Data Train
Training in Research Data Management and Data Science

1. Towards a cross-disciplinary doctoral education program on research data management and data science

As a consequence of the digital change, a novel scientific discipline has been developed – Data science. Data science allows new approaches for interdisciplinary (big) data analyses and, therefore, promises additional knowledge. It is applied in nearly all research fields and often termed as a key discipline (Society of Informatics e.V., 2019). The fundament of data science, however, is a good stewardship of data (following the FAIR principles, Wilkinson et al., 2016) and an open data culture (Nosek et al., 2015). Moreover, a profound knowledge about the data analyses methods and innovative technologies is required to produce reliable data science results.

However, Kirchherr (2018) highlights a massive demand for experts in complex data analyses and innovative information technologies. Further, the Council for Information Infrastructure (RfII) stressed that quality (and hence quality control) of scientific data needs to be enhanced urgently and advises to foster education in data literacy and data handling skills (Council for Information Infrastructure, 2019).

Given the massive needs in terms of qualifications in research data management and data science, the U Bremen Research Alliance[1] together with the federal state of Bremen initiated the establishment of a discipline-overarching education program for doctoral students to strengthen competencies in data science, research data management and data literacy (see Data Train website). The program operates in close collaboration with the Data Science Center (DSC) of the University of Bremen as it contributes to the pillar "Qualification" of the DSC. This initiative is further in line with efforts of the National Research Data Infrastructure NFDI-consortia (funded by the DFG) which considers education as a cross-cutting topic (Bierwirth et al. 2020).

---

[1]The U Bremen Research Alliance comprises the University of Bremen and eleven non-university research institutes in Bremen and Bremerhaven. The alliance offers an excellent infrastructure and combines competencies of different disciplines which fosters creativity and innovative ideas. The network benefits from joint activities like the Data Train education program for doctoral candidates.

2. Data Train – Training in Research Data Management and Data Science

The central training program in research data management and data science is an additional offer for doctoral candidates from the member institutions of the U Bremen Research Alliance. To account for field-specific perspectives, a multi-disciplinary working group "Research Data" accompanies the development of a curriculum for the program. The curriculum (see section 3.1) constitutes the major part of the education program. Further components are invited talks given by scientists or industry partners ("Data stories"), the establishment of a platform for exchange about data-related issues ("Hacky hours") and the provision of helpful materials (links to online courses, podcasts, literature, etc.) via the program's website.

A cross-disciplinary curriculum was developed that covers relevant basics of research data management and data science. The selection of such discipline-overarching competences is based on an internal status quo and demand analyses (conducted in a dual method approach through qualitative and quantitative surveys) and an intensive external research. The program offers a set of basic cross-discipline courses regularly (see preliminary basic structure of the curriculum in the Appendix, Part A), while more specific courses could be added to the basic structure on demand. Sophisticated discipline-specific qualifications, however, will be beyond the scope of this program but should be covered by the respective scientific fields or NDFI consortia. Such a concept saves resources and fosters networking as well as scientific exchange throughout disciplines (truly beneficial for early career researchers as well as lecturers).

3.1 The curriculum

Courses are subdivided into tracks (see Appendix, Part A). Firstly, courses are classified according to the level of prior knowledge into "Starter track" (i.e. mostly interactive overview lectures; level 1[2]) and "Operator track" (i.e. mostly hands-on workshops; level 2). Secondly, the operator track is split into a "Data Steward track" and a "Data Scientist track". Tracks will take place one after the other and the courses therein follow a given structure (see Appendix Part A, Figure Ab).
Flexibility in terms of teaching content is an important aspect for the sustainability of the program as both, research data management and data science, are dynamic topics that develop rapidly. Therefore, courses are generally stand-alone and accordingly, easy to modify. Moreover, doctoral candidates can attend courses according to their individual needs.
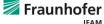
---

[2] Level 1 = "understand level"; level 2 = "apply level"; level 3 = "analyse level"

For this education program, a team of enthusiastic multi-disciplinary lecturers come together and share their expertise within the alliance to build up a joint curriculum (see Data Train website).

References

Baker, M. (2016): 1,500 scientists lift the lid on reproducibility. Nature 533, 452–454.

Kirchherr et al. (2018) Future Skills: Welche Kompetenzen in Deutschland fehlen. https://www.stifterverband.org/medien/future-skills-welche-kompetenzen-in-deutschland-fehlen. (Last access: 15.10.2020)

Maik Bierwirth, Frank Oliver Glöckner, Christian Grimm, Sonja Schimmler, Franziska Boehm, Christian Busse, … Heike Neuroth. (2020, June 15). Leipzig-Berlin-Erklärung zu NFDI-Querschnittsthemen der Infrastrukturentwicklung. Zenodo.

Nosek, B. A. et al. (2015): Promoting an open research culture. Science 348, 1422–1425.

RfII - Council for Information Infrastructure (2019): Herausforderungen Datenqualität – Empfehlungen zur Zukunftsfähigkeit von Forschung im digitalen Wandel, Göttingen, 172 p.

Society of Informatics e.V. (2019): Arbeitspapier: Data Science: Lern- und Ausbildungsinhalte. https://gi.de/fileadmin/GI/Allgemein/PDF/GI_Arbeitspapier_Data-Science_2019-12_01.pdf. (Last access: 15.10.2020)

Wilkinson, M. D. et al. (2016): The FAIR Guiding Principles for scientific data management and stewardship. Science Data 3, 160018.