

Summary Report: “Interoperability of Metadata Standards in Cross-Domain Science, Health, and Social Science Applications II”

Schloss Dagstuhl [Event 19413](#)



A [week-long workshop](#) was held on the subject of standards in cross-domain data use for science, health, and social science at [Schloss Dagstuhl - the Leibniz Center for Informatics in Wadern](#), Germany, 6-11 October 2019. The meeting was sponsored by [CODATA](#), the data-focused arm of the [International Science Council](#) (ISC), and the [DDI Alliance](#), an international member-driven consortium which provides technical standards for research data in the social, behavioral, economic, and health sciences. CODATA is currently working towards a launch of a decadal programme on cross-disciplinary data as part of the [ISC's Science Action Plan](#). The DDI Alliance is now developing an information model for integrating data across domain boundaries. The workshop was subsidized by Schloss Dagstuhl - the Leibniz Center for Informatics.

Workshop series

This workshop was the second addressing this important topic, the [earlier one](#) also taking place at Dagstuhl in October 2018. Both focused on specific real-world use cases: the first was an exploration of specific issues encountered in the use of data across domain boundaries; the second aimed at producing practical guidance for addressing them.

Among other outcomes, the first workshop contributed to the work on the next-generation DDI model, making it a more suitable tool for dealing with cross-domain data integration independent of the social sciences. Other standards and models were also examined (e.g. spatio-temporal aspects of [DCAT](#)). This focus on practical guidance for cross-domain data use is expected to continue into the ISC's decadal programme.

Workshop participants included representatives of use cases and technology and standards experts across several different domains, including the social, behavioral, and economic sciences, geophysical and environmental science, health research, disaster risk reduction, urban planning and policy, and the [UN's Sustainable Development Goals](#) (SDGs). Standards concerned with statistical and research data and metadata were an important consideration (e.g., DDI, [SDMX](#)) and specialized application schemas (e.g. OBO vocabularies from the life sciences, ISO 19115 from geospatial data community). Technical experts ranged from those involved in developing standards to systems implementers. Almost half of the participants overlapped with the first workshop.

Scope of work

Data-sharing across domains is currently a common topic for discussion, but often the more difficult issues of cross-domain data *use* are touched on only lightly. To be usable, data shared across domains must be fit for purpose, and richly described. The context within which data are produced is important, but does not easily travel between domains in a useful form. The vast quantity and complexity of data available often means that automated approaches are the only feasible ones. Harmonizing data across geography, time, demography, and other axes is critical, but it presents many difficult challenges. To address these issues, the workshop produced a series of guidelines for practical implementation.

Workshop goals

The guidelines target three levels of practitioners: a high level, to interface with broad initiatives such as the [FAIR principles](#); a "user" level to help inform practitioners involved in the creation, management, dissemination, and analysis of cross-domain data; and a "technical" level to address the needs of those implementing systems to support cross-domain data use. The intent was to provide meaningful guidance in a practical way, on a set of significant topics. In order to achieve these objectives, it was recognized that a framework within which cross-domain data sharing and use could be meaningfully discussed would need to be developed.

Ultimately, these deliverables might demonstrate how a meaningful community of practice around cross-domain data sharing and use can be developed, to support both science and policy. While the use cases and specific topics addressed in the workshop were exemplary rather than comprehensive, it is hoped that they will point a way forward for larger efforts in the future and help to inform coordination at the international level across all disciplines.

Workshop structure

The work was organized according to four real-world use cases, with aspects of each being considered by a dedicated working group. The use cases included the reporting of SDG indicators, infectious disease outbreaks, disaster risk reduction, and resilient cities projects. They share significant requirements for the use of data coming from several different domains - these needs were viewed in light of a conceptual framework based on the FAIR principles, a model of the data-sharing process, and the various stakeholder perspectives. Significant issues to be addressed across the use cases were identified according to this framework, and these served as the basis for four work groups. The conceptual framework itself also served as a focus for some work, being important for enabling communications about cross-domain data-sharing and use.

Work groups

The first group looked at the issues surrounding indicators used to support the monitoring of policy agendas at the international level, including the SDGs, [Sendai Framework for Disaster Risk Reduction](#), the [Paris Agreement](#), and the [Beijing Declaration and Platform for Action](#), which has its 25th anniversary in 2020). Recommendations were formulated to reduce reporting burden from the national level, enable better planning around the different data streams across policy agendas, and leveraging existing technology efforts to enhance use of the data by those within the reporting chain (national and international agencies), and by those external to it, for purposes such as research (universities, NGOs, etc.) Representatives from the [UN Statistics Division](#) (UNSD) and groups working within [UN Office for Disaster Risk Reduction](#) (UNDRR) presented on-going activities which served as a basis for the guidelines.

The second group focused on how a sustainable data-sharing infrastructure around public health research could be established in Africa, building on current best practice as seen in the [ALPHA Network](#) and similar efforts, and embracing the latest technology and standards-based approaches. Consideration of the issues addressed not only the practical, technical level but also the issues around governance and culture which need to be taken into account to drive any realistic approach.

The third group focused on the issues of semantics and terminology in cross-domain data-sharing scenarios. Building on a number of existing resources and recognizing the wealth of available domain-specific systems, this work produced practical guidance for developers of terminologies, and the users and maintainers of repositories which support their use.

The fourth group worked on regularizing approaches to data harmonization and processing at a practical level, with an emphasis on common approaches and libraries of process steps which implement them. An illustrative example was used to show specifically how this type of harmonization can be implemented, and to identify the specific technical issues (i.e., handling of geographic or temporal coverage) which need to be addressed.

Immediate outputs and future plans

The conceptual framework used in identifying the areas of work was itself seen as a useful product and was the subject of some investigation during the week. More detailed implementation of such a framework came out of the work of the second group, and the third group provided some useful definitions of terms to use when discussing cross-domain semantics and terminology.

It is expected that the work will be carried forward under the auspices of the decadal programme, led by CODATA, and include future intensive workshops as well as meetings with a broader scope. Collaboration with other organizations working toward data use across domain boundaries (e.g., [GO FAIR](#), the [Research Data Alliance](#), etc.) is anticipated. The further development of practical guidelines is intended to lead to broader work under the [ISC decadal programme](#), and collaboration around the DDI data integration model is expected to contribute to this work.

The workshop was organized by [Simon Cox](#) (CSIRO Australia and W3C Dataset Exchange Working Group), Arofan Gregory (DDI Alliance), [Simon Hodson](#) (CODATA), [Steven McEachern](#) (Australian National University and DDI Alliance), [Joachim Wackerow](#) (GESIS - Leibniz Institute for the Social Sciences and DDI Alliance)