model

Abstract

Every five to seven years, the Intergovernmental Panel on Climate Change (IPCC) convenes the climate science community to assess the latest knowledge on climate change relevant to policy-makers. This generally takes the form of Assessment Reports (AR) covering the scientific basis of climate change, its impacts and future risks, and options for adaptation and mitigation. With each cycle, these reports have grown in scope, length, number of referenced papers, and underpinning datasets. During the sixth assessment cycle, a large-scale collective effort went into archiving digital products assessed and generated through the IPCC process. The main objectives driving this initiative are making IPCC’s work more transparent, improving the reproducibility and reusability of the assessment outcomes, better utilization of the services of the IPCC Data Distribution Centre (DDC), and, more generally, compliance with best practices in open science. This paper expands on the motivations for the curation and preservation of digital objects in the IPCC. It gives an overview of how FAIR (Findable, Accessible, Interoperable, Reusable) and open data principles have been implemented in practice and explores some of the successes and setbacks of the AR6 experience. It concludes with recommendations for consolidation and expansion of the approach for AR7. These include a tighter integration of digital curation activities in the IPCC timeline and workflows, better support of IPCC authors and contributors through early training and use of suitable software, improved standardization and harmonization of data and

Intercomparison Project Phase 7; CORDEX, Coordinated Regional Climate Downscaling Experiment; DDC, Data Distribution Centre; DOI, Digital Object Identifier; ES-DOC, Earth System Documentation; ESGF, Earth System Grid Federation; FAIR, Findable, Accessible, Interoperable, Reusable principles for data and software; IAMC, Integrated Assessment Modeling Consortium; ICONICS, The International Committee on New Integrated Climate Change Assessment Scenarios; IEA, International Energy Agency; IIASA, International Institute for Applied Systems Analysis; IPBES, Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services; IPCC, Intergovernmental Panel on Climate Change; NCA, National Climate Assessment; OGC, Open Geospatial Consortium; RDA, Research Data Alliance; SPM, Summary for Policymakers; SSP, Shared Socioeconomic Pathways; TG-Data, Task Group on Data Support for Climate Change Assessments; TRUST, Transparency, Responsibility, User focus, Sustainability, Technology principles for research repositories; TS, Technical Summary; TSU, Technical Support Unit; TSU, Technical Support Unit; WCRP, World Climate Research Programme; WIP, WGCM Infrastructure Panel; WG, Working Group; WGI, Working Group I; WGII, Working Group II; WGIII, Working Group III; WGCM, Working Group on Coupled Modelling.

software handling across Working Groups (WGs), and close collaboration with key external data providers and research organizations.

1. Introduction

The Intergovernmental Panel on Climate Change (IPCC (see List of Abbreviations)) is committed to producing reports on the physical basis, impacts, vulnerability and adaptation and mitigation options of climate change that achieve the highest scientific standards in assessing the scientific literature on a comprehensive, objective, open and transparent basis. Text and figures go through two rounds of reviews, carried out by governments and thousands of experts, typically generating hundreds of thousands of comments, each of which must be responded to by the author teams. Data underlying the report, however, has not benefited from the same level of scrutiny in previous assessments, as they are not always requested for review by experts. In the past, data underlying key IPCC messages and figures were not made publicly available, leading to criticisms of IPCC processes and enabled conditions for data-related controversies to emerge. In its 2010 review of IPCC procedures, the InterAcademy Council underlined the importance of ensuring that the main conclusions of the IPCC reports are underpinned by openly accessible databases [1]. This is in line with open science recommendations [2], arguing among others, for open research data, distribution of IPCC software and code under free and open licenses (FLOSS—Free/Libre and Open Source Software), and in particular, delivering data in compliance with the FAIR data principles (Findable, Accessible, Interoperable, Reusable) [3,4].

Making the data or derived products that underpin IPCC assessments publicly available requires cooperation with a wide array of collaborators. This is because IPCC does not conduct research; it relies on academic journals and external organizations to provide the latest research data for its assessments, for example, the Coupled Model Intercomparison Project (CMIP: <https://wcrp-cmip.org/>) [5,6] for global climate model projections, the Coordinated Regional Climate Downscaling Experiment (CORDEX: <https://cordex.org/>) [7,8] for regional climate projections, the Integrated Assessment Modeling Consortium (IAMC) [9] for emission scenarios, the International Energy Agency (IEA) for energy and emission data, EDGAR emissions data (Emissions Database for Global Atmospheric Research), etc.

The challenges of curating scientific data for the Assessment Reports (AR) are well known to the IPCC Data Distribution Centre (DDC), a network of data-centric institutions offering support to the IPCC since 1997 [10]. At the start of the AR6 cycle, the DDC reached out to the newly formed Technical Support Units (TSUs) of Working Groups (WGs) and laid out plans to improve data handling practices in AR6 [11]. In particular, a new focus was put on archiving data and scripts underpinning key figures and tables from the Summary for Policymakers of the assessment reports.

In 2018, the IPCC established the Task Group on Data Support for Climate Change Assessments (TG-Data), whose mandate [12] includes facilitating the availability and use of IPCC-related data. TG-Data adopted the objectives and perspectives brought by DDC and TSUs and started working to advise the coordination of a large-scale effort to archive and transparently document data underpinning the Sixth Assessment Report (AR6) involving data providers and hundreds of IPCC authors.

This paper lays out why and how FAIR principles were implemented in AR6, discusses the main successes and setbacks, and ends with recommendations for AR7. It touches on the tension between principles and pragmatism, the challenges that come with collaborations across

diverse scientific communities, and the peculiarities of work in the IPCC context. The experiences shared here can be relevant to the newly initiated seventh assessment cycle (AR7) of the IPCC, as well as to other international scientific assessments and coordinated efforts.

2. Implementation across the IPCC AR6

TG-Data issued guidance [13] on incorporating FAIR principles into the IPCC process for AR6. This was followed by concerted efforts from the Working Groups to adopt a uniform approach wherever feasible. Those guidelines distinguish three types of digital information in the IPCC context: input data (external source data for the report, e.g. CMIP6), intermediate data (data products processed and assessed by IPCC authors with a high reuse potential), and final data (data displayed in figures or tables, or interactively, e.g., through the IPCC WGI Interactive Atlas on the physical basis of climate change; <https://interactive-atlas.ipcc.ch/>).

The aim of enhanced transparency of IPCC outputs is guided by best practices in data management, including FAIR, and driven by the following motivations:

- 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.
- Adopting the IPCC Error Protocol on data, in accordance with Annex 3 to Appendix A of the Principles Governing IPCC Work [14].
- Improving the transparency and reproducibility of the assessment process through documentation, including data, metadata and software used to produce figures and tables.
- Preserving the key digital information underlying IPCC assessments over the long-term.
- Enhancing the visibility and accessibility of the digital information assessed therein.
- Giving credit to data creators and developers of analysis tools, for their contribution to the IPCC process, following good scientific practice.

As part of the effort to make IPCC data accessible and reusable, TG-Data has developed data and software licensing guidelines for the IPCC [15]. The guidelines aim to avoid undue restrictions on users, while, at the same time, protecting the legitimate interests of the owners of data. In the interests of preserving clarity and transparency while protecting the rights of data owners, the Creative Commons Attribution License (CC BY 4.0) is the recommended data license for the IPCC. Open-source licenses are recommended for software code that underlies IPCC data products.

The implementation of the FAIR Guidelines focused on content from the Summary for Policymakers (SPM), with significant efforts also made to curate data from reports' Technical Summaries (TS) and chapters. Curation of the TS and chapter contents was not mandatory, resulting in large variations in coverage (see Fig 1). The implementation of FAIR principles relied considerably on IPCC authors and chapter scientists (chapter scientists are typically junior scientists working to support given chapter teams), who are responsible for processing input data and generating intermediate and final data, together with metadata content. Authors and chapter scientists were requested to provide these outputs and related information in a standardized format for curation. These were reviewed by TSU staff and, after iterating with authors, this content was then provided to the DDC for further quality controls, Digital Object Identifier (DOI) creation enabling data citation, long-term preservation and catalog indexing. The DDC catalog (<https://ipcc-browser.ipcc-data.org>) makes these data, and those from previous reports, centrally discoverable.

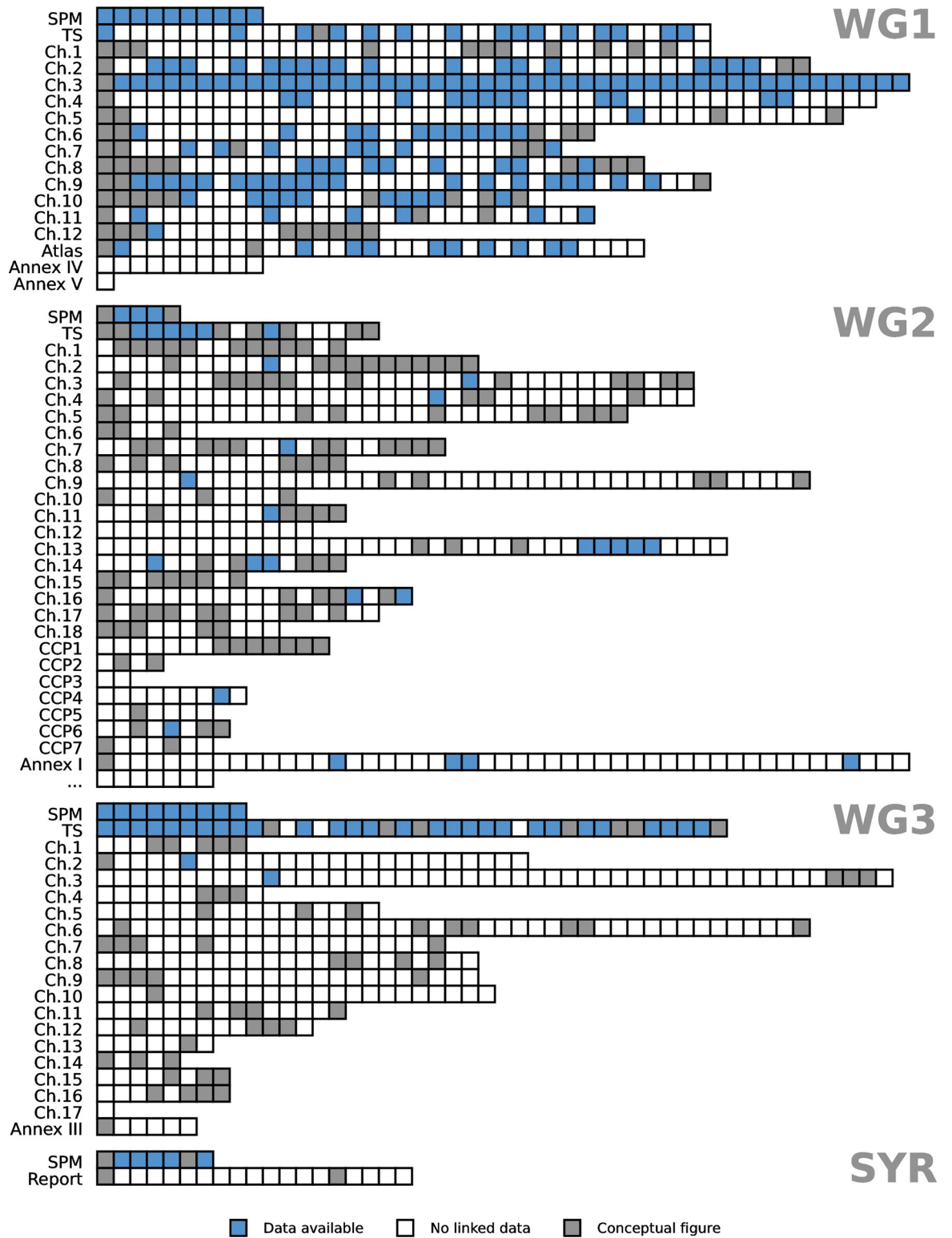


Fig 1. Archived final datasets for AR6.

<https://doi.org/10.1371/journal.pclm.0000533.g001>

Table 1. Overview of how the FAIR principles were put into practice in AR6.

Principle	Material		
	Input data	Intermediate/Final data	Software/Code
Findable	<ul style="list-style-type: none"> • DOI issued for datasets or data collections • Rich metadata • Indexed in the DDC catalog • indexed in DataCite and further portals harvesting DataCite • WGI input data included in the official WGI GitHub repository [16] 	<ul style="list-style-type: none"> • DOI issued for datasets or data collections • Rich metadata • Indexed in the DDC catalog • Indexed in DataCite and further portals harvesting DataCite • WGI final datasets linked from the IPCC figure page • WGI final datasets included in the official WGI GitHub repository (https://github.com/IPCC-WG1) 	<ul style="list-style-type: none"> • DOI issued by Zenodo • Indexed in DataCite and further portals harvesting DataCite • WGI software integrated in the official WGI GitHub repository (https://github.com/IPCC-WG1)
Accessible	<ul style="list-style-type: none"> • Catalog entries link to public data holdings at the DDC or an external provider 	<ul style="list-style-type: none"> • Catalog entries link to public data holdings at the DDC 	<ul style="list-style-type: none"> • Public code repositories
Interoperable	<ul style="list-style-type: none"> • Community data formats • Community metadata standards and vocabularies • Common DDC metadata schema for DDC catalog • Documentation and references accompany the data 	<ul style="list-style-type: none"> • Well known data formats (CSV, netCDF) • Common DDC metadata schema for the DDC catalog • Documentation and references accompany the data 	<ul style="list-style-type: none"> • Recommended use of open end-to-end frameworks: Climate4R and ESMValTool [17]
Reusable	<ul style="list-style-type: none"> • Permissible data usage license • Rich metadata • Metadata-data consistency checks • Provenance information and data-report interlinking provided as available 	<ul style="list-style-type: none"> • Permissible data usage license • Rich metadata • Provenance information provided as available 	<ul style="list-style-type: none"> • Permissible software license • Rich documentation of repository content and software usage including instructions on figure reproducibility

<https://doi.org/10.1371/journal.pclm.0000533.t001>

The following sections describe the experience across the assessment process in implementing the FAIR guidelines in AR6. A synthesis of how the FAIR principles were applied to data and software in AR6 is presented in [Table 1](#).

2.1 WGI—physical science of climate change

Working Group I (WGI) implemented the FAIR Guidelines for the full report. It took a more detailed approach for the provenance documentation compared to WGII and WGIII, including the development of the innovative WGI Interactive Atlas [18]. The implementation in the WGI began in early 2021 when AR6 was already underway. Initially, workflows were designed to spread responsibilities among authors, DDC, and the TSU and communicated to authors. The data curation process was not made mandatory for chapters but recommended, emphasizing the benefits it would have for IPCC products, as well as for authors, in terms of recognition and broadening the uptake of their work. Supplementary materials and Virtual Workspaces were provided for the authors [19]. For final data, a dedicated server, known as the Figure Manager, was custom-made by the TSU and then used to assist authors in meeting the minimum information requirements deemed necessary for proper data storage. This information, along with the data itself, was quality-checked by the TSU and then sent to the DDC for long-term preservation.

In the case of software code, although minimal information requirements were requested, the verification of information quality was directly entrusted to the provider. For intermediate data, the workflow was simplified and customized for specific cases. This involved direct exchange with the authors without the use of automation tools.

For datasets related to international model intercomparison projects, e.g. CMIP6 input datasets, detailed information on dataset usage was gathered from the authors and provided to

the DDC Partner. The DDC quality-checked the information, identified the datasets, gathered data and information (including usage of individual datasets in figures), and long-term archived the data. Authors coordinated directly with the DDC Partner for products related to the Atlas.

WGI made final data and code available for over 200 figures, which represents about a third of the figures in the main report. Additionally, all plotted data for the SPM and the data behind about 20 figures from the TS were provided. Furthermore, selected intermediate datasets resulting from the assessment were curated for some key indicators used in the report (e.g., projections of global surface temperature and sea level rise, where uncertainties were constrained through the assessment of multiple lines of evidence). Moreover, an innovative digital product was produced to support and expand the access to the datasets and assessment of regional information conducted in the WGI AR6, synthesizing key findings for climatic impact drivers: the Interactive Atlas (see subsection 2.4).

2.2 WGII—impacts, adaptation and vulnerability

Working Group II (WGII) assesses the vulnerability of socio-economic and natural systems to climate change, the negative and positive consequences of climate change and options for adapting to the observed and projected impacts of climate change. During AR6, WGII followed a careful curation process often involving personal communications with the authors of each data source, requesting access to data and/or permissions for their inclusion in the DDC catalog. Overall, the FAIR principles were implemented in the final data curation for the SPM and TS figures and tables, and the integrated database of observed and projected climate change impacts, as well as the risk assessment database across Working Groups. The critical input data used across WGs, such as INFORM Global Risk Index were also curated. In total, 42 input and final datasets covering most of the figures and tables from the WGII TS and SPM were successfully cataloged.

2.3 WGIII—mitigation of climate change

Working Group (WGIII) assesses the mitigation options to reduce and remove green-house gas emissions. It utilizes various datasets for AR6, including a synthetic historical emissions dataset (intermediate data) based on the EDGAR database [20] and scenarios data from the Scenarios Explorer hosted by the International Institute for Applied Systems Analysis (IIASA) on behalf of the IAMC [9]. However, due to licensing restrictions, input data, like that from the IEA, could not be curated by WGIII (it was made available to IPCC WGIII authors through a Memorandum of Understanding, granting access to IPCC authors and reviewers, and giving IPCC the right to share numbers derived from IEA datasets on which figures and tables were based). The process of data curation for WGIII's contribution to AR6 was adapted from the TG-Data guidelines [21], and includes providing a clean final dataset in a standard tabular format (covering data class, element, and type), structural metadata explaining the dataset (location in the report, authors' details, data source and underlying research papers, and any post-processing undertaken), and minting a DOI.

2.4 New interactive products supporting the IPCC assessment

The representation of findings (figures, tables, etc.) in a traditional report is limited due to report length restrictions and prevents users from exploring the digital information in more detail such as specific regions, scenarios or variables of interest. This motivated the development of new interactive products with flexible visualization options, supporting and extending the information provided in the report and its figures. The IPCC WGI Interactive Atlas

(<https://interactive-atlas.ipcc.ch/>) was developed as part of the WGI AR6. In addition, the IPCC NASA Sea Level Projection Tool (<https://sealevel.nasa.gov/ipcc-ar6-sea-level-projection-tool>) further extended the accessibility of sea level projections, which were developed in collaboration and endorsed by the IPCC. WGIII AR6 included a call for mitigation scenarios (<https://data.ece.iiasa.ac.at/ar6>) which was issued jointly by the IAMC and the IIASA, and supported by IPCC WGIII under a Collaboration Agreement [22,23].

The main challenge for the Interactive Atlas was to meet the requirements of the formal IPCC review process since interactive products cannot be reviewed in full detail due to the many combinations of choices. The implementation of FAIR principles was a key element for this, since it allowed for an indirect review, focusing on the underlying data (and rich end-to-end metadata) and software [18]. Besides the formal reviews, the Atlas GitHub repository was open for scrutiny, aiming to collect feedback and suggestions, resulting in notable enhancements.

Overall, the repository significantly advanced reproducibility by consistently leveraging existing technical FAIR enablers within the digital infrastructure providing climate data, e.g. for the production of provenance information attached to all final products visualized in the Atlas [24]. This contribution is instrumental in fostering a Community of Practice dedicated to effectively disseminating open and reliable climate change assessment tools and products, facilitating users' active engagement. For further details and implemented solutions, refer to [18].

3. AR6 reflections: Successes, experiences and challenges

In many respects, AR6 has been a step change in how IPCC data is handled. For the first time in IPCC history, final figure data were systematically curated and archived comprehensively, at least for the Summary for Policymakers of reports. Data and reports were interlinked, interactive components became part of the reports for the first time, and a joint data catalog across all reports was established to improve data discoverability. This created new and additional responsibilities, as well as burdens, for authors, TSUs, and DDC. DDC shared responsibilities were clarified in a MoU [25,26], which also established a transparent framework for expanding the DDC to include new members. This section highlights key ideas stemming from the AR6 experience [27].

3.1 Coordination and harmonization

FAIR principles focus on the user experience. In the IPCC context, however, great care has to be taken to minimize the additional burden for IPCC authors and TSU staff. This proved difficult in practice, especially because data and code collection protocols were developed whilst the AR6 assessment process was already underway. TSUs and DDC Partners had to quickly harmonize their practices to exchange data and metadata records and avoid duplicating work. A common schema (<https://github.com/MetadataWorks/Schemata>) describing metadata requirements was agreed across all WGs and DDC Partners, which allowed the creation of a joint DDC data catalog. The schema describes 40 different metadata elements, of which 22 are derived from DCAT (Data Catalog Vocabulary: <https://www.w3.org/TR/vocab-dcat-2/>), while most of the remainder are derived from Schema.org. The metadata elements are grouped into 8 categories as follows: Summary, Documentation, Coverage, Provenance, Accessibility, Enrichment and Linkage, Structural Metadata and Data Status. These groupings allow researchers to find and evaluate datasets quickly and easily before accessing and reusing the data directly. It is worth highlighting that such close collaboration between DDCs and TSUs was unprecedented in the history of the IPCC.

One noteworthy achievement stemming from these new collaborations is the organization of joint training activities like the WGI Training on Data and Software Development [28], and the delivery of international outreach activities, notably regional workshops on the WGI Interactive Atlas [29] hosted by different regional partners, and on the AR6 scenario database [30] co-hosted by IIASA and local governmental and academic organizations within each region. Multiple regional webinars were held with the support of local organizers, including live sessions held in different languages.

3.2 Licensing

Some of the data assessed by the IPCC are published with licensing restrictions. For example, the IEA shared datasets with IPCC authors, but these could not be redistributed publicly by the DDC. Another example relates to CMIP6 climate projections, initially published by climate modeling groups under a “sharealike” license (CC BY-SA 4.0). The “sharealike” clause requires data users to distribute derived products, such as figures, under the same license as the original data. Given that this was at odds with IPCC’s own license (CC BY-NC-ND 4.0), TG-Data and the DDC managers reached out to the World Climate Research Programme Working Group on Coupled Modelling (WGCM) and made recommendations that were adopted by the individual CMIP6 modeling centers to share model outputs under more open licenses (CC BY 4.0 and CC0).

Another licensing challenge arises from input datasets published in various scientific journals, each having its own license and conditions for sharing published data. Authors and TSU staff had to consider each license and, in some cases, make inquiries to journals to be granted permission to redistribute data under IPCC’s terms. This has proven to be extremely time consuming and solutions will need to be explored to streamline the process for AR7.

3.3 Data and software curation

The curation of the majority of data and scripts started after the SPMs were approved. This led to the overall challenge of decreasing author availability and TSU staff capacity during the archival and limited the ability to archive report data. This was especially challenging to finalize the CMIP6 input data archival.

The implemented flexible and inclusive curation approach adopted did not require authors to use specific software to conduct their analysis. This heterogeneity made reviews of software, final and intermediate datasets considerably more difficult and often required time consuming additional solicitation of input from authors to ensure consistency and uniformity to the documentation requirements. Another time-consuming task was to create a record of all the CMIP6 datasets used in the AR. The DDC focuses on datasets (variables and versions) used in the report to generate figures or tables. It turned out, however, to be very difficult to track which datasets and which versions were used by different authors over multiple chapters and hundreds of figures, especially at a time when author teams and TSUs were gradually winding down their activities. Despite considerable efforts, gaps remain in the preservation of those input datasets, and tools and training will have to be developed to ease the process in AR7.

In retrospect, it is remarkable that over 300 final datasets (Fig 1), a handful of intermediate datasets and ca. 65,000 CMIP6 input datasets and a couple of further input datasets have been archived and issued DOIs when needed. Around 130 codes have been created by WGI authors and included in WGI’s official GitHub repository (<https://github.com/IPCC-WG1>) by WGI TSU. Codes for individual figures used the GitHub-Zenodo link to become citable with a DOI. WGI’s GitHub repository additionally includes the scripts used to produce analyses and figures. Together, these efforts increase the transparency of the IPCC process and give credit to

scientists who work behind the scenes to produce the report's content. A visible outcome of these efforts are the links to the dataset that now accompany these report figures on the IPCC website.

3.4 Implementation of the IPCC error protocol

One of the motivations to implement FAIR principles is to support the implementation of the IPCC Error Protocol [14] in case errors are identified in the reports' contents, and datasets or code underpinning report figures or calculations used to underpin estimates in the assessment. The Error Protocol is part of the principles governing IPCC work. It serves to document and correct errors of fact or accuracy that could have been avoided in the context of the information available at the time the report was written. Once a potential error is identified, the Protocol foresees an exchange managed by TSUs with authors of the report to check, confirm, and provide a correction if needed. The outcome of this process is appropriately documented and is made available publicly on the IPCC website.

The Error Protocol has so far not been invoked for data or software produced during the AR6. Should this arise, TSU will need to collaborate with DDC managers to put in place a new process to curate and document any updated, corrected versions of datasets. Corrections made to code that is archived in TSU GitHub repositories will also need to be documented alongside new versions. The implications for the published report (corrected figures or estimates in tables or text) will need to be documented following the standard Error Protocol approach alongside the relevant information on corrections made to the underlying data and/or code.

4. AR7 recommendations and ongoing work

Following AR6, TG-Data prepared a list of recommendations (Table 2) [31] resulting from the collective experience of TSUs and DDCs. Recommendations include widening the scope of the data curation effort to all chapters, ensuring authors cite software code and data used in the preparation of the report, ensuring proper resourcing within TSUs and the DDC, embedding data and metadata preparation tasks in the AR timeline, offering tools and training material to authors, and embed data stewards within data-intensive chapters to support authors and liaise with TSUs. This scaling-up requires more comprehensive planning and coordination across authors, TSU and DDC.

In the relative lull between AR6 and AR7, many individuals involved in AR6 data curation have been busy working on improvements to data and software practices within and around the IPCC. The following samples are a few ongoing initiatives that will streamline the AR7 FAIR implementation. For AR7, the scope of the IPCC implementation of FAIR principles should be expanded to include all WG report figures underpinned by data and, ideally, the code and provenance for generating these figures.

4.1 External collaborations on input data

The World Climate Research Programme (WCRP) coordinates a wide range of climate research activities, including the CMIP and CORDEX model intercomparison projects. Their results have always been one of the most important sources of information and data for the IPCC assessments. For AR7, high-level consultations between representatives of the IPCC and WCRP are ongoing with the aim of aligning WCRP's research priorities with the scope of IPCC's AR7 in order to meet the tight AR7 timeline. Based on this consultation and further community engagement, WCRP CMIP7 (<https://wcrp-cmip.org/cmip7>) redesigned CMIP by identifying and grouping a set of experiments relevant to AR7 in the so-called "CMIP7 AR7

Table 2. Synthesis of recommendations.

Outcome	Recommendation
Start early	Provide guidance and training to authors for the organization of code and data at the beginning of the assessment period. Empathy with the challenge of finding a balance between the practices of authors and the requirements for long term archival.
Flexibility	Show empathy with authors to find a reasonable balance between the practices and time constraints of authors and the requirements for long term archival.
Enhanced integration	Integrate the FAIR guidelines into the IPCC assessment process with clear data policies, enhanced internal collaboration, effective communication and outreach, and improved tool support.
Fully linked data provenance	Apply FAIR data practices to connect input data, through intermediate data to the final data plotted in the figures.
More author engagement	Inform authors about mandatory requirements and provide clear timelines.
More engagement with external partners	Engage and collaborate with external partners providing input data or technical solutions in workflow improvements, FAIR guidelines implementation and training and outreach activities
Reduce time pressures	Provide sufficient resources at the end of the cycle, when the substantial part of archival and curation work will be completed, including the inter-connection of data and code with the report itself.
License compatibility	Encourage authors to pay attention to the licenses that are used to publish their data and strongly encourage the use of open and accessible licenses.
Continuity between assessments	Improve data workflows between assessment cycles and provide hands-on assistance to authors.
Additional positions in TSU	Hire dedicated TSU staff supporting authors and working with the IPCC DDC for the duration of the assessment cycle.

<https://doi.org/10.1371/journal.pclm.0000533.t002>

Fast track". This subset of CMIP7 experiments must conform to a higher level of standardization and adhere to the AR7 schedule.

Requirements for data and information from the IPCC data guidelines are brought into the CMIP infrastructure development and standardization process by TG-Data representatives involved directly in the WGCM Infrastructure Panel (WIP), CMIP Panel and relevant CMIP7 Task Teams. CMIP is strengthening errata, data citation and model documentation services (ES-DOC: <https://es-doc.org/>) with stricter requirements for minimum viable information that accompany model submissions. This work is coordinated by the WIP and implemented in collaboration with the Earth System Grid Federation (ESGF: <https://esgf.llnl.gov/>) and further CMIP data infrastructure providers. ESGF itself is moving to a new architecture meant to improve data discoverability, provenance tracking and the communication/integration of further services like the errata service, among others. Plans also include a messaging service that can be utilized to inform scientists of data updates and errata. We highlight the need for a federated data citation service to ensure homogeneous data citation approaches across ESGF data nodes.

For WGII and WGIII, the International Committee on New Integrated Climate Change Assessment Scenarios (ICONICS: <https://depts.washington.edu/iconics/>) and the IAMC (<https://www.iamconsortium.org/>) coordinate important activities related to scenario development, in particular when it comes to the Shared Socio-economic Pathways (SSPs) and curation of related data products (e.g., SSP Extensions Explorer: <https://ssp-extensions.apps.ece.iiasa.ac.at/>, SSP Update: <https://data.ece.iiasa.ac.at/ssp/>). Up to now, scenario collection and vetting were conducted under the tight deadlines set by the assessment timeline. In the IPCC workshop on scenarios [32], it was generally agreed that the IAMC implements an ongoing community-coordinated process, irrespective of IPCC timelines. This would allow the IPCC

authors to take a snapshot of the data whenever needed, and hence optimize the time to focus on the scientific assessment. In addition, a separate call for collecting national scenarios [33] following a harmonized scenario protocol has recently been opened.

4.2 Development of tools supporting the IPCC process

The IPCC has initiated the implementation of a centralized Figure Manager; the internal server enables the consistent management of the figures included in its reports. Its functionalities are inspired by similar software developed by AR6 WGI TSU that in turn built on a resource developed for the US National Climate Assessment (NCA). AR7 WGIII has decided to use the US NCA Figure Manager. The Figure Managers aim to provide TSU with an overview of figures status, collect metadata and figure datasets, facilitate communication with authors, record versions and link to external assets (input data and software) that have contributed to the generation of a particular figure version. All this information characterizes the provenance of a particular figure.

We foresee that collecting standardized provenance records for the figures' generation through the Figure Manager together with a set of internal analysis tools would enable the systematic citation of input data and the traceability of the reports' figures. Furthermore, the information contained in the provenance records is essential for the identification of input datasets for long-term archival in the DDC and for establishing relations between report (full report, chapter, figure), all data types (input, intermediate, final) and applied software, making these relations accessible to users and to infrastructures such as Knowledge Graphs.

Ideally, collecting figure data and metadata should require low levels of effort from authors. In AR6, some chapter writing teams and the WGI Interactive Atlas [24] have adopted analysis tools that automatically track provenance information. In cases where those tools are not applicable, clear instructions will have to be provided for authors to record the relevant information and submit it to the figure manager database.

4.3 External collaboration on data management practices

There are multiple initiatives where the community actively encourages and works towards the transparent sharing of climate-related data, fostering collaboration among scientists, data providers and, eventually, policymakers. In Europe, the RODEO project is creating services for the retrieval of meteorological and observational climate datasets identified as High Value Datasets (HVD). The distribution process aligns with the FAIR principles, trying to homogenize and accommodate metadata models for discovery and detailed description of the informative content, adhering to an open license. IPCC participation in networking initiatives aiming at raising global awareness of the importance of achieving FAIRness and equity of the information provided by future climate services will be fundamental to steer priorities and support. It will also impact the creation of traceable climate scenarios at a regional scale by national meteorological services, as the IPCC depends on data produced and managed by international organizations. This highlights the significance of these collaborations aimed at concrete implementations of a FAIR climate data infrastructure.

Ultimately, an effective strategy towards the implementation of FAIR can benefit from the emerging FAIR evaluation frameworks, which can be used for the compilation of guidelines, as well as their incremental implementation. This allows advisors and implementers to refer to a common and detailed model which disambiguates terminology and fosters an incremental application of the principles. For instance, the Research Data Alliance published the FAIR Data Maturity Model that organizations can use to assess their progress in FAIR coverage, thanks to a collection of detailed indicators. This could be used in combination with the FAIR

implementation profile, delivered by GOFAIR organization, which aims at gathering and disseminating FAIR implementation choices made by a Community of Practice. These are made openly available to others for reuse: <https://www.go-fair.org/how-to-go-fair/fair-implementation-profile/>.

Further collaborations with groups such as the Research Data Alliance (RDA) and the Open Geospatial Consortium (OGC) on technical solutions for open issues related to the implementation of the AR7 recommendations are ongoing. For example, TG-Data members are engaged in the RDA Complex Citations WG, aiming to provide recommendations for the citation and credit assignment of many datasets, partly subsets, stored across multiple repositories using a single complex citation object. The documentation of the figure creation process in the figure captions has been one of the three guiding use cases for WG [34].

Other RDA groups with case statements relevant for the IPCC guidelines include the emerging 'RDA Coordinating Earth, Space, and Environmental Science Data Preservation and Scholarly Publication Processes Events WG', the 'RDA & ReSA: Policies in Research Organisations for Research Software WG' or the FAIR Digital Object Fabric IG. OGC's Climate Resilience Domain WG provides an open forum for the discussion and presentation of OGC Standard usage in the context of cross-sector climate actions, where TG-Data can contribute its use case to the discussion.

Informal exchanges with other assessment bodies such as the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services (IPBES) and national climate assessments have been initiated to share tools, practices and jointly improve data and code handling in assessments.

4.4 Funding and sustainability

The DDC Partners provided in-house data curation expertise and data storage infrastructures to the IPCC as a national or institutional in-kind contribution up to 2024. With the introduction of the IPCC data and software guidelines, the workload for DDC Partners increased substantially. At the same time, the in-kind contributions of some members were undergoing substantial reductions, jeopardizing the future operation of key DDC services. Given the growing importance of data curation activities within the IPCC, TG-Data recommended the IPCC Panel establish a stable and predictable funding mechanism for DDC activities. At the 60th Session of the IPCC [35], a three year budget conditional on the approval of the TG-Data programmer of work for the Seventh cycle was adopted, paving the way for the planning of DDC activities covering the full AR cycle. These funds support DDC activities and data for the new cycle. The long-term preservation of the data and the maintenance of the digital products from previous cycles still rely on in-kind contributions or alternative funding sources of the DDC Partners.

4.5 Ideas for future work

Beyond the immediate operational requirements for AR7, TG-Data is also thinking more broadly about future developments. One avenue could be to generate a machine-readable version of IPCC Assessment Reports, similar to what has been done by the IPBES. IPBES has made three reports available in the Linked Open Data format (https://ipbes-data.github.io/IPBES_LOD/) [36], with the objective of facilitating its ingestion into databases and machine learning training datasets. We are also investigating the use of complex citations to give credit to all data providers contributing to the figures included in IPCC reports and in support of provenance documentation of the figures. Many figures aggregate results from dozens of different modeling groups, and at the moment they cannot be credited directly.

5. Conclusions

The IPCC AR6 successfully embraced open science principles, particularly FAIR data practices. This enhanced the transparency of the assessment by making data underlying key findings accessible and traceable. Systematic archiving of data and clear author guidelines have been useful in developing streamlined workflows. Further key elements of the IPCC data and software guidelines were the cross-referencing between data, code and report, and data and software long-term preservation by the DDC according to best practices for repositories, such as outlined in the TRUST (Transparency, Responsibility, User focus, Sustainability, Technology) principles [37]. A notable highlight was the creation of the Interactive Atlas for WGI, an innovative digital product that expanded access and allowed for the exploration of the assessment data.

However, challenges arose when implementing the FAIR principles, given the diverse methodologies of different scientific communities. Also, in some cases, navigating complex licensing restrictions on data and proprietary software tools proved difficult. Additionally, the increased workload for authors, TSUs, and the DDCs required careful management, especially given the requirement to publish the data of the SPM figures and tables on the release date of the reports. The unsustainable funding of the DDC, with some partners losing their funding during AR6, was a challenge. It put into question the long-term preservation of data and its continuous FAIRness as an essential part of the aimed transparency of IPCC outcomes.

The IPCC's AR6 experiences offer a number of valuable lessons for AR7. Harmonizing the approaches of the WGs through streamlined workflows, early implementation of a FAIR Data policy, and use of tools support to manage comprehensive data coverage are crucial for the traceability of key outcomes. Applying WGI's FAIR practices to WGs II and III, including intermediate data and code curation, would offer even greater transparency and enable the reproducibility of results. Integrating FAIR principles from the very beginning of the assessment cycle will be key, along with engaging authors early with robust training and support on the data and software request. Close collaboration with external partners is essential to preemptively resolve licensing and workflow issues but also to incorporate the latest standards and best practices in data management, including the CARE principles for indigenous data governance [38]. Finally, securing stable funding for the IPCC DDC will guarantee the curation of valuable assessment data and the preservation of the enhanced transparency of the IPCC assessments, and enable work on the envisioned machine-readable version of the IPCC Assessment Report.

Acknowledgments

The IPCC Data Distribution Centre is jointly operated based on a MoU by the Deutsches Klimarechenzentrum (DKRZ) in Germany responsible for the CMIP data preservation since SAR (1997) and intermediate datasets in AR6, the Center for International Earth Science Information (CIESIN) in the USA responsible for socio-economic data and scenarios preservation and integrated databases across Working Groups and Synthesis Report, the Spanish Research Council (CSIC) in Spain responsible for the WGI AR6 Interactive Atlas, MetadataWorks in the UK responsible for figure data preservation of AR6 WGII and III and the joint catalog, and the Centre for Environmental Data Analysis (CEDA) in the UK responsible for the figure data preservation of AR6 WGI.

The authors would like to thank Robin Matthews, Diego Cammarano, Andrés Alegría, Elvira Poloczanska, Charlotte Pascoe, Charlotte Reynolds, Jyoti Rogers, Courtney Irwin, Tim Waterfield, and Martin Juckes for their contributions to the work presented in this paper.

Author Contributions

Supervision: Martina Stockhause.

Writing – original draft: Martina Stockhause, David Huard, Alaa Al Khourdajie, José M. Gutiérrez, Michio Kawamiya, Volker Krey, David Milward, Anna Pirani, Lina E. Sitz, Alessandro Spinuso, Xiaoshi Xing.

Writing – review & editing: Martina Stockhause, David Huard, Alaa Al Khourdajie, José M. Gutiérrez, Michio Kawamiya, Nana Ama Browne Klutse, Volker Krey, David Milward, Andrew E. Okem, Anna Pirani, Lina E. Sitz, Silvina A. Solman, Alessandro Spinuso.

References

1. Committee to Review the IPCC. Climate change assessments: Review of the processes and procedures of the IPCC InterAcademy Council. 2010 https://www.ipcc.ch/site/assets/uploads/2018/03/doc07_p32_report_IAC.pdf.
2. UNESCO. UNESCO Recommendation on Open Science. 2021 <https://doi.org/10.54677/MNMH8546>.
3. Wilkinson M, Dumontier M, Aalbersberg I, Appleton G, Axton M, Baak A, 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> PMID: 26978244
4. Barker M, Chue Hong NP, Katz DS, Lamprecht A-L, Martinez-Ortiz C, Psomopoulos F, et al. Introducing the FAIR Principles for research software. *Sci Data* 9, 622 2022. <https://doi.org/10.1038/s41597-022-01710-x> PMID: 36241754
5. Meehl GA. The Role of the IPCC in Climate Science. *Oxford Research Encyclopedias*. 2023 <https://doi.org/10.1093/acrefore/9780190228620.013.933>.
6. Eyring V, Bony S, Meehl GA, Senior CA, Stevens B, Stouffer RJ, et al. Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization, *Geosci. Model Dev.*, 9, 1937–1958, 2016 <https://doi.org/10.5194/gmd-9-1937-2016>.
7. Gutowski WJ Jr, Giorgi F, Timbal B, Frigon A, Jacob D, Kang H-S, et al. WCRP COordinated Regional Downscaling EXperiment (CORDEX): a diagnostic MIP for CMIP6, *Geosci. Model Dev.*, 9, 4087–4095, 2016 <https://doi.org/10.5194/gmd-9-4087-2016>.
8. Giorgi F, Coppola E, Teichmann C, Jacob D. Editorial for the CORDEX-CORE Experiment I Special Issue. *Clim Dyn* 57, 1265–1268 2021 <https://doi.org/10.1007/s00382-021-05902-w>.
9. Byers E, Krey V, Kriegler E, Riahi K, Schaeffer R, Kikstra J, et al. AR6 Scenarios Database [Data set]. In *Climate Change 2022: Mitigation of Climate Change (1.1)*. Intergovernmental Panel on Climate Change. 2022 <https://doi.org/10.5281/zenodo.7197970>.
10. Stockhause M and Lautenschlager M. Twenty-five years of the IPCC Data Distribution Centre at the DKRZ and the Reference Data Archive for CMIP data, *Geosci. Model Dev.*, 15, 6047–6058, 2022 <https://doi.org/10.5194/gmd-15-6047-2022>.
11. Stockhause M, Juckes M, Chen R, Moufouma Okia W, Pirani A, Waterfield T, et al. Data Distribution Centre Support for the IPCC Sixth Assessment. *Data Science Journal*, 18(1), p.20.2019 <https://doi.org/10.5334/dsj-2019-020>.
12. IPCC. Terms of Reference and Mandate of the IPCC Task Group on Data Support for Climate Change Assessments (TG-Data). 2020 https://www.ipcc.ch/site/assets/uploads/2020/10/TG-Data_TORs.pdf.
13. Pirani A, Alegria A, Al Khourdajie A, Gunawan W, Gutiérrez JM, Holsman K, et al. The implementation of FAIR data principles in the IPCC AR6 assessment process. *Zenodo*. 2022 <https://doi.org/10.5281/zenodo.6504469>.
14. IPCC. Appendix A: Procedures for the Preparation, Review, Acceptance, Adoption, Approval and Publication of IPCC Reports. In *Principles Governing IPCC Work*. 2013 <https://www.ipcc.ch/site/assets/uploads/2018/09/ipcc-principles-appendix-a-final.pdf>.
15. Huard D, Pirani A, Chen R, Gutiérrez JM, Juckes M, Krey V, et al. IPCC Data and Code Licensing Guidelines (1.1). *Zenodo*. 2022 <https://doi.org/10.5281/zenodo.7431834>.
16. Stockhause M, Wachsmann F, Krüss B. DDC AR6 Reference Data Archival of CMIP6 input datasets (1.2). *Zenodo*. 2023 <https://doi.org/10.5281/zenodo.8301741>.
17. Eyring V, Bock L, Lauer A, Righi M, Schlund M, Andela B, et al. Earth System Model Evaluation Tool (ESMValTool) v2.0 –an extended set of large-scale diagnostics for quasi-operational and

- comprehensive evaluation of Earth system models in CMIP, *Geosci. Model Dev.*, 13, 3383–3438, 2020 <https://doi.org/10.5194/gmd-13-3383-2020>.
18. Iturbide M, Fernández J, Gutiérrez JM, Pirani A, Huard D, Al Khourdajie A et al. Implementation of FAIR principles in the IPCC: the WGI AR6 Atlas repository. *Sci Data* 9, 629 2022 <https://doi.org/10.1038/s41597-022-01739-y> PMID: 36243817
 19. Pirani A, Matthews R, Sitz L. AR6 Working Group I FAIR Supplementary Material (Version 1). Zenodo. 2022 <https://doi.org/10.5281/zenodo.6451137>.
 20. Minx JC, Lamb WF, Andrew RM, Canadell JG, Crippa M, Döbbling N, et al. A comprehensive and synthetic dataset for global, regional and national greenhouse gas emissions by sector 1970–2018 with an extension to 2019 [Data set]. Zenodo. 2022 <https://doi.org/10.5281/zenodo.6483002>.
 21. Al Khourdajie A. AR6 Working Group III FAIR Supplementary Material. 2022 <https://doi.org/10.5281/zenodo.6490164>.
 22. Peters GP, Al Khourdajie A, Sognaes I, Sanderson BM, AR6 scenarios database: an assessment of current practices and future recommendations. *npj Clim. Action* 2, 31 2023 <https://doi.org/10.1038/s44168-023-00050-9>
 23. Skea J, Shukla P, Weyant J, Kabar P. Collaboration Agreement between IPCC Working Group III, IAMC, and IIASA. https://data.ene.iiasa.ac.at/ar6/static/files/collaboration_agreement_ipccwgiii_iamc_iiasa.pdf.
 24. Bedia J, San-Martín D, Iturbide M, Herrera S, Manzanar R, Gutiérrez JM. The METACLIP semantic provenance framework for climate products, *Environmental Modelling & Software*, Volume 119, 2019, Pages 445–457, ISSN 1364-8152, <https://doi.org/10.1016/j.envsoft.2019.07.005>.
 25. Juckes M, Stockhause M, Chen B, Gutierrez Llorente JM. Memorandum of Understanding for Operation of the IPCC Data Distribution Centre. Zenodo. 2021 <https://doi.org/10.5281/zenodo.4889908>.
 26. Xing X, Stockhause M, Gutiérrez Llorente JM, Irwin C. Memorandum of Understanding (MoU) for Operation of the IPCC Data Distribution Centre. Zenodo. 2021 <https://doi.org/10.5281/zenodo.5914483>.
 27. Pirani A, Cammarano D, Fisher E, Krüss B, Matthews R, Pascoe C, et al. Experience in the Implementation of FAIR Data Principles in the WGI AR6 Assessment (Version 1). Zenodo. 2022 <https://doi.org/10.5281/zenodo.6992173>.
 28. IPCC. WGI Training on Data and Software Development. 2019 <https://www.ipcc.ch/event/wgi-training-on-data-and-software-development/>.
 29. IPCC. Interactive Atlas Regional Webinars. 2022 <https://www.ipcc.ch/event/interactive-atlas-regional-webinars/>.
 30. IPCC. IPCC TG-Data Scenario Database and Scenario Explorer Webinars. 2023 <https://www.ipcc.ch/event/ipcc-tg-data-scenario-database-and-scenario-explorer-webinars/>.
 31. TG-Data. TG-Data Recommendations for AR7. Zenodo. 2023 <https://zenodo.org/records/10059282>.
 32. IPCC. Workshop Report of the Intergovernmental Panel on Climate Change Workshop on the Use of Scenarios in the Sixth Assessment Report and Subsequent Assessments [Masson-Delmotte V, Pörtner H-O, Roberts DC, Shukla PR, Skea J, Zhai P, et al. (eds.)]. Working Group III Technical Support Unit, Imperial College London, United Kingdom, 67 pp. 2023 https://www.ipcc.ch/site/assets/uploads/2023/07/IPCC_2023_Workshop_Report_Scenarios.pdf.
 33. IAMC. IAMC SWG on National Scenarios: Open call for country-level scenarios. 2023 <https://www.iamcconsortium.org/iamc-announcements/news-iamc-announcement/iamc-swg-on-national-scenarios-open-call-for-country-level-scenarios/>.
 34. Stockhause M. RDA WG Complex Citation: Use case IPCC. Zenodo. 2023, February 28 <https://doi.org/10.5281/zenodo.7684261>.
 35. IPCC. Sixtieth Session of the IPCC (IPCC-60), Istanbul, Türkiye, 16–19 January 2024. 2024 <https://www.ipcc.ch/meeting-doc/ipcc-60/>.
 36. Dadvar M and Niamir A. IPBES Ontology (Version 04) [Data set]. Zenodo. 2024 <https://doi.org/10.5281/zenodo.13863920>.
 37. Lin D, Crabtree J, Dillo I, Downs RR, Edmunds R, Giaretta D, et al. The TRUST Principles for digital repositories. *Sci Data* 7, 144 2020 <https://doi.org/10.1038/s41597-020-0486-7> PMID: 32409645
 38. Carroll SR, Garba I, Figueroa-Rodríguez OL, Holbrook J, Lovett R, Materchera S, et al. The CARE Principles for Indigenous Data Governance. *Data Science Journal*, 19(1), p.43.2020 <https://doi.org/10.5334/dsj-2020-043>.