Pressemitteilung

U Bremen Research Alliance e.V.

Timo Joost

02.12.2021

http://idw-online.de/de/news783419

Forschungs- / Wissenstransfer, wissenschaftliche Weiterbildung
fachunabhängig
überregional

New Findings with Data Train

They are heterogeneous, not always easy to handle, and are increasing by the second: data. Without it there is no new knowledge. How is it possible for the scientific sector to use data sustainably, yield new, reproducible findings from it, and develop innovations? With “Data Train - Training in Research Data Management and Data Science,” the U Bremen Research Alliance has created an interdisciplinary advanced training program for PhD students that passes on skills for dealing with research data and when drawing value from data.

Katharina Friz has a great deal planned. The economic researcher carries out work on three crises in 29 countries: the finance crisis 2008 / 2009, the Crimea crisis in 2015, and the latently smoldering environmental crisis. Which economic consequences do these crises have, especially on innovation behavior in companies in Russia, but also in other transition countries? Thus, the countries in Central and Eastern Europe and Asia that once belonged to regions influenced by the Soviet Union and experienced the transition from centralized economy to a market-based economic regime? That is what the 27-year-old is investigating in her doctoral dissertation at the University of Bremen.

“I work empirically,” says Katharina Friz. Data forms the basis of her PhD. Said data comes from the European Bank for Reconstruction and Development or the Russian survey institute Nevada, for example. The data comprises Word, Excel, and PDF documents, or sometimes simply rows of numbers that she analyzes using statistics software. “As soon as you enter the world of data, you realize how big it actually is and then it can get complicated quickly,” she explains. How do you find the right data? How do you assess it? How reliable is it? How do you use it sustainably? Those are questions that she addresses during her academic daily life.

--- “As soon as you enter the world of data, you realize how big it actually is and then it can get complicated quickly.” ---

In order to be able to answer those questions better, to talk to other PhD students, and to take a look outside of her own discipline, the research assistant, who works in Prof. Dr. Jutta Günther’s research group, takes part in Data Train. The U Bremen Research Alliance established the program with the support of Bremen State. “Researchers from all disciplines need basic knowledge on how to handle data. With Data Train, we want to pass on sound knowledge to those who take part,” states Prof. Dr. Iris Pigeot, Vice Chairwoman of the U Bremen Research Alliance and program initiator.

Data management is becoming ever more relevant. Increasing data quantities are being produced continually. Every second, 9000 tweets are send, 1000 photos are posted on Instagram, and sensors gather movement, environmental, and health data. It is estimated that the mountain of data will grow from 64 to 175 zettabytes in the year 2025. A zettabyte is a number with 21 zeros.

Not all but a great deal of this data is of enormous value for science. For example, the connection of GPS data that is produced by cell phones with health data allows for an important contribution to be made to the fight against the corona pandemic. Those who have mastered the technologies of research data management and data science
applications not only have an advantage when looking for innovation and new findings in scientific work. These skills are also in high demand outside of universities and research institutes. The economy is desperately searching for people with data competence.

It is not necessary that one is an expert in complex data analysis methods or needs to have mastered the programming of algorithms or machine learning. Said topics are, however, part of the training program. “The good thing is that it is flexible. The participants can choose specializations according to their needs, or just visit individual classes. It can also be well integrated into work for a PhD,” says Katharina Friz.

--- “In addition to the classes, we want to offer the participants a platform where they can talk and network.” ---

Data Train is made up of three interdisciplinary parts that follow each other. “Starter Track” is the first part. It encompasses interactive, two-hour online lectures on basic topics such as data science, big data, data management, and data security. Then comes “Operator Track Data Steward,” which is a multi-day, hands-on workshop on documenting, managing, pre-processing, and the harmonization of data. In the last part - “Operator Track Data Scientist” - it is all about passing on comprehensive data analysis methods from the fields of mathematics, statistics, and computer science, such as machine learning or data visualization.

“In addition to the classes, we want to offer the participants a platform where they can talk and network,” explains Dr. Tanja Hörner, the program coordinator. “We also hold inspiring talks about data from business, society, and all areas of sciences and will make digital advanced training materials accessible.” The program, which started in March 2021, has a duration of 9 months. If a track is completed, the participants receive a certificate stating the classes they took. In spring 2022, the next group will start.

One of the central components of the training is the teaching of FAIR principles: Findable, Accessible, Interoperable, and Re-usable is what the data must be. That sounds easier than it is as the data is extremely heterogeneous and one must act within the legal and ethical frameworks. The data may be photos, text or audio files, films, numbers, or tables in varying formats.

“Everyone talks about data but metadata is what is important,” emphasizes Prof. Dr. Frank Oliver Glöckner. The professor of Earth system data sciences within the Faculty of Geosciences at the University of Bremen and head of Data at the Computing Center of the Alfred Wegener Institute in Bremerhaven is one of the program architects. Another architect is Prof. Dr. Rolf Drechsler, spokesperson of the University of Bremen’s Data Science Center. In total, more than a dozen researchers from the U Bremen Research Alliance are active within Data Train and pass on their knowledge in classes - without additional payment by the way. Also non-alliance research institutes and businesses support Data Train with their contributions.

But let us get back to the metadata: Metadata is information that describes a file. For example, in the case of a photo the metadata states who took the photo when, where, and with which aperture and what can be seen on the photo. In the case of a book, the information is the name of the author, the edition, the year of publication, the publishing house, and the ISBN. Thus, metadata describes a dataset. “As mentioned, it needs to comply with the FAIR data principles, be machine readable, and interpretable,” states Glöckner.

--- “It would have been good for me if the program had been around at the start of my PhD.” ---

Katharina Friz has completed her first “Starter Track” classes and found them to be valuable. One of the classes is a lecture held by Professor Dr. Dr. Normal Sieroka, a philosophy professor, on critical thinking in relation to data handling. “That was a very different, educational perspective.” The early-career researcher will have completed her PhD by the end of the year. There is a great deal to do in the final phase. She does not know whether she will complete all of Data Train, thus the Data Steward and Data Scientist tracks. “It would have been good for me if the program had been
around at the start of my PhD,” she says.

**********

--- The new event format of the Data Train program is starting; “Data Stories” ---
In exciting talks, inspiring stories about data handling, data management, and data science applied in the private sector or in science will be told. The speakers shed light on the importance of data competences with regard to their individual working fields and will discuss current challenges. The talks are complemented by a joint discussion.

--- The new format will kick-off with: ---
Data Story #001 “Data Centric Systems in the Public Sector and Industry”

Speakers: Sonja Sievi & Götz Anspach von Broecker (AIRBUS)
Host: Prof. Frank Oliver Glöckner (Data Train spokesperson)
When: 4 p.m. to 6 p.m. on December 8, 2021
Where: Online via VC

wissenschaftliche Ansprechpartner:
Dr. Tanja Hörner (data-train@ww.uni-bremen.de)

Originalpublikation:
Impact – The U Bremen Research Alliance science magazine
The University of Bremen and twelve non-university research institutes financed by the federal government cooperate within the U Bremen Research Alliance.

Twice per year, the Impact science magazine provides an exciting insight into the effects of cooperative research in Bremen. "New Findings with Data Train" was published in issue 4 (June 2021).


URL zur Pressemitteilung: http://www.uni-bremen.de/en/research-alliance/research-data/data-train Information on Data Train and Data Stories
Anhang Impact-Magazin U Bremen Research Alliance; Ausgabe 4; Juni 2021 http://idw-online.de/de/attachment88096