HELMHOLTZ
Open Science

62nd ONLINE SEMINAR

The IPCC FAIR Guidelines from the perspective of the IPCC Data Distribution Centre (DDC)

Thursday, December 9, 2021

Slides are published under

https://gfzpublic.gfz-potsdam.de/pubman/item/item_5008903

The IPCC FAIR Guidelines from the perspective of the IPCC Data Distribution Centre (DDC)

Helmholtz Open Science Online-Seminare, 2021-12-09

Martina Stockhause (DKRZ/IPCC DDC)

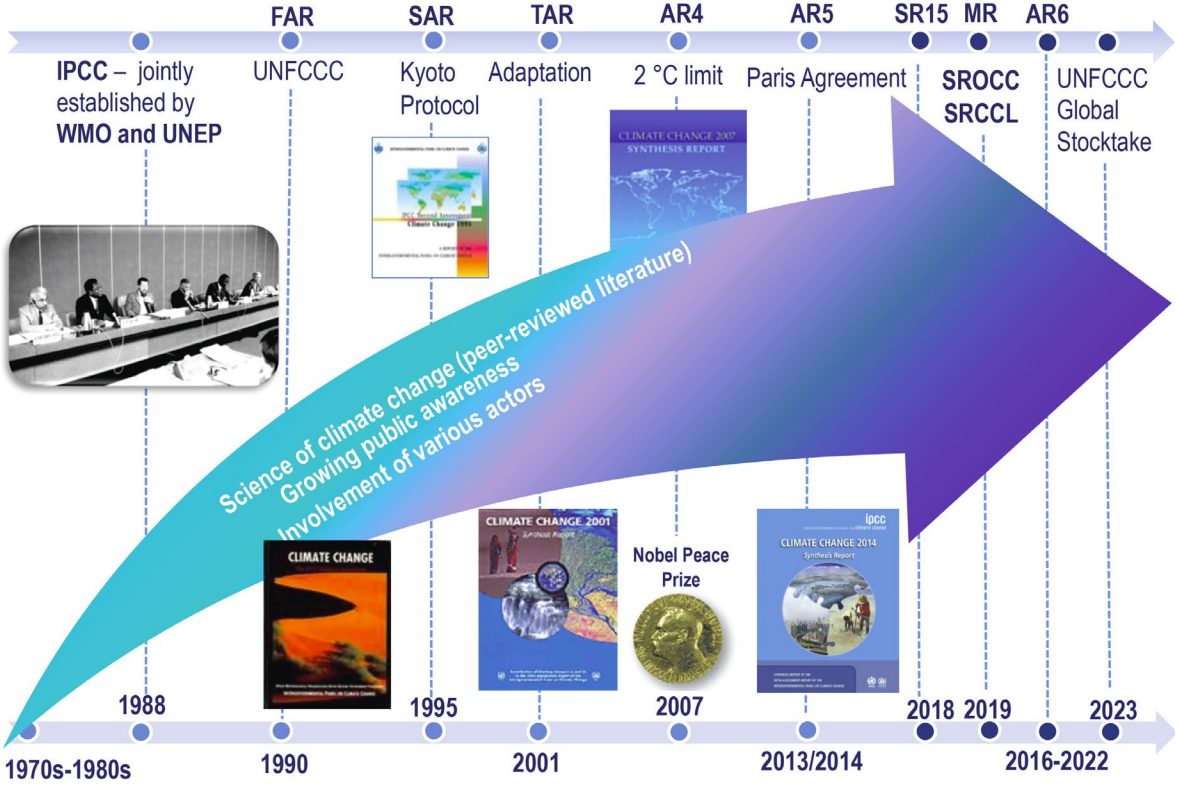
Contributors: Anna Pirani, Martin Juckes, José Gutierrez, Robert Matthews, Charlotte Pascoe, Robert Chen, Xiaoshi Xing, David Huard and many others

Intergovernmental Panel on Climate Change (IPCC)

IPCC reports are neutral, policy-relevant but not policy-prescriptive.
The assessment reports are a key input into the international negotiations to tackle climate change.

- No research conducted
- Scientific consensus
- Voluntary contributions

Established by the UN Environment Programme (UNEP) and the World Meteorological Organization (WMO) in 1988,
the IPCC has 195 Member countries



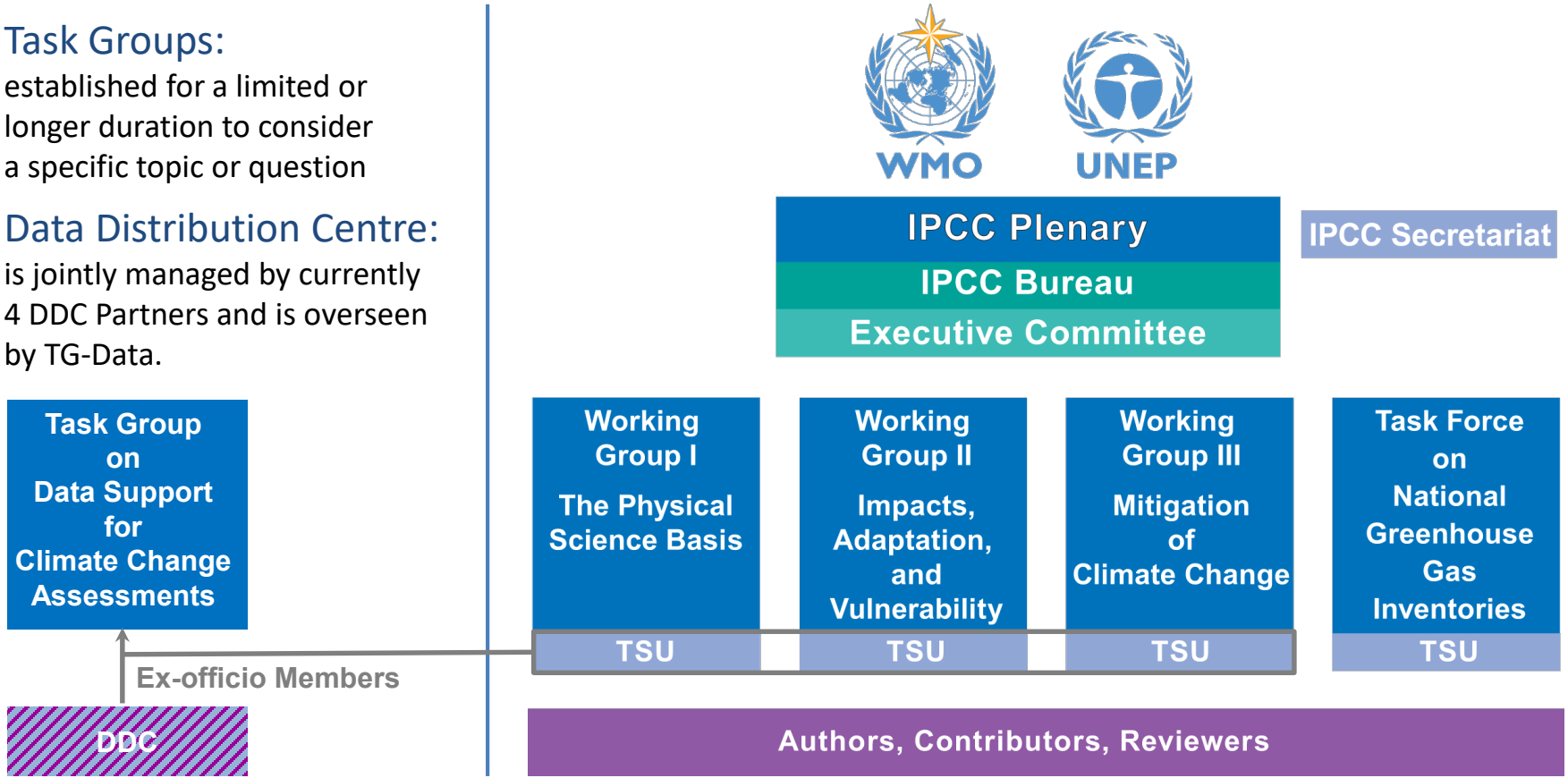
Intergovernmental Panel on Climate Change

Task Groups:

established for a limited or longer duration to consider a specific topic or question

Data Distribution Centre:

is jointly managed by currently 4 DDC Partners and is overseen by TG-Data.



IPCC DDC (Data Distribution Centre) – ipcc-data.org

established in 1997 and jointly managed by:

- Metadataworks:
  Climatologies, final data, web pages, joint catalog
- German Climate Computing Center (DKRZ):
  Reference Data Archive for climate model output
- Center for International Earth Science Information Network (CIESIN) at Columbia University:
  Socioeconomic data and scenarios
- University of Cantabria (CSIC):
  Interactive Atlas

→ Established DDC Partners are Regular Members of the World Data System (WDS)

Coupled Model Intercomparison Project (CMIP6)

provides recent research results on climate future projections

- **Research - CMIP6** is jointly coordinated by the CMIP Panel and the WGCM Infrastructure Panel; its infrastructure includes:
  - Earth System Grid Federation (ESGF) as data infrastructure
  - Data Citation Service providing DataCite data DOIs
  - Long-Term Archival of CMIP6 data in the IPCC Data Distribution Centre (DDC) at DKRZ

- **Assessment - IPCC FAIR Guidelines** are implemented by
  - IPCC WGI TSU gathering metadata/data from authors,
  - CEDA/Metadataworks long-term preserving final datasets, and
  - DKRZ long-term preserving CMIP6 input data subset.

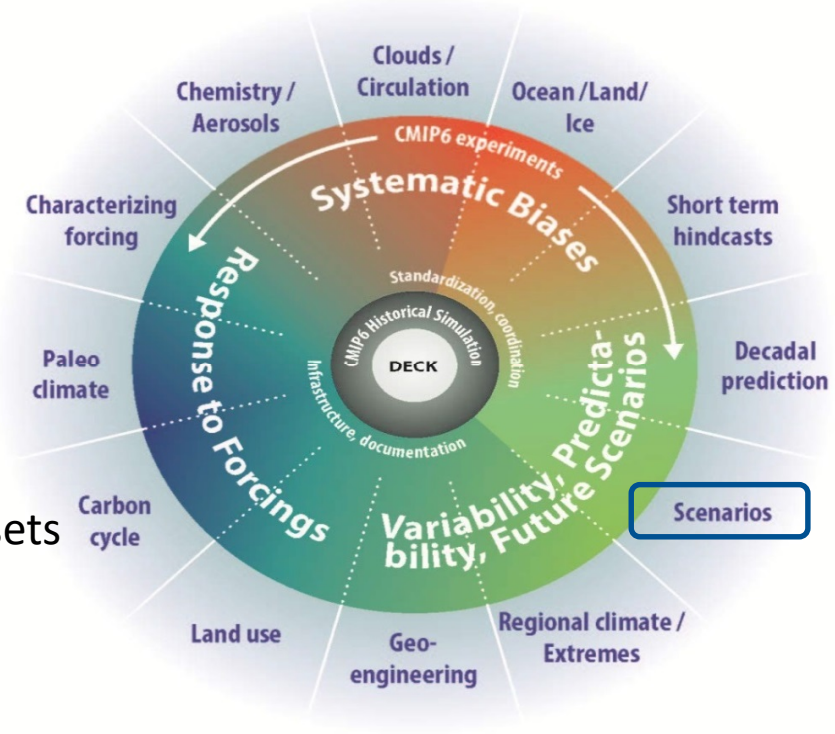
CMIP Phase 6 (CMIP6)

References:
ESGF: http://esgf.llnl.gov
Stockhause, M. et al., 2019. Data Distribution Centre Support for the IPCC Sixth Assessment. http://doi.org/10.5334/dsj-2019-020
Stockhause, M. et al., 2015. CMIP6 Data Citation and Long-Term Archival (WIP white paper). https://doi.org/10.5281/zenodo.35178

CMIP6 data in the IPCC AR6

CMIP6 Numbers (2021-11-15):

- 24 MIPs+DECK with ca. **334** experiments are part of CMIP6 (CMIP5: 1 MIP; 101 → 50 exp.)
- 50 institutions with **157** models are currently registered for CMIP6 (CMIP5: 27 institutions; 60 models)
- **1 309** requested parameters/variables
- Data volume: ca. 21.6 PB / 12 million datasets (CMIP5: ca. 3 PB)
- ca. 25 ESGF data nodes world-wide will disseminate data



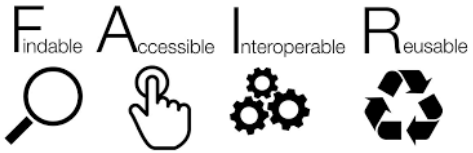
(Source: Eyring, V. et al. (2016). Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization, Geosci. Model Dev., 9, 1937-1958, doi:10.5194/gmd-9-1937-2016.)

IPCC FAIR Guidelines

combine established transparent scientific review process based on peer-reviewed literature with FAIR data and CoreTrustSeal repository standards.

Motivation:

Enhancement of the transparency of IPCC's outputs.



Aims:

- Traceability of key statements of the report
- Credit for input data, created scripts and final data (underpinning figures and tables)
- Preservation of curated core CMIP6 input data, scripts and final data

Development of the IPCC FAIR Guidelines by WG TSUs and DDC

TGICA under review / stopped (26-27 January 2016)

→ IPCC EM on the Future of TGICA (26-27 January 2016, report)

- Informal "coffee-break" discussions on data requirements with WGI Co-Chair

First IPCC AR6 Data Workshop (19-20 September 2017, report):

- Representatives from DDC at CEDA and DKRZ, and WGI and WGII TSUs
- Exchange WG requirements and services provided by the DDC and prioritize collaboration issues

Second IPCC AR6 Data Workshop 20 February 2018, WebEx:

- Additional representatives from DDC at CIESIN and WGIII TSU
- Organizational issues

TG-Data established (13-16 March 2018)

→ Participation at IPCC EM on Assessing Climate Information for Regions (16-18 May 2018, IPCC-XL VIII/INF.5):

- DDC at CEDA and DKRZ participated at WGI EM
- Aim of the EM: collaboration between WGI and WGII

WGI Training on Data and Software Development (6-7 June 2019, information)

- Focus on implementation

TG-Data's first meeting (6-8 November 2019)

→ First Meeting for TG-Data (6-8 November 2019)

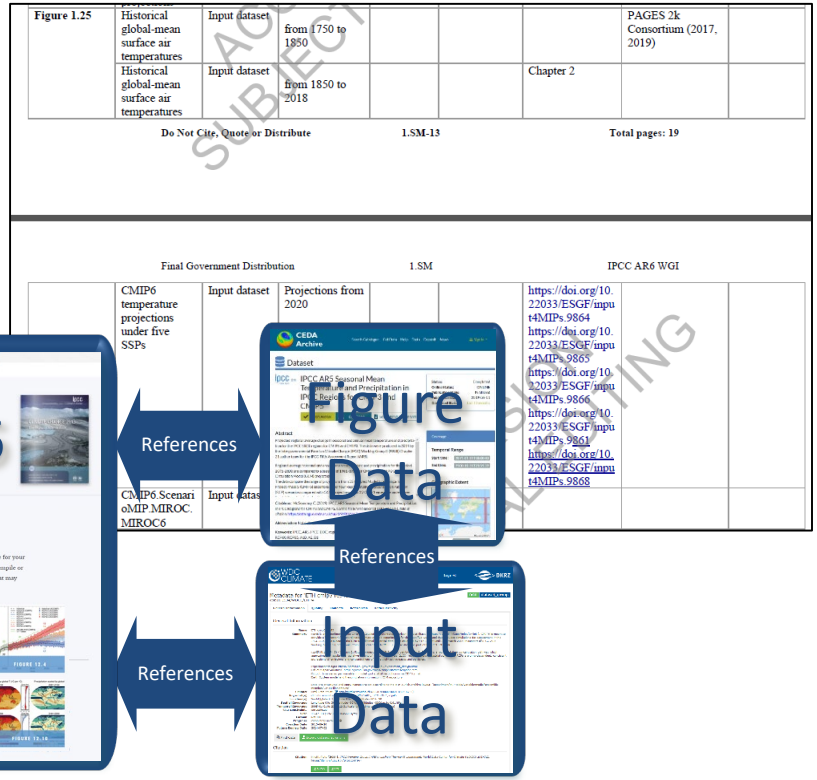
- Formal approval of the IPCC FAIR Guidelines

IPCC FAIR Guidelines - Traceability

Aims: Traceability – Credit - Preservation

Traceability of key statements of the report by documentation of data usage and figure/table creation

- Supplementary Materials document data usage and figure creation enabling figure reproducibility
- References to data from digital AR6
- References to AR6 chapters are added to the DDC Data Archive for CMIP6 input data enabling machine-access (e.g. via Scholix).

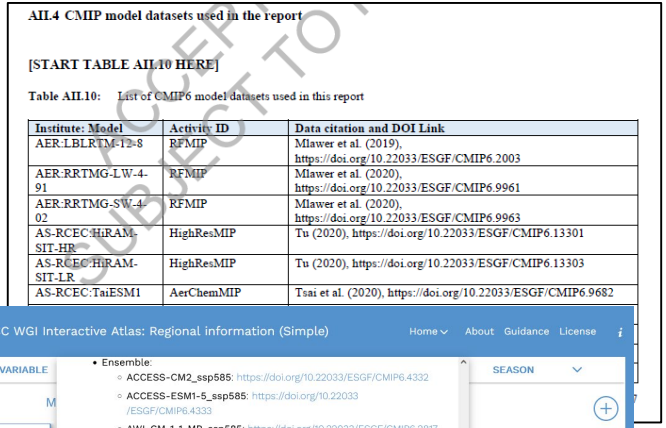


IPCC FAIR Guidelines - Credit

Aims: Traceability – Credit - Preservation

Credit for CMIP6 input data, other input data, created scripts and final data:

- CMIP6 data is cited in AR6 WGI Annex II and the Interactive Atlas
- Created final datasets and scripts (figures, tables) are citable by DOIs

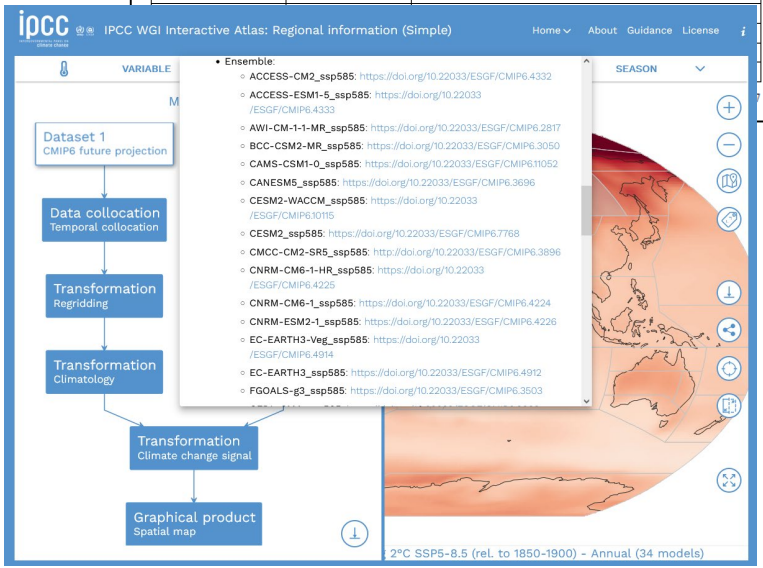


AII.4 CMIP model datasets used in the report

[START TABLE AII.10 HERE]

Table AII.10: List of CMIP6 model datasets used in this report

Institute: Model	Activity ID	Data citation and DOI Link
AER:LBLRTM-12-8	RFMIP	Mlawer et al. (2019), https://doi.org/10.22033/ESGF/CMIP6.2003
AER:RRTMG-LW-4-91	RFMIP	Mlawer et al. (2020), https://doi.org/10.22033/ESGF/CMIP6.9961
AER:RRTMG-SW-4-02	RFMIP	Mlawer et al. (2020), https://doi.org/10.22033/ESGF/CMIP6.9963
AS-RCEC:HiRAM-SIT-HR	HighResMIP	Tu (2020), https://doi.org/10.22033/ESGF/CMIP6.13301
AS-RCEC:HiRAM-SIT-LR	HighResMIP	Tu (2020), https://doi.org/10.22033/ESGF/CMIP6.13303
AS-RCEC:TaiESM1	AerChemMIP	Tsai et al. (2020), https://doi.org/10.22033/ESGF/CMIP6.9682



Data Citation requirements:

- CMIP6 data licenses (BY in CC BY-[NC-]SA) require to give appropriate credit to the creators.
- Citing original data sources, software, materials are part of Good Scientific Practice.

References:
DFG (2019). Guidelines for Safeguarding Good Research Practice. https://doi.org/10.5281/zenodo.3923602

IPCC FAIR Guidelines – Credit requires CMIP6 Citation Service

Aims: Traceability – Credit - Preservation

CMIP6 Citation Service provides DataCite DOIs for data citation

The citation service was very successful: All CMIP6 datasets were citable on the AR6 WGI cut-off date 2021-01-31.

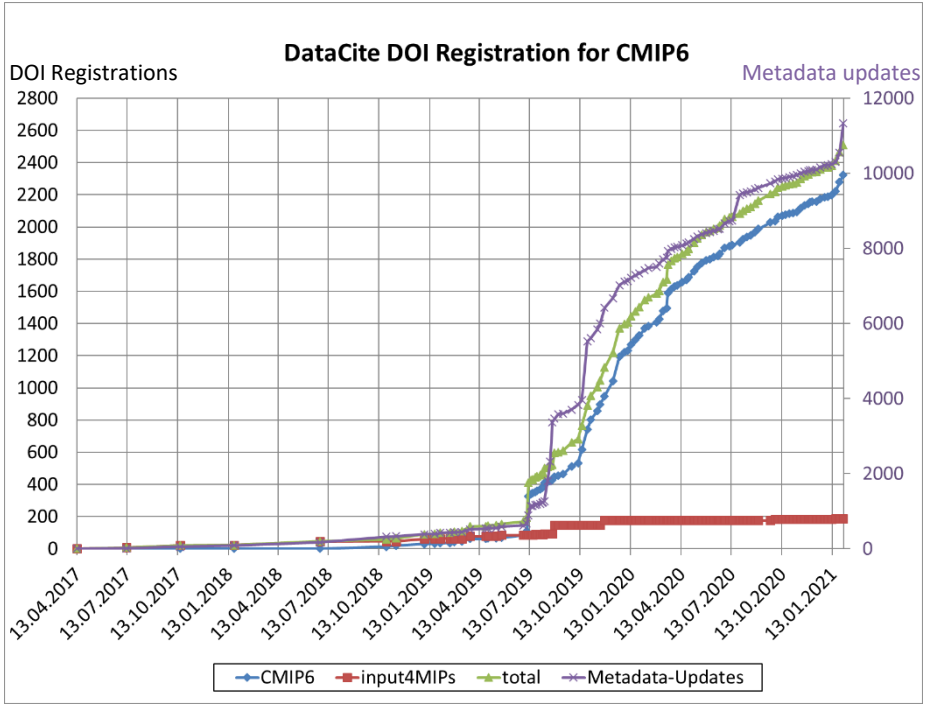
DOI Status on 2021-01-31:

Number of DOIs:	2 509
• CMIP6:	2 322
• input4MIPs:	187
Number of MD updates:	13 876
Number of Scholix references:	63

License Issue:

- CMIP6 data are licensed under CC BY-SA or CC BY-SA-NC
- IPCC's reports are licensed under CC BY-NC-ND with exemptions upon request (see copyright)
- IPCC DDC data are licensed under CC BY

→ IPCC was granted an exemption from WGCM CMIP6

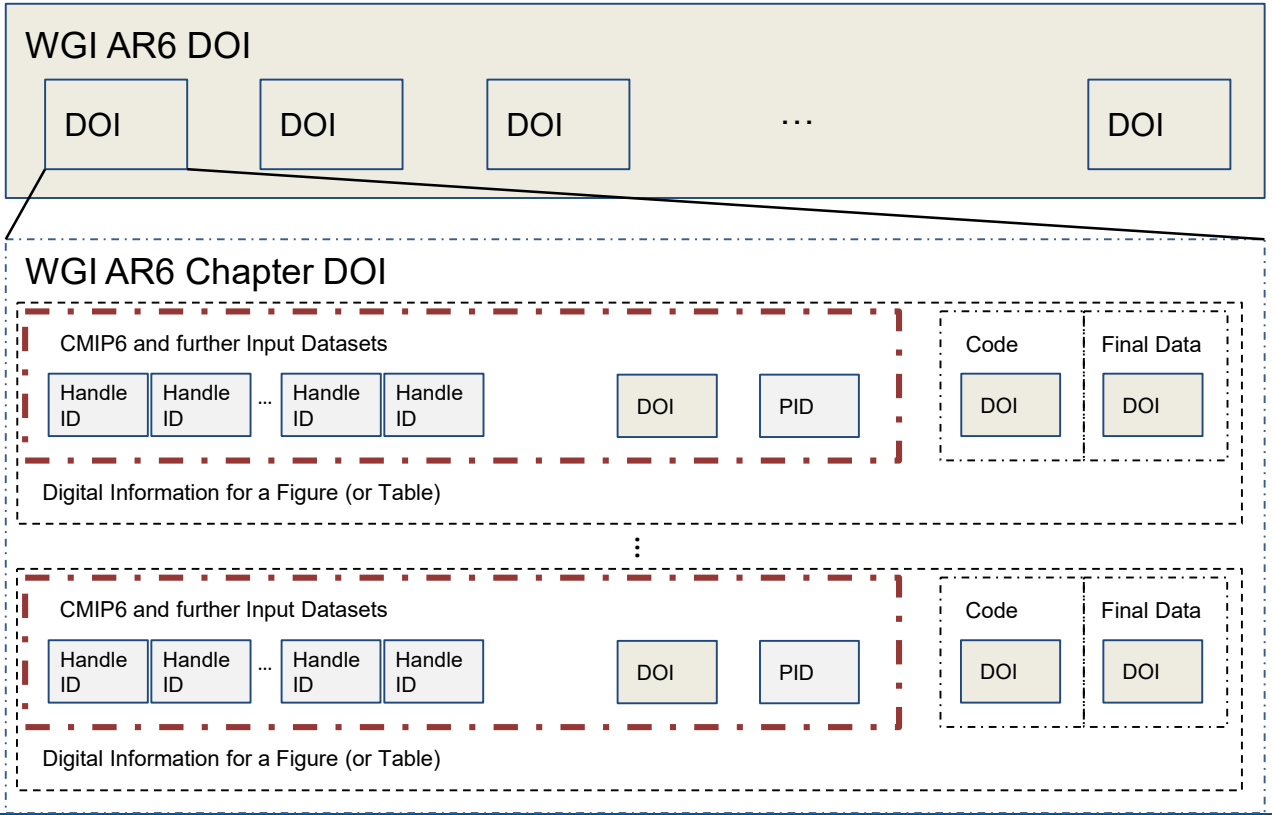


Source: http://bit.ly/CMIP6_DOI_Statistics

IPCC FAIR Guidelines – Credit for whom?

Aims: Traceability – Credit - Preservation

- For IPCC authors
  of the AR6 and its chapters
  via crossref DOIs
- For IPCC script and final data creators via DataCite DOIs (are planned for an individual IPCC Report, e.g. WGI AR6, and for each Chapter in the Report.
- For CMIP6 and other input data creators via data citation in addition to peer-reviewed literature citations

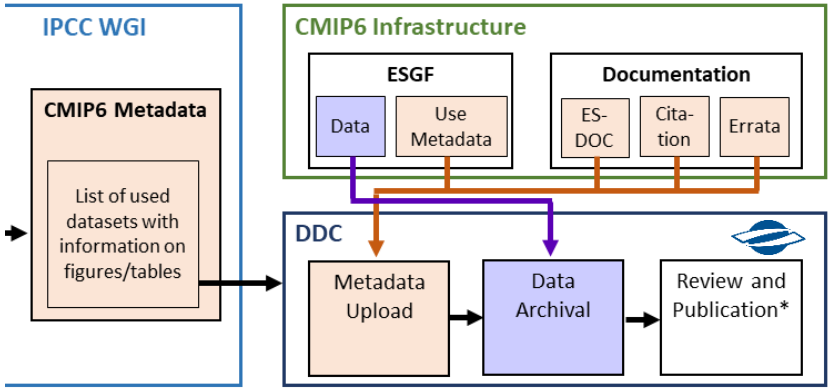


IPCC FAIR Guidelines - Preservation

Aims: Traceability – Credit - Preservation

Preservation of curated core CMIP6 input data, scripts, and final data:

- CMIP6 data underpinning the AR6 is curated by enriching metadata with Controlled Vocabulary, Citation Service (authors, references) and ES-DOC information and preserved on the long-term in the IPCC DDC at DKRZ



* DataCite publication and publication on IPCC webpages

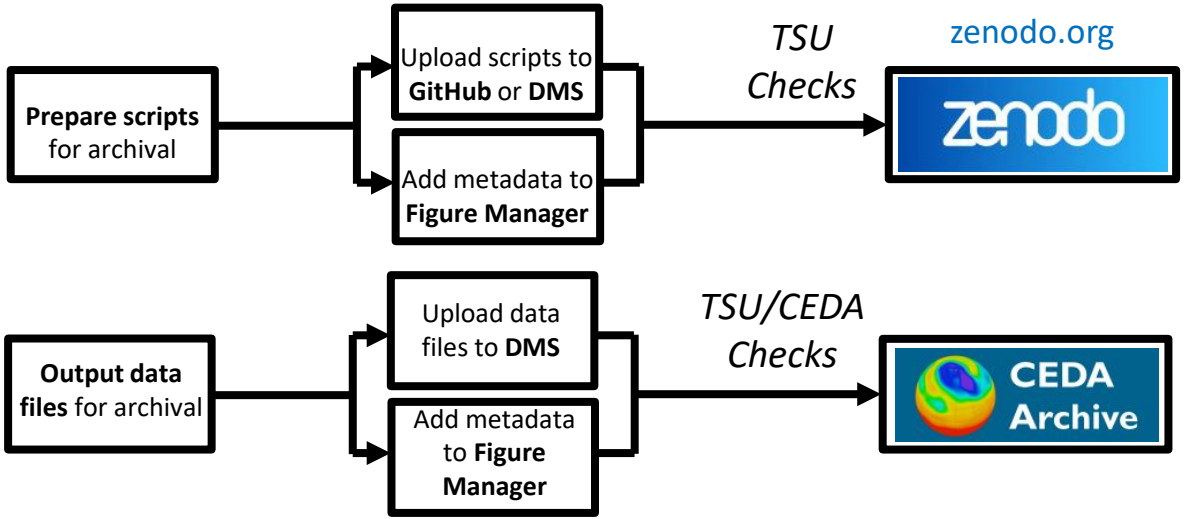
References:
Controlled Vocabulary: https://github.com/WCRP-CMIP/CMIP6_CVs
Citation Service: http://cmip6cite.wdc-climate.de
ES-DOC: http://es-doc.org

IPCC FAIR Guidelines - Preservation

Aims: Traceability – Credit - Preservation

Preservation of curated core CMIP6 input data, scripts, and final data:

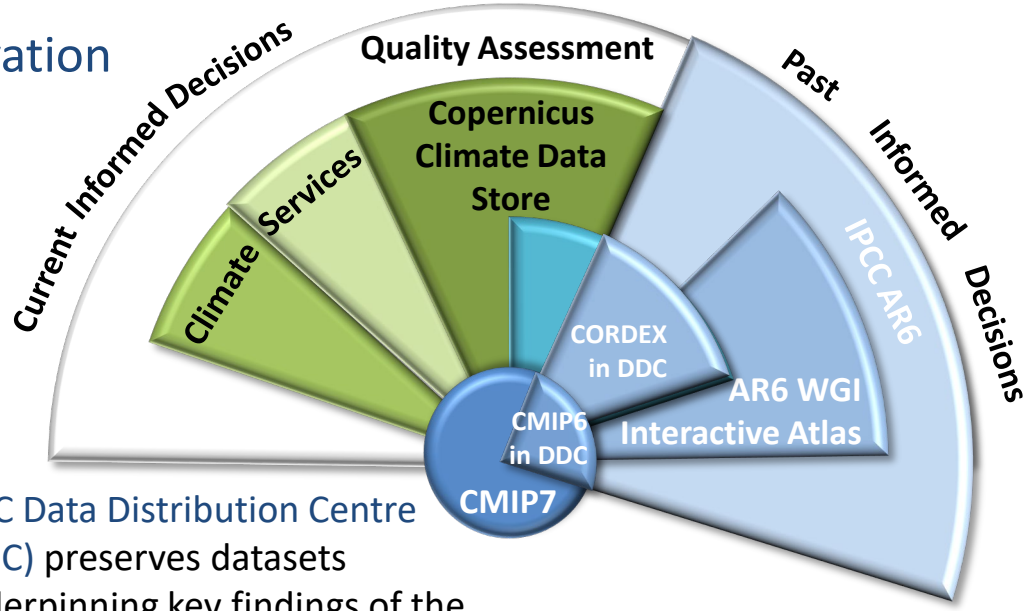
- Scripts used for figure creation are preserved via Zenodo/GitHub
- Final data is preserved on the long-term at CEDA and Metadataworks



IPCC FAIR Guidelines – Preservation in context of informed decisions

Aims: Traceability – Credit - Preservation

Tracing back data used for past decisions relies on long-term preservation of the core FAIR Digital Objects (DO) framework.



Long-term preservation:

- TRUST principles (Traceability, Responsibility, User Focus, Sustainability, Technology) describe the effort of repositories
- Regular Membership of the World Data System (WDS) includes a Core Trust Seal certification of the data center operations.

IPCC Data Distribution Centre (DDC) preserves datasets underpinning key findings of the IPCC AR6.
Established DDC Partners are WDS Regular Members.

Long-term preservation of the FAIR DO framework is required for the sustainable traceability of decisions.

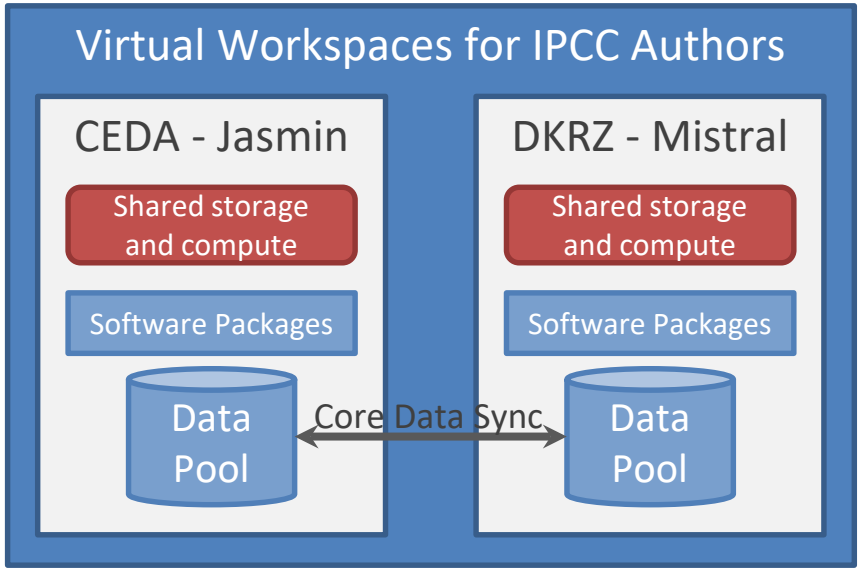
DDC Support for the IPCC Authors

Virtual Workspaces: Data – Software – Collaboration Space

The DDC Partners CEDA and DKRZ support IPCC AR6 authors in the writing process by providing Virtual Workspaces (co-funded by IS-ENES).

Virtual Workspaces provide platforms for data analysis and author collaboration on data and figure creation

Data Pool content is quality controlled for the Copernicus' Climate Data Store (CDS) and synchronized with IPSL as third partner.



IPCC FAIR Data Guidelines Implementation: First Experiences

IPCC AR6 FAIR Data Guidelines Implementation: First Experiences

Challenges:

- Increased workload for Authors, TSU and DDC Partners
- Additional requirements for long-term data preservation
- Enhanced communication between Authors, TSU and DDC Partners required

Benefit:

- Authors: Aids best practice in writing, drafting figures and compiling tables, and improves recognition of authors' work
- TSU: Provides added value for report users creating new work using the report. Streamlines TSU workflow in answering user requests for report data.
- DDC: Best practices in data management are adopted, which strengthen the relation of data and report. Streamlines DDC Partner user support for requests about data usage in the AR6.

IPCC FAIR Data Guidelines Implementation: Current Challenges

IPCC AR6 FAIR Data Guidelines Implementation: Current Challenges

- Recommendations for Improvements:
  of the FAIR Guidelines' implementation and the interaction between the different partners
- Continuation:
  IPCC WGs and their TSUs change every assessment cycle, which requires a hand-over of ideas and tools.
- Exhaustive Data Archival:
  Enhancing the IPCC data archival capability.
- Enhanced Functionalities:
  DDC service gaps and areas for enhancement include uniform data discovery (DDC catalog), improved data access (compute and cloud services), and machine-access via a single API.
- Integration:
  Better integration of FAIR Guidelines into IPCC's assessment cycle
- Procedures:
  Revisiting IPCC procedures and protocols relating to data and the DDC.
- Sustainable Funding:
  This implementation effort currently relies on institutional and national contributions, which are not guaranteed on the long-term.

IPCC AR6 FAIR Data Guidelines are a success and mayor step forward combining Good Scientific Practice with best practices in data management (FAIR and TRUST principles). They can serve as example for other climate information and service providers, e.g. climate services like Copernicus Climate Data Store.

https://www.ipcc-data.org/
https://zenodo.org/communities/ddc-support
stockhause@dkrz.de
https://orcid.org/0000-0001-6636-4972

HELMHOLTZ
Open Science

Contact

- Mail – open-science@helmholtz.de
- Website – https://os.helmholtz.de
- Open Science Newsletter
- Mailing list (only for members of Helmholtz) – Helmholtz Open Science Professionals
- Twitter – @helmholtz_os
- LinkedIn – Helmholtz Open Science Office

HELMHOLTZ
Open Science

Thank you for your interest!

Helmholtz Open Science Office:

Lea Maria Ferguson,
https://orcid.org/0000-0002-7060-3670

Antonia C. Schrader,
https://orcid.org/0000-0001-7080-634X

Nina Leonie Weisweiler,
https://orcid.org/0000-0001-6967-9443