# Interoperability of Metadata Standards in Cross-Domain Science, Health, and Social Science Applications

## Summary

This presentation will give an overview on the outcome of an international expert workshop in October 2018 in Dagstuhl, Germany. The goal of the workshop is a better understanding of how metadata specifications can be aligned to support cross-discipline (or cross-domain) data integration and analysis, with a consideration of standards from the social, health and environmental sciences, along with generic all-purpose metadata standards. It will do so with specific reference on the one hand to the standards listed below and on the other hand to specific case studies in infectious disease, disaster risk and resilient cities.

## Background

Standards are a vital tool enabling integration and semantic linking of data within and between disciplines. However, standards tend to get developed and adopted within disciplines or application domains with little consideration of cross-discipline requirements and technologies, so data integration can often only be easily achieved within and between closely allied fields. Addressing global scientific challenges that depend on cross-discipline integration remains difficult. The challenge is to make cross-discipline data integration a routine aspect of data-driven science.

Metadata support data discovery, selection, access and use, and are critical for data integration. Data from different sources/domains should be described in a way that cross-discipline discovery can detect and access the relevant data collections, and so that transformations and analyses can be automated. The use of cross-discipline data should become efficient, scalable and reproducible, enabling discipline-neutral data processing and analysis tools to be applied. Furthermore it would be possible to apply (meta-)data mining approaches and reasoning. In sum, new opportunities of insights and realization will develop.

A CODATA initiative on interdisciplinary data integration is seeking to explore these challenges and opportunities in relation to three specific case studies in interdisciplinary research into infectious disease outbreaks, disaster risk and resilient cities. These case studies provide a concrete focus for exploring the potential of interoperability and data integration through metadata alignment.

## Focus of the Workshop

The core objective of the workshop will be to investigate and advance alignment between the cross-disciplinary and domain-specific metadata standards, and to bridge from standards focusing on collection-level to variable-level metadata.

There are several different areas where metadata comes into play:

* Description of studies or data collections for discovery purposes.
* Descriptions of provenance and scientific context for purposes.
* Description of data variables or dimensions for analysis purposes.
* Description of data transformation steps for recording purposes (possibly also for reusing the transformation steps on similar data).
* Controlled vocabularies to ensure standardized and agreed concepts (in relation to variables, collections, measurements, techniques and procedures etc.).

The capability to express discoverable and structured metadata must be automatic and achieved as far as possible using tools that are familiar and in common use.

**Topics for Discussion and Possible Outcomes**

Areas of exploration and discussion will identify and describe following:

* Common rules for metadata specifications
* Advantages and limitations of generic approaches
* Techniques for profiling or specializing generic standards for specific applications
* Best practices for setting up domain-specific data/metadata for cross-domain use
* Controlled Vocabularies, domain-independent and useful domain-specific ones
* Contact points/overlaps of specifications, crosswalks and transformations
* Identification of gaps. Possible workarounds, possible areas for future specifications