



#NFDI\_de

# National Research Data Infrastructure (NFDI)

York Sure-Vetter  
February 24th, 2021

#NFDI\_de

# Vision & Mission

More and more research results are achieved  
by using **already existing** research data.

see FAIR principles: <https://doi.org/10.1038/sdata.2016.18>

However, **significant efforts** are still required for  
**finding** the right data set, **understand** it and **use** it  
for the purpose at hand.

see FAIR principles: <https://doi.org/10.1038/sdata.2016.18>

# „Stand on the shoulders of giants“

#NFDI\_de

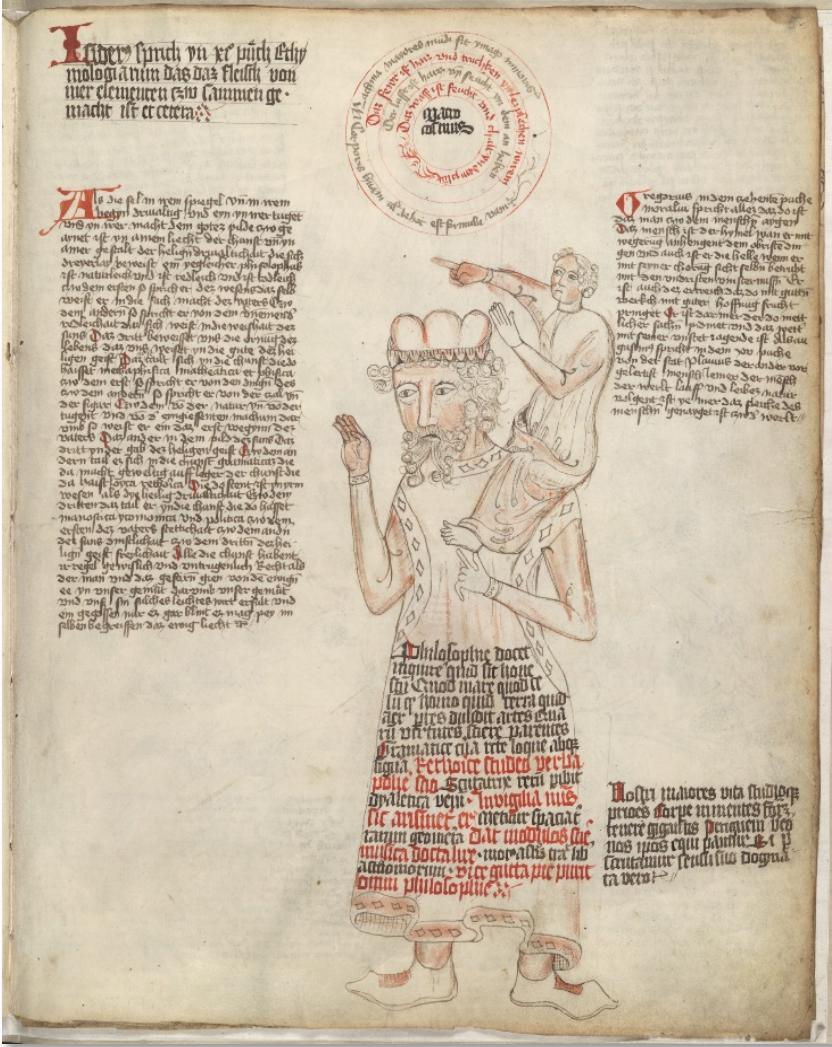


Figure (Southern Germany, approx. 1410)

## Research data sets

Empower research and offer new perspectives to society, nature and life in general.

## Today

Decentral, project-financed and often only temporarily stored data collection.

# Mission statement

#NFDI\_de

- The #NFDI is a living network of active partners ...
- who jointly share goals, values and ...
- an aspiration for high quality, that ...
- in a dialog with and according to the needs of science ...
- deliver research infrastructure services.



Figure: pixabay.com (Pixabay License)

Research data sets are  
FAIR\*.  
For all\*\*. Forever\*\*\*.

\* Findable, Accessible, Interoperable, Reusable

\*\* Depending on data access categories.

\*\*\* Alice: „How long is forever?“

White rabbit: „Sometimes, just one second.“

#NFDI\_de

# From F to R Example for FAIRification

## „Findable“

The first step in (re)using data is to find them. Metadata and data should be easy to find for both humans and computers. Machine-readable metadata are essential for automatic discovery of datasets and services, so this is an essential component of the FAIRification process.

- F1. (Meta)data are assigned a globally unique and persistent identifier
- F2. Data are described with rich metadata (defined by R1 below)
- F3. Metadata clearly and explicitly include the identifier of the data they describe
- F4. (Meta)data are registered or indexed in a searchable resource ...“

Source: <https://www.go-fair.org/fair-principles/>

- In 2010 GESIS founded da|ra – Registry Agency for Social and Economic Data
- The European Values Study (EVS) received the first PID from da|ra
- Cooperation with ZBW and DataCite



Bild: GESIS

**da|ra**  
Registry Agency for  
Social and Economic Data  
<http://www.da-ra.de>

- The idea was simple:
  - **Assign DOIs and rich metadata ...**
  - **to make the data set „citable“**

### Examples:

- Sozio-oekonomisches Panel (SOEP), Version 35, Daten der Jahre 1984-2018 (SOEP-Core v35).  
2019. DOI: 10.5684/soep-core.v35
- EVS (2020): European Values Study 2017: Integrated Dataset (EVS 2017). GESIS Data Archive,  
Cologne. ZA7500 Data file Version 4.0.0, <https://doi.org/10.4232/1.13560>

- The idea was simple:
  - Assign DOIs and rich metadata
  - ... to make the data set „citable“
  - **... and findable**

The screenshot shows the da|ra Data Search interface. The search bar contains the text "european values study". The main result is the "European Values Study 2017: Integrated Dataset (EVS 2017)" dataset. On the left, there are filters for "Metadata" (set to German), "Downloads" (set to English), and "Citation". On the right, the dataset details are shown: DOI (https://doi.org/10.4232/1.13560), Version (4.0.0), and Resource Type (Dataset : Survey and aggregate data). The da|ra logo is in the top left corner of the page.

da|ra

GET A DOI SUPPORT SEARCH NEWS & EVENTS Contact Sign-in

Data Search

european values study

Advanced Search

Back to results

Metadata ▾

German English

Downloads ▾

Citation

DOI

<https://doi.org/10.4232/1.13560>

Version

4.0.0

Resource Type

Dataset : Survey and aggregate data

Source: <https://www.da-ra.de/>

<b>Bibliographic Citation</b>	
Citation 	EVS (2020): <i>European Values Study 2017: Integrated Dataset (EVS 2017)</i> . GESIS Data Archive, Cologne. ZA7500 Data file Version 4.0.0, <a href="https://doi.org/10.4232/1.13560">https://doi.org/10.4232/1.13560</a>
Study No.	ZA7500
Title	European Values Study 2017: Integrated Dataset (EVS 2017)
Current Version	4.0.0, 2020-10-20, <a href="https://doi.org/10.4232/1.13560">https://doi.org/10.4232/1.13560</a>
Date of Collection	19.06.2017 - 01.03.2020
Principal Investigator/ Authoring Entity, Institution	<ul style="list-style-type: none"><li>· Gedeshi, Ilir - Center for Economic and Social Studies (CESS), Tirana, Albania</li><li>· Poghosyan, Gevorg - Institute of Philosophy, Sociology and Law, Armenian National Academy of Sciences, Yerevan, Armenia</li><li>· Kitzinger, Sylvia - Department of Government, University of Vienna, Vienna, Austria</li><li>· Rotman, David - The Center of Sociological and Political Research, Belarus State University, Minsk, Belarus</li></ul>

Source: <https://dbk.gesis.org/dbksearch/sdesc2.asp?no=7500&db=e&doi=10.4232/1.13560>



Bild: GESIS

Thanks to the great team!

... actually, da|ra addressed also the **remaining letters** of FAIR:

- A1. (Meta)data are retrievable by their identifier using a standardised communications protocol
- A2. Metadata are accessible, even when the data are no longer available
- I1. (Meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
- I2. (Meta)data use vocabularies that follow FAIR principles
- I3. (Meta)data include qualified references to other (meta)data
- R1. (Meta)data are richly described with a plurality of accurate and relevant attributes

Source: <https://www.go-fair.org/fair-principles/>

<b>Bibliographic Citation</b>	
Citation 	EVS (2020): European Values Study 2017: Integrated Dataset (EVS 2017). Cologne. ZA7500 Data file Version 4.0.0, <a href="https://doi.org/10.4232/1.13560">https://doi.org/10.4232/1.13560</a>
Study No.	ZA7500
Title	European Values Study 2017: Integrated Dataset (EVS 2017)
Current Version	4.0.0, 2020-10-20, <a href="https://doi.org/10.4232/1.13560">https://doi.org/10.4232/1.13560</a>
Date of Collection	19.06.2017 - 01.03.2020
Principal Investigator/ Authoring Entity, Institution	<ul style="list-style-type: none"><li>Gedeshi, Ilir - Center for Economic and Social Studies (CESS), Tirana, Albania</li><li>Poghosyan, Gevorg - Institute of Philosophy, Sociology and Law, Armenian National Academy of Sciences, Yerevan, Armenia</li><li>Kritzinger, Sylvia - Department of Government, University of Vienna, Vienna, Austria</li><li>Rotman, David - The Center of Sociological and Political Research, Belarus State University, Minsk, Belarus</li></ul>

„R“ (Reuse) of FAIR:  
Versioning allows for  
better reproducibility  
of results

Source: <https://dbk.gesis.org/dbksearch/sdesc2.asp?no=7500&db=e&doi=10.4232/1.13560>

- **The effect has since been pervasive:**
  - research data sets quickly became “findable” in search engines
  - ... without additional efforts

The screenshot shows a Google search results page for the DOI 10.4232/1.10059. The search bar at the top contains the text "10.4232/1.10059". Below the search bar, there is a "Scholar" search result section with the text "Ungefähr 14 Ergebnisse (0,04 Sek.)" and a "Beliebige Zeit" filter button. A tip message in German encourages searching in German. The main search results list includes a link to a paper titled "Subjective Life Satisfaction in the European Union: Determinants and Policy Implications" by M Hlavac, available at SSRN 1853623, 2011, and a link to the GESIS dataset version 1.0.0 (2010-06-30) with DOI 10.4232/1.10059.

Source: <https://www.google.com/> (... in ~2012)

- **in operation since 2010**
- supported by two **DFG grants** (THANKS!)
- evolved into **da|ra Consortium** in 2020
  - AUSSDA, AHRI, AfM, Banco de Portugal, bibb BZgA, WZB, ZOBIS, DIPF, DIW Berlin, DJI, DZA, Deutsche Bundesbank, DeZIM, FELU, bifie, gesis, fdz.DZHW, RatSWD, DIE, SWP, ifu, HSRC, HS Wirtschaft Berlin, IAB, INDEPTH, IQB, IREE, IWH, IOS, UB Mannheim, EKKE, destatis, RKI, RWI, FDZ BOD, SHARE, SOEP, SRDA, SASD, SocialData Network, ADP, SAMRC, ifbq, FFZG, WVS, ZBW, ZEW, ZPID, CSDA, ihi
- ... and part of the **NFDI** via **KonsortSWD**

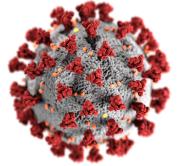


#NFDI\_de

From F to R  
Two recent examples

# CORD-19 The Covid-19 Open Research Dataset

#NFDI\_de



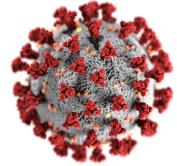
Source: PHIL #23312

„The COVID-19 Open Research Dataset (CORD-19) is a growing resource of scientific papers on COVID-19 and related historical coronavirus research. CORD-19 is designed to facilitate the development of **text mining** and **information retrieval systems** over its rich collection of metadata and structured full text **papers**. Since its release, CORD-19 has been **downloaded over 200K times** and has served as the basis of many COVID-19 text mining and discovery systems.“

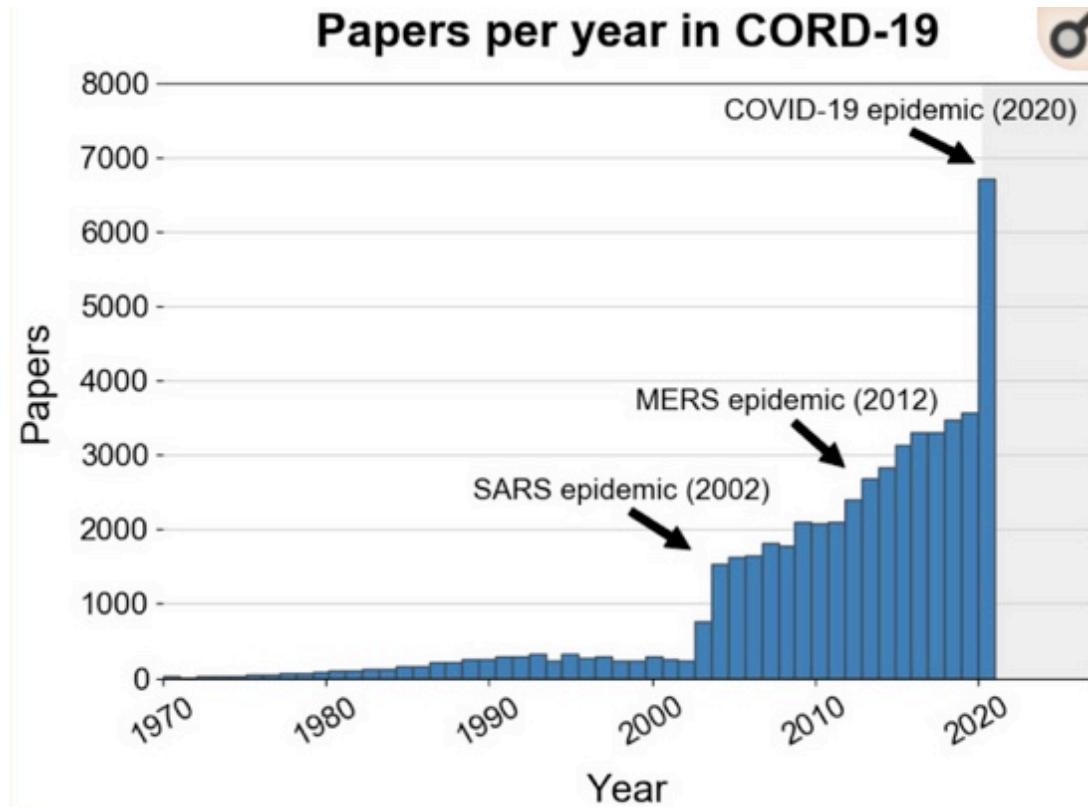
Source: <https://arxiv.org/abs/2004.10706v4>

# CORD-19 The Covid-19 Open Research Dataset

#NFDI\_de



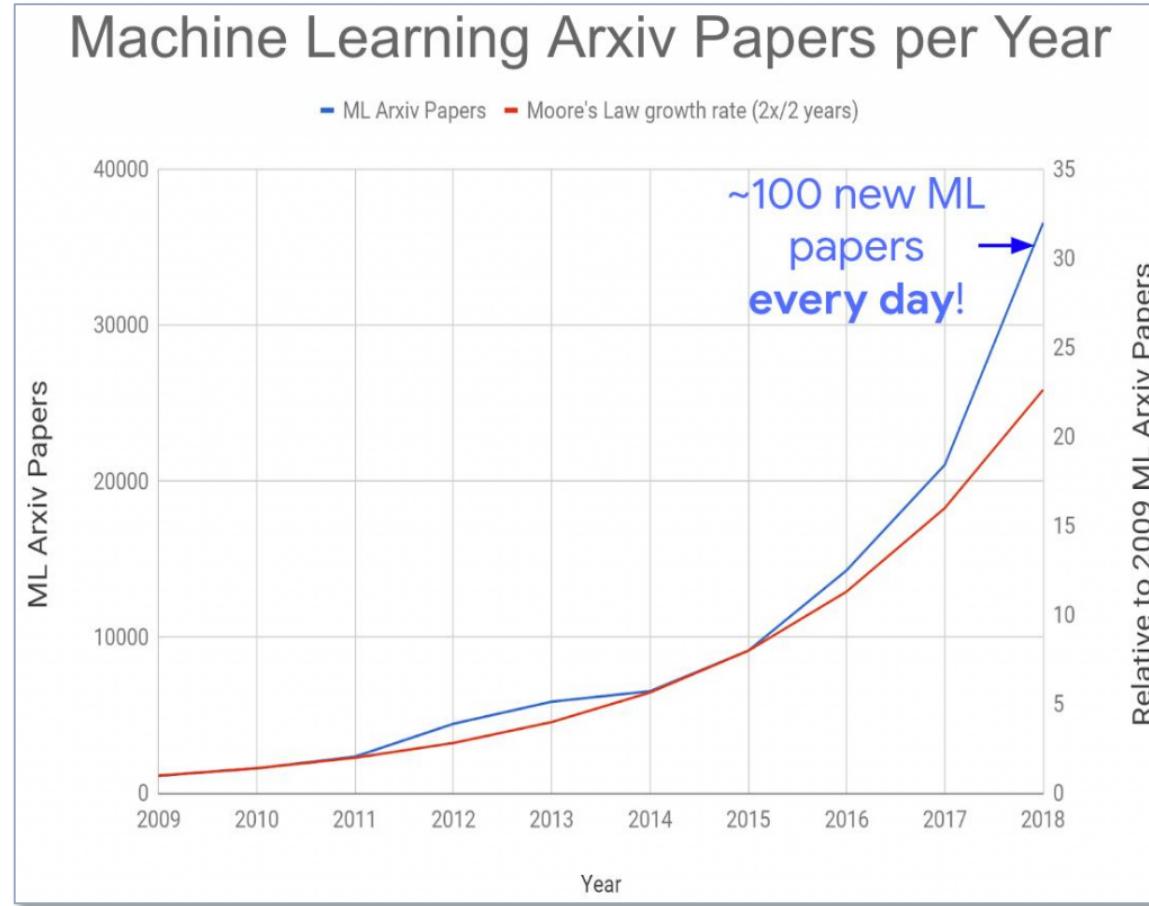
Source: PHIL #23312



Source: <https://arxiv.org/abs/2004.10706v4>

# Machine Learning: explosion of papers

#NFDI\_de



Source: <https://data-mining.philippe-fournier-viger.com/too-many-machine-learning-papers/>

# DBLP: search for „neural network“ (Wednesday, Feb 24th, 2021, 5:28h)

#NFDI\_de

The screenshot shows the DBLP homepage. At the top, there is a logo for SCHLOSS DAGSTUHL Leibniz Center for Informatics, followed by navigation links: home | browse | search | about. Below this, the DBLP logo (stylized blue and yellow geometric shapes) and the text "computer science bibliography" are displayed. A search bar contains the query "neural network". To the left of the search bar, there is a button labeled "[+]" and "Search dblp for Venues" with a question mark icon. Below this, it says "powered by CompleteSearch, courtesy of Hannah Bast, University of Freiburg". At the bottom, there is a breadcrumb trail "> Home > Search" and a "Dagstuhl" dropdown menu.

[+] **Search dblp for Venues** ?  
[-]

powered by CompleteSearch, courtesy of Hannah Bast, University of Freiburg

> Home > Search

 Dagstuhl

[\[-\] Venue search results](#) 

## Likely matches

- Neural Networks
- International Conference on Artificial Neural Networks and Machine Learning (ICANN)  
also: International Conference on Artificial Neural Networks
- International Symposium on Neural Networks (ISNN)
- IEEE Transactions on Neural Networks and Learning Systems  
also: IEEE Transactions on Neural Networks
- The European Symposium on Artificial Neural Networks (ESANN)

All 22 matches

- Artificial Neural Networks (ANN)
- Artificial Neural Networks and Expert Systems (ANNES)
- Artificial Neural Networks and Intelligent Information Processing (ANNIIP)  
also: Artificial Neural Networks (ANNS)
- Brazilian Symposium on Neural Networks (SBRN)
- Conference on Artificial Neural Networks in Medicine and Biology (ANNIMAB)

# DBLP: search for „neural network“ (Wednesday, Feb 24th, 2021, 5:28h)

#NFDI\_de

Publication search results

found 100,545 matches

2021

Alireza M. Javid:  
**Neural Network Architecture Design: Towards Low-complexity and Scalable Solutions.** Royal Institute of Technology, Stockholm, Sweden, 2021

Jan Melchior:  
**On the importance of centering in artificial neural networks.** Ruhr University Bochum, Germany, 2021

Thai-Son Nguyen:  
**High Performance Neural Networks for Online Speech Recognizer.** Karlsruhe Institute of Technology, Germany, 2021

Michael Zgurovsky, Victor Sineglazov, Elena Chumachenko:  
**Artificial Intelligence Systems Based on Hybrid Neural Networks - Theory and Applications.** Studies in Computational Intelligence 904, Springer 2021, ISBN 978-3-030-48452-1, pp. 1-510

Abdulaziz Aldegeishem, Mubbashra Anwar, Nadeem Javaid, Nabil Ali Majahid, Muhammad Shafiq, Hasan Ahmed:  
**Towards Sustainable Energy Efficiency With Intelligent Electricity**

[–] Refine list

refine by author

Jinde Cao (374)  
Tingwen Huang (225)  
Patricia Melin (212)  
Jun Wang 0002 (211)  
Zhigang Zeng (205)  
Huaguang Zhang (149)  
Witold Pedrycz (146)  
Zhanshan Wang (139)  
Yoshua Bengio (136)  
Chuandong Li 0001 (126)  
155,105 more options

refine by venue

CoRR (12,243)  
Neural Networks (4,517)  
IEEE Trans. Neural Networks (3,061)  
IEEE Trans. Neural Networks Learn. Syst. (2,853)  
Neurocomputing (2,837)  
IJCNN (2,737)  
IEEE Access (1,740)  
Neural Comput. Appl. (1,709)  
ISNN (1,502)  
ICASSP (1,108)  
4,423 more options

**refine by year**

- 2021 (1,473)
- 2020 (14,547)
- 2019 (13,355)
- 2018 (10,945)
- 2017 (7,490)
- 2016 (5,087)
- 2015 (3,648)
- 2014 (2,892)
- 2013 (2,565)
- 2012 (2,429)
- 35 more options*

# Making Neural Networks FAIR

## Source code, models, training data sets

#NFDI\_de

The screenshot shows the arXiv.org page for the paper 'Making Neural Networks FAIR'. At the top left is the Cornell University logo. Below it, the arXiv.org URL and category ('Computer Science > Machine Learning') are displayed. The title 'Making Neural Networks FAIR' is prominently shown in large bold black font. Below the title, the authors' names (Anna Nguyen, Tobias Weller, York Sure-Vetter) and the submission date ('Submitted on 26 Jul 2019') are listed. The abstract begins with a paragraph about the growth of neural networks and the lack of FAIR data management.

Cornell University

arXiv.org > cs > arXiv:1907.11569

Computer Science > Machine Learning

**Making Neural Networks FAIR**

Anna Nguyen, Tobias Weller, York Sure-Vetter

(Submitted on 26 Jul 2019)

Research on neural networks has gained significant momentum over the past few years. A plethora of neural networks is currently being developed, both in research and industry. Because training is a resource-intensive process and training data cannot always be made available, there has been a recent trend to attempt to re-use already-trained neural networks. As such, neural networks themselves have become a valuable commodity. In this paper, we present the Neural Network Ontology, an ontology to make neural networks findable, accessible, interoperable and reusable. We also propose well-established FAIR guiding principles for scientific data management and stewardship. We created the new FAIRnets Dataset that

Source: <https://arxiv.org/abs/1907.11569>

# Making Neural Networks FAIR

## Source code, models, training data sets

#NFDI\_de

Cornell University

arXiv.org > cs > arXiv:1907.11569

Computer Science > Machine Learning

**Making Neural Networks FAIR**

Anna Nguyen, Tobias Weller, York Sure-Vetter

(Submitted on 26 Jul 2019)

Research on neural networks has gained significant momentum over the past few years. A plethora of neural networks is currently being developed, both in research and industry. Because training is a resource-intensive process and training data cannot always be made available, there has been a recent trend to attempt to re-use already-trained neural networks. As such, neural networks themselves have become a valuable asset. In this paper, we present the Neural Network Ontology, an ontology to make neural networks findable, accessible, interoperable and reusable. We follow the well-established FAIR guiding principles for scientific data management and stewardship. We created the new FAIRnets Dataset that

FAIRnets

car

All Privacy About Index Repository SPARQL aifb

About results (178 hits)

danielwo/CarND-Behavioral-Cloning-P3  
<https://w3id.org/nno/data#danielwo/CarND-Behavioral-Cloning-P3>  
Behavioral Cloning via Deep Learning (training a simulated car to drive)

scrambleegg7/CarND-Behavioral-Cloning-P3  
<https://w3id.org/nno/data#scrambleegg7/CarND-Behavioral-Cloning-P3>  
UdaCity CarND-Behaviral-Cloning-P3

Kazutaka333/behavioral\_cloning  
[https://w3id.org/nno/data#Kazutaka333/behavioral\\_cloning](https://w3id.org/nno/data#Kazutaka333/behavioral_cloning)  
deep neural net that clones driving a car in a simulator

derzaarsad/behavioral\_clonina

Source: <https://arxiv.org/abs/1907.11569>

#NFDI\_de

# Overview

**NFDI is funded by the German federation and all federal states.**

**NFDI receives ~900 MEUR for a period of 10 years.**

**NFDI consists of a coordinated network of (up to) 30 consortia.**

- **DFG - Deutsche Forschungsgemeinschaft** hosts the reviewing of consortia proposals
- ... large number of independant reviewers involved in reviews
- **Recommended:** two videos provided by DFG
  - „What is the NFDI“:  
<https://www.youtube.com/watch?v=x3Cvn1vNQ98&feature=youtu.be>
  - „The NFDI review process“:  
<https://www.youtube.com/watch?v=XTzwPwMAqHM&feature=youtu.be>

## Consortia

- DataPlant
- GHGA
- KonsortSWD
- NFDI4BioDiversity
- NFDI4Cat
- NFDI4Chem
- NFDI4Culture
- NFDI4Health
- NFDI4Ing

## Some numbers:

- Round 1 (out of 3 rounds)
- 9 consortia (out of 30 consortia)
- ~100 (co-)applicant organisations
- ~150 (co-)spokespersons
- ~270 further participants

NFDI brings together **people** coming from:

- Research and Research Infrastructure

NFDI is covering organisations, as of today (i.e. round 1), coming from:

- (few|**many**|all) German States
- (few|**many**|all) Research Disciplines
- (few|**many**|all) Universities
- (few|many|**all**) Research Organisations (Fraunhofer, Helmholtz, Leibniz, Max-Planck)
- ... and many other related research organisations

**Commercial Director: Eva Lübke**  
01.10.2020

#NFDI\_de



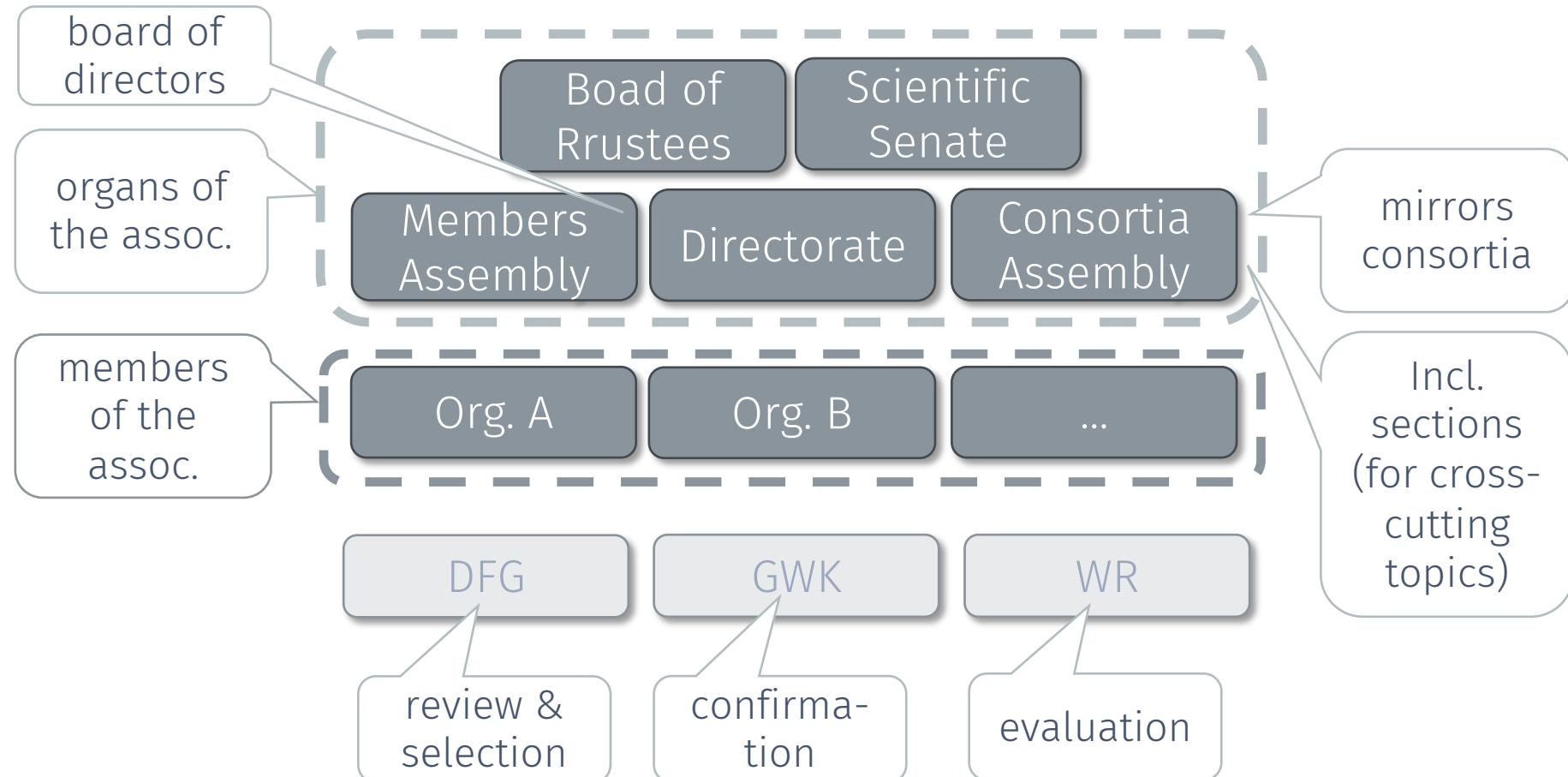
# Founding assembly of NFDI e. V.

12.10.2020 in Hannover, Germany

#NFDI\_de



- Website (incl. overview of open jobs)
  - <https://www.nfdi.de/>
- Community pages:
  - [https://twitter.com/NFDI\\_de](https://twitter.com/NFDI_de)
  - <https://de.linkedin.com/company/nfdi-de>
  - <https://zenodo.org/communities/nfdi/>
- Mailing lists, Rocket.Chat, ...
- Host frequent meetings: spokespersons, coordinators, ...



## Status & next steps

#NFDI\_de

- Q4-2020:
- Funding for consortia starts
  - Foundation of association NFDI e.V.
  - „Baton change“ from FIZ Karlsruhe/KIT
  - Entry at association register
  - First meetings of some organs
  - Start of work in consortia
  - First workshops with EOSC, GAIA-X, ...
  - EOSC membership application (foreseen to be „mandated member“)



Quelle: fotolia

## Status & next steps

#NFDI\_de

- Q1/2-2021:
  - Transfer of holdings into association
  - Establish organs of the association
  - Grow team of directorate
  - Establish links to EOSC and GAIA-X
- Q3/4-2021:
  - Onboarding of Round 2
- 2022:
  - Onboarding of Round 3
- 2025:
  - Evaluation by Wissenschaftsrat



Quelle: fotolia

#NFDI\_de

There is one last thing ...

**Tell us your data story**

#NFDI\_de

- Did you successfully solve an **interdisciplinary research question**?
- Did you successfully **(re-)use data sets**?
- Did your work benefit from **data sets of another community**?

**Tell us your data story**

#NFDI\_de

- Did you successfully solve an **interdisciplinary research question**?
- Did you successfully **(re-)use data sets**?
- Did your work benefit from **data sets of another community**?

If any of the answers is YES:

**Please share your research data story!**

**#RDatastory**

**#NFDIrocks**

**Tell us your data story**

#NFDI\_de

- Did you successfully solve an interdisciplinary research question?
- Did you successfully (re-)use data sets from different research communities?

**If any of the answers is YES:  
Please share your data story!**

[#RDatastory](#)

[#NFDIrocks](#)

#NFDI\_de

WE ❤ DATA

## NFDI Team:

(Source: nfdi.de)





## Questions?

We are looking forward to the next steps of RDA and NFDI!

### Contact

Prof. Dr. York Sure-Vetter

**National Research Data Infrastructure (NFDI)**

Albert-Nestler-Straße 13

76131 Karlsruhe

Germany

[info@nfdi.de](mailto:info@nfdi.de)

[nfdi.de](http://nfdi.de)