Towards a cross-disciplinary Training Programme in Research Data Management and Data Science (Data Train) in Bremen, Germany

Frank Oliver Glöckner, Tanja Hörner, Rolf Drechsler, Iris Pigeot

FAIRsFAIR workshop 2021
"FAIR data competences in doctoral education programmes"

(1) Motivation

(2) About the programme "Data Train"

(3) Development phase

(4) Concept & curriculum

Motivation

Motivation

530%
increase of global data volume
From 33 zettabytes in 2018 to 175 zettabytes

€829 billion
value of data economy in the EU27
From €301 billion (2.4% of EU GDP) in 2018

10.9 million
data professionals in the EU27
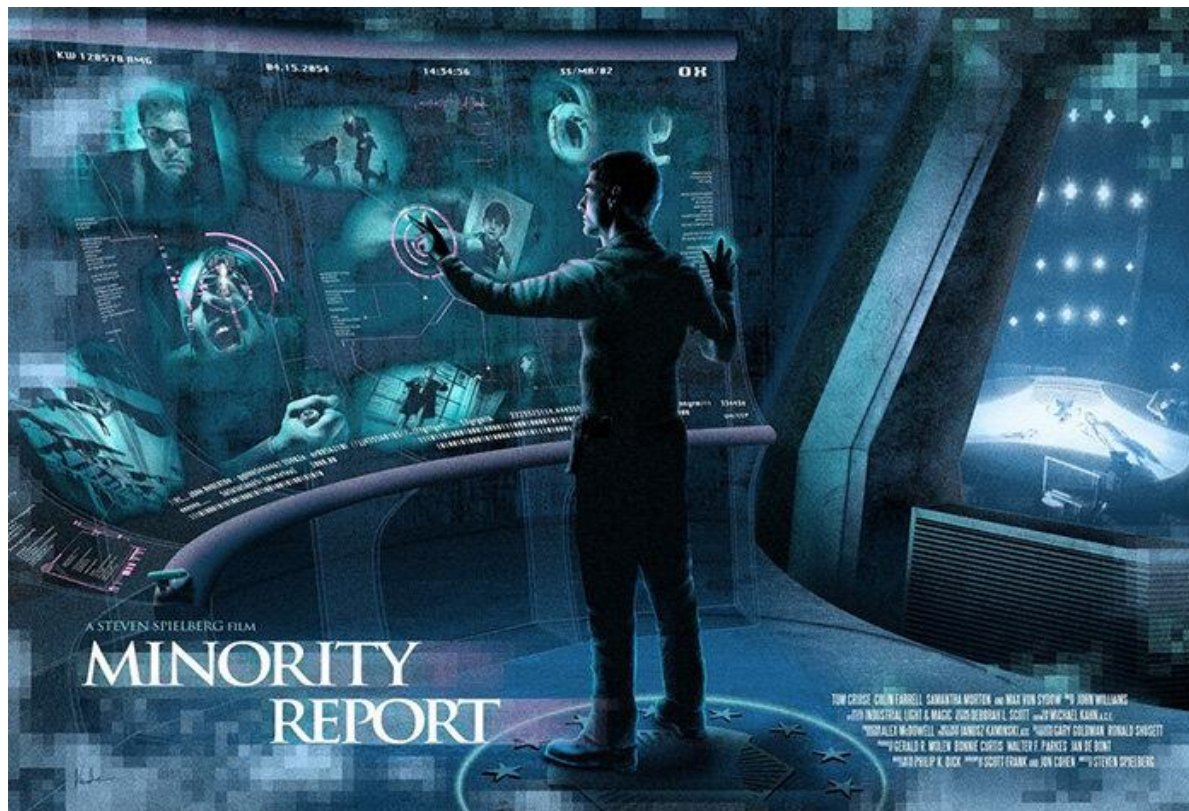From 5.7 million in 2018

65%
Percentage of EU population with basic digital skills
From 57% in 2018

https://ec.europa.eu/info/strategy/priorities-2019-2024/europe-fit-digital-age/european-data-strategy20)

Motivation

Data Availability

The following information was supplied regarding data availability:

This project has been deposited at DDBJ/ENA/GenBank under the BioProject accession no. PRJNA385857 and drafts of genomes are available with accession no. NHBG00000000– NHMJ00000000. All contigs generated using Megahit from each sample are available through ~~Microbes~~ (http://data.~~xxxxxxxxx~~/project/view/261). TARA Oceans has deposited the reads in the NCBI Sequence Read Archive (SRA) with accessions ERS488346, ERS488330, ERS477998, ERS477979, ERS488509, ERS488486, ERS478040, ERS477953, ERS477931, ERS488147, ERS488119, and ERS478017.

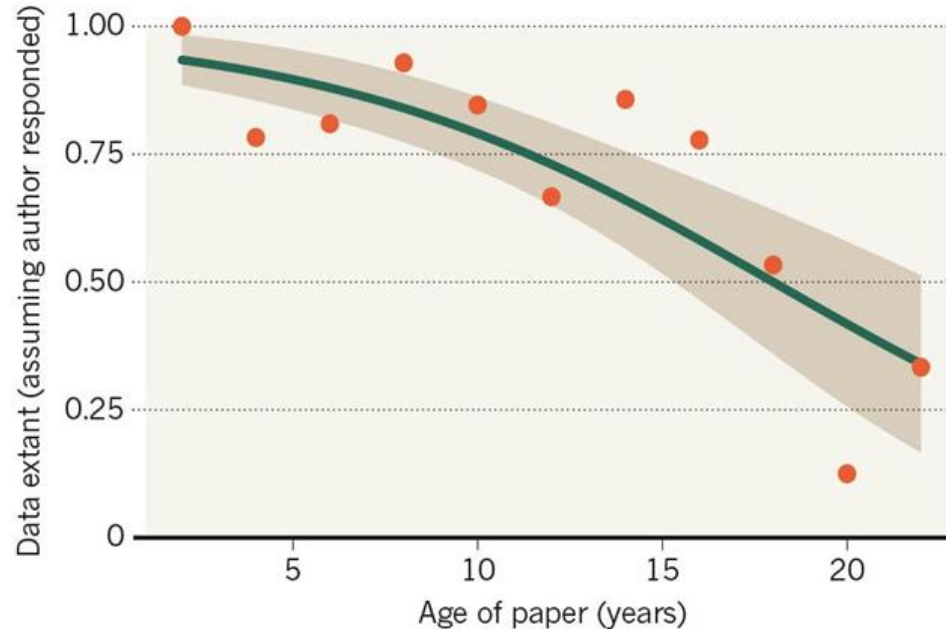https://www.~~xxxxxxxxx~~/project/view/261

404 Not Found

nginx/1.12.2

Motivation

MISSING DATA

As research articles age, the odds of their raw data being extant drop dramatically.

Odds of data being lost are estimated to increase by 17% in every year after publication.

Find a working e-mail address for the first, last, or corresponding author fell by 7% per year.

Overall, we only received 19.5% of the requested data sets, and only 11% for articles published before 2000.

Vines, Timothy H. et al. Current Biology, 2014, Volume 24 , Issue 1 , 94 - 97

Towards a FAIR research DATASPHERE

TARGET: FAIR data

STRATEGY: build up NFDI via consortia

The National Research Data Infrastructure (NFDI)

"… objective to systematically index, edit, interconnect and make available the valuable stock of data from science and research.

… for treating new research issues, generating findings, and making innovations."
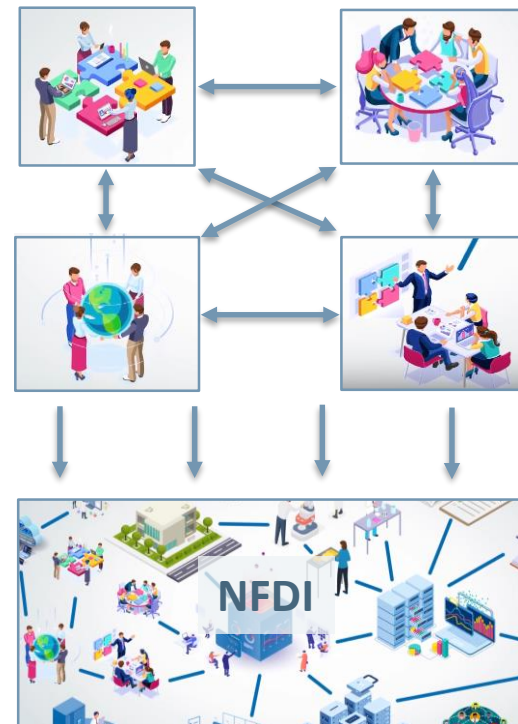
© DFG, German Research Association;
Movie about the NFDI Initiative

© Figures: DFG, German Research Association;
Taken from the movie about the NFDI Initiative

Data Train
Training in Research Data Management
and Data Science

Lack in DATA COMPETENCE
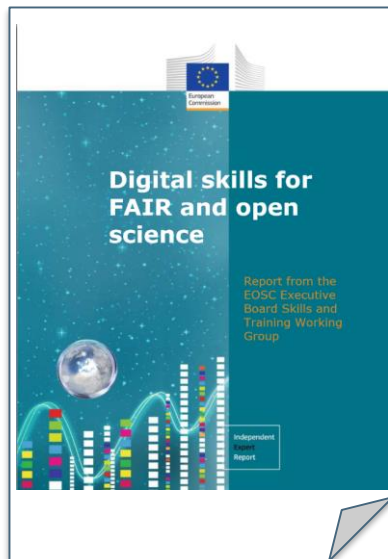
Report from the EOSC Executive Board Skills and Training Working Group: Digital skills for open science

„…the quality of data needs to be enhanced…"
RfII, 2019

„Until 2023, approximately 700,000 persons with technological skills are required, to meet the need for tech-specialists."

Kirchherr et al., 2018

FUTURE SKILLS

- Data literacy

- Data management

- Data science

- Critical thinking

- Innovative technologies

RfII, 2019, 2019; HF Digitalisierung, 2019

Dr. Tanja Hörner

DATA TRAIN

Data Train
Training in Research Data Management and Data Science

RESEARCH DATA MANAGEMENT & DATA SCIENCE

– two highly significant competences that belong together.[1]

OUR TARGET GROUP:

Doctoral researchers

Digital change

Huge amounts of data

Data science as key discipline

Research data management as fundament for data science

▶ ▶ ▶ **Interdisciplinary education** in data science and research data management

[1] see: Turning FAIR into reality – Final report and action plan from the European commision, 2018

Website:
https://www.uni-bremen.de/data-train

Data Train
Training in Research Data Management and Data Science

DATA TRAIN

Data Train
Training in Research Data Management
and Data Science
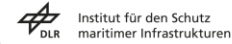
- Joint initiative of the U Bremen Research Alliance[1] and the Federal State of Bremen

DATA TRAIN

Data Train
Training in Research Data Management and Data Science

- Joint initiative of the U Bremen Research Alliance[1] and the Federal State of Bremen

- Linked with NFDI consortia

- In close collaboration with the Data Science Center, University of Bremen (UB)

DATA SCIENCE[6] CENTER

[1] U Bremen Research Alliance; [2] NFDI4Ing; [3] NFDI4Health; [4] KonsortSWD; [5] NFDI4BioDiversity; [6] DSC@UBremen

DATA TRAIN Development phase

RESEARCH

Internal & external

ASSESSMENT

Status quo & demand

- Qualitative approach; more than 70 interviews
- Quantitative approach

DEVELOPMENT

of concept & curriculum

PILOT PHASE

- Cross-disciplinary seminar for students
- Starter Track & Operator Tracks for doctoral researchers
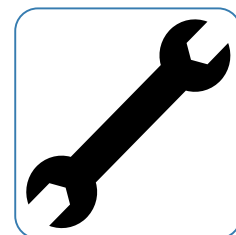
EVALUATION

- Online survey after each lecture
- (topic) tuning
- rethinking

U Bremen
Research
Alliance

Data Train
Training in Research Data Management and Data Science

DATA TRAIN Concept

(1) Voluntary training program for PhDs

(2) Central & inter-disciplinary

(3) Cross-discipline basics in RDM & DS

(4) Flexible curriculum & „Course-picking"

Optimize resources & foster networking across disciplines (at an early career stage)

discipline-specific qualifications via **scientific fields and NFDI consortia**

Features of the Data Train Programme

DATA TRAIN Concept

(1) Voluntary training program for PhDs

(2) Central & inter-disciplinary

(3) Cross-discipline basics in RDM & DS

(4) Flexible curriculum & „Course-picking"

Optimize resources & foster networking across disciplines (at an early career stage)

discipline-specific qualifications via **scientific fields and NFDI consortia**

Features of the Data Train Programme

Requirement: Target group-specific

R(1) Consider limited time resources;

R(2) diverse thematic foci;

R(3) different level of prior knowledge of doctoral researchers

1. Starter Track (L1)
Interactive overview lectures

Course 1.1 — Course 1.2 — Course 1.3 — Course 1.4 — Course 1.5 — ... — Course 1.n
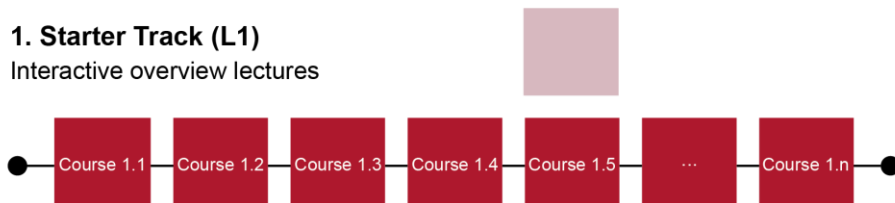
2. Operator Track Data Steward (L2)
Hands-on workshops

Course 2.1 — Course 2.2 — Course 2.3 — Course 2.4 — Course 2.5 — ... — Course 2.n
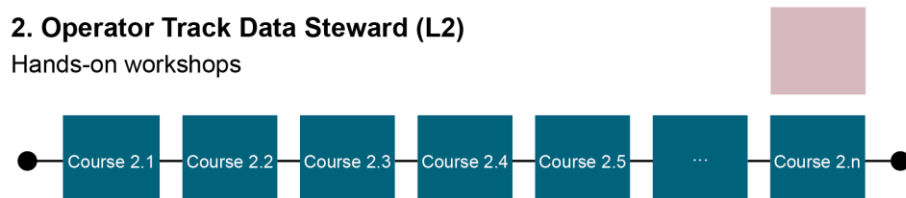
3. Operator Track Data Scientist (L2)
Hands-on workshops

Course 3.1 — Course 3.2 — Course 3.3 — Course 3.4 — Course 3.5 — ... — Course 3.n
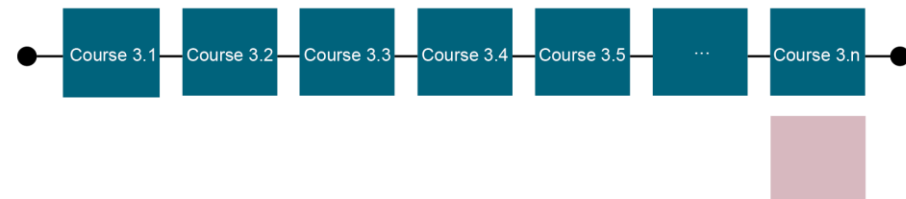
Level
L 1: Level „understand"
L 2: Level „apply"

Courses
Basic structure of the tracks
discipline-sp. courses

U Bremen Research Alliance

Lectures are:

1. Classified according to the level of prior knowledge
- **Starter track, L1**
- **Operator tracks, L2**

2. Operator track separated by topic
- **„Data Steward" track**
- **„Data Scientist" track**

Discipline-specific qualifications via scientific fields and NFDI consortia

> 40 lecturers
- Internal and external institutions
- Industrial and commercial enterprises
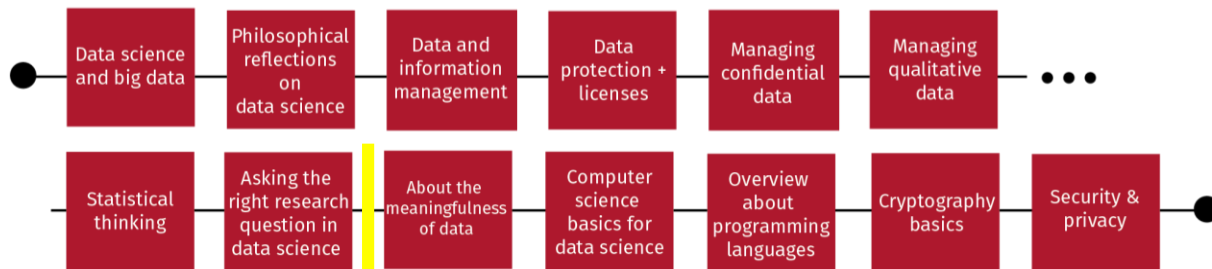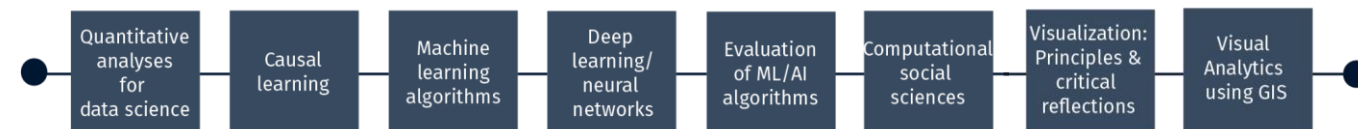
On average > 100 participants

Data Train
Training in Research Data Management and Data Science

DATA TRAIN Tracks

▶ **Starter track, overview lectures (online),** started in March 2021

| Data science and big data | Philosophical reflections on data science | Data and information management | Data protection + licenses | Managing confidential data | Managing qualitative data | ... |

| Statistical thinking | Asking the right research question in data science | About the meaningfulness of data | Computer science basics for data science | Overview about programming languages | Cryptography basics | Security & privacy |

▶ **Data Steward track, hands-on workshops (online),** start in June 2021

| Getting started with R | Erste Schritte mit MATLAB | Getting started with Python | Git/ Git Hub | Tools for FAIR data handling: DMP tool | Reproducibility in science – How and why? | Data preparation | Data provisioning | Data base skills |

▶ **Data Scientist track, hands-on workshops (hybrid if possible),** start in September 2021

| Quantitative analyses for data science | Causal learning | Machine learning algorithms | Deep learning/ neural networks | Evaluation of ML/AI algorithms | Computational social sciences | Visualization: Principles & critical reflections | Visual Analytics using GIS |

UBremen Research Alliance

Data Train
Training in Research Data Management and Data Science

Thank you for your attention!

Whitepaper on the establishment of cooperative research data management in the U Bremen Research Alliance;
https://zenodo.org/record/4775371
> English translation is coming soon.

Contact:
data-train@vw.uni-bremen.de

Websites:
https://www.uni-bremen.de/data-train
>> Research Data Working Group
>> Data Train Team

UBremen Research Alliance

Data Train
Training in Research Data Management and Data Science