

Openness in metadata, dictionaries, and data

Welcome to the first issue of *IASSIST Quarterly* for the year 2022 - *IQ vol. 46(1)*.

IASSIST Quarterly has often published new and further developments in metadata. New submissions in the area are welcome, and the *IQ* expects to continue presenting a flow of interesting articles on this important topic. Often these developments in metadata directly mention their relatedness to the Data Documentation Initiative (DDI). Many IASSIST members have a primary role in supporting users of data in research and education. It is clear that the data item '42' needs metadata to be of any use. Over a long period of time the developments in improving metadata have been rising to higher levels. Among the latest achievements is the work presented in the first paper, on metadata extraction from available documentation using machine learning. The second paper is also drawing on metadata when constructing a dictionary for social terms for searching existing metadata, and for supporting new research's design and development. The last article is on a special type of data - geospatial data or geodata - and with a special look at those data as open data. The insights of this article bring us positive awareness of central concepts and areas that can be used to achieve greater openness. Unfortunately, these days it is not evident in all parts of the world, but I must insist: Openness is a good thing! Closedness has always failed. It is with openness and free speech, openness in metadata, dictionaries, and data with free research, that we create a better world.

The first paper 'Engineering a machine learning pipeline for automating metadata extraction from longitudinal survey questionnaires' is written by a group of researchers and developers at several English institutions. Suparna De and Haeron Pereira are in the Department of Computer Science at the University of Surrey, Harry Moss and Sanaz Jabbari are at the Centre for Advanced Research Computing at University College London (UCL), Jon Johnson and Jenny Li are at CLOSER at UCL Social Research Institute. The article describes the first results from the project of extracting existing questions in survey questionnaires available as PDFs. The machine learning is a central part of automating the extraction for further use in the metadata model of the Data Documentation Initiative-Lifecycle (DDI-L). The supervised machine learning is supported by XML marked-up questionnaires that serves as training and validation. The article describes the details of the technical setup and the design of the processes and experiments, including the more complicated dataset and parameter versioning in machine learning. The experiment workflows are also graphically illustrated in the article.

The two authors of the second article are Ioannis Kallas, Professor at the University of the Aegean, and Dimitra Kondyli, Research Director at the National Centre for Social Research (EKKE) – Institute of Social Research in Greece. Their article presents 'A tool to promote research planning and conceptualization: SoDaNet research infrastructure's scientific dictionary of social terms'. The research infrastructure presented is based on projects developed by SoDaNet, the Greek infrastructure for social sciences, member of the CESSDA consortium. The described Scientific Dictionary of Social Terms was designed to support conceptualization, design, and management of research searched through SoDaNet. The dictionary is a computer application that is developed as a collective hypertext product to ensure continuity and validity. It is designed to meet the needs of the Greek-speaking scientific community. The dictionary is dynamic and develops gradually as it is supplemented with new terms and definitions through the work of many researchers. The organization of the dictionary is on three levels: terms, definitions, and bibliographic records. The terms are in both English and Greek. The article shows the example of the social term 'unemployed person'. As a tool for searching for scientific information, the dictionary also supports the design of new research and has a positive relationship to the DDI – also central in the first article. A follow-up article in the future could be on the project's contributions to the development and design of questions and hypotheses of new research.

A group of four is behind the third article on 'Open geospatial data: A comparison of data cultures in local government'. The authors are Karen Majewicz (Geospatial Project Manager and Metadata Coordinator at the John R. Borchert Map Library, University of Minnesota), Jaime Martindale (Map & Geospatial Data Librarian at the Arthur H. Robinson Map Library, University of Wisconsin-Madison), and Melinda Kernik (Spatial Data Analyst and Curator at the John R. Borchert Map Library, University of Minnesota). The group was supported by Yijing Zhou (GIS & Metadata Programming Intern for the BTAA Geoportal, University of Minnesota). The examples in this article compare the two states of Minnesota and Wisconsin. However, the issue of open data - here open geospatial data - concerns us all. The authors explore the GIS history, programs, organizations, and legislation of the two states, examining how their different approaches have influenced the availability of open data. The article starts with worthy discussions and definitions or qualifications of 'open geodata' and also the many possible sources for creation of public geodata, as well as its availability and barriers. The case studies of Minnesota and Wisconsin are thorough in presenting the historical, political and legal developments and the many stakeholders. The differences between the two states are concentrated in the areas of legislation, funding, workflows, and library involvement. The many central concepts mentioned above are well summed up in the title as culture. The paper won the IASSIST Conference 2020/2021 paper competition with the remarks of being well-written and addressing an important topic.

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 2022