Expanding Integrated Assessment Modelling: Comprehensive and Comprehensible Science for Sustainable, Co-Created Climate Action

D3.6 – OPEN SCIENCE PROTOCOLS

WP3 – Exchanging – Open & FAIR science, mutual learning

28/06/2023

Funded by the European Union

www.iam-compact.eu

Disclaimer

Copyright Message

Grant Agreement Number	101056306		**Acronym**	IAM COMPACT
Full Title	Expanding Integrated Assessment Modelling: Comprehensive and Comprehensible Science for Sustainable, Co-Created Climate Action			
Topic	HORIZON-CL5-2021-D1-01-04			
Funding scheme	HORIZON EUROPE, RIA – Research and Innovation Action			
Start Date	September 2022	**Duration**	36 Months	
Project URL	https://www.iam-compact.eu			
EU Project Advisor	Andreas Palialexis			
Project Coordinator	National Technical University of Athens – NTUA			
Deliverable	D3.6 – Open science protocols			
Work Package	WP3 – Exchanging – Open & FAIR science, mutual learning			
Date of Delivery	**Contractual**	30/06/2023	**Actual**	28/06/2023
Nature	Report	**Dissemination Level**	*Public*	
Lead Beneficiary	ASOCIACION BC3 BASQUE CENTRE FOR CLIMATE CHANGE - KLIMA ALDAKETA IKERGAI (BC3)			
Responsible Author	Clàudia Rodés-Bachs	**Email**	claudia.rodes@bc3research.org	
	BC3	**Phone**	+34 94 401 46 90	
Contributors	Jon Sampedro; Russell Horowitz; Dirk-Jan Van de Ven (BC3)			
Reviewer(s)	Shivika Mittal (Imperial); Konstantinos Koasidis (NTUA); Vassilis Stavrakas (UPRC)			
Keywords	open science; open access; FAIR principles; TRUST principles; harmonisation; output management			

Funded by
the European Union

EC Summary Requirements

1. Changes with respect to the DoA

No changes with respect to the work described in the DoA.

2. Dissemination and uptake

This report will be used internally by all consortium partners to guide all open science practices of the project. It can be also used by users outside consortium as an example for informing their own strategies (especially for other Horizon Europe projects).

3. Short summary of results (<250 words)

This report documents the first version of the Open Science Protocols of IAM COMPACT. First, it includes a detailed description of the state-of-art of open science practices along with a presentation of the FAIR and TRUST principles, highlighting their novelties and limitations while exemplifying their usages. Second, it describes the required infrastructure to facilitate the application of open science principles by enabling code sharing, data storage, and user-friendly documentation. Finally, the report contains a protocol for the consortium members to facilitate the integration of the FAIR and TRUST principles, to promote the smooth interconnection of the models, and to transparently manage the produced outcomes.

4. Evidence of accomplishment

This report.

Preface

IAM COMPACT supports the assessment of global climate goals, progress, and feasibility space, and the design of the next round of Nationally Determined Contributions (NDCs) and policy planning beyond 2030 for major emitters and non-high-income countries. It uses a diverse ensemble of models, tools, and insights from social and political sciences and operations research, integrating bodies of knowledge to co-create the research process and enhance transparency, robustness, and policy relevance. It explores the role of structural changes in major emitting sectors and of political, behaviour, and social aspects in mitigation, quantifies factors promoting or hindering climate neutrality, and accounts for extreme scenarios, to deliver a range of global and national pathways that are environmentally effective, viable, feasible, and desirable. In doing so, it fully accounts for COVID-19 impacts and recovery strategies and aligns climate action with broader sustainability goals, while developing technical capacity and promoting ownership in non-high-income countries.

NTUA – National Technical University of Athens	EL	
Aalto – Aalto Korkeakoulusaatio SR	FI	
AAU – Aalborg Universitet	DK	
BC3 – Asociacion BC3 Basque Centre for Climate Change – Klima Aldaketa Ikergai	ES	
Bruegel – Bruegel AISBL	BE	
CARTIF – Fundacion CARTIF	ES	
CICERO – Cicero Senter for Klimaforskning Stiftelse	NO	
E3M – E3-Modelling AE	EL	
KTH – Kungliga Tekniska Hoegskolan	SE	
POLIMI – Politecnico di Milano	IT	
UPRC – University of Piraeus Research Center	EL	
UVa – Universidad De Valladolid	ES	
WI – Wuppertal Institut fur Klima, Umwelt, Energie GGMBH	DE	
IIMA – Indian Institute of Management	IN	
THU – Tsinghua University	CN	
USMF – University System of Maryland	US	
AAiT – Addis Ababa University	ET	
KEI – International Civic Organisation Kyiv Economics Institute	UA	
RUSL – Raja Rata University of Sri Lanka	LK	
TUM – Technical University of Mombasa	KE	
UNIGE – Université de Geneve	CH	
Imperial – Imperial College of Science, Technology and Medicine	UK	

Funded by the European Union

Executive Summary

Open science is a concept that aims to make scientific research and data available to both the scientific community and non-expert audiences. It has multiple benefits, such as increasing stakeholders' engagement, enhancing scientific collaboration, and ensuring the sustainability of the work over time, and many research fields have gradually embraced open science practices. In this context, we aim to develop an open science protocol to be followed by the different partners within the IAM COMPACT consortium during the project.

First, we present the state-of-art of the open science principles, including the FAIR principles, which aim to assess that data is Findable, Accessible, Interoperable, and Reusable. They have been successful in promoting data exchange and usage in multiple projects, but there are evident limitations to their effectiveness. Additionally, we outline the TRUST principles which focus on the Transparency, Responsibility, User focus, Sustainability, and Technology of data repositories and aim to assess digital repository trustworthiness. Using both principles, researchers can ensure that their work will be the basis of new science outcomes since multiple users will be able to access and reuse it. To apply all these open science principles, powerful and specialised infrastructure is available for researchers, with well-known tools such as GitHub, Zenodo, and OpenAire. These tools are focused on code and data documentation and archiving, ensuring full and permanent accessibility, while featuring scientific reproducibility facilities.

IAM COMPACT includes a wide range of different models and modelling teams. To that extent, the complete transparency and documentation of the models are essential, guaranteeing the full accessibility, understandability, and reproducibility of the outcomes for different audiences. For that purpose, the deliverable proposes a protocol that consists of two stages: Firstly, the model and data harmonisation process, which involves documenting the models and their inputs, as well as interfaces that explain how the models are interconnected; and secondly, the open management of the project outcomes, which deals with validation procedures such as vetting, as well as results visualisation and storage.

Funded by
the European Union

Contents

1 Introduction

Open science is a concept that aims to **make scientific research and data available to everyone**, including scientists, stakeholders, and the general public. The main principle of open science is based on the idea that scientific knowledge should be freely available to everyone and should be easily accessible for reuse, replication, and further development (European Commission, 2019). It rests on a set of core values, including transparency, reproducibility, accessibility, and interdisciplinarity, to ensure that scientific research is conducted in an open and transparent manner, and that the resulting data, methodology, and **findings are made available to the broader scientific community and non-expert audiences** (Munafò et al., 2017). By making scientific research and data available to everyone, open science can help to promote scientific progress and drive innovation (Fecher et al., 2015).

The concept of open science has gained momentum in recent years, with a growing number of researchers, policymakers, and funding agencies recognising the importance of openness in scientific research (Burgelman et al., 2019; Piwowar & Vision, 2013; Tenopir et al., 2015). Researchers from a diverse range of fields agree that the lack of access to data provided by others is a key barrier to scientific advancement and encourage the academic community to embrace better open science procedures (Fecher et al., 2015). A clear example of its growing importance is the **exponential increase of the number of researchers following open science practices all over the world**: the number of users registered in the Open Science Framework (OSF) platform has been duplicated in two years, reaching 400,000 users in 2021 (Center for Open Science, 2021) and, in some platforms to store datasets such as Re3Data[1], the proportion of open data repositories has risen by 94%.

The rapid increase of open science practices in the scientific research community has been driven by different factors. First, multiple organisations and networks have been created worldwide to support the usage of open science methods, such as the Center for Open Science (COS) or the Chinese Open Science Network (COSN) (Jin et al., 2023). These organisations equip scholar communities with the necessary tools to adopt open science practices, while facilitating and steering the data management process. **There are also a number of resources and platforms available to support open science**, including open-access journals, indexed and browsable in the Directory of Open Access Journals[2], open data repositories, such as the World Bank open data[3] or the World Health Organisation data[4], and software for open-source computing and development, for instance, Unix[5] and Git[6], all of which can help to ensure that clear and accessible research methods are used. In addition, a number of initiatives and policies have been developed. In 2015, the first "Transparency and Openness Promotion (TOP) guidelines for journals" was published (Nosek et al., 2015). Since then, multiple initiatives have been launched, for example, the European Union's Horizon 2020 program (European Commission, 2019) includes a policy that requires all research funded by the program to be open access (European Commission, 2017; Jahn et al., 2023) and tracks open data and open access publications[7]. The European Commission also created the EU Open Science Policy Platform (OSPP) as a high-level advisory group to advise the European Commission on how to develop its Open Science Policy and support policy implementation by reviewing best practices, drawing policy guidelines, and encouraging their active uptake by stakeholders (Mendez et al., 2020); as well as the Open Research Europe[8], an open access publishing service to conduct open peer review and quick publishing for all types of research funded by Horizon 2020, Horizon Europe, and Euratom. Similarly, the US National Institutes of Health (NIH) have implemented a policy that requires all research funded by NIH to be open access (National Institutes of Health, 2021). Moreover, several guidelines have been developed to help repository managers to increase the exposure of publications, datasets, and metadata following open access protocols (OpenAire, 2022).

[1] https://www.re3data.org/
[2] https://doaj.org/
[3] https://data.worldbank.org/
[4] https://www.who.int/data/gho/
[5] https://unix.org/
[6] https://github.com/
[7] https://research-and-innovation.ec.europa.eu/strategy/strategy-2020-2024/our-digital-future/open-science/open-science-monitor_en
[8] https://open-research-europe.ec.europa.eu/

There is strong evidence that **open science practices have multiple benefits**. First, they improve research efficiency and reduce total costs through multiple factors such as labour cost savings, productivity improvements, access cost savings, or duplication avoidance. Importantly, they enable broad impactful activities by using new products, services, companies, and collaborations, which would have been less likely to occur in closed environments and allow work that could not otherwise be undertaken (Fell, 2019). All these benefits **increase public engagement by making research more accessible and understandable to non-experts**, which expands reliability and funding opportunities. A representative example of the benefits of open science practices is the scientific research during COVID-19 pandemic. In order to boost the scientific collaboration to fight the pandemic, several publishers and researchers sped up their adoption of open science practices (Besançon et al., 2021; Hayashi, 2021). Reviewing processes were accelerated and preprints were shorter and more quickly available (Fraser et al., 2020). In the same way, data —especially COVID-19 related— was shared among different institutions and countries, which enhanced the transparency and cooperation between scientists. While this new research paradigm and the new sharing and publication rhythm raised concerns about the potential impact on the quality of research output (Besançon et al., 2021), it also encouraged the development of additional guidelines and specific protocols to guarantee the successful deployment of open science practices, which have been maintained in the post-pandemic era.

Nowadays, the practices of **open science have been progressively extended to all research topics** and fields of study in the Science, Technology, Engineering and Mathematics (STEM) disciplines (Fecher et al., 2015). Some examples are sustainable transport (Nielsen et al., 2023), criminology (Branney et al., 2022), psychology (Chin et al., 2021), Artificial Intelligence (AI) (Brinkhaus et al., 2023), and sustainability and climate policies (Iturbide et al., 2022). To provide a relevant example for the context of the project, the Integrated Assessment Modeling Consortium[9] (IAMC) community is adopting open science practices to increase the accessibility and reusability of scenarios from Integrated Assessment Models (IAMs). All models participating in IAMC activities need to be accompanied by a detailed documentation of their input parameters, assumptions, and uncertainties (IAMC, 2022), which is then hosted in open-access websites such as the IAMC documentation platform and the I²AM PARIS[10] data exchange platform (Nikas et al., 2021). Furthermore, in order to facilitate data analysis, comprehension, and the interactive visualisation of model outputs, various tools have been created by different teams within the IAM community. These include the 1.5ºC scenario explorer (Huppmann et al., 2018), the *pyam* package (Gidden & Huppmann, 2019), the Global Hotspots Explorer[11], and the Sixth Assessment Report (AR6) scenario explorer (Byers et al., 2022).

In line with the steps taken by the IAM community, **<u>the aim of this deliverable is to develop an open science protocol that will be followed by the different partners within the IAM COMPACT consortium during the course of the project</u>**. Section 2 provides an overview of the practices for open science, including the guiding principles and the most relevant software infrastructures. Based on this information, Section 3 details the protocols for transparency of the consortium models and modelling exercises and accessibility of the produced outputs for expert and non-expert audiences.

[9] https://www.iamconsortium.org/
[10] https://www.i2am-paris.eu/
[11] https://www.hotspots-explorer.org/

2 Open science practices in scientific research

2.1 Principles for open science

2.1.1 FAIR guidelines for open data management

With open data increasing in importance, a group of stakeholders from academia, industry, and government developed the **FAIR Principles as a measurable guideline to assess if scientific data is Findable, Accessible, Interoperable, and Reusable** (Wilkinson et al., 2016). Promoting the FAIR principles enables researchers to reuse and build upon existing knowledge, leading to new discoveries and advancements. Rather than prescribing specific technical requirements, the FAIR principles provide a set of guidelines that support a spectrum of increasing reusability through various implementations. They outline the qualities and objectives for systems and services to facilitate the production of research outputs that can be rigorously assessed and widely utilised with proper credit, **benefiting both the author and the user** (Mons et al., 2017). The FAIR principles have been progressively adopted by a number of organisations, including the European Commission and the NIH among others.

Different data maturity assessment approaches have been developed to assess the FAIRness of research data and provide **practical solutions to facilitate the application of the FAIR principles to research repositories** (Bahim et al., 2020; RDA FAIR Data Maturity Model Working Group, 2020). For instance, the FAIRsFAIR Data Object Assessment Metrics (Anusuriya Devaraju et al., 2022) introduces a protocol to standardise data and metadata, which includes descriptive core elements to both support data findability and define persistent data identifiers. Moreover, they include examples on how to implement FAIR principles in different research domains, encouraging researchers from different disciplines to adopt them. Other examples are the FAIR data self-assessment tool (Kerry Levett et al., 2022), which evaluates "how FAIR" a research dataset is and suggests practical tips on how to enhance its "FAIRness", and the FAIR Data Maturity Model, which aims to harmonise FAIR assessments (Bahim et al., 2020). In the same line, the IPCC's Working Group I (WGI) applied the FAIR principles to the Atlas engine in the recently published AR6 (Masson-Delmotte et al., 2021), which offers a region-by-region temporal and spatial assessment of climate change effects, as well as an interactive tool to enable users to carry out flexible analysis with regional data. Based on these principles, the Atlas engine includes an open-source repository with code, data provenance, aggregated datasets, and documentation about the interactive user interface (Gutiérrez et al., 2021; Iturbide et al., 2022). Furthermore, the whole process and methodology to meet the FAIR principles were both documented and published, aiming to help other researchers that want to adhere to open science (Iturbide et al., 2022).

The core aspects of the FAIR guidelines are detailed in Box 1.

Box 1: FAIR principles for open science

Findability: Findability ensures that data can be easily located and accessed by interested parties. In order for data to be findable, it should be assigned a unique and persistent identifier (PID). A PID is a standardised identifier that provides a persistent reference to the data, even if its location or access method changes over time. PIDs are a critical component of findability, as they enable data to be discovered and accessed by anyone who has the identifier, regardless of where the data is stored (Jacobsen et al., 2020). Although a PID is necessary to ensure findability, it is not sufficient. Extensive machine-actionable metadata are essential for the automatic discovery of relevant datasets and services, and are therefore an essential component of the "FAIRification" process (Jacobsen et al., 2020).

Accessibility: Accessibility guarantees that scientific data is freely and easily available to researchers who need it. To achieve accessibility, scientific data should be stored in open repositories without any unnecessary restrictions or barriers. Moreover, the data should be accessible not only by human beings but also by computers, as in many cases they need to access a data publication autonomously, unaided by a human operator (Mons et al., 2017). This means that data should be made available in open formats, using standard protocols and interfaces that are widely supported by different tools and software. Additionally, repositories should provide clear and comprehensive documentation about the data they host, including information about their quality, provenance, and access conditions. In addition, metadata should be accessible, even when the data are no longer available. However, accessible does not necessarily mean open. In fact, accessible stands for "accessible under well-defined conditions", which legitimatises reasons to shield data for personal privacy or national security and assuring the proper data protection (Mons et al., 2017).

Interoperability: Interoperability ensures that scientific data can be integrated and analysed across different systems and tools seamlessly. To achieve interoperability, scientific data should be structured using open and standardised formats that can be easily understood and used by different software and applications. This helps ensuring that data can be shared and reused across different disciplines and domains, facilitating collaboration and interdisciplinary research. By using standardised formats and protocols, data can be easily integrated into analysis workflows, enabling researchers to extract meaningful insights and knowledge from diverse datasets (Ravi et al., 2022). To do so, different software packages communicate with each other by exchanging data and metadata through application programming interfaces (APIs), while meeting domain-relevant community standards and including qualifying references to other objects (Barker et al., 2022). Consequently, metadata should be structured following the FAIR principles, including qualified references to other data, metadata, or publications (Campos et al., 2023).

Reusability: Reusability aims to maximise the value and impact of scientific data by enabling their reuse for different purposes. To achieve reusability, scientific data should be well-documented, organised, and annotated in a way that makes it easy to understand and use (da Silva Santos et al., 2023). This includes providing detailed information about the data's provenance, quality, and access conditions, as well as adopting clear and standardised data structures and formats that are widely understood and accepted. By making data reusable over time, researchers can build upon existing knowledge, generate new insights, and accelerate the pace of scientific discovery (Jacobsen et al., 2020).

The FAIR principles have been successful in promoting data exchange and usage in multiple projects. Different scientific disciplines have seen the necessity of the FAIR principles, especially *reproducibility* (Gibney, 2022). This is the case in the AI research area, where reproducibility hinderance threatens the reliability of Machine Learning (ML) results (Haibe-Kains et al., 2020), particularly in prominent applications such as health care diagnosis (Sohn, 2023) or natural language processing (NLP) (Cohen et al., 2018).

However, there are limitations to their effectiveness and although they have gained widespread acceptance in the scientific community, they may not be enough to fully address the challenges of data management and sharing in the scientific enterprise (Boeckhout et al., 2018; Stodden et al., 2018; Tenopir et al., 2011). Some limitations of the FAIR principles are their focus on individual practices and on providing a binary measure of the final project state of whether the principle is satisfied or not (Peng et al., 2022). Moreover, FAIR principles focus mostly on short-term data management and **less on how data are managed and preserved for the long-term**, or **how to address the incentives and motivations that drive researchers to share their data**, or the cultural and social dimensions of data sharing, as thereis high evidence that nationality, age, seniority, and researcher working dynamics play an important role in the decision to share data (Enke et al., 2012; Tenopir et al., 2011). Finally, **FAIR principles do not address the technical aspects of *how to share data***, which leads to unstandardised methods to generate, combine, and access datasets and metadata. Without proper schemes and roadmaps, researchers usually need more time and skills to make their data available, which discourages them to do so (Hughes et al., 2023). In fact, consistent methods to preview and access metadata and data within repositories are not widely available. The standard Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) (Lagoze, 2002; Lagoze et al., 2001) has been adopted by generalist repositories, but their implementations vary per repository and the resulting metadata formats differ. Downloading methods are similarly diverse, pushing researchers to require institutional approvals and data usage agreements before obtaining the data (Hughes et al., 2023).

In conclusion, **the FAIR principles have been successful in promoting data sharing and standardising practices, as well as to emphasise machine-actionability by facilitating data reliance and reusability. Despite all these improvements, they may not be enough to fully address the challenges of data management and exchange in the scientific community** (Boeckhout et al., 2018). The adoption of further principles in data sharing and the provision of support from more organisations and institutions might help enhancing open science practices among researchers.

2.1.2 TRUST principles for digital repositories

The **TRUST principles provide a set of guidelines to demonstrate digital repository trustworthiness**, as well as a common framework to facilitate discussion and implementation of best practices in digital preservation (Lin et al., 2020). The rationale behind these principles is **to ensure that data not only follows the FAIR principles for management but will also be preserved over time**. The TRUST principles include Transparency, Responsibility, User focus, Sustainability, and Technology and were defined in 2020 in a collaborative way. These TRUST principles were further elaborated in 2022, when the Research Library Group and the Online Computer Library Center[12] (OCLC) pushed by the concerns of the previous decade about the potential impact of digital technologies on the future of information, published their eponymous report on Trusted Digital Repositories (RLG/OCLC Working Group on Digital Archive Attributes, 2002). This report assessed the vulnerability of digital information to undetectable alterations and deletion, and posed challenges for long-term preservation (Bak, 2016).

The core aspects of the TRUST principles are detailed in Box 2.

[12] https://www.oclc.org/research/home.html

Box 2: TRUST principles for open science

Transparency: Transparency ensures that users can easily identify and choose the most appropriate repository for their specific needs through access to comprehensive information about the repository's scope, target audience, policies, and capabilities. This level of transparency allows users to evaluate the repository's suitability for their specific requirements (Lin et al., 2020), while providing secure, persistent, and reliable services for the long-term (Delgado et al., 2021). In order to comply with this principle, repositories should provide clear statements regarding their mission, scope, and terms of use for both the repository and data holdings. In addition, repositories should declare the minimum digital preservation timeframe for their data holdings, in addition to any additional features or services, such as the ability to responsibly steward sensitive data (Lin et al., 2020).

Responsibility: Responsibility incorporates the commitment of the repository in relation to the stewardship of data holdings and to serving their user community (Lin et al., 2020). This is exemplified through various actions, including adhering to metadata and curation standards designated by the community, ensuring technical validation, documentation, quality control, authenticity protection, and long-term persistence of the data holdings, as well as providing data services such as portal and machine interfaces, data download or server-side processing. Additionally, the repository is responsible for managing the intellectual property rights of data producers, protecting sensitive information resources, and maintaining the security of the system and its content. Responsibility can be established through legal means such as the right to preserve, or through voluntary compliance with ethical standards (Crabtree, 2020).

User Focus: User focus ensures that data repositories prioritise their target user community. Different communities have different expectations and needs for their repositories depending on their level of experience with data management and sharing. A TRUSTworthy repository must be fully immersed in its user community's data practices and be able to adapt to changing community needs. One key aspect is that repositories should encourage data depositors to thoroughly describe their data and provide mechanisms for users to report any issues with the data's quality or suitability for use (Lin et al., 2020). Additionally, TRUSTworthy repositories must enforce their target user community's norms and standards to ensure data interoperability and reusability. This includes adhering to metadata schemas, data file formats, controlled vocabularies, ontologies, and other semantics commonly used in the user community. To demonstrate compliance with this principle, a TRUSTworthy repository may implement relevant data metrics and make them available to users, contribute to community catalogues to aid data discovery, and monitor and address evolving community expectations as needed (Löbe et al., 2022).

Sustainability: Sustainability strives to provide uninterrupted access to valuable data holdings for current and future user communities over time and respond to evolving user community requirements with new or improved services. To demonstrate sustainability, certain conditions must be satisfied, including planning for risk mitigation, business continuity, disaster recovery, and succession (Delgado et al., 2021). Furthermore, securing funding for ongoing usage and maintenance of the data for the necessary long-term preservation of data is crucial to ensure that data resources remain discoverable, accessible, and usable in the future (Lin et al., 2020).

Technology: Technology implies that repositories must rely on updated software, hardware, and technical services. A TRUSTworthy repository must demonstrate the fitness of its technological capabilities by implementing relevant and appropriate standards, tools, and technologies for data management and curation while having plans and mechanisms to prevent, detect, and respond to cyber or physical security threats (Lin et al., 2020). In this way, data will be reliable and preserved for the long-term.

Nowadays, repositories can become certified as Trustworthy Digital Repositories (TDRs) to "*demonstrate to both their users and their funders that an independent authority has evaluated them and endorsed their trustworthiness*" (CoreTrustSeal, 2019). A TDR's mission is **to provide reliable, long-term access to managed digital resources to its designated community, now and in the future** (RLG/OCLC Working Group on Digital Archive Attributes, 2002). However, Trustworthiness is demonstrated through evidence; it cannot be taken for granted without regular audits and certification (Lin et al., 2020). Thus, **several certification programs** such as the Data Seal of Approval (DSA) and, later, the Core Trust Seal (CTS) have emerged. Both certification programs require a self-audit report that is reviewed and approved by the standards' representatives. Each audit report describes a repository's level of compliance with a set of 16 guidelines covering a repository's background information, organisational infrastructure, digital object management, and technology, among others. The certifications must be renewed over the years, and there is strong evidence that participating repositories increase their compliance with DSA or CTS TDR standards in time (Donaldson, & Russell, 2023).

Although the CTS or DSA certifications are valuable tools, **they can also create barriers**. For instance, these certificates might **create a division between certified and non-certified data repositories**, causing the latter group to be perceived as less trustworthy, even if they meet the same standard for data management. In addition, these certificates can hinder smaller or less well-stablished repositories by placing a financial burden on them, leading only larger and mature repositories to seek certification. Finally, these certificates are based on standards that may not be applicable to all types of data, which can **make it more difficult for interdisciplinary research teams to share and collaborate on data if the CTS or DSA certificates are required** (CoreTrustSeal Standards And Certification Board, 2022; Crabtree, 2020). Nevertheless, an increasing number of repositories have adopted the CTS certificate, making clear their willingness to further professionalise and be perceived as trustworthy (Lin et al., 2020), and multiple studies have advocated for more specific policies that encourage certification and use of TDRs (Anusuriya Devaraju et al., 2022; De Wilde, 2017; Donaldson, & Russell, 2023).

Finally, we note that, in addition to the TRUST principles and the CTS standards, **efforts are underway to develop a more comprehensive framework for data management and sharing**, such as the Data Stewardship Maturity Matrix (DSMM), which provides a tool for self-assessing the maturity of data management and sharing practices in organisations and institutions. It provides a roadmap for improving these practices over time, and it is focused on the individual dataset level (Dunn et al., 2021). So far, DSMM has been only used for the Earth and Environmental Sciences, as it was developed jointly by the National Centers for Environmental Information (NCEI), the official archive of the U.S. National Oceanic and Atmospheric Administration (NOAA), and the Cooperative Institute for Climate and Satellites-North Carolina (Peng et al., 2022).

2.2 Infrastructure

To support open science, **researchers need adequate, specialised software, which should adapt to the needs of their research field and community**. Currently, there are multiple options for publishing data, code, documentation, software packages, and programming environments.

There are many existing software packages and platforms for researchers to share their research materials and source code and to collaborate with others in real-time (Crystal-Ornelas et al., 2021). This can save time and effort by eliminating the need to develop custom software infrastructure, allowing researchers to focus on their research. Moreover, using established software for open science provides a **standardised framework for sharing research materials**, making it easier for other researchers to access and understand the data. This promotes collaboration and transparency and makes it easier to verify and reproduce research results. In this line, open-source software also makes teamwork easier across organisations and allows scientists to have a larger collaborative network, as well as reduces time working with data, lowers duplication of effort, and ensures the longevity of the code (Ramachandran et al., 2021).

Leveraging existing software infrastructure can also help **amplify the spread and impact of research**. By depositing research materials on a widely used sharing platform, researchers can increase the chances of their work being discovered and cited by others. This can increase visibility and recognition for the researchers and their institutions, thereby increasing funding opportunities (Piwowar et al., 2007).

2.2.1 Document and archive code and data

One of the pillars of open science is the documentation and archiving of code and data. Currently, there are multiple platforms available for data and code archiving and curation, many of them featuring similar standards, practices and processes. Thus, it is important that researchers understand the features, standards, and practices of the different platforms and choose the best option for their particular project and community to effectively manage their data and increase the visibility of their work (Peng et al., 2022).

The **most popular platform for version control and collaboration on software development is GitHub** (Finley, 2011). The number of users in GitHub has substantially increased in recent years from 5.8 million in 2016 to 94 million in 2022. Likewise, the number of active repositories in 2022 surpassed 200 million, compared to 60 million in 2020. Researchers can use GitHub to document and archive their code, making it publicly accessible and facilitating collaboration with other researchers. GitHub offers a range of features based on the Git version control system[13], including issue tracking, pull requests, and code review, that can improve the quality of research code. Moreover, it offers social features to its users to maintain collaborative networks with event updates, users' interconnections, and auto-promotion options (Lima et al., 2014; Oliveira et al., 2023).

While tools like GitHub are designed to enable collaborative software development, there are **other popular platforms targeting at long-term access and preservation of research data, code, and publications** (Sicilia et al., 2017). One well-stablished platform is the **Zenodo**[14] online repository hosted by the European Council for Nuclear Research (CERN). Researchers can deposit their code and data in Zenodo, which assigns a unique Digital Object Identifier (DOI) to each item included in the platform, making it easy to cite and reference in future research projects. Currently, Zenodo hosts 2 million records and a petabyte of data, serving 15 million users all around the world (Lars Holm Nielsen, 2023). Additionally, Zenodo has an integration with GitHub, which enables researchers to store and maintain the code in one tool, while archiving the data and the publications in a linked platform. This software was established through the European Commission's **OpenAIRE** program. Spread over 35 countries, OpenAIRE[15] is a platform for sharing research data and publications that is designed to support open science in Europe. Researchers can use OpenAIRE to document and archive their code and data,

[13] http://git-scm.com/
[14] https://zenodo.org/
[15] https://www.openaire.eu/

making them publicly accessible and facilitating collaboration with other researchers. OpenAIRE also offers a range of tools and services to support open science, allowing research communities and the relative e-infrastructures to fully publish, interlink, package, and reuse their research artefacts. End-users can then search and consult a rich and up-to-date knowledge graph of research results, and subscribers have the option to get notified of any potential change or update (Kaiser & McNeill, 2019; Manghi et al., 2017).

The **European Commission has also developed additional platforms** to promote open access to research data resulting from publicly funded research under Horizon 2020 projects (Schouppe & Burgelman, 2018). The **European Open Science Cloud[16] (EOSC)** is a platform for sharing and reusing research data, applications, and other resources across Europe. EOSC provides a range of services, including data management, storage, and analysis, that can support researchers in documenting and archiving their code and data. EOSC also offers a range of tools and resources to support open science, including training and education resources, funding opportunities, and policy guidance. Moreover, the European Commission has set forth action plans to increase the "FAIRness" of managed data, increasing the circulation and exploitation of knowledge. Following that example and encouraged by the EOSC requirements, **some countries and agencies introduced new internal protocols** or updated their existing ones. They started establishing standardised practices to document the steps, troubles, and reflections resulting from research activities, making them publicly available to help others implement open science more effectively (Peroni, 2021).

2.2.2 Emerging tools to guarantee full accessibility

One of the most common roadblocks to reproducing code and reusing data is the **inability to reproduce the execution environment**. This is a related issue to the overall crisis in scientific reproducibility, as reported by scientists in multiple fields (Baker, 2016). Inconsistencies in software source code, package versions, extra add-ons, personalised libraries, and dependencies make it necessary **to go a step beyond data and code archiving**. Integrative data and code management platforms are key to solve existing difficulties associated with the software environment, while additional tools are needed to overcome other aspects related to experiment reproducibility, such as cross-platform portability and modular re-usable elements (Beg et al., 2021; Boettiger, 2015).

For a **more integrated code and data management approach**, many institutions or consortiums make use of cloud computing environments to host code, archive data, and share projects easily and without location restrictions. These engines usually allow users to search data and add tags to uploaded items, as well as deploy AI and ML algorithms on large pools of data. One of the most common options is **INVENIO**[17], which is a free open-source digital library software developed by CERN that enables organisations and institutions to create digital repositories for managing and sharing research data and other digital content. One of its main features is that it provides a free infrastructure for libraries. INVENIO includes features such as document management, search functionality, and data visualisation tools that can help to make research outputs more accessible and discoverable. This software also supports integration with other tools and platforms, such as OpenAIRE, to further improve the accessibility and discoverability of research outputs (Sinhababu & Chakravarty, 2022). Some similar software platforms to create free institutional repositories are DSpace[18] and Eprints[19], but neither provides free hosted repository solutions for the libraries. Additionally, novel software has stemmed from these environments, such as MSD-LIVE[20], which is a paid option based on the INVENIO software. On the other side, there are many well-reputed options for cloud computing that are not open source such as the **Amazon Web Services**[21] (AWS), which is a **cloud computing platform that provides a range of services for storing, processing, and**

[16] https://eosc-portal.eu/
[17] https://inveniosoftware.org/
[18] https://duraspace.org/dspace/
[19] https://www.eprints.org/uk/
[20] https://msdlive.org/
[21] https://aws.amazon.com/

analysing data. AWS offers a range of tools and services that can help to improve the accessibility of research, including tools for managing and sharing data, ML services for analysing data, and cloud-based computing resources for running research software (Avinash Bandaru, 2020).

Furthermore, **other software tools designed to facilitate the reproducibility of experiments** are becoming important to ensure the transparency of scientific research. One example is **Docker[22], a containerisation platform that ensures that software runs consistently by users across different computing environments**, making it more accessible to users who may not have the technical expertise to set up and configure the necessary dependencies (Malan, 2022). The software is designed to support the deployment of websites, databases, APIs, and ML models using just a few lines of code. The Docker ecosystem consists of multiple components, allowing for application packaging with its dependencies and execution environment into a standardised, deployable unit (Openja et al., 2022). Docker uniformises programming environments and reduces technical difficulties when setting up computer infrastructure, allowing researchers to focus directly on the code and data itself. **Another way to create reproducible environments is through JupyterHub**[23], which allows users to deploy Jupyter notebooks containing code that can be interactively executed. Through JupyterHub, a user can create either virtual machines (Littlest JupyterHub) or to deploy containerised environments (JupyterHub with a Kubernetes). JupyterHub enables organisations to provide Jupyter notebook servers to multiple users, managing a separate Jupyter environment for each user. However, the server must be self-hosted and self-maintained (JupyterHub, 2023). **Interactive, executable environments for research code and data can also be developed using Binder**[24], which automatically generates a Docker image of an existing repository containing Jupyter notebooks which is then uploaded to JupyterHub. By using Binder, researchers can create a shareable and interactive environment for their code and data that can be accessed and run in a web browser. Although it does not have the ability to customise one's hardware set-up and scalability, this can improve accessibility by allowing users to run and interact with research code and data without needing to install any software locally (Berg et al., 2021). It is important to highlight that both Docker and Binder have an integration with GitHub, which fosters collaboration between researchers.

One nice way to acknowledge the application of these tools and promote full reproducibility, is the **Reproducibility Badge Initiative** (RBI)[25] which has been adopted by some journals in collaboration with the Code Ocean (CO)[26]. This last organisation certifies that the submitted code and data is reproducible, by making sure it runs, delivers results, and it is self-contained. Then, it gives a badge to the publication recognising this feature.

Overall, **there are multiple tools to help researchers to make their code reproducible and transparent**. Despite some technical barriers, as they require specific knowledge and set-up time, their benefits have proved to be tremendously relevant. These tools can help boost the collaboration between teams and increase the integration of various software applications to analyse and model data.

[22] https://www.docker.com/
[23] http://jupyterhub.org/
[24] https://mybinder.org/
[25] https://www.elsevier.com/physical-sciences-and-engineering/computer-science/journals#:~:text=Open%20archive%20journals-,Reproducibility%20Badge%20Initiative,-Reproducibility%20Badge%20Initiative
[26] https://codeocean.com/

3 IAM COMPACT open science protocol

IAM COMPACT is an international and transdisciplinary project that includes a **wide range of different models and modelling teams** to analyse global climate goals and inform the design of the next round of post-2030 NDCs and long-term strategies. The main modelling activities will be based on diverse multi-model analyses. Scientific literature has demonstrated that multi-model activities can provide more robust insights for alternative scenario analysis, compared to single-model studies (Guivarch et al., 2022; Nikas et al., 2021). However, most models used differ significantly in terms of structure, input requirements, and/or solving algorithms. Thus, **the complete transparency and documentation of the models are essential to ensure that produced outcomes are fully accessible, understandable, and reproducible for different audiences, including scientific and non-expert communities**.

For that purpose, the consortium members will follow the principles of open science, making this a pioneer project in the adoption of strict protocols, given the size and scope of the consortium. To that end, we will make use of the most relevant software infrastructures to store data, document models, and publish outcomes and results. By following this protocol (summarised in Figure 1), **consortium members can ensure that compatible models can work together, and the results are transparent and validated**. This will enable the consortium to provide reliable results, which can be used to inform decision-making. To efficiently apply open science practices, we present a protocol for harmonising and validating models and results obtained by the consortium. The protocol has been structured into two stages, both in line with the FAIR and TRUST principles: **a) the model and data harmonisation process** and **b) the open management of the project outcomes**. The former involves harmonising modelling assumptions and input data, as well as interfaces that explain how the models are interconnected. The latter stage deals with validation procedures, including vetting, visualising, and storing results.

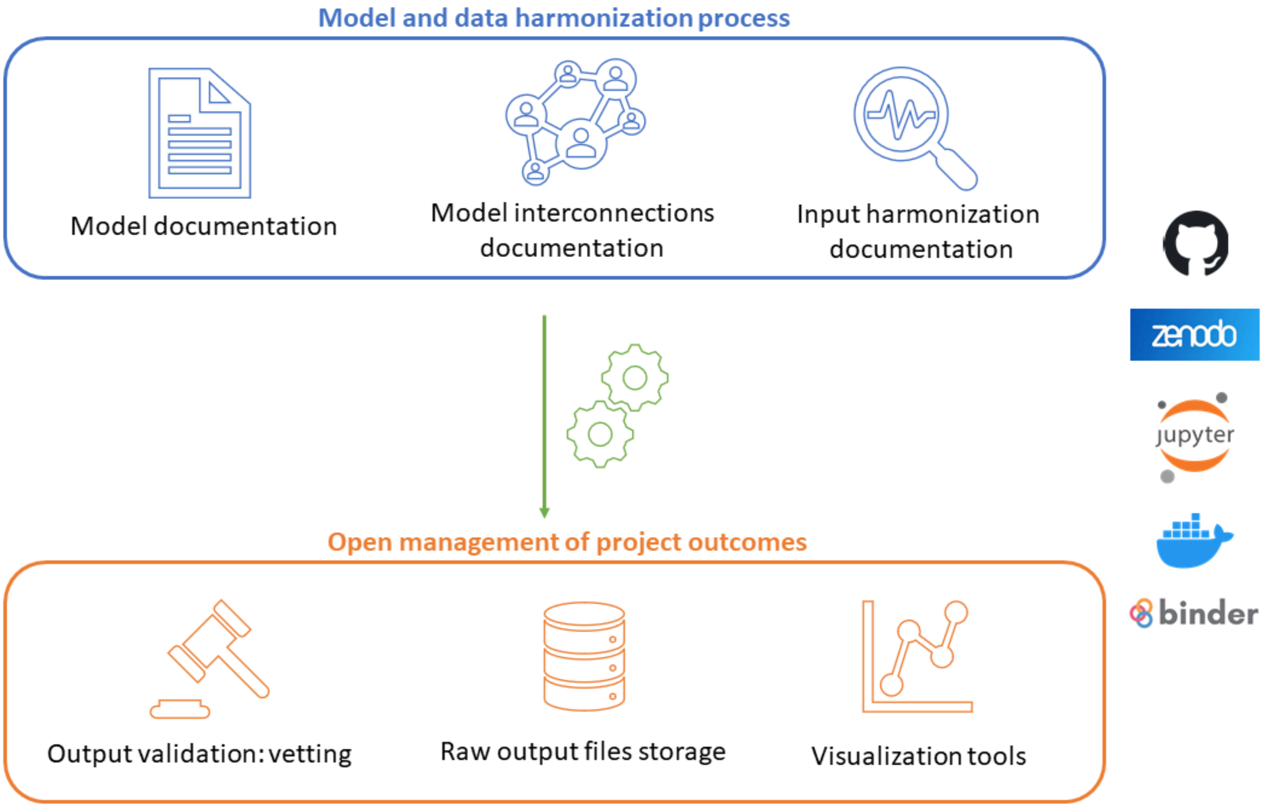

Figure 1: Scheme of the open science protocol that will be used in IAM COMPACT

The first stage of the protocol focuses on the **documentation and harmonisation of models and data**. It is essential to ensure that models developed by different consortium members can work together and their outputs can be compared. To achieve this, we will thoroughly document every model used in the project, including the input data used, relevant links, and specifications such as the model version stored in GitHub and/or Zenodo. The **documentation of the models will be hosted in the I^{2}AM PARIS platform**, which was initially developed in the PARIS REINFORCE project (coordinated by NTUA). Based on its initial design, the platform "*seeks to enable modellers to communicate with one another and stakeholders to interact with modelling assumptions, scenarios and results in an informative way and to understand which decarbonisation pathways are the most relevant and realistic, ultimately enhancing the legitimacy of the scientific processes and improving the transparency of the employed methods, models and tools"*. Having such an interactive and user-friendly documentation platform will **help other consortium members understand how the models work and how they can use them**. In addition to model documentation, tailor-made documentation that explains the **interconnections between the models** will be added to the appropriate site. This documentation should specify how each model interacts with the others as well as inputs and outputs required and assumptions made, with special focus on explaining the achievement of the links and the data parsing between models (in coordination with Task 3.2). The documentation will be shared with all consortium members to ensure that they can understand how the models are interconnected and to explore the possibilities of creating new integrations. Finally, during the multi-model exercises carried out during the project, different modelling teams are expected to follow a common "broad scenario logic" (Task 4.2) that will ensure that all models and tools will harmonise their time resolution, input information, or the baseline specifications to the extent possible. Beyond the documentation of the common assumptions used by the models, **it is also important to document adjustments made by each model during the harmonisation procedure**, for instance through harmonisation heatmaps provided by the I^{2}AM PARIS platform.

The second stage of the protocol is focused on **ensuring the quality and the transparency of produced outcomes**, which will be achieved through a validation process to vet, store, and communicate the results (see Table 1). **The vetting process will include an automated procedure** that assesses if/how the model results in early modelled periods compare to reliable data from global and/or regional data sources, such as the global primary energy balances or the global emission inventories. For the validation of model projections in future time periods, the validation process should involve independent ("within-consortium") experts who can validate the results of the models. This process should include a review of the model inputs, assumptions, and outputs. The review should be done in a transparent and collaborative way, with feedback provided to the consortium members to improve the models. Furthermore, the communication of the results should include both the storage of raw output files and the use of visualisation tools to disseminate the results to non-expert audiences. To store results, **raw output files should be uploaded to a secure and accessible location** (e.g., Zenodo, etc.), with a unique DOI associated. The accessibility of the raw output files will be particularly relevant for the scientific community, as the files can be directly downloaded by researchers and incorporated into further model intercomparison exercises. Additionally, these raw output datasets will be accompanied by detailed metadata documentation, explaining what is contained within each dataset (e.g., inputs, outputs, etc.), as well as the specific step-by-step data handling processes used to develop it. To facilitate model intercomparison within and outside the consortium, the model outcomes will also be formatted to the standardised IAMC style. In addition, **results will also be shown through interactive or user-friendly visualisation tools** (e.g., R scripts, R-Shinny apps, etc.). This will enable to explore the results of the models in a clear and understandable way.

Table 1: Summary of the output management

Open management of project outcomes	Validation	Vetting automated procedure comparing the model results in early modelled periods with reliable data from global and/or regional data sources
		Future projections validation involving independent experts and full review of model inputs, assumptions, and outputs
	Storage	Raw output files stored in a secure and accessible location with a unique DOI
	Visualisation	Outputs shown through interactive and user-friendly tools
		Outputs using a standardised style following the time-series data template of IAMC

Finally, we note that, in order to ensure the full accessibility and reproducibility of the results, **we will explore the potential use of additional software platforms** (e.g., Docker, Jupyter Kubernetes, etc.) to solve potential problems related to the reproducibility of the execution environment of the models (when licences allow it) and of the model intercomparison and post-analysis studies**. In general terms, each modelling team should meet the requirements of open science principles**, as described in Table 2. However, the existing differences across the models within the project makes it challenging to develop a single protocol that is adequate for the entire suite of tools. The potential use of these emerging software applications will be assessed in a model-by-model case, and applied to the extent that this is possible.

Table 2: Overview of the actions taken to meet the FAIR and TRUST open science principles

FAIR	Findability	• Documentation hosted in the I^{2}AM PARIS platform • Code storage and version track • Storage of raw output files with DOI
	Accessibility	• Documentation of data harmonisation • Metadata documentation • Reproducibility tools usage
	Interoperability	• Models' interconnection documentation • Follow a common "scenario logic" • Standardised outputs following the IAMC style
	Reusability	• Metadata documentation of input and output files • Use of reproducibility tools
TRUST	Transparency	• Inner links to model versions and repositories • Vetting procedure
	Responsibility	• Protocol adherence
	User Focus	• Documentation hosted in the I^{2}AM PARIS platform • Visualisation tools: notebooks, R-Shinny apps, etc.
	Sustainability	• Documentation hosted in the I^{2}AM PARIS platform • Data and code storage in common repositories such as Zenodo
	Technology	• Documentation hosted in the I^{2}AM PARIS platform • Use of reproducibility tools • Code and data storage

Overall, **this protocol provides guidance to the consortium members and outlines the procedures to be followed to guarantee that scientific breakthroughs are accessible and reusable over time.** In the same line, the Open Data Management Plan (D3.2) will provide further information about timelines and platforms/software linked to every item and phase of the protocol, facilitating a clear and well-defined protocol application.

Bibliography

Acord, S. K., & Harley, D. (2013). Credit, time, and personality: The human challenges to sharing scholarly work using Web 2.0. *New Media & Society*, *15*(3), 379–397. https://doi.org/10.1177/1461444812465140

Anusuriya Devaraju, Robert Huber, Mustapha Mokrane, Patricia Herterich, Cepinskas, Linas, Jerry de Vries, Herve L'Hours, Joy Davidson, & Angus White. (2022). *FAIRsFAIR Data Object Assessment Metrics*. https://doi.org/10.5281/ZENODO.6461229

Avinash Bandaru. (2020). *Amazon Web Services*.

Bahim, C., Casorrán-Amilburu, C., Dekkers, M., Herczog, E., Loozen, N., Repanas, K., Russell, K., & Stall, S. (2020). The FAIR Data Maturity Model: An Approach to Harmonise FAIR Assessments. *Data Science Journal*, *19*, 41. https://doi.org/10.5334/dsj-2020-041

Bak, G. (2016). Trusted by whom? TDRs, standards culture and the nature of trust. *Archival Science*, *16*(4), 373–402. https://doi.org/10.1007/s10502-015-9257-1

Baker, M. (2016). 1,500 scientists lift the lid on reproducibility. *Nature*, *533*(7604).

Beg, M., Taka, J., Kluyver, T., Konovalov, A., Ragan-Kelley, M., Thiery, N. M., & Fangohr, H. (2021). Using Jupyter for Reproducible Scientific Workflows. *Computing in Science & Engineering*, *23*(2), 36–46. https://doi.org/10.1109/MCSE.2021.3052101

Berg, J., Sorensen, S. A., Ting, J. T., Miller, J. A., Chartrand, T., Buchin, A., Bakken, T. E., Budzillo, A., Dee, N., Ding, S.-L., Gouwens, N. W., Hodge, R. D., Kalmbach, B., Lee, C., Lee, B. R., Alfiler, L., Baker, K., Barkan, E., Beller, A., … Lein, E. S. (2021). Human neocortical expansion involves glutamatergic neuron diversification. *Nature*, *598*(7879), 151–158. https://doi.org/10.1038/s41586-021-03813-8

Besançon, L., Peiffer-Smadja, N., Segalas, C., Jiang, H., Masuzzo, P., Smout, C., Billy, E., Deforet, M., & Leyrat, C. (2021). Open science saves lives: Lessons from the COVID-19 pandemic. *BMC Medical Research Methodology*, *21*(1), 117. https://doi.org/10.1186/s12874-021-01304-y

Boeckhout, M., Zielhuis, G. A., & Bredenoord, A. L. (2018). The FAIR guiding principles for data stewardship: Fair enough? *European Journal of Human Genetics*, *26*(7), 931–936. https://doi.org/10.1038/s41431-018-0160-0

Boettiger, C. (2015). An introduction to Docker for reproducible research, with examples from the R environment. *ACM SIGOPS Operating Systems Review*, *49*(1), 71–79. https://doi.org/10.1145/2723872.2723882

Branney, P., Brooks, J., Kilby, L., Newman, K., Norris, E., Pownall, M., Talbot, C. V., Treharne, G. J., & Whitaker, C. (2022). *Three Steps to Open Science for Qualitative Research in Psychology*.

Brinkhaus, H. O., Rajan, K., Schaub, J., Zielesny, A., & Steinbeck, C. (2023). Open data and algorithms for open science in AI-driven molecular informatics. *Current Opinion in Structural Biology*, *79*. https://doi.org/10.1016/j.sbi.2023.102542

Burgelman, J.-C., Pascu, C., Szkuta, K., Von Schomberg, R., Karalopoulos, A., Repanas, K., & Schouppe, M. (2019). Open Science, Open Data, and Open Scholarship: European Policies to Make Science Fit for the Twenty-First Century. *Frontiers in Big Data*, *2*, 43. https://doi.org/10.3389/fdata.2019.00043

Byers, Edward, Krey, Volker, Kriegler, E., Schaeffer, R., Riahi, K, Kikstra, J, Lamboll, R, Nicholls, Z, & et al. (2022). AR6 Scenarios Database. *Climate Change 2022: Mitigation of Climate Change, Intergovernmental Panel on Climate Change (978-92-9169-160-9)., ISBN: 978-92-9169-160-9*. https://doi.org/10.5281/zenodo.7197970

Center for Open Science. (2021). *Accelerating open sceince together: Impact Report 2021*. https://www.cos.io/impact/report-2021

Chin, J. M., Pickett, J. T., Vazire, S., & Holcombe, A. O. (2021). Questionable research practices and open science in quantitative criminology. *Journal of Quantitative Criminology*, 1–31.

Cohen, K. B., Xia, J., Zweigenbaum, P., Callahan, T. J., Hargraves, O., Goss, F., Ide, N., Névéol, A., Grouin, C., & Hunter, L. E. (2018). *Three Dimensions of Reproducibility in Natural Language Processing*.

CoreTrustSeal. (2019). *CoreTrustSeal Standards And Certification Board: CoreTrustSeal Trustworthy Data Repositories Requirements 2020–2022*. https://doi.org/10.5281/ ZENODO.3638211

CoreTrustSeal Standards And Certification Board. (2022). *CoreTrustSeal Requirements 2023-2025*.

https://doi.org/10.5281/ZENODO.7051011
Crabtree, J. D. (2020). *Evidence for Trusted Digital Repository reviews: An analysis of perspectives*.
Crystal-Ornelas, R., Varadharajan, C., Bond-Lamberty, B., Boye, K., Burrus, M., Cholia, S., Crow, M., Damerow, J., Devarakonda, R., & Ely, K. S. (2021). A guide to using GitHub for developing and versioning data standards and reporting formats. *Earth and Space Science*, *8*(8), e2021EA001797.
De Wilde, Z. G. (2017). *Recommended versus Certified Repositories: Mind the Gap* [Data set]. Mendeley. https://doi.org/10.17632/ZX2KCYVVWM.1
Donaldson, D. R., & Russell, S. V. (2023). Trustworthy Digital Repository Certification: A Longitudinal Study. In *In: , Et al. Information for a Better World: Normality, Virtuality, Physicality, Inclusivity* (pp. 552–562). Springer. https://doi.org/10.1007/978-3-031-28032-0_42
Dunn, R., Lief, C., Peng, G., Wright, W., Baddour, O., Donat, M., Dubuisson, B., Legeais, J.-F., Siegmund, P., Silveira, R., Wang, X. L., & Ziese, M. (2021). Stewardship Maturity Assessment Tools for Modernization of Climate Data Management. *Data Science Journal*, *20*, 7. https://doi.org/10.5334/dsj-2021-007
Enke, N., Thessen, A., Bach, K., Bendix, J., Seeger, B., & Gemeinholzer, B. (2012). The user's view on biodiversity data sharing—Investigating facts of acceptance and requirements to realize a sustainable use of research data—. *Ecological Informatics*, *11*, 25–33.
European Commission. (2017). *H2020 Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020*. chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf
European Commission. (2019). *The European Union's Research and Innovation Performance—Facts and Figures 2019*. https://research-and-innovation.ec.europa.eu/system/files/2019-12/ec_rtd_factsheet_2019_0.pdf
Fecher, B., Friesike, S., & Hebing, M. (2015). What Drives Academic Data Sharing? *PLOS ONE*.
Fell, M. J. (2019). The economic impacts of open science: A rapid evidence assessment. *Publications*, *7*(3), 46.
Finley, K. (2011). Github has surpassed sourceforge and google code in popularity. *ReadWrite. Com. Recuperado de Http://Readwrite. Com/2011/06/02/Github-Has-Passedsourceforge*.
Fraser, N., Brierley, L., Dey, G., Polka, J. K., Pálfy, M., & Alexis, J. (2020). *Preprinting a pandemic: The role of preprints in the COVID-19 pandemic*.
Gibney, E. (2022). Could machine learning fuel a reproducibility crisis in science? *Nature*, *608*. https://www.nature.com/articles/d41586-022-02035-w
Gidden, M., & Huppmann, D. (2019). pyam: A Python Package for the Analysis and Visualization of Models of the Interaction of Climate, Human, and Environmental Systems. *Journal of Open Source Software*, *4*(33), 1095. https://doi.org/10.21105/joss.01095
Guivarch, C., Le Gallic, T., Bauer, N., Fragkos, P., Huppmann, D., Jaxa-Rozen, M., Keppo, I., Kriegler, E., Krisztin, T., & Marangoni, G. (2022). Using large ensembles of climate change mitigation scenarios for robust insights. *Nature Climate Change*, *12*(5), 428–435.
Gutiérrez, J. M., Jones, R. G., Narisma, G. T., Alves, L. M., Amjad, M., Gorodetskaya, I. V., Grose, M., Klutse, N. A. B., Krakovska, S., Li, J., Martínez-Castro, D., Mearns, L. O., Mernild, S. H., Ngo-Duc, T., van den Hurk, B., & Yoon, J.-H. (2021). *Atlas. 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.)].* Cambridge University Press. http://interactive-atlas.ipcc.ch/
Haibe-Kains, B., Adam, G. A., Hosny, A., Khodakarami, F., Massive Analysis Quality Control (MAQC) Society Board of Directors, Shraddha, T., Kusko, R., Sansone, S.-A., Tong, W., Wolfinger, R. D., Mason, C. E., Jones, W., Dopazo, J., Furlanello, C., Waldron, L., Wang, B., McIntosh, C., Goldenberg, A., Kundaje, A., … Aerts, H. J. W. L. (2020). Transparency and reproducibility in artificial intelligence. *Nature*, *586*(7829), E14–E16. https://doi.org/10.1038/s41586-020-2766-y
Hayashi, K. (2021). How Could COVID-19 Change Scholarly Communication to a New Normal in the Open Science Paradigm? *Patterns*, *2*(1), 100191. https://doi.org/10.1016/j.patter.2020.100191

Hughes, L. D., Tsueng, G., DiGiovanna, J., Horvath, T. D., Rasmussen, L. V., Savidge, T. C., Stoeger, T., Turkarslan, S., Wu, Q., & Wu, C. (2023). Addressing barriers in FAIR data practices for biomedical data. *Scientific Data*, *10*(1), 98.

Huppmann, D., Rogelj, J., Kriegler, E., Krey, V., & Riahi, K. (2018). A new scenario resource for integrated 1.5 °C research. *Nature Climate Change*, *8*(12), 1027–1030. https://doi.org/10.1038/s41558-018-0317-4

IAMC. (2022). *IAMC Documentation Wiki*. https://www.iamcdocumentation.eu/index.php/IAMC_wiki

Iturbide, M., Fernández, J., Gutiérrez, J. M., Pirani, A., Huard, D., Al Khourdajie, A., Baño-Medina, J., Bedia, J., Casanueva, A., Cimadevilla, E., Cofiño, A. S., De Felice, M., Diez-Sierra, J., García-Díez, M., Goldie, J., Herrera, D. A., Herrera, S., Manzanas, R., Milovac, J., … Yelekçi, Ö. (2022). Implementation of FAIR principles in the IPCC: The WGI AR6 Atlas repository. *Scientific Data*, *9*(1), 629. https://doi.org/10.1038/s41597-022-01739-y

Jahn, N., Laasko, M., Lazzeri, E., & McQuilton, P. (2023). *Study on the readiness of research data and literature repositories to facilitate compliance with the Open Science Horizon Europe MGA requirements*.

Jin, H., Wang, Q., Yang, Y.-F., Zhang, H., Gao, M., Jin, S., Chen, Y., Xu, T., Zheng, Y.-R., Chen, J., Xiao, Q., Yang, J., Wang, X., Geng, H., Ge, J., Wang, W.-W., Chen, X., Zhang, L., Zuo, X.-N., & Chuang-Peng, H. (2023). The Chinese Open Science Network (COSN): Building an Open Science Community From Scratch. *Advances in Methods and Practices in Psychological Science*, *6*(1). https://doi.org/10.1177/25152459221144986

JupyterHub. (2023). *Project JupyterHub*. https://jupyter.org/hub

Kaiser, O., & McNeill, G. (2019). OpenAIRE für Repository ManagerInnen—Wie Repository ManagerInnen Open Science unterstützen können. *Mitteilungen der Vereinigung Österreichischer Bibliothekarinnen und Bibliothekare*, *72*(2), 373–382. https://doi.org/10.31263/voebm.v72i2.2815

Kerry Levett, Keith Russell, Martin Schweitzer, Kathryn Unsworth, & Andrew White. (2022). *FAIR data assessment tool*. https://ardc.edu.au/resource/fair-data-self-assessment-tool/

Lagoze, C. (2002). The open archives initiative protocol for metadata harvesting-v2.0. *Http://Www. Openarchives. Org/OAI/Openarchivesprotocol. Html*. http://oai.dlib.vt.edu/OAI/2.0/openarchivesprotocol.htm

Lagoze, C., Van de Sompel, H., Nelson, M., & Warner, S. (2001). The open archives initiative: Building a low-barrier interoperability framework. *Proceedings of the 1st ACM/IEEE-CS Joint Conference on Digital Libraries*, 54–62. https://doi.org/10.1145/379437.379449

Lars Holm Nielsen. (2023). *NIH Virtual GREI Workshop*. https://blog.zenodo.org/

Lima, A., Rossi, L., & Musolesi, M. (2014). Coding Together at Scale: GitHub as a Collaborative Social Network. *Proceedings of the International AAAI Conference on Web and Social Media*, *8*(1), 295–304. https://doi.org/10.1609/icwsm.v8i1.14552

Lin, D., Crabtree, J., Dillo, I., Downs, R. R., Edmunds, R., Giaretta, D., De Giusti, M., L'Hours, H., Hugo, W., & Jenkyns, R. (2020). The TRUST Principles for digital repositories. *Scientific Data*, *7*(1), 144.

Malan, D. J. (2022). Standardizing Students' Programming Environments with Docker Containers: Using Visual Studio Code in the Cloud with GitHub Codespaces. *Proceedings of the 27th ACM Conference on on Innovation and Technology in Computer Science Education Vol. 2*, 599–600. https://doi.org/10.1145/3502717.3532164

Manghi, P., Artini, M., Atzori, C., Baglioni, M., Bardi, A., La Bruzzo, S., De Bonis, M., Dimitropoulos, H., Foufoulas, I., & Iatropoulou, K. (2017). *OpenAIRE: Advancing open science*. 23–24.

Masson-Delmotte, V., Zhai, P., Pirani, A., Connors, S. L., Péan, C., Berger, S., Caud, N., Chen, Y., Goldfarb, L., & Gomis, M. (2021). Climate change 2021: The physical science basis. *Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*, *2*.

Mendez, E., Lawrence, R., MacCallum, C. J., Moar, E., & Van Nieuwerburgh, I. (2020). *Progress on Open Science: Towards a shared research knowledge system. Final report of the Open Science Policy Platform*.

Mons, B., Neylon, C., Velterop, J., Dumontier, M., Da Silva Santos, L. O. B., & Wilkinson, M. D. (2017). Cloudy, increasingly FAIR; revisiting the FAIR Data guiding principles for the European Open Science Cloud. *Information Services & Use*, *37*(1), 49–56. https://doi.org/10.3233/ISU-170824

Munafò, M. R., Nosek, B. A., Bishop, D. V. M., Button, K. S., Chambers, C. D., Percie Du Sert, N., Simonsohn, U.,

Wagenmakers, E.-J., Ware, J. J., & Ioannidis, J. P. A. (2017). A manifesto for reproducible science. *Nature Human Behaviour*, *1*(1), 0021. https://doi.org/10.1038/s41562-016-0021
National Institutes of Health. (2021). *NIH Grants Policy Statement*. https://grants.nih.gov/grants/policy/nihgps/HTML5/introduction.htm
Nielsen, A. F., Michelmann, J., Akac, A., Palts, K., Zilles, A., Anagnostopoulou, A., & Langeland, O. (2023). Using the future wheel methodology to assess the impact of open science in the transport sector. *Scientific Reports*, *13*(1), 6000. https://doi.org/10.1038/s41598-023-33102-5
Nikas, A., Gambhir, A., Trutnevyte, E., Koasidis, K., Lund, H., Thellufsen, J. Z., Mayer, D., Zachmann, G., Miguel, L. J., Ferreras-Alonso, N., Sognnaes, I., Peters, G. P., Colombo, E., Howells, M., Hawkes, A., Van Den Broek, M., Van De Ven, D. J., Gonzalez-Eguino, M., Flamos, A., & Doukas, H. (2021). Perspective of comprehensive and comprehensible multi-model energy and climate science in Europe. *Energy*, *215*, 119153. https://doi.org/10.1016/j.energy.2020.119153
Nosek, B. A., Alter, G., Banks, G. C., Borsboom, D., Bowman, S. D., Breckler, S. J., Buck, S., Chambers, C. D., Chin, G., & Christensen, G. (2015). Promoting an open research culture. *Science*, *348*(6242), 1422–1425.
Oliveira, G. P., Moura, A. F. C., Batista, N. A., Brandão, M. A., Hora, A., & Moro, M. M. (2023). How do developers collaborate? Investigating GitHub heterogeneous networks. *Software Quality Journal*, *31*(1), 211–241. https://doi.org/10.1007/s11219-022-09598-x
OpenAire. (2022). *OpenAIRE Guidelines*. https://guidelines.openaire.eu/en/latest/
Openja, M., Majidi, F., Khomh, F., Chembakottu, B., & Li, H. (2022). Studying the Practices of Deploying Machine Learning Projects on Docker. *The International Conference on Evaluation and Assessment in Software Engineering 2022*, 190–200. https://doi.org/10.1145/3530019.3530039
Peng, G., Gross, W. S., & Edmunds, R. (2022). Crosswalks among stewardship maturity assessment approaches promoting trustworthy FAIR data and repositories. *Scientific Data*, *9*(1), 576. https://doi.org/10.1038/s41597-022-01683-x
Peroni, S. (2021). Review of: 'FAIR RDM (Research Data Management): Italian initiatives towards EOSC implementation'. *Qeios*. https://doi.org/10.32388/LAKK5Q
Piwowar, H. A., Day, R. S., & Fridsma, D. B. (2007). Sharing Detailed Research Data Is Associated with Increased Citation Rate. *PLoS ONE*, *2*(3), e308. https://doi.org/10.1371/journal.pone.0000308
Piwowar, H. A., & Vision, T. J. (2013). Data reuse and the open data citation advantage. *PeerJ*, *1*, e175.
Ramachandran, R., Bugbee, K., & Murphy, K. (2021). From Open Data to Open Science. *Earth and Space Science*, *8*(5). https://doi.org/10.1029/2020EA001562
RDA FAIR Data Maturity Model Working Group, B. (2020). FAIR Data Maturity Model: Specification and guidelines. *Res. Data Alliance*, *10*. https://www.rd-alliance.org/group/fair-data-maturity-model-wg/outcomes/fair-data-maturity-model-specification-and-guidelines-0
RLG/OCLC Working Group on Digital Archive Attributes. (2002). *Trusted Digital Repositories: Attributes and Responsibilities*. Mountain View, CA: Research Libraries Group. http://www.oclc.org/research/activities/past/rlg/trustedrep/repositories.pdf
Schouppe, M., & Burgelman, J. C. (2018). *Relevance of EOSC and FAIR in the Realm of Open Science and Phases of Implementing the EOSC*. 1–4.
Sicilia, M.-A., García-Barriocanal, E., & Sánchez-Alonso, S. (2017). Community Curation in Open Dataset Repositories: Insights from Zenodo. *Procedia Computer Science*, *106*, 54–60. https://doi.org/10.1016/j.procs.2017.03.009
Sinhababu, A., & Chakravarty, R. (2022). *Building Digital Repositories with the Open- Source Software Invenio: Use of SaaS Model Zenodo*. *7*(1).
Sohn, E. (2023). The reproducibility issues that haunt health-care AI. *Nature*, *613*(7943), 402–403. https://doi.org/10.1038/d41586-023-00023-2
Stodden, V., Seiler, J., & Ma, Z. (2018). An empirical analysis of journal policy effectiveness for computational reproducibility. *Proceedings of the National Academy of Sciences*, *115*(11), 2584–2589.
Tenopir, C., Allard, S., Douglass, K., Aydinoglu, A. U., Wu, L., Read, E., Manoff, M., & Frame, M. (2011). Data Sharing by Scientists: Practices and Perceptions. *PLoS ONE*, *6*(6), e21101.

https://doi.org/10.1371/journal.pone.0021101
Tenopir, C., Dalton, E. D., Allard, S., Frame, M., Pjesivac, I., Birch, B., Pollock, D., & Dorsett, K. (2015). Changes in data sharing and data reuse practices and perceptions among scientists worldwide. *PloS One*, *10*(8), e0134826.
UNESCO. (2021). *Open Science for a Global Transformation: CODATA*.
Wilkinson, M. D., Dumontier, M., Aalbersberg, Ij. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., & Bourne, P. E. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, *3*(1), 160018. https://doi.org/10.1038/sdata.2016.18