

The reality of implementing FAIR principles in the IPCC context to support open science and provide a citable platform to acknowledge the work of authors.

Emily Anderson<sup>1</sup>, Molly MacRae<sup>1</sup>, Diego Cammarano<sup>2</sup>, Charlotte Pascoe<sup>1</sup>, Anna Pirani<sup>3</sup>, Lina Sitz<sup>4</sup>, Martina Stockhause<sup>5</sup>

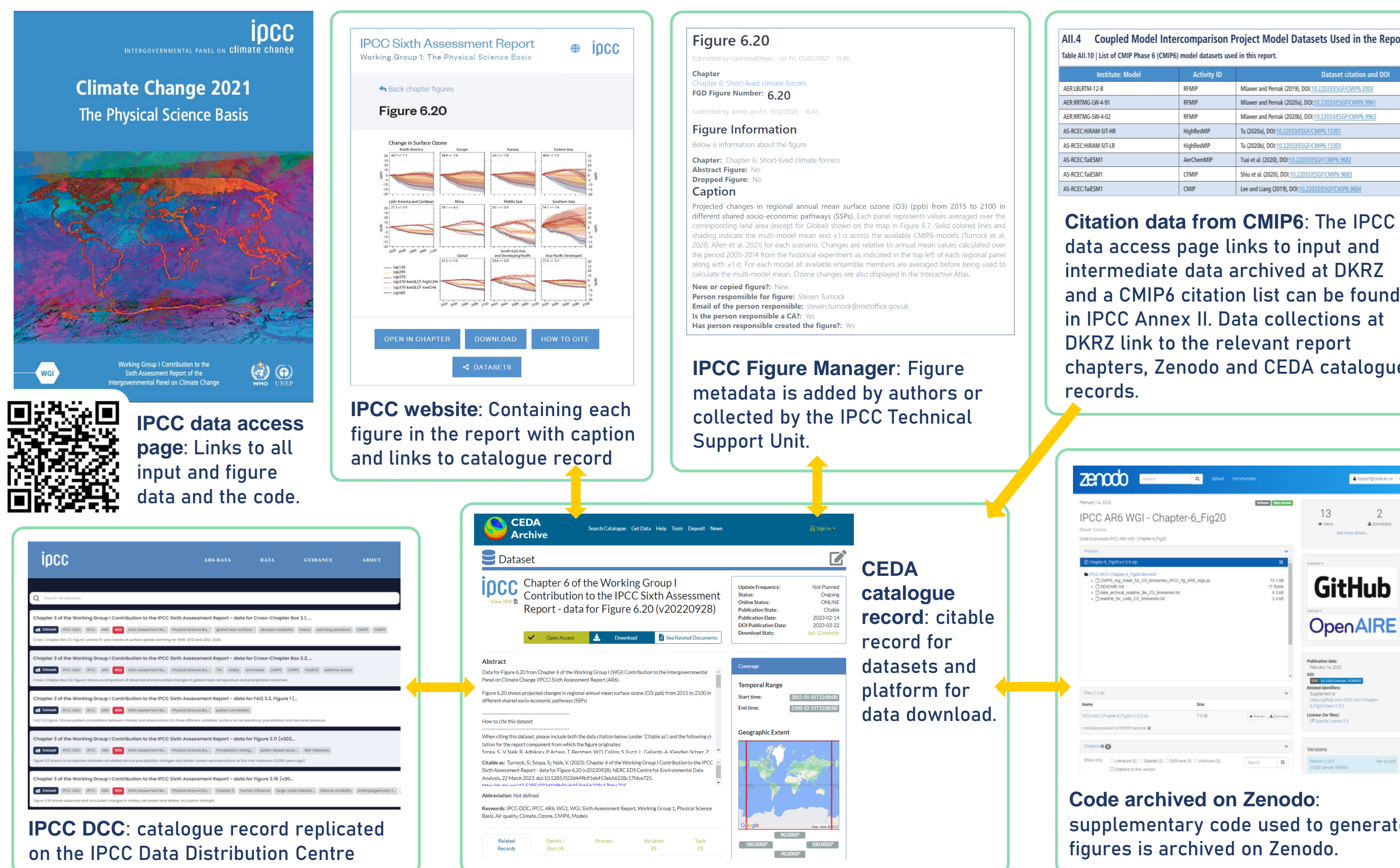
1. Centre for Environmental Data Analysis, Science and Technology Facilities Council, UK, 2. University Paris Saclay, France, 3. ICTP Trieste, Italy, 4. LEDaS, Italy, Research Data Alliance, European Open Science Cloud 5. German Climate Computing centre, Germany

IPCC Data Distribution Centre (DDC) partners at the Centre for Environmental Data Analysis (CEDA) have worked with the IPCC Technical Support Unit (TSU) for WGI to publish figure data from the Sixth Assessment Report (AR6). As the leading voice of climate science, the IPCC reports represent the current understanding of climate change and its global impacts, guidance on international policy and the projected future of our planet. Therefore, ensuring data presented in IPCC reports are findable, accessible, interoperable and reusable (FAIR) increases credibility and permits reuse of IPCC data and results. Additionally, adhering to FAIR principles adds value to the work of the authors and sets an example for the stewardship of valuable data. The IPCC FAIR guidelines are advised by the Task Group on Data Support for Climate Change Assessment (TG-Data). The diversity of figure data included in the IPCC report posed challenges for archival and required careful collaboration between authors, the TSU and DDC partners. Data from the IPCC report can be split into three categories: input (source) data, intermediate (post-processed) data, and final (plotted figure) data. The IPCC data is archived between DDC partners, including CEDA for figure data and the German Climate Computing Centre (DKRZ) for CMIP6 input and selected intermediate data. Here we outline workflows developed, lessons learned and recommendations moving forward for the Seventh Assessment Report.

**AR6** – The Sixth Assessment Report of the IPCC  
**CEDA** – Centre for Environmental Data Analysis  
**CMIP6** – Coupled Model Intercomparison Project Phase 6  
**DDC** – Data distribution centre for the IPCC  
**DKRZ** – German Climate Computing Centre  
**FAIR** – Findable, Accessible, Interoperable, Reusable  
**IPCC** – Intergovernmental Panel on Climate Change  
**TG-Data** – The IPCC Task Group on Data Support for Climate Change Assessments  
**TSU** – Technical support unit for WGI  
**WGI** – Working Group 1

## IPCC Working Group I (WGI) Data Publication Ecosystem

Throughout the process IPCC FAIR guidelines were followed by all DDC partners. There is a two way link between the figures on the IPCC website and the plotted figure data archived at CEDA. There is also a two-way link between the data archive and the code associated with each figure on Zenodo, and between the catalogue record and the joint IPCC DCC catalogue via metadataworks. DKRZ follows the same principle of linking to other available AR6 results, report, figure data and code. The data is also included in the joint DDC catalogue.



## CEDA Catalogue Records are FAIR

The screenshot shows a CEDA Catalogue Record for Figure 6.20. It includes the following information:

- Dataset:** Chapter 6 of the Working Group I Contribution to the IPCC Sixth Assessment Report - data for Figure 6.20 (v20220928)
- Update Frequency:** Not Planned
- Status:** Ongoing ONLINE
- Online Status:** Citable
- Publication State:** Citable
- Publication Date:** 2023-02-14
- DOI Publication Date:** 2023-03-22
- Download Date:** last 12 months
- Abstract:** Data provided in relation to figure. All the data files provided are used to create the time series plots for each region. The numbers in each panel for each region are obtained from 'Surf\_O3\_data\_05\_14\_mean\_for\_IPCC\_figure\_V1\_Smoods.csv' with the time series line for each scenario from 'Surf\_O3\_data\_1st\_mean\_for\_IPCC\_figure\_V1\_Smoods.csv' and the shading obtained by using the values in 'Surf\_O3\_SD\_data\_1st\_mean\_for\_IPCC\_figure\_V1\_Smoods.csv'.
- Keywords:** IPCC-DDC, IPCC, AR6, WGI, Working Group I, Physical Science Basis, Air quality, Climate, Ozone, CMIP6, Models
- Access rules:** Public data: access to these data is available to both registered and non-registered users. Use of these data is covered by the following licence: <http://creativecommons.org/licenses/by/4.0/>. When using these data you must cite them correctly using the citation given on the CEDA Data Catalogue record.
- Related Documents:** CODE FOR FIGURE (archived on Zenodo), LINK TO FIGURE, SUPPLEMENTARY MATERIAL FOR REPORT CHAPTER: Soega, S., V. Naik, B. Adhikary, P. Artaxo, T. Bernsten, W.D. Collins, S. Fuzzi, L. Gallardo, A. Kienler-Scharr, Z. Klimont, H. Liao, N. Unger, and P. Zanis, 2021: Short-Lived Climate Forcers (SLCF) Supplementary Material, In: Climate Change 2021: The Physical Science Basis, Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change.

Figure data from the IPCC Sixth Assessment Report (AR6) is archived at the Centre for Environmental Data Analysis (CEDA). Catalogue records for figure datasets were created providing a platform for easy access to the data, metadata and links to any related documentation. This supports the data to be findable, accessible, interoperable and reusable (FAIR).

Findable - The CEDA catalogue record for each figure dataset enhances findability. Keywords can be easily searched, and records are organised into collections for each AR6 chapter. There is a two-way link between the catalogue record and the figure on the AR6 website. CEDA catalogue records are duplicated on the IPCC-DDC.

Accessible - Scientific language is understandable, with acronyms and specific terminology fully explained. CEDA services provide tools for users to access and download the data.

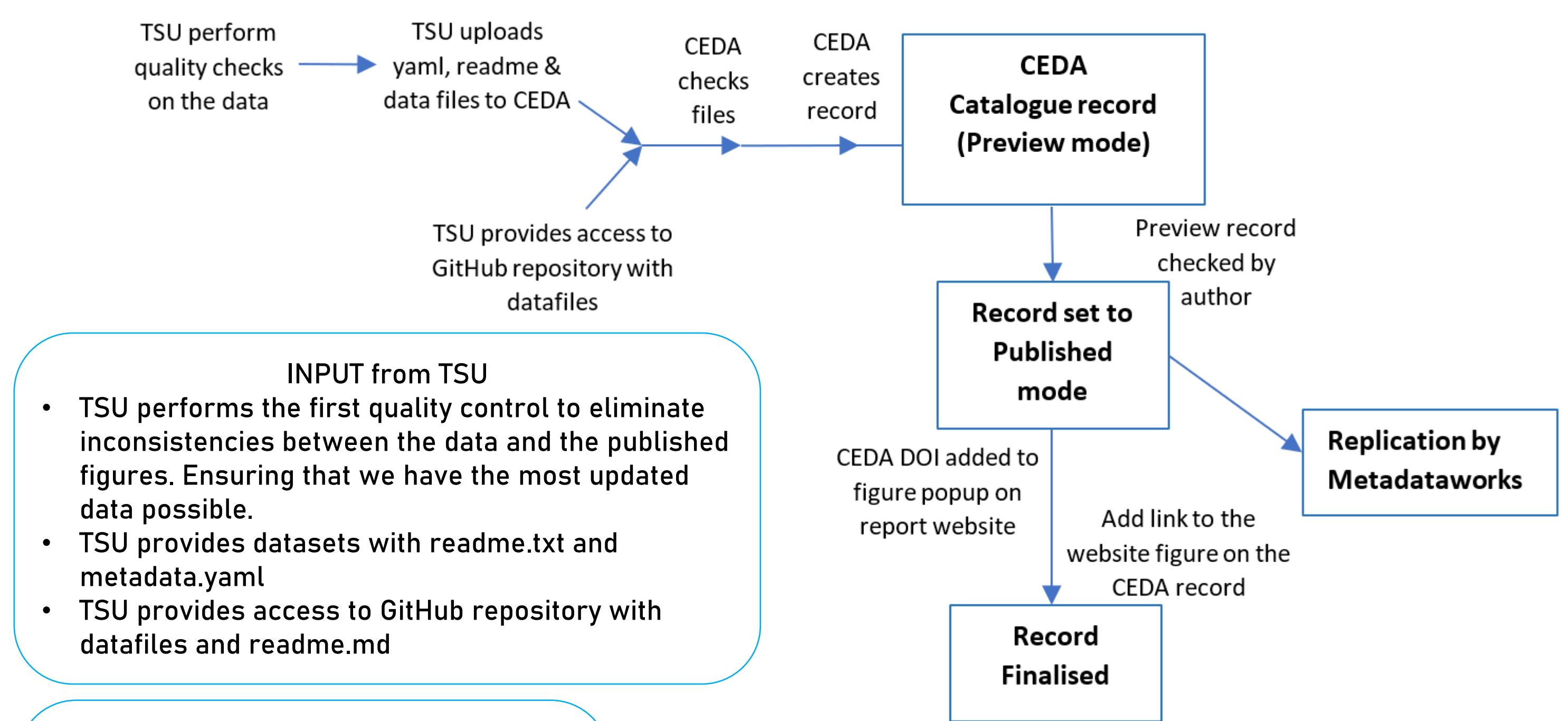
Interoperable - Where possible data variables follow standard file format conventions such as CF-netCDF and have standard names. Where this is not feasible readme files describe the file structure and content.

Reusable - The data can be reused, shared and adapted elsewhere, with credit, under a Creative Commons Attribution 4.0 licence (CC BY 4.0). Catalogue records link to relevant documentation such as the Digital Object Identifier (DOI) for the code and other supplementary information. This allows users to reproduce the figures from the report independently.

Author and data creator recognition  
 CEDA catalogue records provide a platform to acknowledge the specific work of IPCC authors and dataset creators whose work supports the scientific basis of AR6.

## Adapting the Workflow of the Data

The data workflow was adapted as the archival process went on, to reflect the varying ways that the TSU and CEDA received data. Maintaining this workflow was achieved by developing a spreadsheet to easily view the stage of any figure, and to enable easy communication between CEDA and the TSU.



**INPUT from TSU**

- TSU performs the first quality control to eliminate inconsistencies between the data and the published figures. Ensuring that we have the most updated data possible.
- TSU provides datasets with readme.txt and metadata.yaml
- TSU provides access to GitHub repository with datafiles and readme.md

**PROCESSING by CEDA**

- Version id added, yaml adapted for .py script, data files quality checked
- Python script run and catalogue record generated
- Additional information from TSU Figure Manager added to record
- If data is provided via GitHub then the record is created manually

**REVIEW facilitated by TSU**

- Contact TSU and confirm final draft
- TSU seeks approval/edits from author
- CEDA officer makes any corrections and final checks
- Record is published

Spreadsheet created to communicate figure status between CEDA and the TSU

Figure	Data in arrivals	Data in processing	Catalogue Record	Data in archive	Published	Catalogue DOI	Link to Code
1	ch6_fig3	y	y	y	n	n	n
2	ch6_fig9	n	n	n	n	n	n
3	ch6_fig12	y	y	y	y	y	y
4	ch6_fig16	y	y	y	y	y	y
5	ch6_fig17	y	y	y	y	y	y
6	ch6_fig18	n	n	n	n	n	n
7	ch6_fig19	n	n	n	n	n	n
8	ch6_fig20	n	n	n	n	n	n
9	ch6_fig20	y	y	y	y	y	y

## Highlights

**Rewarding experience:** Although time consuming and resource intensive, the process of archiving IPCC data in a FAIR way has enabled this data to be open and accessible to users, and has promoted the reuse of IPCC report results - including for authors of the next cycle.

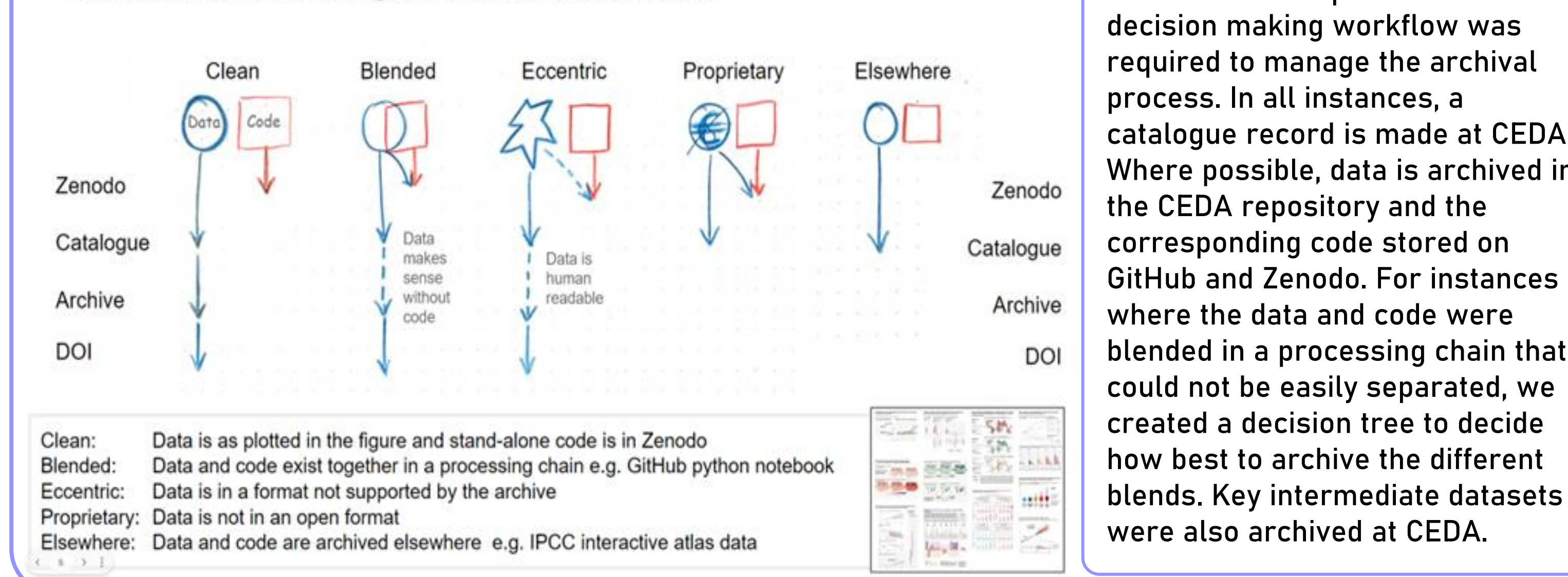
**Team effort of the task:** We would like to highlight the collaborative nature of this task and all the individuals involved in the process. Without this collaboration many of the decisions relating to the workflow and practicalities of archiving the data would not have been possible. The undertaking of this new and complex task provided many challenges, however the specialised experience and knowledge within the team made the process achievable.

**Greatest challenge:** A significant obstacle for us arose in situations where authors had moved on from the IPCC report and were focusing on other projects, so often did not reply to queries regarding data archival. Ideally authors would provide data at the start of the data preparation timeline (rather than at the end of the IPCC report writing process) to streamline the data management process and minimise any delays in archiving data.

The screenshot shows the CEDA Catalogue Record for Figure 6.20, including the abstract, keywords, and related documents.

## Decision Making

Decision Tree for IPCC Figure Data and Code Archival



The diversity of figure data and code from the report meant a decision making workflow was required to manage the archival process. In all instances, a catalogue record is made at CEDA. Where possible, data is archived in the CEDA repository and the corresponding code stored on GitHub and Zenodo. For instances where the data and code were blended in a processing chain that could not be easily separated, we created a decision tree to decide how best to archive the different blends. Key intermediate datasets were also archived at CEDA.

## Recommendations for the Seventh Assessment Report

The novelty of the implementation of FAIR Data Principles in the WGI AR6 Assessment presented many challenges. Lessons learned and recommendations for the seventh assessment report of the IPCC are being developed by the IPCC WGI Technical Support Unit (TSU) and Data Distribution Centre (DDC) representatives.

- Key recommendations include:
- At the beginning of the assessment authors should be provided guidelines on how to organise the code and the input data to reduce the volume of missing data, code and metadata at the end of the cycle.
  - Integrate FAIR data practices into the report procedures and writing process to allow direct links to be made from input data, through intermediate data, to the final data plotted in the report.
  - Authors also need to be provided with clear instructions on mandatory requirements and with a clear timeline to promote author involvement.
  - Dedicated TSU staff is needed to support authors and work with the Data Distribution Centres for the duration of the cycle.
  - Attention to resources for the end of the cycle timeline is needed since this is when a substantial part of the archival and curation work will be completed, including fully interconnecting the code and data products with the report itself.
  - Between assessment cycles workflows can be improved and hands-on assistance provided to support authors in order to decrease time pressures during cycles.
  - It was difficult to find balance between the practices of authors and the requirements and restrictions for long term archival. Support and guidance provided to authors from the start would help resolve these differences.
  - Authors should be encouraged to pay attention to the licence that is used to publish their data and ensure it is open and accessible.

<https://doi.org/10.5281/zenodo.6992173>