

OPEN
ARTICLE

Implementation of FAIR principles in the IPCC: the WGI AR6 Atlas repository

Maialen Iturbide¹, Jesús Fernández¹✉, José M. Gutiérrez¹✉, Anna Pirani², David Huard³, Alaa Al Khourdajie⁴, Jorge Baño-Medina¹, Joaquin Bedia^{5,6}, Ana Casanueva^{5,6}, Ezequiel Cimadevilla⁵, Antonio S. Cofiño¹, Matteo De Felice⁷, Javier Diez-Sierra^{1,5}, Markel García-Díez⁸, James Goldie⁹, Dimitris A. Herrera^{10,11}, Sixto Herrera⁵, Rodrigo Manzanas^{5,6}, Josipa Milovac¹, Aparna Radhakrishnan¹², Daniel San-Martín⁸, Alessandro Spinuso¹³, Kristen M. Thyng¹⁴, Claire Trenham¹⁵ & Özge Yelekçi¹⁶

The Sixth Assessment Report (AR6) of the Intergovernmental Panel on Climate Change (IPCC) has adopted the FAIR Guiding Principles. We present the Atlas chapter of Working Group I (WGI) as a test case. We describe the application of the FAIR principles in the Atlas, the challenges faced during its implementation, and those that remain for the future. We introduce the open source repository resulting from this process, including coding (e.g., annotated Jupyter notebooks), data provenance, and some aggregated datasets used in some figures in the Atlas chapter and its interactive companion (the Interactive Atlas), open to scrutiny by the scientific community and the general public. We describe the informal pilot review conducted on this repository to gather recommendations that led to significant improvements. Finally, a working example illustrates the re-use of the repository resources to produce customized regional information, extending the Interactive Atlas products and running the code interactively in a web browser using Jupyter notebooks.

Introduction

The accessibility and reproducibility of scientific results is a major concern in all scientific disciplines¹. This was identified as a key aspect to be addressed during the preparation of the Sixth Assessment Report (AR6) of the Intergovernmental Panel on Climate Change (IPCC) in order to ensure the transparency of the products underpinning the report and enhance their use. This was particularly relevant for the digital products developed as part of the Atlas chapter and the Interactive Atlas (<https://interactive-atlas.ipcc.ch>) of Working Group I (WGI). Best practices and standards for open science were defined and implemented in collaboration with the IPCC Task Group on Data Support for Climate Change Assessments (TG-Data, <https://www.ipcc.ch/data>). These are summarized in the TG-Data guidance document on FAIR (Findable, Accessible, Interoperable, Reusable) principles for IPCC AR6².

FAIR data principles^{3,4} aim to facilitate open science by ensuring that the data and code used are findable and accessible and can be reused for reproducibility and for further developments using interoperable

¹Instituto de Física de Cantabria (IFCA), CSIC-Universidad de Cantabria, Santander, Spain. ²International Centre for Theoretical Physics (ICTP), Trieste, Italy. ³Ouranos, Montréal, Québec, Canada. ⁴Centre for Environmental Policy, Imperial College London, London, UK. ⁵Dept of Applied Mathematics and Computer Science (MACC), Universidad de Cantabria, Santander, Spain. ⁶Grupo de Meteorología y Computación, Universidad de Cantabria, Unidad Asociada al CSIC, Santander, Spain. ⁷European Commission, Joint Research Centre (JRC), Petten, The Netherlands. ⁸Predictia Intelligent Data Solutions S.L., Santander, Spain. ⁹360info, Monash University, Melbourne, Australia. ¹⁰Instituto Geográfico Universitario, Universidad Autónoma de Santo Domingo, Santo Domingo, 10103, Dominican Republic. ¹¹Department of Geography & Sustainability, University of Tennessee, 1000 Philip Fulmer Way, Knoxville, TN, 37996-0925, USA. ¹²Princeton University, Princeton, USA. ¹³Koninklijk Nederlands Meteorologisch Instituut (KNMI), De Bilt, The Netherlands. ¹⁴Axiom Data Science, Anchorage, USA. ¹⁵Commonwealth Scientific & Industrial Research Organisation (CSIRO) - Oceans & Atmosphere, Canberra, Australia. ¹⁶IPCC WGI TSIU, Université Paris Saclay, Gif-sur-Yvette, France. ✉e-mail: Jesus.Fernandez@unican.es; gutierrezjm@unican.es

infrastructures. Guidance and training were provided to all authors, recommending that FAIR principles were adopted in the preparation of the WGI report. This is still ongoing work and has been successful to varying degrees. All the data underpinning the Summary for Policymakers (SPM) figures is now publicly available⁵, but data, metadata and scripts supporting the Technical Summary and chapter figures and key results are still being assembled and are available only in some cases, such as in the Atlas chapter here described.

These activities benefited from the adoption of documentation protocols, open-access computing frameworks (such as R or Python) and software management platforms (such as GitHub), which facilitate the collection of metadata (including data provenance) and code, as well as some post-processed datasets. This makes it possible to reproduce the key findings and figures of the report, in particular those of the Atlas chapter. These activities are also being undertaken by the other working groups, namely WGII on climate adaptation, WGIII on climate mitigation and in the Synthesis Report (SR).

FAIR principles were particularly relevant for the Atlas chapter and the Interactive Atlas to facilitate reproducibility of complex figures based on multiple lines of evidence and reproducibility and reusability of the digital products and code, respectively. In particular, the Interactive Atlas is a novel online tool for flexible analyses of observed and projected climate change information for about 25 variables and indices underpinning the WGI contribution to AR6. The implementation of FAIR principles was particularly critical to develop a rigorous and transparent interactive tool as part of the WGI report, as well as an important contribution to a traceable implementation of the IPCC Error Protocol for report figures. Therefore, the Atlas is a comprehensive test case for the development and implementation of FAIR principles, exploring frameworks, protocols and best practices to be expanded to other chapters and the IPCC more broadly. For the Atlas chapter, these results have been collected and uploaded to an open source Atlas repository⁶. These have been made available for review and scrutiny of the scientific community and users more broadly, and this has improved the quality of the final Atlas products.

In this paper, we provide an overview of the Atlas repository scope and contents, emphasizing the data and annotated notebooks readily available to reproduce Atlas figures. We describe the FAIR guiding principles of the Atlas repository, including the informal review process in the context of the IPCC review process, and the challenges faced during their implementation. Overall, the informal review greatly improved the Atlas repository in all aspects and strengthened community trust in the available tools and products. Reusability is a key principle of the Atlas repository and a new working example is provided to illustrate the re-use of the available resources to create and extend the Interactive Atlas products. In particular, we show how to create customized Global Warming Level (GWL) scaling plots using a well-documented Jupyter notebook and freely available online computational resources from MyBinder⁷. Note that GWL is a prominent analysis dimension for climate change assessment in AR6, and these plots allow exploring the sensitivity of regional climate change signals in this context, making it an ideal case for reusability using the Atlas repository resources. Finally, we discuss the challenges that remain for the future, building on the experience of this initial implementation of the FAIR principles into the IPCC review process.

Results

The atlas repository. The Atlas repository contains the code (including Jupyter notebooks), data provenance and some aggregated datasets underpinning key figures of the Atlas chapter and its interactive companion (the Interactive Atlas). Figure 1 shows a schematic representation describing its structure and highlighting its contents. This repository is maintained on GitHub (<https://github.com/IPCC-WG1/Atlas>), which is the current *de facto* standard for code sharing with over 200 million repositories and over 65 million registered users. GitHub provides a collaborative environment for code development, with integrated version control and issue tracking among many other features. The repository is self-documented by Markdown-formatted files exploiting GitHub capabilities to browse folders and link to internal and external resources. Moreover, GitHub enables long-term archiving of repository snapshots, interoperating with services such as Zenodo (<http://zenodo.org>). The final release of the Atlas repository, containing data, metadata, code and documentation, is available not only through GitHub, as described next, but also as a Zenodo entry⁶ with its own persistent Digital Object Identifier (DOI).

Reproducible environment. Full reproducibility and reusability of scientific results requires not only the data and code, but also the whole software environment in which the code was executed. There are several solutions to achieve reproducible execution environments⁸. We adopted incremental Reproducible Execution Environment Specifications (REES) to build a reproducible environment for the Atlas repository. First, we provide the software versions used as a Conda (<https://docs.conda.io>) environment configuration file and a Docker container configuration file, as well as the resulting Docker image to be readily used. In a local computer, the Conda software and environment manager provides a reliable, multi-platform solution, handling versions and dependencies for most of the software stack. The Docker⁹ container configuration file and image extend this by allowing system library dependencies to also be encapsulated. Docker images can be deployed by the user in a local computer, but they also pave the way to deploy the execution environment on cloud computing resources, where virtual machines can be created for particular operating systems running a particular Docker image containing the specific environment used to generate a given scientific result.

In addition, the contents of the Atlas repository can be run interactively on a browser through the MyBinder.org⁷ links provided. Binder provides a simple way to compile a Docker image from a wealth of different REES, including Conda configuration files. Moreover, the MyBinder.org initiative provides cloud resources to run interactive notebooks based on Docker images. This provides an execution environment for users to try and modify notebooks, without installing software on their computers. This environment can optionally include additional content from a GitHub repository through the nbgitpuller extension, which in this case is used to retrieve the Atlas repository containing the notebooks and the data they use.

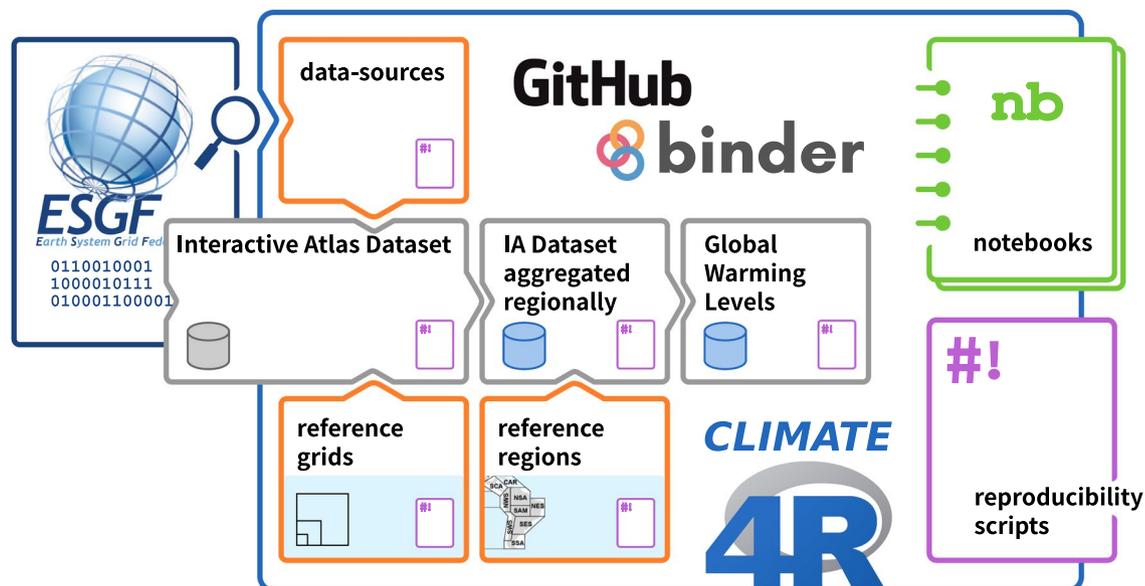


Fig. 1 Schematic representation of the contents of the GitHub Atlas repository backing the Interactive Atlas and Atlas chapter figures. Data at different stages of processing are shown in grey boxes, representing different directories in the GitHub repository. Blue (grey) cylinders represent datasets (not) readily available in the repository. Orange boxes are auxiliary information used at different stages. ESGF provides metadata to build the data sources catalogue, which dictates the data to be processed, also retrieved from ESGF. Notebooks and scripts, mainly relying on the *climate4R* framework, feed from intermediate datasets at different stages to illustrate their use and reproduce specific Atlas chapter figures, respectively. All directories contain a subdirectory with scripts allowing to reproduce the corresponding contents. To enable a reproducible execution environment, the repository contents are loaded in a MyBinder cloud environment with all the required software pre-installed.

This approach ensures practical reproducibility for most of the contents of the Atlas repository. The approach can also be replicated to reuse the data and the execution environment, as illustrated in the reusability example below.

Repository structure. The Atlas repository is structured in folders including the auxiliary products (standard grids, reference regions, etc.) and intermediate datasets (see Fig. 1). Most of the data inputs come from ESGF¹⁰, which distributes the global and regional climate simulations available from CMIP and the Coordinated Regional Climate Downscaling Experiment (CORDEX), respectively. The first step for findability and accessibility is to build an inventory of all data sources used (described in the *data-sources* folder). This involves parsing metadata for all datasets for scenario and historical experiments in CMIP5, CMIP6 and CORDEX. The final selection of model simulations takes into account the availability of all required variables and analysis periods, discarding erroneous datasets reported in errata registries (e.g. in Earth System Documentation, ES-DOC, errata search service) along with other issues found during the subsequent data processing as documented in the *data-sources* folder of the repository. The final inventory is therefore the result of a semi-automated and iterative process. All issues and decisions made are stored in the *data-sources* folder in the working spreadsheet files used during this process. The final inventory of data sources used, including their version, is also stored as machine-readable CSV files. Moreover, the repository provides unique pointers to the exact data and metadata considered, either as DOIs¹¹, ESGF search URLs or handle.net¹² persistent identifiers. Thus, this folder provides both data provenance information and a machine-readable auxiliary product (the inventory of data sources) to be used in subsequent processing steps.

Another auxiliary product stored in the Atlas repository is a set of reference grids used to bring all data sources to a common spatial resolution. The *reference-grids* folder contains global grid definitions, regular in longitude-latitude, stored in netCDF files containing also common land-sea masks. Commensurable grids (including land-sea masks) are defined starting at 0.25 by 0.25 degrees (used to interpolate ERA5 and observational products) and, by aggregating 2 by 2 grid-cells, grids at 0.5 (CORDEX), 1 (CMIP6), and 2 (CMIP5) degrees resolution were obtained. In addition to these reference grids and masks, there are also other masks used in the Atlas products. These include mountain range masks and missing data masks for several observational products which provide global gridded data but interpolate data from distant stations in empty cells.

A third auxiliary product available at the Atlas repository is the specifications of the new set of AR6 reference regions¹³. This is version 4 of the IPCC WGI reference regions and it builds on the previous version used in AR5¹⁴ with slight modifications to split the most climatically heterogeneous regions, and with the addition of 16 new oceanic regions. The polygons defining the regions are provided in the *reference-regions* folder as CSV files containing corner coordinates and also as shapefile, GeoJSON and R data files. Other sets of regions used in the

Atlas are also provided. Namely, IPCC WGII continental regions, 6 land regions subject to monsoons, 28 major river basins, 9 small island regions and another set focused on the analysis of ocean biomes.

The Interactive Atlas Dataset (IAD) consists of a set of 25 climate variables and derived indices used in the Interactive Atlas¹⁵ computed over common grids from several observational datasets plus CMIP5, CMIP6 and CORDEX historical and future projections (under different scenarios) aggregated at a monthly temporal resolution. This dataset will be stored by the IPCC Data Distribution Center (IPCC-DDC; <https://ipcc-data.org>) for long-term archival and is not available as part of the Atlas repository. However, the repository provides the scripts used for the preparation of the IAD (*datasets-interactive-atlas* folder), retrieving the data sources from ESGF, computing derived indices, post-processing (including bias adjustment when necessary) and, finally, interpolating each dataset to its reference grid. These scripts rely on the *climate4R* framework¹⁶, version 2.5.3, an open R framework for climate data access and processing. Only interpolation was delegated to the Climate Data Operators (CDO¹⁷), which provides efficient, well-tested, first-order conservative remapping routines, commonly used in climate science.

One of the key resources of the Atlas repository is the direct availability of climate datasets aggregated regionally over the new AR6 WGI reference regions¹³ (*datasets-aggregated-regionally* folder). This dataset consists of monthly precipitation and near-surface temperature from the IAD spatially averaged over the reference regions for CMIP5, CMIP6 and CORDEX model output datasets, separately for land, sea, and all (land-sea) grid cells in the region. Regional averages are area-weighted by the cosine of latitude in all cases. This product extends the regional averages presented in Iturbide *et al.*¹³ by including (1) data from CORDEX for all reference regions overlapping the area of the corresponding CORDEX domain by more than 80%, (2) an extended set of CMIP6 models, and (3) for reference, an observation-based product (W5E5¹⁸) is also provided in the same format, enabling evaluation studies. These datasets are provided in CSV format as plain data matrices (monthly time series in columns for each AR6 reference region). This format has been extended to include the corresponding metadata as a human- and machine-readable header. Moreover, a sample script (in the R language) used to compute the regional averages and produce these CSV files is also included, along with other functions facilitating their exploitation (e.g. computation of delta changes or depiction of boxplots and scatterplots).

An example of use of the regionally-averaged datasets provided in the repository are the scripts to compute time periods corresponding to different GWLs in the *warming-levels* folder. GWLs are global surface temperature anomalies at a given level (e.g. +1.5°C) with respect to the average values corresponding to the period 1850–1900 (a proxy for pre-industrial conditions)¹⁹. GWLs are widely used in policy-making as goals for greenhouse gas (GHG) concentration stabilization. In AR6, the use of GWLs has become common practice to complement scenario-based information on future climate, such as at the end of the century (2081–2100).

GWLs can be an effective means of integrating available scenario-based information, independent of when the level of global warming is reached. The warming-level folder within the Atlas repository also provides tables as CSV files and plots with the central year of 20-year periods when different GWLs (+1.5, +2, +3 and +4°C) are reached for all CMIP5 and CMIP6 Global Climate Model (GCM) simulations considered in the Atlas. Moreover, the scripts and notebooks used to generate these tables and plots are also included. They also allow the creation of new tables and plots for other GWLs or moving-window lengths, using only the repository resources. See also the new reusability example below, which computes non-overlapping decadal GWLs to produce GWL scaling plots.

Finally, there are two key resources for reproducibility and reusability of the results. The first is the *reproducibility* folder, which contains scripts to reproduce the exact figures shown in the AR6 Atlas chapter. They provide end-to-end sample code to produce the final Atlas figures, providing all processing and plotting steps in all their complexity. The second resource is the *notebooks* folder which provides simpler analyses of the data, illustrating the basic processing workflow on datasets available in the Atlas repository (either the aggregated datasets which are part of the repository or samples of the Interactive Atlas dataset). They are implemented as Jupyter notebooks²⁰, which are open-source and run on a web application which can be accessed either locally or from a remote client using any modern web browser. Jupyter supports over 40 different programming languages. Most of the sample notebooks are provided in R and some in Python. A new notebook illustrating how to re-use these data is also provided as a companion to this article (see sections on Reusability example and Code availability).

Implementation of FAIR principles. In the case of the WGI Atlas, FAIR principles are implemented through the Atlas repository described above. In particular:

Findability of the repository itself is promoted by its deployment on the widely-used GitHub platform. Additionally, Zenodo snapshots provide a unique, persistent identifier DOI. For all climate data inputs used in the Atlas, the repository lists the unique URLs (and DOIs, when available) for both data and metadata. This collection of unique identifiers is searchable, using either the automatic GitHub filter on comma-separated values (CSV) files or their version rendered as a searchable HTML table. Metadata are rich regarding both the data file content description, as well as the data provenance. The former is enforced by the CF conventions (<https://cfconventions.org>), which is the climate science standard for metadata. The latter was achieved by means of DOIs pointing to World Data Center for Climate (WDCC) entries describing model configuration and references.

Accessibility to the data and metadata in the Atlas repository is open. The full contents can be retrieved anonymously via HTTP, either from GitHub or Zenodo. All the content is licensed under a Creative Commons Attribution (CC-BY 4.0) license, except for some CORDEX datasets with non-commercial restrictions (CC-BY-NC 4.0 license). The original data and metadata sources listed are accessible through trusted repositories, in particular via authenticated HTTP connections to ESGF, the Copernicus Climate Change Service (C3S), and the Potsdam Institute for Climate Impact Research website.

Interoperability is achieved through the use of human- and machine-readable file formats such as CSV, with additional header information (metadata) to ensure coherency and joint data/metadata transfer. These files

can be read by most programming languages and software utilities. A few data tables requiring text formatting are distributed in Microsoft's Office Open XML (OOXML) format, which provides open specifications currently implemented in a number of multi-platform office suites. Other data file formats are also open (netCDF, GeoJSON, etc.) or *de facto* standards (e.g. shapefile) provided to facilitate user access. The code is also in the multi-platform, free software programming languages R and Python, widely adopted by the climate science community. Code usage examples are provided in Jupyter notebooks, which rely on an open source JSON plain text file format to combine both narrative explanations and code. They can be executed interactively by the open source Jupyter notebook server software in any modern web browser. Their static content is also directly rendered by GitHub. The rest of the files provided in the Atlas repository are also stored in open formats, namely PNG (raster images), PDF or SVG (vectorial graphics), or Markdown (formatted plain text files).

Reusability is at the core of the Atlas repository design and construct. For this reason, special care has been taken to describe the scope of the provided data, indicating the processing that has been applied to the original data source. This goes to the extent of providing the processing code along with all software versions used, as explained above. All data source versions are provided, and versioning has also been applied to the data and code produced within the repository. Most of the Atlas repository is licensed under CC-BY 4.0, which promotes maximum reusability allowed by the underlying data sources. CMIP6 participants licensing data under share-alike or non-commercial clauses granted a waiver to WGI and the IPCC-DDC to distribute CMIP6-derived report products, such as the Atlas regionally-aggregated datasets, under a CC-BY-4.0 license.

In addition to the scripts available for reproducibility, several notebooks provide reusable code snippets along with background information, explanations of the code and results, guidance to extend the analyses shown, and references to further information. A new sample notebook with additional reusable code in Python is provided as a companion to this article (see sections below on Reusability example and Code availability).

Atlas FAIR review. As an integral part of the IPCC AR6, the WGI Atlas chapter was subject to the formal IPCC review process²¹. Additionally, in order to support the integration of FAIR principles in the AR6 and enhance scrutiny of the digital products being developed in the report, the IPCC WGI Technical Support Unit (TSU) launched an informal review process of the data and codes underpinning the Atlas digital products, including the Interactive Atlas. The review was open from 25th November, 2020 until 20th December, 2020 (later extended until 10th January, 2021). This was a pilot to explore, on the one side, how to build on the rigorous review process that IPCC reports undergo and, on another side, how to incorporate open science practices of community scrutiny of digital-based material into the implementation of FAIR principles in the preparation of the Atlas. A selection of 30 reviewers were contacted for this informal review process, gathering expertise from IPCC WGI, WGII and WGIII authors, volunteers from the IPCC TG-Data, and experts from the open data community.

The objective of this review was to gather comments and recommendations about all aspects involved in the implementation of FAIR principles in the Atlas, including the use of platforms and tools (such as GitHub) for open collaboration and version control, the use of open community tools (mainly R) for data processing, the use of metadata standards and model provenance for reproducibility, and the use of Jupyter notebooks to facilitate reusability and provide user guidance on technical aspects. Feedback was also sought on how to improve the repository and the review process, considering how such a review could be extended to all report chapters.

The review was conducted directly on the Atlas repository using the GitHub issue system to collect the reviewers' comments and reply to them. The guidance provided to reviewers suggested addressing specific questions on key aspects (Table 1). A total of 34 issues were opened as a result of this process. The issues are currently closed, but still accessible through GitHub under the "review" label. This exchange is public and a snapshot of the repository related to the review is accessible through the GitHub tag "v1.6-review".

The main weak points unveiled by this review, that were addressed in the final version of the Atlas, as described in the previous Sections, were:

- Insufficient content description: README files were missing or needed to be extended to provide additional information, such as external references for the methodologies applied in the repository, and detailed description of the repository structure in the root README file, which is crucial to guide users.
- Lacking details on the description of the interactive execution environment and execution problems: information on the software versions and the platform requirements were not explicitly included, and the Atlas Hub at some point was not accessible, which was finally deployed through MyBinder.
- Unnecessary content: obscure content of the repository that was not necessary was removed.
- Code cleaning: repetition to be removed, additional explanation comments and headers to be included, and unnecessary and disrupting output elements, such as long warning messages, to be cut down in the notebooks.
- File formats: proprietary formats (e.g. shapefiles) needed to be replaced with alternative open formats.
- Lacking metadata and processing steps: sources of CORDEX and CMIP5 datasets were lacking unique version identifiers, and in the header of CSV files, information was missing to fully describe their contents. Some steps in the data processing, such as the interpolation software and version, were not included.
- Lacking observational data: suggestion was to include also the observational data in the repository, which would be useful for many users to evaluate model results, especially if regridded to the common grid.
- Notebook location and description: it was suggested that notebooks previously scattered across the repository to be gathered in one common location which would be easily accessed by users, and to include sections with more guidance for users to be able to extend their work beyond the examples available in the repository.
- Alternative programming language: As most of the code and notebooks are given in the R programming language, their adaptation to alternative programming languages such as Python would be an added value.

Technology	Do you find the used community tools appropriate/useful (in particular GitHub and R)?
Structure	Do you find the structure of the repository appropriate in terms of navigation and functionality?
Scope	Is the scope of the repository (datasets, reference regions and reference periods – including warming levels) clearly described?
Guidance	Does the Atlas repository describe clearly the structure and use? What information is missing and what improvements could be made?
Provenance	Do you find the provenance and intermediate information provided (including, e.g. model metadata and warming levels) suitable?
Usefulness	Do you find this repository useful as a resource to the scientific community?
Reusability	Do you find the R scripts and notebooks well documented and easy to follow (in particular for code reusability)?
Notebooks	Do you find the auxiliary Jupyter notebooks informative and useful?
License	Do you find the license for the code appropriate?
Development	Do you have ideas for further development of the repository?

Table 1. Questions and key aspects to be addressed that were suggested in the revision process of the Atlas repository.

- **Licensing and authorship:** unclear and mixed licenses needed to be better specified and organized. The suggestion was to release the scripts and notebooks under a Creative Commons license (as the data), instead of the specific General Public License (GPL) software license originally set. Furthermore, the author information was also requested in scripts and notebooks, to be able to comply with the attribution license.

Overall, this informal review was a very successful and enriching experience that has provided us with a broad perspective from a wider community allowing us to greatly improve the Atlas repository in all aspects of the FAIR principles, improving also the usability and user experience. Open code helps to build community trust, and the review process hopefully translates to a more robust and thorough product. The lessons learned from this process, documented in this paper, constitute valuable information for the community.

Finally, we want to remark that the review did not include a technical validation of the content (code review). This was done internally by performing basic tests and double checking key results with independent workflows. External code review would be facilitated in future work by including and documenting an automated test suite (unit and integration tests).

Errata and further development. When the WGI AR6 report was published on August 9th 2021, a final version of the repository was published with the tag “v2.0-final” and a Zenodo snapshot was produced for long-term archival⁶. No further development will take place on this repository since the report content is now frozen with its publication. This means that modifications to the existing resources will take place only for maintenance and to correct for any errata, and all changes will be fully documented.

Errata in IPCC reports, such as for chapter figures, need to be documented according to the IPCC Error Protocol. Following the experience of the review described above, error or problem reporting for the Atlas repository related to code and data underpinning figures in the Atlas has been implemented using the GitHub issue tracking system. Moreover, since the Atlas repository supports the implementation of FAIR principles for the Interactive Atlas, the same error or problem reporting system was extended to cover issues reported for this WGI product. This is implemented including a special template with instructions which is linked at the Interactive Atlas landing page. A special tag (“IA”) is automatically assigned to these issues. All issues reported are managed by the Atlas support team and classified and labeled as either *errata*, *problem*, *suggestion*, *question* including relevant comments and actions taken (if any). The label *errata* is assigned only to problems affecting the published IPCC reports (e.g. Atlas figures and Interactive Atlas maps and graphics) that also need to be reported to IPCC following the official errata reporting procedure²².

The Atlas GitHub repository will remain active with basic maintenance and improvement activities, fixing and/or documenting the problems reported (with full tracking on the actions taken) and enhancing reusability with expanded notebooks and examples. These will always be linked to an issue filed in GitHub, explaining the reason to modify the repository and keeping track of these minor changes. The idea is to provide basic maintenance, rather than further develop a repository which is considered final.

Further development of the underlying open tools supporting the Atlas code, in particular the *climate4R* framework for climate data access and post-processing¹⁶, will take place in the original repositories. New data, metadata and sample scripts or notebooks exploiting them can be contributed to this alternative repository, which is also hosted by GitHub.

Reusability example: GWL scaling plots. In this section, we provide a working example on the use of the Atlas repository resources (code and data) to reproduce and extend some of the results provided by the Interactive Atlas. In particular, we focus on the *Global Warming Level (GWL) scaling plots*, a variant of the empirical scaling relationship, that can be used to quantify the sensitivity of regional changes in different impact-relevant indices as a function of global warming²³. The use of GWLs as an analysis dimension for climate change studies has become key in the last IPCC report as a means to overcome the uncertainties inherent to future socio-economic pathways. GWLs allow focusing on the regional impacts for a given level of global warming, regardless of when or how this

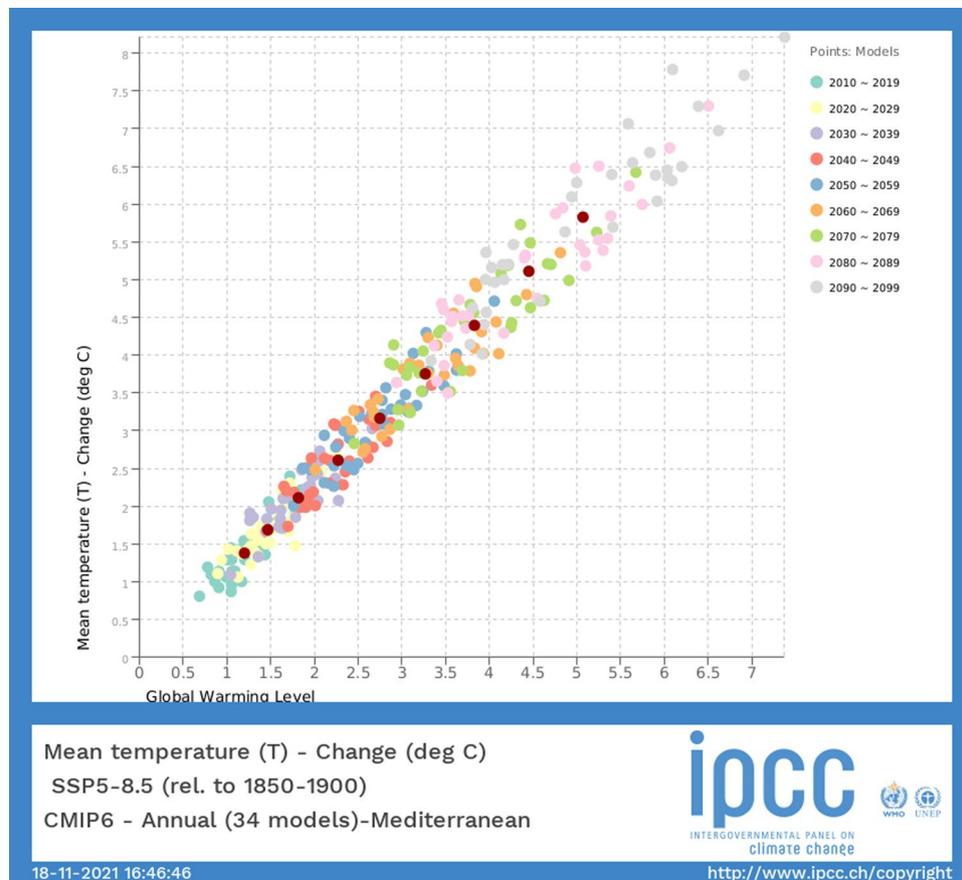


Fig. 2 Near surface temperature CMIP6 GWL scaling plot for the Mediterranean region (MED). Source: IPCC WGI AR6 Interactive Atlas¹⁵.

warming level will be reached. Despite its potential limitations, this approach has been widely adopted by the climate change impacts and adaptation communities²⁴.

The Interactive Atlas includes the possibility to analyze regional information using GWL plots (Fig. 2). However, these plots were not included in the Atlas chapter nor in the repository and constitute a good example of reusability, since they can be easily produced building on the Atlas repository resources. The analysis in this section has been coded as a new annotated notebook in Python, which can be executed using the software environment provided in the repository. To restrict our analyses to the data available in the Atlas repository, we will focus on regional mean temperature and precipitation. For a given ensemble of climate model projections, the GWL scaling plot represents the decadal mean signals of the regional variable/index of interest versus the corresponding level of global warming, decade by decade (e.g. 2020–2029, 2030–2039, etc.) and ensemble member by member. This visual representation is included in the Interactive Atlas for the analysis of regional information; an example using CMIP6 is provided in Fig. 2 displaying regional temperature increases (relative to the 1851–1900 period) in the Mediterranean as a function of the corresponding global warming (for the SSP5-8.5 scenario). In principle, this relationship could be non-linear but, in practice, most variables are shown to linearly scale with global warming in most of the regions and robustly across scenarios²³. Ensemble averages are also shown in this plot for each decade as overlaid dark red dots. These allow us to easily read the evolving mean ensemble global warming and regional response from the corresponding axes.

For this exercise, we recreated the Interactive Atlas GWL plot using Python code (Fig. 3) to exploit the resources provided by the Atlas repository. Moreover, we extend the plot provided by the Interactive Atlas by adding the linear scaling factor (β) and the coefficient of determination (R^2), as a measure of the linearity of the relationship. These are derived from a standard least squares fit, also shown in the new figures as a solid line. Finally, the line representing a unit scaling factor is added for reference (dashed line). Figure 3 compares mean temperature GWL plots for CMIP6 and CMIP5 for 3 different regions. We provide a new notebook with the code and guidance to reproduce these figures and to reuse the code for other variables (e.g. precipitation), regions and model ensembles.

Overall, Fig. 3 shows similar scaling properties (similar slope β) between CMIP5 and CMIP6. The main difference being the known²⁵ tendency of CMIP6 towards higher climate sensitivity. For the last decade of the century, CMIP6 models reach global surface air temperatures beyond 7°C, while CMIP5 remains below 6°C. Also, for some regions such as northeastern North America (NEN) there is a difference in multi-model spread between the two CMIP ensembles, affecting the goodness of fit for the linear relationship (R^2). Finally, northern Europe (NEU), seems to be prone to outlier models in the regional response, which could be due to different

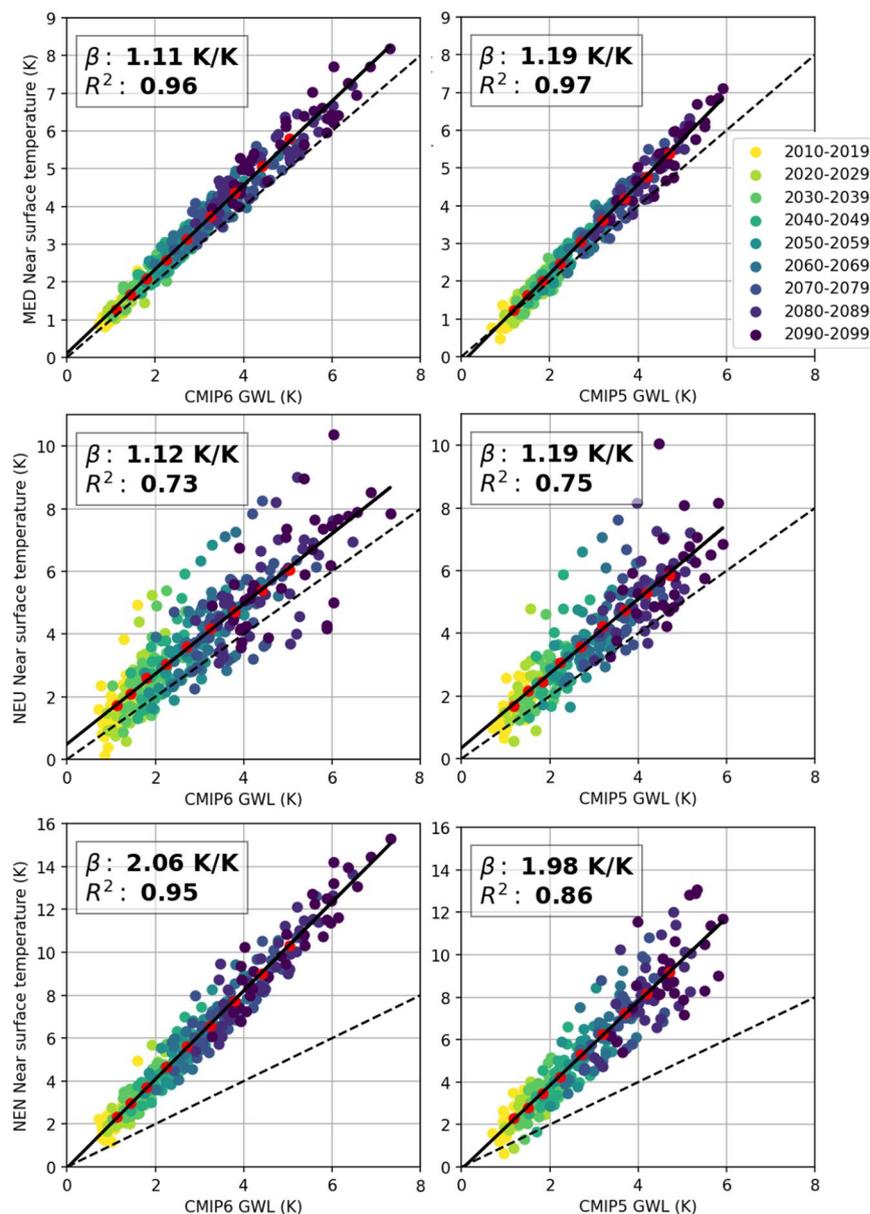


Fig. 3 Near surface temperature GWL scaling plots for (top) the Mediterranean, (middle) Northern Europe, and (bottom) Northeastern Northamerica regions for CMIP6 (left) and CMIP5 (right) ensemble projections (SSP5-8.5 and RCP8.5 scenarios, respectively). Solid and dashed lines correspond to the regression line and the reference of a unit scale factor, respectively.

sea ice extent. The Interactive Atlas can be used to easily identify those outliers, EC-Earth3-Veg and NESM3 for CMIP6, and IPSL-CM5B-LR for CMIP5, since it displays on hover the model corresponding to each point in the scatter diagram.

Discussion

The Atlas repository is a test case developed in the framework of WGI for the comprehensive implementation of FAIR principles into the IPCC report preparation process. This example, together with recommendations issued by the WGI TSU on best practices, has raised awareness amongst authors and the IPCC more widely on the value of adopting FAIR principles within the IPCC. Authors of other WGI chapters have also followed the guidance to varying degrees and the outcomes are available on the WGI GitHub repository, with more resources becoming available over time. This implementation experience has helped to refine guidance to support FAIR principles in the IPCC context, develop standard documentation templates and FAIR practices as part of the technical support provided by the TSU to authors.

FAIR principles should be considered as an integral part of future IPCC reports and the efforts required to achieve this should not be underestimated, with adequate human and technical resources allocated to support this task in all chapters, within both the TSU, the IPCC-DDC and the chapter author teams. Transparency and reproducibility could be favored by supporting and promoting common infrastructures and resources as part

of the preparation of an IPCC report. Promising examples are the new workspace resources provided for the first time to IPCC AR6 authors by the IPCC-DDC to undertake server-side analyses of global and regional climate datasets, bringing analysis closer to data². In addition, leading scientific journals are increasingly requiring authors to make their data and code public, and the IPCC could channel these efforts to be at the forefront of scientific transparency. For instance, for ancillary packages that implement specific tasks that do not reveal confidential aspects of the assessment, IPCC could promote publication of code in open journals that include checklist-driven review (including FAIRness); this would be aligned with the publication of scientific results which constitutes the basis of the assessment reports.

A further recommendation is to consolidate the review procedure for code and interactive products that is aligned with the application of FAIR principles and with the IPCC review process. The formal IPCC review process is designed for written reports and it is a challenge to implement it for code and interactive products. Experts with relevant data and software development knowledge together with expertise in climate science applications need to be involved. Code review could be done in parallel with the formal review of the chapters, but using modern software review tools such as those used in AR6 WGI (GitHub/GitLab), and involving software development communities, such as Pangeo (<https://pangeo.io>), in the review process. Quality assurance, including technical validation of code, should be carefully planned early in the process, including automated testing, with both unitary and integration tests, implemented as a key component of a continuous integration workflow. The use of well-established data processing frameworks, e.g. climate4R¹⁶ or ESMValTool²⁶, could help in this process and help simplify the code in those tasks for which solutions already exist.

Methods

The application of FAIR principles to the Atlas repository of the WGI Sixth Assessment Report (AR6) was conducted developing a GitHub repository including information about data provenance (with links/DOIs for findability of the complex datasets used), intermediate datasets (for reproducibility of key results) and code/notebooks (for reusability). An informal review was conducted to gather expertise from IPCC WGI, WGII and WGIII authors, volunteers from the IPCC TG-Data, and experts from the open data community. They used Table 1 as guidance, but were otherwise free to openly comment on the GitHub repository issue tracking system. The full discussion, leading to the current form of the repository, is available on GitHub.

Custom code to re-use the Atlas repository resources as shown in the reusability example above is available on Zenodo²⁷ (see also the Code Availability section).

Data availability

All datasets in the Atlas repository are available at <https://github.com/IPCC-WGI/Atlas> and on Zenodo⁶ under the Creative Commons Attribution license, CC-BY 4.0 (with the exception of the aggregated data corresponding to a few CORDEX simulations which are distributed under CC-BY-NC 4.0).

Code availability

All code and notebooks in the Atlas repository are available at <https://github.com/IPCC-WGI/Atlas> and on Zenodo⁶ under the Creative Commons Attribution license, CC-BY 4.0. The new notebook reusing the Atlas repository dataset to produce GWL plots (Fig. 3) is available at https://github.com/SantanderMetGroup/2022_Isturide_FAIRprinciplesIPCC and also on Zenodo²⁷.

Received: 16 March 2022; Accepted: 5 October 2022;

Published online: 15 October 2022

References

- Baker, M. Reproducibility: Seek out stronger science. *Nature* **537**, 703–704 (2016).
- Pirani, A. *et al.* The implementation of FAIR data principles in the IPCC AR6 assessment process. Task Group on Data Support for Climate Change Assessments (TG-Data) guidance document. *Zenodo* <https://doi.org/10.5281/zenodo.6504468> (2022).
- Wilkinson, M. D. *et al.* The FAIR Guiding Principles for scientific data management and stewardship. *Sci. Data* **3**, 160018 (2016).
- Stall, S. *et al.* Advancing FAIR Data in Earth, Space, and Environmental Science. *Eos* **99** (2018).
- IPCC. IPCC Data Distribution Center catalogue. *Discover and Understand Climate Change Data*, https://ipcc-data.org/data_catalogue.html (2021).
- Isturide, M. *et al.* Repository supporting the implementation of FAIR principles in the IPCC-WGI Atlas *Zenodo* <https://doi.org/10.5281/zenodo.5171760> (2021).
- Jupyter, P. *et al.* Binder 2.0 - Reproducible, interactive, sharable environments for science at scale. *Proc. 17th Python Sci. Conf.* 113–120, <https://doi.org/10.25080/Majora-4af1f417-011> (2018).
- Meng, H. & Thain, D. Facilitating the Reproducibility of Scientific Workflows with Execution Environment Specifications. *Procedia Comput. Sci.* **108**, 705–714 (2017).
- Merkel, D. Docker: lightweight linux containers for consistent development and deployment. *Linux J.* **2014**, 2 (2014).
- Williams, D. N. *et al.* The Earth System Grid: Enabling Access to Multimodel Climate Simulation Data. *Bull. Am. Meteorol. Soc.* **90**, 195–206 (2009).
- Stockhause, M. & Lautenschlager, M. CMIP6 Data Citation of Evolving Data. *Data Sci. J.* **16**, 30 (2017).
- Sun, S., Lannom, L. & Boesch, B. Handle System Overview. <https://www.hjp.at/doc/rfc/rfc3650.html> (2003).
- Isturide, M. *et al.* An update of IPCC climate reference regions for subcontinental analysis of climate model data: definition and aggregated datasets. *Earth Syst. Sci. Data* **12**, 2959–2970 (2020).
- Christensen, J.H. & Kanikicharla, K.K. IPCC AR5 reference regions. *Centre for Environmental Data Analysis* <https://catalogue.ceda.ac.uk/uuid/a3b6d7f93e5c4ea986f3622e002b96f> (2021).
- Gutiérrez, J. M. *et al.* Atlas (Available at <https://interactive-atlas.ipcc.ch>). 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* [Masson-Delmotte, V., P. Zhai, A. Pirani, S.L. Connors, C. Péan, S. Berger, N. Caud, Y. Chen, L. Goldfarb, M.I. Gomis, M. Huang, K. Leitzell, E. Lonnoy, J.B.R. Matthews, T.K. Maycock, T. Waterfield, O. Yelekçi, R. Yu, and B. Zhou (eds.)] (2021).

16. Iturbide, M. *et al.* The R-based climate4R open framework for reproducible climate data access and post-processing. *Environ. Model. Softw.* **111**, 42–54 (2019).
17. Schulzweida, U. Climate Data Operators (CDO) User Guide. *Zenodo* <https://doi.org/10.5281/zenodo.3539275> (2019).
18. Lange, S. WFDE5 over land merged with ERA5 over the ocean (W5E5). *GFZ Data Services* <https://doi.org/10.5880/PIK.2019.023> (2019).
19. Vautard, R. *et al.* The European climate under a 2 °C global warming. *Environ. Res. Lett.* **9**, 034006 (2014).
20. Kluyver, T. *et al.* Jupyter Notebooks – a publishing format for reproducible computational workflows. in *Positioning and Power in Academic Publishing: Players, Agents and Agendas* (eds. Loizides, F. & Schmidt, B.) 87–90, <https://doi.org/10.3233/978-1-61499-649-1-87> (IOS Press, 2016).
21. IPCC. IPCC Factsheet: How does the IPCC review process work? https://www.ipcc.ch/site/assets/uploads/2018/02/FS_review_process.pdf (2015).
22. IPCC. Errata. <https://www.ipcc.ch/report/ar6/wg1/#errata> (2021).
23. Seneviratne, S. I. & Hauser, M. Regional Climate Sensitivity of Climate Extremes in CMIP6 Versus CMIP5 Multimodel Ensembles. *Earths Future* **8**, e2019EF001474 (2020).
24. IPCC. *Climate Change 2022: Impacts, Adaptation, and Vulnerability. Contribution of Working Group II to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change* [H.-O. Pörtner, *et al.* (eds.)]. vol. In Press (Cambridge University Press, 2022).
25. Zelinka, M. D. *et al.* Causes of Higher Climate Sensitivity in CMIP6 Models. *Geophys. Res. Lett.* **47**, e2019GL085782 (2020).
26. Eyring, V. *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).
27. Fernandez, J. Repository supporting the results presented in the manuscript on Implementation of FAIR principles in the IPCC: The WGI AR6 Atlas repository. *Zenodo* <https://doi.org/10.5281/zenodo.7102034> (2022).

Acknowledgements

We acknowledge partial funding from projects ATLAS (PID2019-111481RB-I00) funded by MCIN/AEI/10.13039/501100011033 and IS-ENES3 which is funded by the European Union's H2020 programme under grant agreement No 824084. We also acknowledge the World Climate Research Programme's Working Group on Coupled Modelling and Working Group on Regional Climate, responsible for CMIP and CORDEX, respectively. We also thank the climate modeling groups for producing and making available their model output, as described in the data-source folder of the repository. We also acknowledge the Earth System Grid Federation infrastructure, an international effort led by the U.S. Department of Energy's Program for Climate Model Diagnosis and Intercomparison, the European Network for Earth System Modelling and other partners in the Global Organisation for Earth System Science Portals (GO-ESSP). The opinions expressed are those of the author(s) only and should not be considered as representative of the European Commission's official position. JF and ASC acknowledge support from the CORDyS project (PID2020-116595RB-I00) funded by MCIN/AEI/10.13039/501100011033. JM acknowledges support from MDM-2017-0765 funded by MCIN/AEI/10.13039/501100011033. JBM acknowledges support from Universidad de Cantabria and Consejería de Universidades, Igualdad, Cultura y Deporte del Gobierno de Cantabria via the project “instrumentación y ciencia de datos para sondear la naturaleza del universo”. Finally we want to thank the reviewers participating in the Atlas FAIR review described in the paper and the editor and the two anonymous referees for their work and constructive comments, helping us to improve the manuscript.

Author contributions

Anna Pirani, Özge Yelekçi, José M. Gutiérrez, Alaa Al Khouardjia and David Huard contributed to the framing of the FAIR principles in the context of the Technical Support Units and Task Group on Data Support for Climate Change Assessments (TG-Data) and the definition of the review protocol. Maialen Iturbide, Jesús Fernández, José M. Gutiérrez, Joaquin Bedia, Ezequiel Cimadevilla, Javier Diez-Sierra, Rodrigo Manzananas, Ana Casanueva, Jorge Baño-Medina, Josipa Milovac, Sixto Herrera, Antonio S. Cofiño, Daniel San-Martín and Markel García-Diez have contributed to the creation of the contents of the Atlas repository. Alaa Al Khouardjia, Matteo De Felice, Javier Diez-Sierra, Jesús Fernández, James Goldie, Dimitris A. Herrera, Rodrigo Manzananas, Aparna Radhakrishnan, Alessandro Spinuso, Kristen Thyng and Claire M. Trenham reviewed and shaped the final version of the Atlas repository. Jesús Fernández created the figures. Maialen Iturbide, Jesús Fernández and José M. Gutiérrez wrote the first draft. All authors contributed to the final writing and revision of the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to J.F. or J.M.G.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022