



CODATA International Data Policy Committee (IDPC) Action Plan 2023 to 2025

Contents

Introduction.....	2
Vision.....	2
Mission.....	2
Rationale.....	2
Mandate.....	3
Context and background.....	3
The IDPC and the CODATA Strategic Plan 2023-2027.....	4
IDPC Topics for Action: 2023 to 2025.....	5
1. Data policy for data quality, reliability, and integrity.....	5
2. Data policy for science in crisis situations.....	5
3. Data policy for education.....	6
4. Data policy for AI.....	6
5. Data policy for Open Science.....	7
6. Data policy for the publication and communication of science.....	7
Recent engagements.....	8
Priority future activities.....	8
Milestones and measures of success.....	9

Introduction

The CODATA International Data Policy Committee (IDPC) is the key instrument by which CODATA pursues its mission of promoting effective and appropriate policies for Open Science and FAIR data. A subsidiary task is to support the other strategic objectives of CODATA.

The IDPC is composed of diverse and knowledgeable individuals who are dedicated to advancing responsible and effective data governance through data policy on a global scale. IDPC members play a vital role in shaping the development of policy studies, research, projects, publications, guidelines, and standards that promote data quality, reliability, and integrity, while upholding the values of Open Science and ethical principles and practices.

Vision

The vision of the CODATA IDPC is to foster global collaboration and establish ethical, transparent, and inclusive data policies that enable the reliable and responsible collection, curation, analysis, access, sharing, and use of data across borders, disciplines, and sectors.

The IDPC promotes a culture of Open Science, innovation, and evidence-based decision-making in its development of data policy, recognizing the value of data as a global public good. By developing and implementing forward-thinking data policies, the IDPC facilitates knowledge exchange, drives scientific breakthroughs, addresses societal challenges, and contributes to the well-being of people around the world.

Mission

The mission of the CODATA IDPC is to serve as a global platform for collaboration, research, and advocacy in shaping effective data policies that address the complex challenges of today's increasingly digital societies.

The IDPC fosters international dialogue, knowledge sharing, and consensus-building among diverse stakeholders, including policymakers, researchers, industry leaders, civil society organizations, and the public. By conducting rigorous analysis, engaging in broad consultation, providing evidence-based recommendations, and promoting best practices, the IDPC develops robust and inclusive data policies that foster innovation, protect privacy, contribute to data security, promote data accessibility, and advance the responsible and ethical use of data for the benefit of humanity.

Rationale

This is a critical time for developing data policy in the context of advancing frameworks for data sharing, artificial intelligence, data diplomacy, digital economies, the revision of research assessment in universities, and the drive toward Open Science. Data policy lies at the intersection of science, public policy, and enterprise in the rapid advancement of the digitalization of societies across the world. The achievements of science in society increasingly rely on the purposes to which data is put and how it is managed for the attainment of private and public goods.

The IDPC focuses on developing pathways to data policy that promote data integrity, data ethics, FAIR data, data diplomacy, and leadership in data. The IDPC also identifies and addresses new areas of concern for data policy, such as AI and the use of Large Language Models (LLMs) and Generative Pre-trained Transformers (GPTs), and data and AI diplomacy.

The IDPC serves the CODATA community by bringing policy to data science while advancing international collaboration in the use of data to understand and respond to the most pressing challenges of today's world, be those challenges local, national, regional, or global. Data must serve science while conjointly serving the needs of all sectors in society. Data policy provides avenues for exercising the voice of science in societal debates that also engage governments, enterprises, the media, and citizens on critical issues of fundamental importance. Well-developed and documented data policy for crisis situations is of vital importance to ensuring that the voice of science plays a strategic role in local, national, regional, and global preparedness, response, and recovery for health emergencies, natural hazards, geo-political disruptions, or (generally) crisis situations. Ultimately, the IDPC develops policy research, strategy, and guidance for CODATA in support of its role in advancing the International Science Council (ISC)'s vision of science as a public good.

The IDPC works across disciplines and sectors, engaging partners at the national and international levels while also bringing together academia, the private sector, government and intergovernmental agencies, and representatives of civil society. Additionally, the IDPC is mindful of the role it plays in education.

Mandate

The IDPC is mandated to support CODATA and its mission to the ISC through the following modes of activity:

1. Helps CODATA respond to emerging issues and requests for advice, particularly from the ISC community. An important recent example is [Open Science for a Global Transformation: CODATA-coordinated submission to the UNESCO Open Science Consultation](#) (2020).
2. Provides a source of expertise for funded projects and reports. Recent examples include the [CODATA Twenty-Year Review of GBIF](#) (2020) and [Opening Big Data for a Sustainable Future: Review of the CASEarth Program's Data Policy](#) (2022).
3. Generates its own initiatives through sub-group activities. For example the [Beijing Declaration on Research Data](#) and the related [Workshop on Open Research Data Policy and Practice](#) (Sept 2019).

Context and background

The CODATA International Data Policy Committee (IDPC) was formed in 2014 following a recommendation from the Executive Committee and an ISC Review of CODATA. The development of policies promoting open data for research had been a longstanding and prominent feature of CODATA activities. In order to ensure greater continuity and focus for these core activities — and to amplify CODATA's voice in these matters — the [CODATA Strategic Plan 2015-18](#) identified the IDPC as the principal means by which CODATA should further pursue an international FAIR and open data policy agenda for research.

Contributions to data policy are made, not only by the IDPC, but throughout CODATA in nearly all its engagements. CODATA has contributed substantially, for example, to the [2007 OECD Principles and Guidelines for Access to Research Data from Public Funding](#) and to successive iterations of the [GEO Data Sharing Principles and Data Management Guidelines](#). From 2015, CODATA has produced a series of significant [data policy reports](#). As well as deploying specific expertise through the IDPC, the CODATA Secretariat and parts of CODATA contribute to data policy activities. Among these activities, the CODATA Executive Director co-chaired the OECD-CODATA study on [Business Models for Sustainable Data Repositories](#), chaired the EC Expert Group that produced the [Turning FAIR into Reality](#)

report, and was Vice-chair of the Expert Advisory Group to UNESCO's [Recommendation on Open Science](#).

The IDPC supports and seeks to advance the implementation of [UNESCO's Recommendations on Open Science and on the Ethics of Artificial Intelligence](#). In the context of specific legal and policy frameworks for data privacy, data security, and intellectual property, the IDPC contributes to policy that reflects the position that data created by research and of use to research should be 'as open as possible, as closed as necessary'. The IDPC advocates for open data policies and Open Science policies, facilitates discussion around best practices, and supports the development of robust policy frameworks and tools that reflect the goal of making research outputs open and data work for the benefit of society, including in the context of its uses in AI.

The IDPC and the CODATA Strategic Plan 2023-2027

The CODATA Strategic Plan 2023-2027 provides an important overall framework for guiding the work of the organization in relation to the fast-developing importance of data in our evolving digital societies. The Strategic Plan outlines four thematic priorities. Below are these priorities as well as a short description of the role the IDPC plays in achieving the goals of each priority.

Thematic Priority 1: Making Data Work

Through its work on data policy, the IDPC focuses on achieving the highest utility of data for the benefit of society across disciplines and sectors. Each of the 'Topics for Action' listed below will engage in promoting the greatest and most beneficial uses of data in the activities engaged. Of principal importance here for the IDPC is Topic 1. Data policy for data quality, reliability, and integrity. The IDPC promotes data integrity for ensuring the accuracy, completeness, quality, and trustworthiness of data, which is essential for conducting valid analyses, maintaining the overall integrity of systems and processes reliant on data, and making informed decisions.

Thematic Priority 2: Data Policy

Data policy is at the core of the activities of all the IDPC work. This underlies all the activities and engagements as well as the research and outcomes of the IDPC.

Thematic Priority 3: The Science of Data and Data Stewardship

A fundamental role of the IDPC is to promote data science. The IDPC contributes to trust in data science as well as in the applications of data science, including AI, to society. Particular attention is given to the rights of individuals and communities. The IDPC also promotes the importance of data stewardship for achieving reliable and trustworthy outcomes in data science and its applications.

Thematic Priority 4: Data Skills and Education

The IDPC promotes the development and correct application of data skills through its work on data policy. It is particularly engaged in education under Topic 3: Data policy for education, while also seeing an education component in each of the topics.

IDPC Topics for Action: 2023 to 2025

The CODATA IDPC focuses on the following six topics for action in the years 2023-2025 while recognizing that each of these topics requires an engagement beyond 2025 and not all can be equally addressed within the period foreseen. Nonetheless, these topics all contribute to the longer-term vision of CODATA and the IDPC.

1. Data policy for data quality, reliability, and integrity

This topic addresses the responsibility of researchers and data infrastructures to ensure the quality, reliability, and integrity of the data they curate and process. Data policy contributes to guidelines and procedures to ensure that data is accurate, reliable, consistent, and secure throughout its lifecycle and in its usage and communication. This includes the roles, responsibilities, and processes for managing data quality, reliability, and integrity in the areas of data management, decision-making, and the governance of data. Data quality measures (such as data cleaning, deduplication, anonymization, redaction, and other curation methods) also play an essential role in enhancing the performance, reliability, and integrity of AI systems.

Because AI models rely heavily on high-quality interoperable data for training and inference, data policy can contribute guidelines and standards for data collection, data curation, and data processing that help ensure that the data used for AI development is accurate, relevant, and representative of the problem domain. Addressing data quality, reliability, and integrity through data policy frameworks fosters the transparent, accountable, and ethical use of AI.

Initial activity

The initial scope for this activity will be set by an article being prepared for the Data Science Journal on the extension of the concept of data quality towards reliability of data, with a possible extension to reliability in times of crisis.

2. Data policy for science in crisis situations

This topic focuses on the collection, processing, and use of data in situations of natural hazards, health crises, geo-political conflicts, and other disruptive circumstances. It examines the data policy frameworks necessary to ensure that scientific projects, particularly regarding data collection and processing, are viable and relevant to crisis situations while also contributing to scientific results in preparing for and responding to crisis situations. Data policy also plays a crucial role in shaping the effectiveness and ethical use of AI in crisis situations. Accessible and interoperable data allows AI models to be trained on comprehensive and diverse datasets, leading to more accurate and informed decision-making during crises and protecting confidentiality. Data policy can help protect sensitive information during crisis response and thus allows AI systems to handle sensitive information in trusted data ecosystems, fostering public confidence in their use during crises.

The objectives of this topic are (a) to examine the scientific, political, and societal frameworks needed to develop data policy addressing crisis situations; (b) to consider the underlying ethical, human rights, and humanitarian frameworks needed to support data policy during crisis situations; and (c) to support the development of tools that promote the responsible practice and use of data when generating scientific evidence in crisis situations and guiding decision making in preparedness and response.

Initial activities

- A. A special collection on "Data and AI policy, systems, and tools for crises" is being prepared for the Data Science Journal.
- B. A project to prepare a contribution to the UNESCO Open Science Toolkit on 'Data Policy for Open Science in Times of Crisis' has been approved by UNESCO and is underway.

3. Data policy for education

This topic addresses the need for education in data policy as well as how data policy can contribute to education programs involving data science and data applications, including the use of AI. It addresses the role of data policy in education and training for data, data curation, and Open Science, and specifically for data policy. At the same time, this topic contributes to

the discussions on data management in education and research settings, including in the discussions around research assessment reform. Well-designed data policy in education promotes transparency, accountability, privacy, and the responsible use of data. It enables educational institutions to harness the power of data to improve teaching and learning outcomes, support evidence-based decision-making, and protect the privacy rights of students and stakeholders. Data policy also contributes to the current revision of research assessment practices by promoting the adoption of more comprehensive and responsible evaluation approaches that consider the diverse impacts and contributions of research beyond traditional metrics, fostering a more robust, equitable, and accountable research ecosystem.

Initial activities

- A. The IDPC will build on its draft curriculum for teaching data policy and explore how this can be refined and delivered via the CODATA-RDA Data Schools and other venues.
- B. The IDPC will collaborate with the annual International Training Workshop on Open Science and SDGs organized by CAS, GOSC, and CODATA with partner organizations. It will also collaborate with FSIC on the course 'The Role of AI Ethics in Scientific Publications on Open Science Platforms: Examining the Use of AI in Global Publication Governance Models for Impactful Knowledge Sharing and Social Benefit'.

4. Data policy for AI

This topic addresses the role of data policy in existing and potential concerns surrounding the access, usage, sharing, and storage of data for AI. It outlines existing data policies and guidelines on AI and identifies potential gaps and areas for improvement, considering ethical, legal, and social implications for handling data about AI, including data quality, data privacy, and data security. Data policy for AI addresses the common concerns surrounding data in AI focused on promoting the responsible collection, usage, storage, and sharing of data to be used in AI systems. Data governance policies ensure that AI systems have access to high-quality, reliable, interoperable, and relevant data while establishing protections for individual and community privacy, preventing bias and discrimination and fostering transparency and accountability, and promoting data governance in AI applications.

These policies define rules and procedures for accessing and sharing data, particularly in cases where AI systems require diverse and representative datasets. AI has a high dissemination capacity that increases its role in generating scientific knowledge and data analysis in diverse sectors and situations. The diversity of sectors and situations may pose some specific ethical or social problems to the use of AI. This topic examines how these common and specific issues are addressed in data policy. It discusses how data policy can promote open data initiatives and encourage data sharing among organizations, fortifying the role of AI in research, development, and data management. Data policy for AI can also contribute to the development and adoption of explainable AI techniques, ensuring accountability, understanding, and trustworthiness in AI applications.

Of crucial importance is the role of data policy in promoting ethical considerations in AI development and deployment, helping to mitigate potential risks and ensure that AI systems respect individual and collective rights as well as societal values, particularly in terms of Open Science, while obtaining maximum benefit through its use. The transformational importance of AI and the close relation of AI issues to data policy means that this topic cuts across several other action areas.

Initial activities

- A. The initial scope of this activity will be set by an article that overviews existing data policies for AI, mapping gaps and areas for improvement of data curation for use in AI.

- B. A further activity is to identify how the data policy can be applied to AI in order to contribute to other IDPC topics, including crises, education, open science, and the publication and communication of science.

5. Data policy for Open Science

This topic focuses on the role of data policy in advancing Open Science with the aim of ensuring that scientific research and data are openly accessible, transparent, reusable, reliable, and ethical. Data policy provides guidance on rules, standards, best practices, and conditions for sharing research data in an Open Science context. It contributes to guidelines for researchers to make their data openly available to the scientific community and the public. Data policy encourages data sharing through repositories, data archives, and other platforms, facilitating collaboration, reproducibility, and the advancement of scientific knowledge. It outlines practices and standards for effective data management and documentation while encouraging researchers to adopt best practices for data organization, metadata creation, data versioning, and long-term preservation. This contributes to supporting data management and ensuring that research data is findable, accessible, interoperable, and reusable (FAIR), trustworthy, and ethical while enabling other researchers to discover and utilize the data effectively. It promotes data sharing and integration across disciplines, enabling researchers from different domains to leverage diverse datasets for novel insights and discoveries. Data policy is also critical for establishing guidelines, standards, and principles for the ethical, responsible, and transparent use of AI in the context of Open Science. Data policy may encourage interdisciplinary collaboration, data harmonization, and the development of data interoperability standards to facilitate knowledge exchange and interdisciplinary research endeavors.

Initial activity

The IDPC will collaborate with the colleagues at the University of Virginia School of Data Science in the development of OS data policies, leading to the publication of evidence-based recommendations for the OS community and data science institutes.

6. Data policy for the publication and communication of science

By promoting open and transparent data and AI practices, the IDPC will contribute to enhancing science communication by promoting researchers', science communicators', and the public's access to clear and interpretable data and transparency in the use of AI that fosters collaboration, knowledge dissemination, and evidence-based discussions. In this context, the IDPC will promote data and AI diplomacy to facilitate international collaboration in scientific research and communication. Data and AI diplomacy will be used to contribute to establishing agreements and frameworks for data sharing, fostering the harmonization of data standards, and promoting cooperation in data and AI endeavors. Data diplomacy is also of importance in helping to shape policies on data and AI governance.

The IDPC will help to develop data policies that promote the accurate and full representation of available data and use of AI, encouraging scientists, authors, reporters, and policymakers to report fully on the data they have accessed and the limitations of the data. By promoting transparency, accountability, accuracy, and the free flow of information, the IDPC will contribute to the reliability of data used in science, the media, and political decision-making and dissuade the use of scientific data for mis/disinformation, censorship, and propaganda.

Initial activity

The IDPC will engage the topic of research assessment reform by contributing to a CoARA working group on 'Ethics and Research Integrity Policy in Responsible Research Assessment for Data and Artificial Intelligence'.

Recent engagements

The CODATA IDPC has most recently developed activities around Topic 1. Data policy for data quality, reliability, and integrity; Topic 2. Data policy for science in crisis situations; Topic 3. Education for data policy; and Topic 6. Data policy for the publication and communication of science principally active of these, so far, has been that of data policy for science in crisis situations. With regard to Topic 2, two workshops were carried out ([Data Policy in Times of Crisis](#) at the FAIR Convergence Symposium, Leiden, 24 October 2022; [Data Policy for Open Science in Disaster Situations](#) at the EOSC Symposium, Prague, 16 November 2022) as well as a thematic strand at the UNESCO, ISC, CODATA and WDS event [Towards a FAIRer World: Implementing the UNESCO Recommendation on Open Science to address global challenges](#), Paris, 29 March 2023; the Roundtable Session on ‘Data policy for addressing data dilemmas in crisis situations: Preparing for digital sustainability in science’ at the [United Nations World Data Forum](#) satellite meeting at the United Nations University in Macau: ‘[Dealing with Data Dilemmas](#): Towards a human-centered systems approach to sustainable data and digital technology development’ 25 April 2023; and the Global Open Science Cloud (GOSC) SDG-13 [Take urgent action to combat climate change and its impacts] Case Study Working on 16-18 May 2013, Bangkok, Thailand. See this [link](#). The IDPC is now developing a collection of articles for the Data Science Journal on Data Policy in Times of Crisis and will contribute to the development of a UNESCO Toolkit on the same topic in the context of UNESCO’s Declaration on Open Science (2021).

Priority future activities

The CODATA IDPC will give priority to these future activities:

1. developing recommendations for data quality, reliability, and integrity;
2. continuing the current work on data policy for science in crisis situations including the project to contribute to the UNESCO Open Science Toolkit with a factsheet, guidance and a checklist to support data policy in times of crisis;
3. refining, testing, and publishing tools for education in data and AI policy;
4. exploring the issues of data policy for open data, particularly in relation to research assessment and AI;
5. examining how CODATA can usefully contribute to the issue of data policy for specific populations.

In addition, the CODATA IDPC will:

- A. engage with and support, where appropriate, the CODATA Data Ethics Working Group, and the GOSC Data Policy Working Group as well as collaborating closely with the CODATA Institution and National Member Associations;
- B. develop an IDPC fellowship program;
- C. respond on behalf of CODATA to external data policy enquiries and consultations; and
- D. explore opportunities for funded activities and reports in the data policy space.

Milestones and measures of success

The IDPC will work to achieve the following outputs:

- publish recommendations on data quality, reliability, and integrity;
- deliver training on data policy, including the development of a fellowship program in the CODATA IDPC;
- develop a strategy for data policy for open data, particularly in relation to research assessment;

- develop a strategy for a CODATA contribution to the issue of data policy for specific populations; and
- develop recommendations and a framework for data policy for science in crisis situations, including a collection of articles on the topic.