WorldFAIR

Project Title	Global cooperation on FAIR data policy and practice
Project Acronym	WorldFAIR
Grant Agreement No	101058393
Instrument	HORIZON-WIDERA-2021-ERA-01
Topic, type of action	HORIZON-WIDERA-2021-ERA-01-41 HORIZON Coordination and Support Actions
Start Date of Project	2022-06-01
Duration of Project	24 months
Project Website	http://worldfair-project.eu

D2.1 'FAIR Implementation Profiles (FIPs) in WorldFAIR: What Have We Learnt?'

Deliverable	'FAIR Implementation Profiles (FIPs) in WorldFAIR: What Have We Learnt?' report on D2.1 'Completed FAIR Implementation Profiles for each Case Study'
Work Package	WP02 - Recommendations, Synthesis, FAIR Assessment
Lead Author (Org)	Arofan Gregory (CODATA) and Simon Hodson (CODATA)
Contributing Author(s) (Org)	Lesley Wyborn (Australian National University) Laura Molloy (CODATA)

Due Date	30.11.2022
Date	29.11.2022
Version	1.0 DRAFT NOT YET APPROVED BY THE EUROPEAN COMMISSION
DOI	https://doi.org/10.5281/zenodo.7378109

Dissemination Level

- [X] PU: Public
- [ ] PP: Restricted to other programme participants (including the Commission)
- [ ] RE: Restricted to a group specified by the consortium (including the Commission)
- [ ] CO: Confidential, only for members of the consortium (including the Commission)

Versioning and contribution history

Version	Date	Authors	Notes
0.9	22.11.2022	Arofan Gregory, Simon Hodson	Draft for internal review
1.0	29.11.2022	Arofan Gregory, Simon Hodson	Content ready

Disclaimer

WorldFAIR has received funding from the European Commission's WIDERA coordination and support programme under the Grant Agreement no. 101058393. The content of this document does not represent the opinion of the European Commission, and the European Commission is not responsible for any use that might be made of such content.

Abbreviations and Acronyms

FAIR	Findable, Accessible, Interoperable, Reusable
DDI	Data Documentation Initiative
FIP	FAIR Implementation Profile

'Global cooperation on FAIR data policy and practice' (WorldFAIR) has received funding from the European Union's Horizon Europe project call HORIZON-WIDERA-2021-ERA-01-01, grant agreement 101058393.

Executive Summary

This report gives a brief overview of the experience of the WorldFAIR project in using FAIR Implementation Profiles (FIPs). It describes the WorldFAIR project, its objectives and its rich set of Case Studies; and it introduces FIPs as a methodology for listing the FAIR implementation decisions made by a given community of practice. Subsequently, the report gives an overview of the initial feedback and findings from the Case Studies, and considers a number of issues and points of discussion that emerged from this exercise. Finally, and most importantly, we describe how we think the experience of using FIPs will assist each Case Study in its work to implement FAIR, and will assist the project as a whole in the development of two key outputs: the Cross-Domain Interoperability Framework (CDIF), and domain-sensitive recommendations for FAIR assessment.

We hope this report will be of interest to data experts who want to find out more about the WorldFAIR project, its remarkable and diverse array of Case Studies, and about FIPs. It is important to stress that this report does not set out to give a comprehensive appraisal of the FIPs approach and could not do so. All the WorldFAIR Case Studies have developed an initial FIP, but the process of reflection on practice will continue throughout the project. Each Case Study will complete at least one further FIP, and in some cases more than one, towards the end of the project and this will enrich our understanding of the utility of the approach. At that stage, we intend to be able to incorporate some robust prospective and aspirational considerations, and we need to consider how best to represent this in the FIPs.

As noted above, the final section of this report looks forward to the development of the CDIF and domain-sensitive recommendations for FAIR assessment. On both these counts, we consider that the FIPs approach has helped considerably:

- For the CDIF, through helping refine our initial functional analysis of the requirements for cross-domain FAIR, and—as predicted—helping identify *some* candidate cross-domain standards.
- For the FAIR assessment recommendations, through demonstrating that the FIPs can provide an empirical basis for such recommendations, reflecting both the current practice, and the aspirations of a given community or research domain.

'Global cooperation on FAIR data policy and practice' (WorldFAIR) has received funding from the European Union's Horizon Europe project call HORIZON-WIDERA-2021-ERA-01-01, grant agreement 101058393.

FAIR Implementation Profiles (FIPs) in WorldFAIR: What Have We Learnt?

Executive Summary 3

1. Introduction and overview of this report 6

2. WorldFAIR and FIPs 7

2.1. What is WorldFAIR? 7

2.2. What are FIPs and why is WorldFAIR interested in them? 8

2.3. What are the WorldFAIR Case Studies and what is their scope? 9

Chemistry (WP03) 10

Nanomaterials (WP04) 10

Geochemistry (WP05) 11

Social Surveys (WP06) 11

Population Health (WP07) 11

Urban Health (WP08) 12

Biodiversity (WP09) 12

Agricultural Biodiversity (WP10) 13

Ocean Science (WP11) 14

Disaster Risk Reduction (WP12) 14

Cultural Heritage (WP13) 15

2.4. What did WorldFAIR do with FIPs? 15

3. Feedback from the Case Study experiences 17

3.1. General considerations 17

'Global cooperation on FAIR data policy and practice' (WorldFAIR) has received funding from the European Union's Horizon Europe project call HORIZON-WIDERA-2021-ERA-01-01, grant agreement 101058393.

3.2. Definition of 'Community of Practice' 18

3.3. Clarity and focus around FAIR 19

3.4. Standards, conventions, and specific implementations 19

4. Issues and discussions 21

4.1. Overview 21

4.2. Support across the FAIR principles 21

4.3. Identifiers and controlled vocabularies: elements of convergence 22

4.4. Communities and community standards 24

5. Impact on cross-domain FAIR 26

5.1. FAIR within and across domains 26

5.2. Convergence on standards and agreement on implementation 27

5.3. Addressing gaps in functional FAIR support 29

5.4. Describing and assessing FAIR 30

6. Conclusions and next steps 33

'Global cooperation on FAIR data policy and practice' (WorldFAIR) has received funding from the European Union's Horizon Europe project call HORIZON-WIDERA-2021-ERA-01-01, grant agreement 101058393.

1. Introduction and overview of this report

This report gives a brief overview of the experience of the WorldFAIR project in using FAIR Implementation Profiles (FIPs). It describes the WorldFAIR project, its objectives and its rich set of Case Studies; and it introduces FIPs as a methodology for listing the FAIR implementation decisions made by a given community of practice. Subsequently, the report gives an overview of the initial feedback and findings from the Case Studies, and considers a number of issues and points of discussion that emerged from this exercise. Finally, and most importantly, we describe how we think the experience of using FIPs will assist each Case Study in its work to implement FAIR, and will assist the project as a whole in the development of two key outputs: the Cross-Domain Interoperability Framework (CDIF), and domain-sensitive recommendations for FAIR assessment.

We hope this report will be of interest to data experts who want to find out more about the WorldFAIR project, its remarkable and diverse array of Case Studies, and about FIPs. It is important to stress that this report does not set out to give a comprehensive appraisal of the FIPs approach and could not do so. All the WorldFAIR Case Studies have developed an initial FIP, but the process of reflection on practice will continue throughout the project. Each Case Study will complete at least one further FIP, and in some cases more than one, towards the end of the project and this will enrich our understanding of the utility of the approach. At that stage, we intend to be able to incorporate some robust prospective and aspirational considerations, and we need to consider how best to represent this in the FIPs.

As noted above, the final section of this report looks forward to the development of the Cross-Domain Interoperability Framework (CDIF), and domain-sensitive recommendations for FAIR assessment. On both these counts, we consider that the FIPs approach has helped considerably:

- For the CDIF, through helping refine our initial functional analysis of the requirements for cross-domain FAIR, and—as predicted—helping identify *some* candidate cross-domain standards.
- For the FAIR assessment recommendations, through demonstrating that the FIPs can provide an empirical basis for such recommendations, reflecting both the current practice, and the aspirations of a given community or research domain.

The remainder of this short report will set out the parameters and purpose of the WorldFAIR project, the FAIR Implementation Profile approach, and how the former used the latter to better understand current research data practices in relation to the FAIR data principles.

'Global cooperation on FAIR data policy and practice' (WorldFAIR) has received funding from the European Union's Horizon Europe project call HORIZON-WIDERA-2021-ERA-01-01, grant agreement 101058393.

2. WorldFAIR and FIPs

2.1. What is WorldFAIR?

'WorldFAIR: Global cooperation on FAIR data policy and practice' is a two-year project to advance implementation of the FAIR principles,[1] particularly in relation to interoperability. Funded by the European Commission HORIZON-WIDERA-2021-ERA-01-41 project with an explicit mission to advance global collaboration and include partners from outside the European Union, WorldFAIR is coordinated by CODATA, the Committee on Data of the International Science Council, with the Research Data Alliance (RDA) association as a major partner.

The project is conceived as responding to Recommendation 4 of the *Turning FAIR into Reality* report, which identified the need to 'Develop interoperability frameworks for FAIR sharing within disciplines and for interdisciplinary research'. The Recommendation states that 'Research communities need to be supported to develop interoperability frameworks that define their practices for data sharing, data formats, metadata standards, tools and infrastructure. To support interdisciplinary research, these interoperability frameworks should be articulated in common ways and adopt global standards where relevant.'[2]

WorldFAIR is working with a set of eleven domain and cross-domain Case Studies[3], carefully chosen from existing CODATA and RDA activities to provide maximum impact. The Case studies, whilst drawn from diverse domain and cross-domain research areas, are clustered in connected themes in order to maximise scope while retaining a critical mass of activity and allowing learning and cross-fertilisation of ideas. Each Case Study (in project terms, a 'Work Package') will develop an interoperability framework, recommendations and/or a FAIR implementation for their discipline or interdisciplinary research area. Led by CODATA, a coordinating and synthesis activity (Work Package 2) has been supporting each Case Study in understanding their requirements through the completion of FAIR Implementation Profiles, the subject of this report. In turn these insights will be incorporated into the development of a Cross-Domain Interoperability Framework and more domain-sensitive recommendations for FAIR assessment and benchmarks.

[1] Wilkinson, M., Dumontier, M., Aalbersberg, I. et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data 3, 160018 (2016). https://doi.org/10.1038/sdata.2016.18

[2] European Commission, Directorate-General for Research and Innovation, *Turning FAIR into reality : final report and action plan from the European Commission expert group on FAIR data*, Publications Office, 2018, https://data.europa.eu/doi/10.2777/1524; p.29.

[3] Listed at https://worldfair-project.eu/case-studies-of-worldfair/

'Global cooperation on FAIR data policy and practice' (WorldFAIR) has received funding from the European Union's Horizon Europe project call HORIZON-WIDERA-2021-ERA-01-01, grant agreement 101058393.

2.2. What are FIPs and why is WorldFAIR interested in them?

FAIR Implementation Profiles (FIPs) are a methodology, developed by GO FAIR, through which a research community expresses its practices and decisions around the FAIR principles. The approach involves a series of questions on how the community makes data and metadata FAIR and what 'FAIR Enabling Resources' (FERs)[4] are used.

The original insight that led to the FIPs approach emerged from discussions at a meeting of the GO FAIR Implementation Networks in January 2019, and was further developed by the GO FAIR Convergence Matrix and FIPs Working Group, including through co-funded GO FAIR - CODATA workshops around the First FAIR Convergence Symposium in October 2020. The approach has been further developed, and refined with the addition of the online FIP Wizard tool, with support and extensive testing by the ENVRI-FAIR project starting 2019 and continuing to the time of writing.

At the heart of the methodology are a set of questions designed to allow a community to state what FAIR Enabling Resources it uses in response to each of the FAIR principles. The term 'FAIR Enabling Resource' (FER) encompasses any technology, service or tool, specification or standard, that is used in response to one or more of the FAIR principles: depending on the principle concerned the FER could be an identifier service, a metadata schema, a terminology or vocabulary, a language for encoding metadata, an ontology, or any other of the many things that have a role in making data and metadata FAIR. The question set can be used for a process of self-inquiry by a given community to better understand current practice and—potentially—to identify gaps and priority activities where useful to improve their FAIR practices. Such a process can be conducted in a workshop or a series of meetings and can use a simple spreadsheet to record the necessary information and conclusions.

In the FIP Wizard, the responses can be entered into an online tool. The FERs and the FIPs can be expressed as a 'nanopublication', a short RDF description of the resource or the profile which comprises the many resources used by a given community.[5] In turn this means that the graph of FERs and their use by various communities can be visualised and analysed. With use, the creation of more and more FIPs and FERs will furnish a resource which can give useful insights into FAIR practices, help identify FERs that are used across domains, and—it is suggested—encourage greater convergence in FAIR practices.

WorldFAIR's interest in using FIPs is first, as an exercise to prompt enquiry and reflection within each Case Study and discussion among them; second, as a tool to provide a reasonable

[4] See e.g. https://peta-pico.github.io/FAIR-nanopubs/fip/index-en.html#https://w3id.org/fair/fip/terms/FAIR-Enabling-Resource
[5] See https://www.w3.org/TR/trig/

'Global cooperation on FAIR data policy and practice' (WorldFAIR) has received funding from the European Union's Horizon Europe project call HORIZON-WIDERA-2021-ERA-01-01, grant agreement 101058393.

benchmark of current practices. The completed FIP(s) will help the Case Studies identify any gaps and areas requiring further attention. Towards the end of the project, a second FIP(s) will help characterise any progress made in understanding and implementation of FAIR. It is also likely that the WorldFAIR Case Studies will provide useful feedback to improve and refine the FIPs methodology and potentially identify important issues which may not be adequately covered by the current question-set based on the FAIR principles.

The FIPs exercises will also feed into the other WorldFAIR objectives mentioned above: the development of the Cross-Domain Interoperability Framework (CDIF) and recommendations for more domain-sensitive FAIR assessment. In relation to the former, the functional analysis implicit in the FAIR principles and the FIPs question set contributes to our understanding of the issues to be addressed. It seems likely that analysis across the WorldFAIR FIPs will help us identify standards or other FERs that are currently used, or could be used, across domains. In relation to the latter, the FIPs provide an empirical statement of practice in a given community, which can and should be taken into account in guidance, assessment, and benchmarking around FAIR.

2.3. What are the WorldFAIR Case Studies and what is their scope?

The FIP Wizard User Guide describes a FIP as 'a list of declared technology choices, also referred to as FAIR Enabling Resources (FERs), that are intended to implement one or more of the FAIR Guiding Principles, made as a collective decision by the members of a particular community of practice.'[6] A community of practice, in this context, is any group 'whose members agree to the same set of FAIR implementation choices'. In practical terms, this means that FIPs could be usefully completed by different types of communities or entities: by projects, infrastructures, by defined communities of practice, disciplines or sub-disciplines. It is clear that the range of data used in many disciplines (as commonly understood) are too diverse and complex to be usefully described in a single FIP. It became quickly apparent to the WorldFAIR project that the Case Studies should focus on describing the practices of an relatively circumscribed community of practice: those of a project, a community of practice around a given infrastructure, data collection or type or a relatively-focussed research area.

Information about the Case Studies is provided on the WorldFAIR website; the table below summarises this and the focus of the FIPs:

[6] https://osf.io/5ygzx, p. 1.

'Global cooperation on FAIR data policy and practice' (WorldFAIR) has received funding from the European Union's Horizon Europe project call HORIZON-WIDERA-2021-ERA-01-01, grant agreement 101058393.

Chemistry (WP03)

Leads and partners	The International Union of Pure and Applied Chemistry (IUPAC, USA); leveraging an international community of experts through its Committee on Publications and Cheminformatics Data Standards.
Focus of Case Study	IUPAC is the world authority on chemical nomenclature, terminology, and standardised methods of measurement, and is engaging in a concerted effort through collaboration with the broader chemistry and data science communities to translate a range of assets and activities into the digital domain. Aligning standards development and implementation with the FAIR data principles will facilitate development of guidelines, tools and validation services that support scientists to share and store data in a FAIR manner and support the ability to compile and interpret data across scientific disciplines.
Focus of FIP	IUPAC Standards and IUPAC Gold Book (through outcomes from WorldFAIR, IUPAC will make considerable steps to make IUPAC standards and other information assets FAIR; the Gold Book is the IUPAC Compendium of Chemical Terminology, making the Gold Book FAIR will significantly advance FAIR and interoperability in chemistry).

Nanomaterials (WP04)

Leads and partners	University of Birmingham (UK), 7past9 (Germany), Novamechanics (Cyprus); partners in a number of nanomaterials and nanosafety activities, including NanoCommons, and the CODATA Task Group on NanoInChI.
Focus of Case Study	This Case Study will enable the further adoption of the FAIR principles by the international nanomaterials community and encourage greater alignment with neighbouring disciplines and communities. It builds on the partners' successful collaboration in NanoCommons (a research infrastructure for nanoinformatics and FAIR nanomaterials data) and their leadership of the IUPAC InChI Trust efforts to develop a standard extension of the InChI for nanomaterials.
Focus of FIP	European Nanotechnology Data, Knowledge and Informatics Community (Started by the NanoCommons infrastructure project, the NanoCommons community infrastructure is meant to be the data and knowledge hub of the EU NanoSafety Cluster coordinating nanosafety research in Europe. Led by three core partners being part first in the NanoCommons project and now in the WorldFAIR project, it is connecting all stakeholders from current

'Global cooperation on FAIR data policy and practice' (WorldFAIR) has received funding from the European Union's Horizon Europe project call HORIZON-WIDERA-2021-ERA-01-01, grant agreement 101058393.

nanomaterials and new advanced materials projects to foster knowledge exchange and harmonisation of the data ecosystem.)

Geochemistry (WP05)

Leads and partners	AuScope (the Australian Geoscience Research Infrastructure, Australia); leveraging an international community of experts through OneGeochemistry.
Focus of Case Study	Through OneGeochemistry, an informal international network of national geochemical data infrastructure organisations, the geochemistry community seeks to define the minimum common variables for a set of geochemical data types and build them into FAIR Implementation Profiles, that can also be used by laboratories/ repositories/publishers for QA/QC validation of data.
Focus of FIP	OneGeochemistry (Geochemistry community resource collection enabling creation of FAIR implementation profile and other FAIR enabling resources). Draft FIPs have also been prepared for a range of Geochemistry Research Infrastructures.

Social Surveys (WP06)

Leads and partners	Australian Data Archive (Australia) and Norwegian Agency for Shared Services in Education and Research (SIKT, Norway); the organisations that curate the Australian Social Survey (AUSSI-ESS) and the European Social Survey (ESS), respectively.
Focus of Case Study	This Case Study will undertake a comparative study of the data management, harmonisation and integration practices of AUSSI-ESS and the ESS, an ERIC social science infrastructure. It will then leverage the DDI (Data Documentation Initiative) metadata standards to understand how such multi-national collections could be made increasingly interoperable and reusable through shared procedural and technical development, and establish a set of guidelines and tools for the development of cross-national collections into the future.
Focus of FIP	AUSSI-ESS and ESS. Social survey community resource collection.

Population Health (WP07)

Leads and partners	London School of Hygiene and Tropical Medicine (LSHTM, UK), and CODATA, (France) are partners in the INSPIRE Project.
Focus of Case Study	The Implementation Network for Sharing Population Information from Research Entities (INSPIRE) project is assembling technologies and standards in support of a data hub that facilitates federated and/or shared research capable of interoperating across often-neglected low-resource

'Global cooperation on FAIR data policy and practice' (WorldFAIR) has received funding from the European Union's Horizon Europe project call HORIZON-WIDERA-2021-ERA-01-01, grant agreement 101058393.

	settings: it aims to provide a platform-as-a-service, which can make data of disparate types available to many different styles of analysis, among which AI systems are increasingly prominent.
Focus of FIP	INSPIRE Network (supports health data from population cohorts and public health responses).

Urban Health (WP08)

Leads and partners	Drexel University (USA), lead of the SALURBAL Project.
Focus of Case Study	The SALURBAL project (Urban Health for Latin American cities) is a five-year project based at the Urban Health Collaborative, Drexel University, and with partners throughout Latin America and in the United States that studies how urban environments and urban policies impact the health of residents from almost 370 cities in 11 Latin American countries. To pursue this goal, the SALURBAL project 1) has systematised a process for city definition and operationalization that integrates multiple ways in which a city can be delimited; 2) has created a data structure that allowed the incorporation of data from different sources, making it shareable across several cores and disciplines; and 3) has developed procedures and standards that systematically documented issues related to data access, quality, and completeness during the process of data harmonisation. The Case Study will explore and further refine this approach to provide recommendations for urban health data that reflect the FAIR and CARE principles and contribute to promote best practices in data sharing and use within and beyond the Urban Health field.
Focus of FIPs	The Urban Health Case Study created two FIPs: the first describing the SALURBAL Data Portal, serving a community derived from LAC-Urban Health Network. It is based on the work done at the SALURBAL project (Urban Health in Latin America) focused on collecting, harmonising, and using data on the natural, built, and social environment as well as population health in 370 cities of 100,000 or more resident in 11 countries (Argentina, Brazil, Chile, Colombia, Costa Rica, El Salvador, Guatemala, Mexico, Nicaragua, Panama, Peru). The second FIP covers the overall Urban Health Discipline.

Biodiversity (WP09)

Leads and partners	GBIF (Denmark), the Global Biodiversity Information Facility.
Focus of Case Study	As a leader in open and reusable data before the FAIR principles were developed, GBIF has an influential role in generating FAIR data throughout

'Global cooperation on FAIR data policy and practice' (WorldFAIR) has received funding from the European Union's Horizon Europe project call HORIZON-WIDERA-2021-ERA-01-01, grant agreement 101058393.

	biodiversity sciences. All data mobilised through the GBIF network is completely free and open access with all metadata easily available. Nevertheless, the biodiversity community has determined that it is time to revisit GBIF's core data model so that it can better implement FAIR data integration, especially in connection to the growing Digital Extended Specimen initiative. It is also becoming critically important to meet the growing need to enable the integration of new data types, especially DNA-derived data, and to better connect the growing long-term biodiversity monitoring data that isn't currently well connected to the global data network. Through FAIR recommendations and assessments, and a new FAIR interoperability framework at the heart of the GBIF data model, our goal is to improve our processes and thereby help the global biodiversity community to implement FAIR data standards.
Focus of FIP	GBIF Data Model (the Global Biodiversity Information Facility is an international network and data infrastructure funded by the world's governments and aimed at providing anyone, anywhere, open access to data about all types of life on Earth).

Agricultural Biodiversity (WP10)

Leads and partners	Led by Brazilian Agricultural Research Corporation (EMBRAPA, Brazil), with partners the Kenya Agricultural and Livestock Research Organisation (KALRO, Kenya), Meise Botanic Garden (Belgium), HiveTracks (USA); leveraging the RDA IGAD Community of Practice on Agricultural Data[7].
Focus of Case Study	Plant-pollinator interactions are recognized for their key role in ecosystem functioning and sustainable agriculture. However, plant-pollinator data are currently stored in silos across multiple networks and country-specific initiatives. The capacity to integrate those data at regional and global levels is crucial to enable pattern analysis and understanding at biologically-relevant scales. In this context, adoption of community data standards on pollination and good practices is urgently needed. This Case Study will ensure broad participation and alignment with other agricultural data initiatives in Europe and at the global level to facilitate the implementation of the FAIR data principles.
Focus of FIP	Plant-pollinator interaction data.

[7] Improving Global Agricultural Data (IGAD) Community of Practice - https://www.rd-alliance.org/groups/igad-community-practice

'Global cooperation on FAIR data policy and practice' (WorldFAIR) has received funding from the European Union's Horizon Europe project call HORIZON-WIDERA-2021-ERA-01-01, grant agreement 101058393.

Ocean Science (WP11)

Focus of Case Study	Leads and partners Alfred Wegener Institute for Polar and Marine Research (AWI, Germany), leveraging the international community around the Ocean Data Information System (ODIS).
Focus of Case Study	With the onset of the UN Decade of Ocean Science for Sustainable Development, a surge of activity focusing on ocean observation, operations, commerce, socio-economics, and culture will generate data with greater complexity, depth, and volume than ever before. There is a considerable need to align the policies and practices across independent technologies and systems. To this end, this Case Study will leverage the progress made by the Ocean InfoHub and Ocean Data and Information System (OIH, ODIS[8]), launched by International Oceanographic Data and Information Exchange (IODE) of the Intergovernmental Oceanographic Commission (IOC) of UNESCO. The AWI/Helmholtz partner (whose personnel chaired the technical implementation of ODIS) will examine how the ODIS Interoperability Architecture (ODIS-Arch) being piloted with regional partners can be coordinated with other Case Studies and central guidelines of CODATA and RDA to support digital policy alignment. The key objective will be to ensure policies support regional and local specificity, but allow the concrete implementation of global FAIRness around key (meta)data types. Through these actions, this Case Study aims to sustainably interface the ODIS digital ecosystem with many others.
Focus of FIP	ODIS Architecture.

Disaster Risk Reduction (WP12)

Focus of Case Study	Leads Tonkin+Taylor (New Zealand), leveraging partners in UN Disaster Risk Reduction-International Science Council (UNDRR-ISC) Expert Groups and the CODATA Task Group on FAIR Data for Disaster Risk Research.
Focus of Case Study	The UNDRR-ISC Sendai Hazard Definition and Classification Review Technical Report supports these agreements by providing a common set of hazard definitions for monitoring and reviewing implementation which calls for ‘a data revolution, rigorous accountability mechanisms and renewed global partnerships’. An important step will be to make this vocabulary FAIR following the guidelines in the Ten Simple Rules.[9] Advances in technology

[8] See https://oceaninfohub.org/, https://www.odis.org/

[9] Cox SJD, Gonzalez-Beltran AN, Magagna B, Marinescu M-C (2021) Ten simple rules for making a vocabulary FAIR. PLoS Comput Biol 17(6): e1009041. https://doi.org/10.1371/journal.pcbi.1009041

‘Global cooperation on FAIR data policy and practice’ (WorldFAIR) has received funding from the European Union’s Horizon Europe project call HORIZON-WIDERA-2021-ERA-01-01, grant agreement 101058393.

	have enabled a dramatic increase in the availability of satellite imagery and the power of geospatial services for DRR, yet significant challenges remain for the effective operationalization of these data for practical purposes, including societal use, policy making, rapid response etc. The FAIR principles are critical to facilitating the use of advanced technologies to extract pertinent information for DRR and climate adaptation and resilience.
Focus of FIP	Disaster risk reduction data focusing on Fiji and Tonga Case Studies.

Cultural Heritage (WP13)

Leads and partners	Digital Repository of Ireland (DRI, Ireland), leveraging a wider community of experts on cultural heritage data.
Focus of Case Study	Cultural Heritage data emerges primarily from the cultural sector (not the research sector), but provides the input into research for a range of humanities disciplines, making this Case Study itself multidisciplinary and multisectoral. The sharing of visual sources in particular has challenges around copyright, but also increasingly around formats, with the emergence of 3D data. More generalised issues around metadata standards, vocabularies, digital preservation and persistent identifiers also exist, and humanities disciplines have comparatively less-developed data sharing cultures. Several global image-sharing communities/platforms exist online. These communities provide massive (but not very FAIR) datasets and crucial networks for coordination. The Digital Repository of Ireland, a CTS-certified TDR for AHSS data that plays a leading role in FAIR globally, will work with these communities to understand what practices exist, and how they could collectively be made more FAIR.
Focus of FIP	Cultural heritage image data (image sharing practices of cultural heritage institutions and services with the goal of making recommendations for improving and aligning existing image sharing platforms and standards, and developing a framework for FAIR assessment and benchmarking in cultural heritage).

2.4. What did WorldFAIR do with FIPs?

The CODATA WP02 team participated in an in-person knowledge transfer workshop run by the GO FAIR Foundation on 5 July 2022. The following day, the CODATA and GO FAIR teams jointly gave an online briefing to the whole WorldFAIR project and Case Studies. This provided an introduction to FIPs, an overview of the activity and links to documentation. Five intensive online workshops were run with the Case Studies in August and September 2022, providing opportunity for worked examples with the FIPs spreadsheet, demonstrations of the FIP Wizard, discussions and questions.

'Global cooperation on FAIR data policy and practice' (WorldFAIR) has received funding from the European Union's Horizon Europe project call HORIZON-WIDERA-2021-ERA-01-01, grant agreement 101058393.

All Case Studies completed their FIPs spreadsheets by mid-October, prior to a workshop held on 25 October as part of the Second FAIR Convergence Symposium in an intensive week of events in Leiden, which also included the First FAIR Digital Objects Conference.

The workshop 'FAIR Implementation Profiles (FIPs) in WorldFAIR: what have we learnt?' was attended by 50 people in person and 120 online. The event was introduced by Simon Hodson, Executive Director, CODATA; six of the eleven Case Studies described their experiences of preparing a FIP and the remainder participated in the discussions both in person and online.

The Case Study presenters were:

1. Chemistry, IUPAC (WP03): Leah McEwen (online), Ian Bruno and Stuart Chalk (in person)
2. Nanomaterials (WP04): Iseult Lynch and Thomas Exner (in person)
3. Social Surveys (WP06): Steve McEachern and Hilde Orten (online)
4. Agricultural Biodiversity (WP10): Maarten Trekels (in person) and Debora Drucker (online)
5. Disaster Risk Reduction (WP12): Bapon Fakhruddin and Jill Bolland (online)
6. Cultural Heritage (WP13): Beth Knazook (in person).

In their presentations, each Case Study was asked to respond the following questions:

1. What have you learnt from the process?
2. Has using FIPs helped you describe practices around FAIR in your Case Studies?
3. Has it helped identify any gaps or areas which would benefit from further attention?
4. Has the process identified ways in which the FIPs methodology and the tools around it can be improved?
5. What have you learnt about the FAIRness of your community or domain?
6. Have you identified any next steps in response to what you have learnt?

Prompted by a presentation from Arofan Gregory, Standards Expert with CODATA, outlining our findings and the implications for our work on the Cross-Domain Interoperability Framework (CDIF), the discussion covered the following topics:

1. How have the outcomes of the FIPs assisted the project in the development of a Cross-Domain Interoperability Framework and recommendations for more domain-sensitive FAIR assessment?
2. What commonalities have we identified?
3. Has the process helped our identification of components of CDIF and candidate standards?
4. Have we identified any specific needs in domains that should be part of domain sensitive FAIR assessment?
5. What are the key findings about FIPs as a methodology?
6. What improvements would we recommend?

The outcomes of these discussions and the initial findings from WorldFAIRs FIPs activity are summarised in the remainder of this report.

'Global cooperation on FAIR data policy and practice' (WorldFAIR) has received funding from the European Union's Horizon Europe project call HORIZON-WIDERA-2021-ERA-01-01, grant agreement 101058393.

3. Feedback from the Case Study experiences

3.1. General considerations

This section will summarise the range of responses from different Case Studies within WorldFAIR regarding the creation of FAIR Implementation Profiles (FIPs). The range is quite broad, based on the relative focus and scope of the groups. Specific issues are described in the next section, but it is useful to have a general sense of the types of responses coming from the different Case Studies.

It is useful to provide a rough characterisation of the Case Studies in order to better understand their responses. Some of the Case Studies are fairly focused in their remit: they covered a community conducting research in a well-circumscribed field, and thus were considering practice within the data and resources of concern to a relatively narrow range of users. One good example of this is SALURBAL (Urban Health, WP08), where—although the data being used comes from a variety of disciplines (public health, economics, demographics)—the integrated data of interest to the community is reasonably well-bounded. For our purposes we will describe such focused communities and the information systems they share as individual 'infrastructures', for which a single accepted set of standards, models, and interfaces can reasonably be supported.

At the other end of the spectrum, some of the Case Studies were looking at large-scale interoperability and reuse of data and metadata across a large group of infrastructures. Here, the range of different types of data and metadata is much broader, reflecting the needs of a wider scope in terms of research. There will already be a broad range of standards, models, and interfaces in use by the various infrastructures, and these will themselves need to be harmonised to support interoperability across them. The single best example of this can be seen in the Oceans InfoHub ODIS work (Ocean Sciences, WP11): the approach is one which covers a very broad range of infrastructures, and the 'meta' approach of the Case Study reflects this.

This breadth of scope is only one dimension in considering the variation across Case Studies. Another is the relative state of maturity in terms of data reuse and FAIR more generally. For some Case Studies, issues of standardisation and reuse are relatively new, while in others they have received a lot of attention, with different projects focused specifically on these issues for many years in the past. Darwin Core[10] is a good example of an effort which was significant to more than one of the Case Studies (Biodiversity, WP09, and Agricultural Biodiversity, WP10): it is the result of long-standing effort to better support interoperability and exchange across infrastructures in the biological domains. The existence of a strong community working on biological information

[10] https://www.tdwg.org/standards/dwc/

'Global cooperation on FAIR data policy and practice' (WorldFAIR) has received funding from the European Union's Horizon Europe project call HORIZON-WIDERA-2021-ERA-01-01, grant agreement 101058393.

standards leads to continuous improvement of existing standards in order to better describe the several aspects related to biodiversity data and, ultimately, to allow for cross-disciplinary research related to the sustainability of life on earth.

Across both of these dimensions, the Case Studies represent a spectrum of positions, and consequently reacted to the production of FIPs differently. The relatively limited depth of FIPs—a single set of questions asking for specific types of resources—guarantees that different communities will find them more or less suitable (and challenging) according to their own experiences, and this is naturally reflected in the responses from different groups.

3.2. Definition of 'Community of Practice'

One of the most common responses from the various Case Studies regarding the FIPs exercise was a request for guidance regarding who was covered in the scope of the activity. FIPs are posited on the idea that communities of practice are self-defining: they consist of that set of researchers and those who support them concerned with a given pool of resources. It is part of the FIP process to define (briefly) how a given community being described in the FIP would be bounded.

In some cases, a Case Study decided that it could usefully describe more than one community. Some chose to focus quite narrowly (as the SALURBAL example) and some very broadly (as for OIH/ODIS). In all cases, this bounding affected what was possible within the questions provided by the FIP: for a narrower community, specific answers can be provided with more confidence; for a broader community, this is less a description of fact than an indication of desired practice.

In the FIPs documentation and Wizard, the term FAIR Enabling Resource (FER) is used to cover chosen technologies which may range from widely adopted 'standards', to more narrowly-used, or even internal project, solutions. Yet, within and across domains, for widespread, global interoperability to be realised, a significant degree of harmonisation and the adoption of 'standards' will be necessary. No specific guidance is given in the FIPs documentation as to what makes a 'standard', nor how the degree of adoption can be measured. The Case Studies provide indications of the status of FERs, and whether they can be considered 'standards' in their FIPs implementation notes. Nevertheless, the issue of how to indicate that there were ongoing discussions and processes of harmonisation in some communities came up. It was felt by some of the Case Studies that a lot of effort encouraging adoption of, and assuring compliance with, existing guidelines around interoperability were being glossed over by the FIPs approach, and would require more focus in subsequent work to reflect the actual state of play within the community.

In some Case Studies, it became clear that some existing infrastructures have a mandate to serve a much broader community outside of academic researchers in a particular discipline (e.g. the general public, the national interest) and were therefore likely to stick with existing solutions for the problems

'Global cooperation on FAIR data policy and practice' (WorldFAIR) has received funding from the European Union's Horizon Europe project call HORIZON-WIDERA-2021-ERA-01-01, grant agreement 101058393.

of reuse and exchange that serve those designated communities well. Those same infrastructures may also be unlikely to pursue alternatives for reusability without considerable funding, training, or practical support within the professional culture in which they operate. This was visible in the Cultural Heritage Case Study (WP13), where simultaneous support for many different approaches was seen as the best solution to the range of different infrastructure implementations. In the Nanomaterials Case Study (WP04) a similar situation existed: even if the need for sharing is globally acknowledged, the history of individual infrastructures developed independently with specific information required to describe the holdings built for specific purposes or projects, and partly incompatible approaches for FAIRification, actually presented a barrier to broader sharing. It is difficult to know how to define a 'community' beyond the level of individual infrastructures in these cases, and yet that was seen as a goal from the perspective of the Case Studies broadly, as consonant with the FAIR principles.

3.3. Clarity and focus around FAIR

Some of the Case Studies found that the FIPs exercise was a good way to bring the issues around data sharing into clearer focus. While many research support organisations have long shared the underlying aspirations reflected in FAIR, the FIP activity provided them with a structure which allowed a systematic approach to the discussions and reflection on practices which had been lacking before.

In those areas which were relatively less mature in terms of data interoperability and reuse, this was more pronounced. The more advanced Case Studies had already either used FAIR or their own methods for addressing issues of interoperability and reuse. In most cases, however, even if FIPs were seen as only a starting point, the focus and clarity provided by the exercise was welcome. The basic question was 'Where do we start when we are trying to implement FAIR' and FIPs provide a relatively simple and easy answer.

3.4. Standards, conventions, and specific implementations

There were many questions about what constituted an FER for the purposes of answering specific questions. In some cases there is a complex relationship between standards, services, and technology approaches which made it difficult for the Case Studies to be confident that their responses were appropriate for specific questions. Many of the standards or technologies (FERs) used in the Case Studies cover more than one of the functions identified in the FIPs typology of FERs: for example, a semantic model and a schema are often combined in a single specification, and this can produce a lack of clarity for those who are not specialists in these matters.

'Global cooperation on FAIR data policy and practice' (WorldFAIR) has received funding from the European Union's Horizon Europe project call HORIZON-WIDERA-2021-ERA-01-01, grant agreement 101058393.

The orientation of FIPs seems to be fundamentally based on a specific technology perspective, and given the range of different experiences across the Case Studies, the terms used in the documentation for the FIPs and the Wizard tool were sometimes understood in different ways. Specifically, there appears to be, in some parts of the FAIR community, a presumption of the ubiquity of Linked Data technologies (RDF, etc.), which is not necessarily the case in all of the research communities in WorldFAIR.

In some Case Studies the dominance of a single infrastructure has driven the adoption of their interfaces and models by others in the community, producing a de facto standard. In some cases, competing solutions would provide more than one possible candidate, without any mechanism for determining which of these was preferable. This was more of an issue with the Case Studies where the relative maturity of data sharing was lower, as evidenced by the lack of formal organisations or agreements for the governance of a recognized standard. In one case study (Geochemistry, WP05) there are no agreed community standards compliant with FAIR.

'Global cooperation on FAIR data policy and practice' (WorldFAIR) has received funding from the European Union's Horizon Europe project call HORIZON-WIDERA-2021-ERA-01-01, grant agreement 101058393.

4. Issues and discussions

4.1. Overview

The feedback from the Case Studies regarding FIPs reflected their experience of addressing the specific issues raised according to the FAIR principles themselves. The information provided by the Case Studies in their responses provides some insight into the breadth of maturity and the specific issues which will be faced when large-scale implementation of systems supporting the FAIR principles is undertaken. This section summarises some of the topics raised in thinking not about the FIPs methodology, but about what the responses tell us about the current state of play across the Case Studies.

It should be noted that FIPs can be used for different purposes, and that some of the early FIPs work focused on trying to drive convergence of approaches to data management, dissemination, and reuse (notably in the GO FAIR and ENVRI-FAIR work with FIPs). WorldFAIR did not use this approach, but instead used the first round of FIPs as more of an empirical tool for exploring the state of FAIRness within and across the Case Studies, in a regular (and thus comparable) fashion. The respondents to the FIPs questionnaire were encouraged to describe the current situation as accurately as possible, with the prospect of a second FIP being performed at the end of the project to focus prospectively on convergence and other aspects of increasing FAIRness within their communities.

While the results of this exploration were not particularly surprising, they did provide specific, detailed input which is very useful. This section summarises some of the findings, and highlights important issues.

4.2. Support across the FAIR principles

It is generally accepted that some of the FAIR principles are easier to implement in practical terms than others, and this is very much reflected in the FIPs produced by the WorldFAIR Case Studies. In the broadest terms, Findability is well supported, while other aspects of FAIR less so. Interoperability and Reuse are generally seen as a ‘later stage’ (although this is distinctly not true across all the Case Studies and the exceptions are notable). Accessibility is often reasonably well supported, not because of any particular focus by the communities of practice, but because it is a precondition for many forms of dissemination independent of FAIR.

‘Global cooperation on FAIR data policy and practice’ (WorldFAIR) has received funding from the European Union’s Horizon Europe project call HORIZON-WIDERA-2021-ERA-01-01, grant agreement 101058393.

Identification is an important component of FAIR, and it is one of the areas where the state of play is the furthest advanced. This, combined with higher-level metadata needed to support search and cataloguing forms the basis of many current efforts around FAIR. These areas also highlight some similarities of approach which form a solid foundation for promoting cross-domain FAIR (see below).

Some of the Case Studies—notably those with a more-mature culture around data reuse and consequent standardisation—are addressing the more challenging areas of FAIR implementation. The Case Studies on Social Surveys (WP06) and Population Health (WP07) both show sophisticated approaches to data reuse at a granular level, across institutional boundaries. Notably, these communities have dedicated significant resources over many years to realise the possibility for such re-use (Social Surveys around the Data Documentation Initiative (DDI) standards and Public Health around the Observational Health Data Sciences and Informatics (OHDSI) Observational Medical Outcomes Partnership Common Data Model (OMOP CDM)).

The use of generic standards (for Findability) and community specific standards for Interoperability and Reuse is a significant finding. Even more important, the way these intersect with FAIR maturity within these domains, and the awareness of communities of practice as the domains across which standards are adopted, suggests a stepwise approach to FAIR implementation. The specifics of the landscape revealed in the WorldFAIR FIPs can provide the basis for outlining practical approaches to the implementation of FAIR both within and across domains. This impacts both the formulation of the Cross-Domain Interoperability Framework (CDIF) and the recommendations regarding FAIR assessment which will be produced by the coordinating and synthesis activity (WP02).

4.3. Identifiers and controlled vocabularies: elements of convergence

Two of the most important aspects of FAIR are identifiers and controlled vocabularies. Identifiers are critical for the management and access of the many resources needed for effective data sharing; controlled vocabularies describe the concepts/units of meaning needed to qualify and understand different aspects of data. These two elements of FAIR were important across all the Case Studies, and provide the best example of convergence (identifiers) as well as highlighting some of the subtler issues for achieving interoperability between systems.

The most common form of identification was the Digital Object Identifier (DOI), and this was used in every Case Study at least to some extent. The majority of Case Studies used the DataCite service for assigning DOIs to data, and a cluster of related metadata. Typically, the resolution of a DOI takes you to a ‘landing page’, which provides links to the data and high-level metadata, sufficient for discovery and cataloguing, at least in a human-readable form, and sometimes a machine-actionable one.

‘Global cooperation on FAIR data policy and practice’ (WorldFAIR) has received funding from the European Union’s Horizon Europe project call HORIZON-WIDERA-2021-ERA-01-01, grant agreement 101058393.

Other common identifiers used similar mechanisms for identifying researchers (ORCIDs) and the related ROR mechanism for identifying research organisations. Publications were in some cases identified with DOIs using the service (and metadata schema) provided by Crossref.

Although the use of these services and identifiers does not guarantee interoperability in a machine-actionable fashion (see below), it is a solid basis on which to base FAIR data exchange. Because DataCite particularly is accompanied by a 'minimum set' of metadata which is sufficient to support Findability, this is also a very positive finding.

In many cases, controlled vocabularies exist within the WorldFAIR communities in a standardised fashion – they are published and maintained in a useful form. This includes different types of classifications, code lists, ontologies, thesauri, and so on. Issues remain regarding the standardised publication of controlled vocabularies, but in cases where these are disseminated in a machine-actionable way, the standard used for this is often the Simple Knowledge Organization System (SKOS), with the Web Ontology Language (OWL) being also employed for ontologies in domains where these are popular. The use of SKOS is in line with best practice.[11] Again, this forms a solid basis for further progress in the implementation of FAIR, even while significant work remains to be done.

The use of controlled vocabularies in navigating collections of data and metadata is, however, potentially problematic. The single best example of the potential for confusion was seen in those Case Studies which rely on the InChI system for identifying molecular entities in their data models.

In some systems, the InChI "identifier" is used as a way of understanding which data sets are concerned with a specific compound - they can be used as search terms, etc. They are *not* intended to be used as "identifiers" for the data (or related metadata records), however, and this can be a source of confusion. InChI is best understood as a systematic notation or machine-processable nomenclature for distinguishing chemical structures - it functions in this respect like a controlled vocabulary. When we are considering the importance of identifiers in FAIR, this is a fundamental distinction: identifiers for data and metadata are distinct from those identifying the subject of study.

This confusion with InChI is an example of a phenomenon which also occurred in some other areas with other controlled vocabularies: the use of such resources within domains (or in many cases infrastructures) is understood by current users, but is not self-evident. This type of expert knowledge within an infrastructure – required for effective reuse of data – presents a barrier to broader reuse, especially machine-actionable reuse.

[11] Cox SJD, Gonzalez-Beltran AN, Magagna B, Marinescu M-C (2021) Ten simple rules for making a vocabulary FAIR. PLoS Comput Biol 17(6): e1009041. https://doi.org/10.1371/journal.pcbi.1009041

'Global cooperation on FAIR data policy and practice' (WorldFAIR) has received funding from the European Union's Horizon Europe project call HORIZON-WIDERA-2021-ERA-01-01, grant agreement 101058393.

Although identifiers and controlled vocabularies perform distinct roles within data systems, the way in which they are used to support FAIR sharing must be both explicit and clear, and currently this is not always the case. Despite this, the existing level of standardisation in these areas is promising.

4.4. Communities and community standards

As noted above, the Case Studies needed to overcome the challenge of defining exactly who was represented by the 'community' for which they were developing a FIP. In other efforts (ENVRI-FAIR is a good example) we have seen the tendency for FIPs—and the use of standards they describe—to follow the organisation of the community according to infrastructures which manage and disseminate many of the resources. While having such standards in place is important (although many may really be de facto standards, rather than more open and recognised standards) for the implementation of FAIR it is not sufficient to realise scalable data reuse within the domain. Ideally, broader data-sharing communities should exist: the size of the community that can interact with a standard is as large as the size of the community that knows about and uses that particular standard.

The scope of a 'community', employing the same standards, services and technologies (FERs) for the same data and metadata, is what defines the practical limits of data sharing. While convergence on similar approaches is a positive development across disparate infrastructures, it is not in itself sufficient for scalable, machine-actionable reuse of data and related resources. Below, we describe four specific challenges identified from the WorldFAIR work on FIPs.

What we see in the WorldFAIR Case Studies is that in many cases there is a degree of convergence on a common system for identifiers (DOIs, and especially DataCite), and the existence of high-level metadata descriptions sufficient to support Findability. Some elements of the information needed to support Interoperability and Reuse (especially around controlled vocabularies) exist, but these are less well-developed. Accessibility is in general terms reasonably well-supported, but not in a fashion which would support machine-actionability.

If we look towards what is required to support machine-actionable exchange of data not across individual infrastructures, but across broader communities, we see that much remains to be done. The challenges can be understood as:

1. Better defining what constitutes a 'community of practice' for FAIR implementation, to ensure that improvements in the shareability of data can be realised;
2. Identifying meaningful levels of FAIRness, based on existing practice and building toward more complete support for all aspects of FAIR (starting with Findability and proceeding from there), in concrete terms;

'Global cooperation on FAIR data policy and practice' (WorldFAIR) has received funding from the European Union's Horizon Europe project call HORIZON-WIDERA-2021-ERA-01-01, grant agreement 101058393.

3. Establishing strategies for building on existing metadata holdings and data management and dissemination practices, and working step-wise toward more complete support;
4. Identifying or organising ‘standards bodies’ which are recognised as authoritative can develop, endorse and maintain standards for the communities identified, in cases where these do not already exist.

The interplay between the definition of a ‘community of practice’ and the organisation and standardisation within that community is central to the ability to realise the promise of the FAIR principles. Existing infrastructures are the key ingredient in establishing the ‘community’ at a meaningful level, and in providing—in concrete terms—what steps can be taken, and what standards may best be used to realise FAIR in practice. This will require dialogue across infrastructure boundaries in those cases where this is not already occurring. In some cases (for example, OIH/ODIS (WP11), Social Surveys (WP06), Population Health (WP07), Disaster Risk Reduction (WP12)) we see this approach already in action. In some others the dialogue may be characterised as remaining primarily within the domain in question. In every case, however, specific guidance around FAIR implementation could help these efforts moving forward.

‘Global cooperation on FAIR data policy and practice’ (WorldFAIR) has received funding from the European Union’s Horizon Europe project call HORIZON-WIDERA-2021-ERA-01-01, grant agreement 101058393.

5. Impact on cross-domain FAIR

This section will look at some of the specific issues and actions which have emerged from the FIPs activity thus far in WorldFAIR, and what the implications are for FAIR in practical terms both within each of the domain-specific Case Studies and also at the level of cross-domain FAIR. Identifying what is needed for FAIR sharing in the broadest sense—across domain and institutional boundaries—can help establish and guide practice within the domains themselves. The outputs from the WP02 coordination and synthesis activity in WorldFAIR is aimed at providing this level of guidance, and this section will describe the lessons learnt in the development of this vision of cross-domain FAIR, both in terms of FAIR assessment and as recommendations for practice, to be described in the CDIF guidelines.

5.1. FAIR within and across domains

While different in scope, the essential aspects of FAIR implementation within and across domains are fundamentally similar. A sufficient set of information (data and—critically—metadata) must be exchanged in a standard form between the suppliers and users. It should be noted that often this level of information provision is performed by infrastructure players (repositories, registries, data producers) rather than by individual researchers.

To understand what is practically required, the FAIR principles themselves must be broken out into a set of specific activities, and these then reduced to an agreed set of functions or services which can then be implemented. The FIPs represent a first step in this direction, by identifying key resources in developing this type of a framework.

The activities needed to support FAIR resemble—but do not follow point-for-point—the FAIR principles themselves. Currently, a starting set of activities has been proposed:

1. Findability
    a. Search
    b. Cataloguing
    c. Assessment of fitness for purpose
2. Accessibility
    a. Determination of access privileges and extent
    b. Provision of data access
3. Interoperability and Reuse
    a. Import the resource into the local system
    b. Understand the resource (semantics)

'Global cooperation on FAIR data policy and practice' (WorldFAIR) has received funding from the European Union's Horizon Europe project call HORIZON-WIDERA-2021-ERA-01-01, grant agreement 101058393.

    c. Understand provenance and context
    d. Use the resource (integration, harmonisation)
4. Manage the data-sharing process
    a. Establish links to previous versions, related publications and resulting data/metadata
    b. Control versions and ownership of resources

This is presented as an example of the types of activities involved in FAIR sharing, rather than as a mature proposal, but it suggests the way in which FAIR implementation can be approached. Currently, activities in WorldFAIR are pursuing this idea further, to understand the functions and services needed to support these activities, and to identify the information required, and the standards with which that information can be expressed and modelled.

It should be noted that because of differences in the subjects researched, different domains will have different requirements or priorities in many areas. Not all types of data and metadata are of equal interest across domains—in some cases, for example, physical samples and objects are of primary interest, while in others, what is most important are quantitative data sets resulting from the administration of questionnaires, while in others sensor readings may be of the greatest importance, and so on. In some domains, the ability to describe and reproduce the specifics of computational environments is critical, while in others this is of secondary or no importance.

These differences across domains must be understood both for the purposes of assessing the FAIRness of that domain, and for the accurate reuse of the data which it shares across domain boundaries. Determining the priority of activities in FAIR implementation within a domain can be informed by an understanding of what resources are in highest demand within that domain, but also across domain boundaries. Especially in cross-domain scenarios, it cannot be assumed that knowledge about data and metadata exists: explicit documentation of the purpose and conditions of data collection and generation, and subsequent processing, is essential for informed reuse.

5.2. Convergence on standards and agreement on implementation

The FIPs produced within WorldFAIR give us some specific examples of how we can approach these challenges, and also point to some of the specific issues which will need to be addressed. This section looks at some exemplary points which emerged from the first round of FIP development.

We have mentioned that DOIs—and especially DataCite—represented a strong point of convergence across the WorldFAIR Case Studies. When we consider what would be needed to leverage this convergence to the point where machine-actionable exchange becomes possible, we encounter some specific points which can help us understand the issues which will occur at every point where standards are employed for FAIR exchange.

'Global cooperation on FAIR data policy and practice' (WorldFAIR) has received funding from the European Union's Horizon Europe project call HORIZON-WIDERA-2021-ERA-01-01, grant agreement 101058393.

While DataCite mandates the use of DOIs as a form of persistent identification, aligned with the handle system, and also provides a minimum metadata set, this in itself is not sufficient for machine actionability. When we arrive at a landing page, by resolving a DOI, we know what information we need (a resolvable link to the data, a set of known metadata fields with known data types) but we do not know exactly how this information will be encoded, nor how it can reliably be located within that page.

There are several different ways in which this is currently done, typically with particular infrastructures employing their own conventions. One popular technique is to embed metadata within the landing page using a JSON-LD[12] file, linked from the HTML as a JavaScript <script> element. This technique is one which has been used successfully by some of the WorldFAIR Case Studies, and indeed is recommended also by DataCite itself.

What is missing is a predictable model for exactly how the metadata items and the link to the data itself would be described. One approach here would be the adoption of the FAIR Digital Object Framework (FDOF), although the specification is still in development. The FDOF provides a set of relationships and classes which can be expressed in JSON-LD, and which would establish the basis for machine-actionable services.

The specific models—that is, the metadata fields and their types—still need to be agreed, in a form sufficient to support the identified activities in support of FAIR sharing, but this basic mechanism is a requirement to leverage even existing convergence around the use of DOIs and related high-level metadata suitable for discovery purposes.

If we look further into the activities related to Findability, above, we can see that there are some additional models in common use, which may also be significant. There are two popular standards used in several of the WorldFAIR Case Studies which suggest themselves: the first is Schema.org, which we see used in Public Health and in the OIH/ODIS implementations, and the second is DCAT, which is used in several Case Studies. While to some extent overlapping in functionality, Schema.org is primarily focused on supporting search capabilities, while DCAT is aimed at describing the catalogue entries for data sets, streams, and services. In some cases, the use of both standards is anticipated, in order to leverage the full range of functionalities they provide. Thus, they should not be understood as direct competitors.

A closer look at how these standards are implemented, however, shows us that in many cases domain-specific extensions have been developed for use within the domain. This is necessary as

[12] https://json-ld.org/

'Global cooperation on FAIR data policy and practice' (WorldFAIR) has received funding from the European Union's Horizon Europe project call HORIZON-WIDERA-2021-ERA-01-01, grant agreement 101058393.

both Schema.org and DCAT are generic: they provide mechanisms for exactly this type of extension because they understand that specific domains have specific requirements.

This use of extensions is not so much an issue for FAIR implementation within domains, but does present a challenge for cross-domain use: domain-specific extensions must be understood by any domain wishing to use that information to support reuse of the data. This detailed understanding of the way in which even common standards must be used and aligned is an important input for the work of establishing the CDIF guidelines.

Other aspects of FAIR will employ different standards for different functions, but the needed level of agreement to support machine-actionability across domain boundaries will not change. What we see in this example around Findability will exist also for all of the FAIR functions requiring support.

5.3. Addressing gaps in functional FAIR support

One of the aspects of FAIR implementation which requires close examination is the extent to which the needed functions can in fact be fully automated. For Findability, it is felt that this is an achievable goal, and indeed is a practical first step in many domains. Automating data access is far more challenging, and the ability to integrate data from disparate sources is a bigger challenge still.

In order to understand these more challenging aspects of FAIR implementation, we need to look at how these activities are done manually today, and effort thus far to support them through automation. The FIPs and discussions around them have shown us several important things.

When it comes to data access, there are emerging models for describing the conditions of access, and even performing automated checks. To some extent, the efforts which today are manual and extremely resource-intensive could be better supported with machine-actionable policies. One major barrier to automating data access is the lack of granular descriptions of data: often, some individual fields within a data set are potentially disclosive or otherwise confidential, while the majority of a data set is not. Unless the individual fields are subject to machine operations, entire data sets will be inaccessible due to the properties of a small subset of their useful content. It is notable that this granular level of metadata regarding individual fields within the data is also required for many operations around data harmonisation and integration.

The degree of support which can be provided through automation for FAIR data access is still an area which will require more exploration, but it is clear that improvements are possible, especially if more granular metadata is available to systems.

'Global cooperation on FAIR data policy and practice' (WorldFAIR) has received funding from the European Union's Horizon Europe project call HORIZON-WIDERA-2021-ERA-01-01, grant agreement 101058393.

The topic of data integration and harmonisation is a complicated one. Current thinking attempts to separate some of the different functions needed to support these activities through automation. The ability to load data into processing systems requires an understanding not only of its formatting (as, say a CSV file, etc.) but also of its structure: which fields are identifiers? Which fields are measures? (Etc.) This type of 'structural' metadata can be expressed separately from the semantic and conceptual information needed to understand and use the data.

An entirely different set of models and resources must be applied to the difficult problems of semantic harmonisation: what concepts are used and how are they defined? What role do they play in relation to the data in question (as values for specific measurements? As the definitions for populations or for fields/columns within a table? Etc.) There is today a lot of exploration going into how semantics can be formally described (as ontologies, classifications, etc.) and how standard sets of concepts can be usefully mapped to support machine actionability.

A further area that also requires support is the description of data processing and provenance. Again, we have some popular standards which are used in some of the Case Studies (PROV is common) but there is no single approach to their use, nor is complete coverage of the topic of data provenance well-supported by a single standard. This also represents an area where more work is needed.

5.4. Describing and assessing FAIR

Analysis of the FIPs across the WorldFAIR Case Studies suggests that a staged approach to FAIR implementation is needed. It is clear that some aspects of FAIR—notably Findability—are more practical near-term targets for most of the Case Studies and other domains. At the same time, some Case Studies are progressing in other areas in significant ways, especially in the realisation of Accessibility (e.g. Social Surveys (WP06)) and toward Interoperability and Reuse (e.g. Population Health (WP07), SALURBAL (Urban Health, WP08), Biodiversity (WP09), and Agricultural Biodiversity (WP10)).

Different domains are at different levels of maturity in terms of data reuse, ranging from practices which approach the machine-actionable vision of the FAIR principles, to those which are very manual and siloed. Any framework for assessing FAIRness or understanding the priorities and strategies for realising FAIRness need to take this dimension of the challenge into account. Further, in order to apply such levels to specific domains, the needs of each domain in terms of data and metadata must be understood. The WorldFAIR FIPs give us a systematic and scaled view of this requirements landscape.

What emerges from this analysis is that for each of the needed activities identified to support FAIR—with their information requirements and standard expressions—a series of levels could be

'Global cooperation on FAIR data policy and practice' (WorldFAIR) has received funding from the European Union's Horizon Europe project call HORIZON-WIDERA-2021-ERA-01-01, grant agreement 101058393.

described ('basic', 'intermediate' and 'advanced', or some similar scale) which would be specifically informed by the current, prospective and aspirational statements of communities expressed in FIPs. Most importantly, for any given community, such a framework—based on consultation and dialogue through the FIPs process—could be used as the basis of FAIR assessment and planning. Further, a separate assessment could be done for each domain in terms of the additional criteria for cross-domain FAIR (since cross-domain FAIR is likely to require the use of related-but-different standards, and to have higher requirements in terms of topics such as provenance and contextual information).

It should be noted that FAIR implies not only the existence of human-readable documentation, but also a focus on machine-actionability. This aspect of FAIRness would be taken into consideration in the proposed assessment framework.

An example of a possible assessment matrix is shown in the figure below.

EXAMPLE OF PROPOSED FAIR ASSESSMENT MATRIX (AS CURRENTLY BEING CONSIDERED)

FAIR Principle/Activity	Community A	Community B	Community C	Community D	Community E
Findability: Search					
Findability: Catalogue					
Interoperability: Describe Semantics					
Interoperability: Provide Structural Metadata					

Criteria: [priority, standards/practices]
Level: Basic, Intermediate, Advanced
Rationale: _____

It should be noted that this is not yet a mature proposal, but merely illustrates an approach which could be used to achieve an appropriate degree of domain-sensitive understanding in FAIR assessment, while also retaining the ambition to address cross-domain FAIR data. This type of assessment is only really useful in the context of a clear definition of what comprises a community of practice. For any given domain, this is likely to require the establishment or recognition of an

'Global cooperation on FAIR data policy and practice' (WorldFAIR) has received funding from the European Union's Horizon Europe project call HORIZON-WIDERA-2021-ERA-01-01, grant agreement 101058393.

organisation or initiative which is seen as playing a consensual, representative role and is able to define for the domain what standards and practices are to be recommended.

While the existence of recommended standards—or, in many cases, community appropriate profiles of standards—is a necessary precondition, it is not sufficient to guarantee interoperability of data and metadata. Adherence to standards and the determination of fitness for purpose must also be in operation. While enforcing strict conformance can be problematic in many communities, more easily acceptable approaches such as self-conformance exist for ensuring that the recommended standards and guidelines are being followed.

How FAIR assessment and the determination of adherence to standards can best be implemented is an open topic, but establishing a framework which is sensitive to the needs of specific communities, but which also recognises the potential needs of cross-domain data re-use is a first step toward the development of a practical solution to implementation of the FAIR principles and appraisal of good practice.

'Global cooperation on FAIR data policy and practice' (WorldFAIR) has received funding from the European Union's Horizon Europe project call HORIZON-WIDERA-2021-ERA-01-01, grant agreement 101058393.

6. Conclusions and next steps

The FIPs exercise has been invaluable in helping the CODATA team in WorldFAIR understand better the complex data practices in each of the Case Studies. We also have convincing testimony that it has been useful for the Case Study / Work Packages themselves.

The points of discussion that emerged and that are described above—including, but not limited to, the definition of communities, the role and assessment of standards, the interplay between different functions, particularly in some instances identifiers and categories— will inform our work and will be discussed further in WorldFAIR and with the GO FAIR team behind the FIPs.

As noted in the Introduction, it is important to stress that we have not set out to give a comprehensive appraisal of the FIPs approach and could not do so at this stage of our work. All the WorldFAIR Case Studies have developed an initial FIP, but the process of reflection on practice will continue throughout the project. Each Case Study will complete a further FIP towards the end of the project and this will enrich our understanding of the utility of the approach. At that stage, we intend to be able to incorporate some robust prospective and aspirational considerations, and we need to consider how best to represent this in the FIPs.

WorldFAIR is now looking forward to the development of the Cross-Domain Interoperability Framework, and domain-sensitive recommendations for FAIR assessment. On both these counts, we consider that the FIPs approach has helped considerably. For the CDIF, through helping refine our initial functional analysis of the requirements for cross-domain FAIR, and—as predicted—helping identify *some* candidate cross-domain standards. For the FAIR assessment recommendations, through demonstrating that the FIPs can provide an empirical basis for such recommendations, reflecting both the current practice, and the aspirations of a given community or research domain.

'Global cooperation on FAIR data policy and practice' (WorldFAIR) has received funding from the European Union's Horizon Europe project call HORIZON-WIDERA-2021-ERA-01-01, grant agreement 101058393.