

## Data management for students, researchers, and data science projects

Welcome to the second issue of *IASSIST Quarterly* 2021 (*IQ* vol. 45(2) 2021).

Data management is the focal point of the articles in this issue of the *IQ*. Aspects of data management have often been the central topic of many earlier *IQ* articles. Metaphorically simply said: Data is what IASSIST members breathe - data management is how we breathe it. The first two articles concern raising the attention and knowledge of data management among students and researchers. In both articles it turned out that more than the target group could benefit from the efforts. When instructing students, faculty at the university gained as well because the data management workshop was well integrated with the teaching. And when nudging researchers and graduate students in Illinois with data management hints, the Data Nudge also extended to other people. The positive reception - also from persons not directly targeted by the efforts - can be viewed as a sign that there is in general a great need for data management skills. Some of the skills are highly specialized - as exemplified in the third article on DATABOOK - but the general skills are necessary at all levels of the data society.

In the first article Elizabeth Blackwood from the Channel Islands campus of California State University describes how undergrad students at a university without 'very high research activity' benefit from knowledge of data management in the article 'Outside the R1: Equitable data management at the undergraduate level'. The project addressed the problem that students and faculty at these educational institutions often lack data management instruction. The article describes in much detail how a workshop was prepared and planned. You will find meticulous descriptions with headings and subpoints of the flow and structure of the data management workshop, with headings like 'what data', 'why management', 'data management plans' and more. Because of its success, the workshop and student assignments have been integrated as permanent parts of the undergraduate course. It was furthermore demonstrated - when reality kicked in with Covid-19 - that the workshop was transferred to remote delivery without difficulty. The term 'equitable' in the title is because knowledge of data management is part of the essential preparedness for access to both graduate school and ultimately to the digital society with interesting career paths also for students that were 'outside the R1'.

The second article is titled 'Better data management, one nudge at a time'. A good question - and the answer in the same sentence! The data management nudging was implemented and described by a team at the University of Illinois at Urbana-Champaign consisting of Daria Orłowska, Colleen Fallaw, Yali Feng, Livia Garza, Ashley Hetrick, Heidi Imker, and Hoa Luong. The team had noticed that although researchers at Illinois are aware of data services, the services are under-utilized. By releasing a monthly Data Nudge email the researchers became more alert to best practices and also to the resources on campus. The Data Nudge has high ratings and many email compliments, and now has more than 500 subscribers. The article presents the background of data services at some universities, and details on the distribution and growth in the number of subscribers as well as of some of the topics used in the Data Nudge. At the top of the most popular topics is 'file organization'. All past nudges can be found on the web. I am a great fan of using the ISO date, so I like the Data Nudge of 2021-05-25 (not 5/25, 2021). The leftmost carries most importance. I am not

talking politics here, just what we expect from numbers. Good data management respects expectations. Among the tables in the article, you will learn that 'data sharing' is a topic often promoted in the Data Nudge – important data management topics need to be addressed regularly.

The third article is 'DATABOOK: a standardised framework for dynamic documentation of algorithm design during Data Science projects' – an extensive description of a system to secure documentation and reproducibility, which are fundamental concepts in data management. Anna Nesvijevskaia proposes a framework called Databook for Data Science projects, based on points of critique of other platforms. The author finds that traditional data management tools are unable to absorb the final algorithmic model while data science platforms tend to structure only technical aspects, which excludes some project stakeholders. As data science is often developed in the context of business, projects involve people with many different skills, and the division between data skills and business skills when data scientist meets decision-maker can be difficult to bridge. The article is extensive and includes many references and Databook builds upon earlier attempts, including the most successful – the Cross-Industry Standard Process for Data Mining (CRISP\_DM). The framework developed has been tested in projects at the French company Quinten during 2017 to 2020. Further details of the Databook are shown in an appendix. Anna Nesvijevskaia is researcher at the laboratory DICEN Ile de France and a partner at the private data science company Quinten.

Enjoy the reading!

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](mailto: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 - September 2021