

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

[Schloss Dagstuhl](#), Leibniz Center for Informatics, 1 – 5 October 2018

A workshop on the practical application of computer science to enable data sharing and data interoperability across disciplinary boundaries was hosted at the internationally renowned computer science institute at Schloss Dagstuhl in Germany. Sponsored by CODATA (the [Committee on Data](#) of the [International Science Council](#)) and the Data Documentation Initiative Alliance ([DDI](#)), it brought together 24 participants from many different domains. These included representatives of a number of metadata specifications and researchers involved in pilot projects currently being pursued as part of the ISC and CODATA [Data Integration Initiative](#). The workshop lasted for 5 days, encouraging intense involvement on the part of all participants, largely due to Dagstuhl's unique dynamics (as described on the DDI site [here](#)); Dagstuhl subsidized this event. The workshop organizers included [Simon Cox](#) (CSIRO Australia and W3C Dataset Exchange Working Group), [Simon Hodson](#) (CODATA), [Steven McEachern](#) (Australian National University and DDI Alliance), [Joachim Wackerow](#) (GESIS - Leibniz Institute for the Social Sciences and DDI Alliance).



The workshop examined how modern web-friendly computer science techniques and standards could better enable data-sharing in the context of the ISC and CODATA [Data Integration Initiative](#) pilot projects. These are major cross-disciplinary data integration projects to advance solutions for three important global challenges: infectious disease outbreaks, resilient cities, and disaster risk reduction. The infectious disease pilot built on work by the [Infectious Diseases Data Observatory \(IDDO\)](#) to support both research and humanitarian efforts, with Ebola used as the primary example for discussion. The resilient cities pilot focused on the work in Medellín, Columbia, in partnership with [Resilience Brokers](#). Examples involved air quality measurement, location of hospitals, and geo-spatial data. The disaster risk reduction pilot, led by Public Health England in partnership with the [Integrated Research on Disaster Risk](#) looked at

how data could support the Sendai Framework, especially in cases where the SDG indicators would not be sufficient. Different approaches for obtaining data both from within and from outside the realm of official statistics were explored, with an emphasis on research data. In each case, data integration presented significant challenges.

Metadata standards are a part of the computer science landscape which can facilitate data integration and use within a particular scenario. Representatives of many of these were present, helping to understand the data integration challenges faced by each of the projects. These standards included many of the W3C Linked Data vocabularies (DCAT, SSN, Data Cube, PROV-O, etc.), DDI, HL7 FHIR, CDISC, and many others.<sup>i</sup> Some of these standards are focused on the data within a particular discipline or domain. Others are more general in scope. The workshop examined the relationships between these standards in the context of their real-world application (the pilot projects). This required an understanding of the granularity of the metadata being expressed by each standard (at the level of a study or dataset, at the variable and observation level, etc.)

Working groups consisted of combined teams of business experts involved in the pilot projects, and experts in the relevant technology and domain standards. Some additional technical topics which arose during the exploration of the pilot projects were also addressed separately by small groups of the appropriate experts.

The workshop was extremely productive, with outputs to include working papers describing the findings as regards each of the pilot projects. An article will also be produced describing the overall goals of the effort and the relationship of various standards and technology approaches to the cross-disciplinary data integration projects. The intention is that these will be published in peer-reviewed journals appropriate to their content. In addition, it is anticipated that specifically technical outputs will be presented (for example, a DCAT profile to support granular description of data in online catalogues). The workshop will be presented at the upcoming [SciDataCon](#) conference (at the 2<sup>nd</sup> [International Data Week](#) organized by CODATA together with the Research Data Alliance and the World Data System) in Gaborone, Botswana in early November of 2018. Further collaborative work between CODATA, the DDI Alliance, and other interested organizations is anticipated in the future, including more intense, focused workshops of this kind.<sup>ii</sup>

---

<sup>i</sup> A list of [metadata specifications for the workshop](#) is available at the workshop site.

<sup>ii</sup> A list of recent [CODATA and DDI Workshops](#) on relevant topics is available at the workshop site.