



WDS response to the WorldFAIR Policy Brief

Hodson, S., & Gregory, A. (2024). WorldFAIR (D1.4) Second Policy Brief (Version 1). Zenodo. <https://doi.org/10.5281/zenodo.11242702>

[Introduction](#)

[Brief Responses](#)

[Policy Issues: Evidence, Analysis and Recommendations](#)

[Recommendations from WorldFAIR Synthesis Reports](#)

[Recommendations in relation to FIPS](#)

[Recommendations in relation to CDIF](#)

[Recommendations in relation to FAIR assessment](#)

[Recommendations for a framework for FAIR assessment](#)

[Working Group and Advisory Group for CDIF](#)

Introduction

World Data System IPO and ITO presented the following notes on WorldFAIR to the WDS Scientific committee for their input and feedback at the August 2024 WDS SC meeting. The following comments were provided for the SC to add to or to discuss. We stated in the CODATA/WDS monthly check in meetings, we acknowledge that this format is difficult to receive feedback on.

These efforts, including participation in the CDIF advisory and workgroups will be discussed at the forthcoming WDS Scientific Committee meeting to be held in October to see if they mesh with the next WDS Action Plan and the organization's goals and objectives.

WDS International Program Office

c/o WDS-IPO @ UT-Oak Ridge Innovation Institute

University of Tennessee, Knoxville

Tel. 865-974-7909

wds-ipo@utk.edu | www.worlddatasystem.org



Brief Responses

Policy Issues: Evidence, Analysis and Recommendations

1. **Data Engineering** (p. 7) Recommendation: Policy makers and funders need to encourage and enable a data engineering approach in the data infrastructures that are the most important to address major societal and planetary challenges. Specifically, this requires supporting long-lasting data aggregation and data integration services as part of EOSC and globally.
 - Agreed, although there are existing examples of data repositories that are already leading the way in terms of this data engineering approach; in particular, this can be seen in cases of real-time data acquisition for earth and oceans monitoring (e.g., Ocean Networks Canada, NEON).
 - The forthcoming RDA Data Granularity WG output ([RDA P22 Session](#)) also argues for the need to provide discovery & access services based on metadata that characterizes datasets at varying levels of granularity, such that researchers (and machines) can obtain only the data they need.
 - Citing precise subsets of data can be done as per the RDA Data Citation WG guidance.
 - i. Rauber, A., Asmi, A., van Uytvanck, D., & Pröll, S. (2016). Identification of Reproducible Subsets for Data Citation, Sharing and Re-Use. Bulletin of the IEEE Technical Committee on Digital Libraries, 12(1). <https://doi.org/10.5281/zenodo.4048304>
 - ii. Rauber, A., Gößwein, B., Zwölf, C. M., Schubert, C., Wörister, F., Duncan, J., Flicker, K., Zettsu, K., Meixner, K., McIntosh, L. D., Pröll, S., Miksa, T., & Parsons, M. A. (2021). Precisely and Persistently Identifying and Citing Arbitrary Subsets of Dynamic Data. Harvard Data Science Review, 3(4). <https://doi.org/10.1162/99608f92.be565013>



- Investments in new paradigms of training and disciplinary data repositories are essential to curating datasets and building infrastructures that meet these data engineering objectives.
2. **Metadata Uplift** (p. 9) Recommendation: Sufficiently detailed, standardised, and interoperable metadata are a precondition for the data products that are essential for high priority research areas. Research infrastructures and data infrastructures need to put into practice ‘metadata uplift’. They must be enabled to do so by policy makers and funders.
- Agreed, but advise that the summary recommendation includes the notion of provenance metadata and quality indicators.
 - A good start on information needs has been developed as part of the draft ESIP AI Data Readiness checklist (<https://github.com/ESIPFed/data-readiness/blob/main/checklist-published/ai-ready-data-checklist-v.1.0.md>) which will undergo community testing, refinement and automation as collaborative work between ESIP and the World Data System.
3. **Interdisciplinary Research for Global Challenges** (p. 10) Recommendation: There is a need for investment in technologies and approaches that facilitate data aggregation and data integration for interdisciplinary, grand challenge research areas. Such investment should prioritise work that automates the integration approach and allows it to be performed year on year with time series data.
- Suggest to include federated data infrastructures as a viable approach in the fulfillment of this recommendation.
4. **Reproducible and Transparent Research** (p. 10) Recommendation: The global data stewardship and metadata standards community should direct concerted effort to refine and improve standards for describing data provenance and processing, as well as technologies that enable such standards to be used to provide a full and machine-actionable account of data lineage. Such work should be enabled by funders and policy makers.
- Agreed with the need for better provenance information, and the commentary on that topic.

WDS International Program Office

c/o WDS-IPO @ UT-Oak Ridge Innovation Institute

University of Tennessee, Knoxville

Tel. 865-974-7909

wds-ipo@utk.edu | www.worlddatasystem.org



- Certainly agree that data be deposited in an “appropriate repository”, but would further emphasize that preference is for repositories that adhere to the TRUST Principles, and even better if certification (e.g., CoreTrustSeal) is achieved.
 - Suggest to revise the language of “data should be deposited in a recognised regional or international data aggregator, or in a local or national service that can be harvested internationally.” It would be better to recommend that data be deposited into a trusted repository, utilizing disciplinary repositories when available, and that these repositories facilitate harvesting into federated systems (regional and disciplinary).
- 5. Increasingly Automated and Controlled Access to Sensitive Data** (p. 11)
- Recommendation: In partnership with the global data stewardship and metadata standards community, data services looking after sensitive data should direct concerted effort to developing and implementing systems that can support negotiation of access to data in a dynamic and more automated way. This effort should be supported by funders and policy makers.
- An example where very granular access can be implemented for sensitive datasets is at Ocean Networks Canada. Access can be restricted at the level of particular data products, variables and time ranges (both fixed, and relative to the current time - e.g., last 2 weeks) for instrument data. That said, the management of these cases is cumbersome and prone to long-term sustainability challenges as the technical infrastructure evolves or as documentation (e.g., written agreements) associated to the permissions evolve (language, structure, etc); moreover, staff that are implementing technical administration may not always have access or the same interpretation to these agreements to support them properly. Whilst more use cases can be supported with this approach, robust technical and organizational structures need to be well thought out to avoid mishaps.
 - Automation will also rely on more seamless user authentication technologies. One inspiring example is by Quartech System Ltd and the

WDS International Program Office

c/o WDS-IPO @ UT-Oak Ridge Innovation Institute

University of Tennessee, Knoxville

Tel. 865-974-7909

wds-ipo@utk.edu | www.worlddatasystem.org



Province of British Columbia; they partnered to develop a sophisticated and streamlined solution for authenticated access to data and digital documentation in the justice system (e.g., lawyers' access to digital evidence), and are currently extending the system to more use cases.

- Alongside these infrastructures, techniques to monitor for usage complaints and approaches to react to violations are necessary. Perhaps shared legal resources within the data repository community could alleviate financial barriers. Even the case where data are openly available, but with licenses such as CC-NC introduce challenges for ensuring non-commercial clauses are not violated.
6. **Responsible Use of AI** (p. 12) Recommendation: AI technologies (particularly generative tools based on LLMs) present a number of challenges, notably the potential misuse of sensitive information, a lack of transparency and reproducibility, and the risk of hallucinations and imprecision. The use of detailed, accurate and structured metadata should be explored as one of the means of enhancing the utility and precision of these technologies and to help in imposing guardrails.
- Agreed with this recommendation and supporting arguments.
 - Stress the need for reproducibility to build trust with the data and outcomes
7. **Support the Further Development of the Cross-Domain Interoperability Framework (CDIF)** (p. 13) Recommendation: Given the pressing need for a data engineering approach to support interdisciplinary research, policy makers and funders should enable the further development of CDIF and its extension to more profiles and recommendations. The data stewardship community, metadata standards organisations and data infrastructures should collaborate to implement, test, refine and improve the current and forthcoming CDIF profiles and recommendations.
- A fuller understanding of the developments thus far is needed by the WDS to fully support this activity. It is suggested to include the WDS (and its membership) more explicitly in future collaborations to refine, improve and extend the CDIF profiles and recommendations.

WDS International Program Office

c/o WDS-IPO @ UT-Oak Ridge Innovation Institute

University of Tennessee, Knoxville

Tel. 865-974-7909

wds-ipo@utk.edu | www.worlddatasystem.org



8. Invest in Research Infrastructures (p. 14) Recommendation: Research infrastructures should be encouraged and enabled to transition to a data engineering approach and provide researchers with pre-integrated, 'science-ready', interdisciplinary, data products. This will require investment to build capacity and to adopt new technologies and approaches. This transition should be enabled by funders and policy makers. It should be embraced by research infrastructures and data repositories as a necessary evolution of their function and mission.

- In a way, it may be a 'return to data engineering' rather than a 'shift from bibliographic to data engineering' that is needed. Many of the early repositories emanating from the World Data Centres and pioneering disciplinary data repositories (e.g., space sciences, physics, earth observations) have always had a data engineering approach. It is more the institutional library data repositories that introduced the increasingly prevalent bibliography approach as that is more consistent with the librarian approach. We should be careful to acknowledge that rich legacy that already exists, and is now validated even more.
- Interdisciplinary research infrastructures is a key challenge worthy of ample investments via interoperability advancements that serve priority research questions.
- Agreed that researchers should not be expected to become data stewards, but rather emphasize the value of data stewards/engineer expertise and curation.
- It will be important to prioritize research questions (e.g., SDGs) for which relationships can be established between researchers and infrastructure providers; that way use cases can be developed and supported through partnership with tangible outcomes.

9. Enable FAIR Practice in Research Communities (p. 15)

- Recommendation 9.1: Enable standards bodies and international and representative organisations. To advance FAIR implementation, it is important to take a multi-stakeholder approach. In particular, policy makers and funders should help support the role played by representative

WDS International Program Office

c/o WDS-IPO @ UT-Oak Ridge Innovation Institute

University of Tennessee, Knoxville

Tel. 865-974-7909

wds-ipo@utk.edu | www.worlddatasystem.org



and authoritative organisations as well as standards bodies on an international scale.

- i. Repositories must be explicitly named as a key stakeholder alongside standards-bodies in defining standards.
 - o Recommendation 9.2: Support WorldFAIR+. The WorldFAIR approach should be further tested and refined as a means of enabling research communities to develop and implement good practice, through a multi-stakeholder approach. Policy makers and funders should support the WorldFAIR+ initiative and similar activities.
 - o Recommendation 9.3: Support the FIPs approach and infrastructure. Funders and policy makers, through investment and appropriate policy direction, should support the global data stewardship community to socialise the FIPs approach, and to improve and sustain the necessary infrastructures.
 - i. Agreed upon the value of FIPs.
 - o Recommendation 9.4: : Support CODATA and RDA to enable research communities. The Research Data Alliance and CODATA should be supported in their activities to enable research disciplines and interdisciplinary research areas to advance their FAIR practices.
 - i. Would suggest to include World Data System alongside CODATA and RDA, especially given the direct connection and mandate to the data repositories that need to implement FAIR practices.
10. **Support the Sustainability of Semantic Artefacts** (p. 16) Recommendation: Policy makers and funders should help to ensure the sustainability of the standards organisations that develop and maintain metadata schema, terminologies, and other semantic artefacts. An immediate step would be to fund a project to conduct survey and analysis of business models, governance and juridical status, and make recommendations to improve the sustainability and effectiveness of such organisations.
- o Agreed that sustainable approaches to funding internationally-relevant entities is critical. It would be fantastic if the EU takes a lead in identifying

WDS International Program Office

c/o WDS-IPO @ UT-Oak Ridge Innovation Institute

University of Tennessee, Knoxville

Tel. 865-974-7909

wds-ipo@utk.edu | www.worlddatasystem.org



pathways for that. The RDA Funders Forum and the Biodata Coalition are other useful entities to include in those efforts.

11. **Strengthen International Partnerships** (p. 17) Recommendation: EOSC should actively build international partnerships with standards bodies, inter-governmental initiatives and global representative bodies. Contributing to the development of a strong global ecosystem for data and metadata exchange will benefit EOSC and is necessary for its success.
 - Recommend to include World Data System in this vision of orchestrating international collaborations, alongside CODATA and the RDA.

Recommendations from WorldFAIR Synthesis Reports

Recommendations in relation to FIPS

1. Provide policy support and investment for FIPs. Funders and policy makers, through investment and appropriate policy direction, should support the global data stewardship community to socialise the FIPs approach, and to improve and sustain the necessary infrastructures. (Policy; policy makers and funders).
2. Publish FERs: Representative and authoritative organisations that are responsible for particular FAIR Enabling Resources (FERs) should describe and publish these FERs in a way that they can be easily referenced in FIPs. They should be supported and enabled in doing so by other stakeholders (funders, policy makers and the data stewardship community at large). (Organisational; authoritative/representative organisations, standards organisations, data repositories).
3. Make it easier to publish FERs and FIPs: Supported by the global data stewardship community, the GO FAIR Foundation (GFF) should explore solutions to lower the barriers to information entry to the FIPs ecosystem, including a wider range of import and export profiles. The GFF should be enabled to do so by well-targeted funding to enable the development and sustainability of the necessary infrastructures. (Organisational; GFF, data stewardship community, funders, policy makers).
 - a. Specifically, JSON-LD import and export should be developed and enabled in the FIPs

WDS International Program Office

c/o WDS-IPO @ UT-Oak Ridge Innovation Institute

University of Tennessee, Knoxville

Tel. 865-974-7909

wds-ipo@utk.edu | www.worlddatasystem.org



Wizard and FAIR Connect. 109 b. Specifically, the GFF (and any other stakeholders using systems that will interoperate) should enable a distinction to be made between 'no answer given' (e.g. due to lack of maturity of the community) and 'not applicable' (e.g. because the community has determined that no declaration is necessary, no resource is needed for this sub-principle).

4. Make it easier to visualise FIPs: Collectively, the global data stewardship community and the GFF should explore numerous avenues to improve the way in which FIPs and FERs can be visualised and analysed. This endeavour should be enabled by well-targeted funding and direction from policy makers. (Organisational; GFF, data stewardship community, funders, policy makers).
5. Enable the wholesale use of FAIRsharing entries as FERs in FIPs: GFF and FAIRsharing, supported by the global data stewardship community and other stakeholders, should explore mechanisms to enable the wholesale use of FAIRsharing information as FERs in FIPs. Such a process should be supported by well-targeted funding and direction from policy makers. (Technical; GFF, FAIRsharing, data stewardship community, funders, policy makers).
6. Review optimal encoding of FIPs and explore alignment: GFF, ODIS, the CDIF WG and AG, and global data stewardship community, should constructively review the pros and cons of the technical underpinnings to the current FIPs approach and explore whether an alignment of the nanopublications approach and the suggested JSON-LD/schema.org approach could be achieved. Such a process should be supported by well-targeted funding and direction from policy makers. (Technical; GFF, FAIRsharing, data stewardship community, funders, policy makers).

General Comment: We support maturing the technical and policy aspects of FIPs and FERs, but would strongly encourage a wider stakeholder engagement internationally. While the GO FAIR Foundation is an important partner, there are other major organizations internationally progressing FAIR practices and implementations that can also pursue these advancements.

Recommendations in relation to CDIF



1. **General:** CDIF metadata should be embedded in landing pages or linked stand-alone files, encoded in JSON-LD. The supported profiles will be indicated as part of the metadata.
2. **Discovery profile:** This profile recommends the use of a set of key Schema.org fields for describing static data sets and queryable data sources, with the DCAT equivalent recognised as an acceptable alternative. **Access profile:** This profile recommends that ODRL Actions and Entities be used to describe policies and conditions for the use of data. At this time, the utility of this approach is limited by the lack of shared vocabularies for conditions of use, user qualifications, legal constraints, and similar important items. ODRL is thus limited to describing policies in terms of the disseminating institution but provides a basis for expansion in future when the needed vocabularies are developed.
 - a. Given the number of groups working on schema.org profiles, duplication of effort or conflicting outcomes needs to be mitigated through liaisons/collaborations (e.g., ESIP Science on Schema.Org cluster)
3. **Controlled vocabularies profile:** This profile recommends the use of SKOS for describing controlled vocabularies, understood to mean any terminological resource. The use of OWL as a linked extension to what is presented in SKOS is also recommended, as is the use of XKOS for formal statistical classifications.
4. **Data description profile:** This profile recommends the use of DDI-CDI to provide a granular description of the structure of data sets, and how the logical content of those datasets relates to their physical encoding. Text-based data is supported (CSV and other delimited formats, fixed-width ASCII, etc.), with the intention of expanding support for other types of data in future. The recommendations cover description of individual data sets to make them 'integration-ready'.
 - a. The RDA Data Granularity WG has done a fair bit of investigations into dataset granularity which is very nuanced; standardizing the granular structure of data in a domain agnostic way or a wide variety of use cases may be quite challenging.
5. **Universals profile:** This section recommends the information which should be provided when describing time, geography, and units of measurement in other metadata sets. Some standards for this purpose are recommended in each area.



Recommendations in relation to FAIR assessment

1. Funders and policy makers should encourage and enable research communities to develop FAIR Implementation Profiles so that these can be used as the basis of alignment and ultimately assessment. (Policy recommendation; for funders and policy makers, research communities).
 - a. Agreed.
2. Funders and policy makers should encourage and enable authoritative organisations — those that can meaningfully represent specific research communities — to formalise agreements on FAIR practice, in particular by ensuring FERs are published and that the community curates FIPs to describe its practice. (Policy recommendation; for funders and policy makers, research communities).
 - a. Agreed.

Recommendations for a framework for FAIR assessment

1. A framework for FAIR assessment should include domain-agnostic and domain-specific criteria. Cross-domain criteria form a third category in certain use cases. (Technical recommendation; for developers of FAIR assessment criteria and tools).
 - a. Agreed
2. A framework for FAIR assessment should take into account the practices expressed in specific disciplines: FAIR assessment should be weighted accordingly, and the information should be used to improve guidance. (Technical recommendation; for developers of FAIR assessment criteria and tools).
 - a. “Weighted accordingly” is not sufficiently contextualized
 - b. To include machine-to-machine interoperability, a framework for FAIR assessment will need to reference profiles for how FERs are implemented. CDIF shows us the level of specificity which may be necessary to achieve machine-to-machine interoperability. Developments in the standards space in relation to ‘shapes’ and ‘profiles’ provide an indication for how this can be done and direction for future work. (Technical recommendation; for developers of FAIR assessment criteria and tools).



- c. Agreed with the first sentence, but the utility of CDIF to deliver upon that is still not totally clear.