Editor's notes:

FAIR BOT. As metadata is data is metadata is data ...

Welcome to the first issue of IASSIST Quarterly for the year 2023 - IQ vol. 47(1).

The last article in this issue has in the title the FAIR acronym that stands for Findable, Accessible, Interoperable, and Reusable. These are the concepts most often focused on by our articles in the IQ and FAIR has an extra emphasis in this issue. The first article introduces and demonstrates a shared vocabulary for data points where the need arose after confusions about data and metadata. Basically, I find that the most valuable virtue of well-structured data – I deliberately use a fuzzy term to save you from long excursions here in the editor's notes – is that other well-structured data can benefit from use of the same software. Similarly, well-structured metadata can benefit from the same software. I also see this as the driver for the second article, on time series data and description. Sometimes, the software mentioned is the same software in both instances as metadata is treated as data or vice versa. This allows for new levels of data-driven machine actions. These days universities are busy investigating and discussing the latest chatbots. I find many of the approaches restrictive and prefer to support the inclusive ones. Likewise, I also expect and look forward to bots having great relevance for the future implementation of FAIR principles.

The first article is on data and metadata by George Alter, Flavio Rizzolo, and Kathi Schleidt and has the title 'View points on data points: A shared vocabulary for cross-domain conversations on data and metadata'. The authors have observed that sharing data across scientific domains is often impeded by differences in the language used to describe data and metadata. To avoid confusion, the authors develop a terminology. Part of the confusion concerns disagreement about the boundaries between data and metadata; and that what is metadata in one domain can be data in another. The shift between data and metadata is what they name as 'semantic transposition'. I find that such shifts are a virtue and a strength and as the authors say, there is no fixed boundary between data and metadata, and both can be acted upon by people and machines. The article draws on and refers to many other standards and developments, most cited are the data model of Observations and Measurements (ISO 19156) and tools of the Data Documentation Initiative's Cross Domain Integration (DDI-CDI). The article is thorough and explanatory with many examples and diagrams for learning, including examples of transformations between the formats: wide, long, and multidimensional. The long format of entity-attribute-value has the value domain restricted by the attribute, and in examples time and source are added, which demonstrates how further metadata enter the format. When transposing to the wide format, this is a more familiar data matrix where the same value domain applies to the complete column. The multidimensional format with facets is for most readers the familiar aggregations published by statistical agencies. The authors argue that their domain-independent vocabulary enables the cross-domain conversation. George Alter is Research Professor Emeritus in the Institute for Social Research at the University of Michigan, Flavio Rizzolo is Senior Data Science Architect for Statistics Canada. Kathi Schleidt is a data scientist and the founder of DataCove.

The format discussion in the first article is also the point of the second paper on 'Modernizing data management at the US Bureau of Labor Statistics'. The US Bureau of Labor Statistics (BLS) has a focus on time series and Daniel W. Gillman and Clayton Waring (both from the BLS) view time series data as a combination of three components: A measure element; an element for person, places, and things (PPT); and a time element. In the paper Gillman and Waring also describe the conceptual model (UML) and the design and features of the system. First, they go back in history to the 1970s and the Codd relational model and to the standards developed and refined after 2000. You will not be surprised to find here among the references also the Data Documentation Initiative's Cross Domain Integration (DDI-CDI). The mission is: 'to find a simple and intuitive way to store and organize statistical data with the goal of making it easy to find and use the data'. A semantic approach is adopted, i.e. the focus is on the meaning of the data based upon the 'Measures / People-Places-Things / Time' model. Detailed examples show how PPT are categories of dimensions, for instance 'nurse' is in the Standard Occupational Classification and 'hospital' in the North American Industry Classification System. The paper – like the first paper – also refers to multidimensional structures. The modernization described at BLS is expected to be released in early 2023.

The third paper is by João Aguiar Castro, Joana Rodrigues, Paula Mena Matos, Célia Sales, and Cristina Ribeiro where all authors are affiliated with the University of Porto. Like the earlier articles this also references the Data Documentation Initiative (DDI) with a focus on the concepts behind the FAIR acronym: Findable, Accessible, Interoperable, and Reusable. The title is: 'Getting in touch with metadata: a DDI subset for FAIR metadata production in clinical psychology'. Clinical psychology is not an area frequently occurring in IASSIST Quarterly, but it turns out that the project described started with interviews and data description sessions with research groups in the Social Sciences for identifying a manageable DDI subset. The project also draws on other projects such as TAIL, TOGETHER, and Dendro. The TAIL project concerned the integration metadata tools in the research workflow and assessed the requirements of researchers from different domains. TOGETHER was a project in the psycho-oncology domain and family-centered care for hereditary cancer. As most researchers showed to be inexperienced with metadata, they concentrated on a DDI subset that meant that FAIR metadata would be available for deposit. Support for researchers is essential as the they have the domain expertise and can create highly detailed descriptions. On the other hand, data curators can ensure that the metadata follow the rules of FAIR. This was achieved by embedding the Dendro platform in the research workflow, where creation of metadata is performed in an incremental description of the data. The article includes screenshots of the user interface showing the choice of vocabularies. The approach and the adoption of a DDI subset produced more comprehensive metadata than is usually available.

Submissions of papers for the IASSIST Quarterly are always very welcome. We welcome input from IASSIST conferences or other conferences and workshops, from local presentations or papers especially written for the IQ. When you are preparing such a presentation, give a thought to turning your one-time presentation into a lasting contribution. Doing that after the event also gives you the opportunity of improving your work after feedback. We encourage you to login or create an author profile at https://www.iassistquarterly.com (our Open Journal System application). We permit

authors to have 'deep links' into the IQ as well as deposition of the paper in your local repository. Chairing a conference session or workshop with the purpose of aggregating and integrating papers for a special issue IQ is also much appreciated as the information reaches many more people than the limited number of session participants and will be readily available on the IASSIST Quarterly website at https://www.iassistquarterly.com. Authors are very welcome to take a look at the instructions and layout:

https://www.iassistquarterly.com/index.php/iassist/about/submissions

Authors can also contact me directly via e-mail: kbr@sam.sdu.dk. Should you be interested in compiling a special issue for the IQ as guest editor(s) I will also be delighted to hear from you.

Karsten Boye Rasmussen - March 2023