

# Why science needs to be FAIR?

Institut für Medizinische Informatik

Prof. Dr. Oya Beyan

# FAIR principles as an answer to Reproducibility Crises



**nature** International weekly journal of science

Home | News & Comment | Research | Careers & Jobs | Current Issue | Archive | Audio & Video

Archive > Specials and supplements archive > Challenges in irreproducible research

## How quality control

It may not be sexy, but quality

Monya Baker

27 January 2016

See all s

SPECIAL

## Studies show only 10% of published science articles are reproducible. What is happening?

Posted on May 3, 2012 by Moshe Pritsker

Studies show a very low reproducibility for articles published in scientific journals, often as low as 10-30%. Here is a partial list:

- The biotech company Amgen had a team of about 100 scientists trying to reproduce the findings of 53 “landmark” articles in cancer research published by reputable labs in top journals. [Only 6 of the 53 studies were reproduced](#) (about 10%).
- Scientists at the pharmaceutical company, Bayer, examined 67 target-validation projects in oncology, women’s health, and cardiovascular medicine. Published results were reproduced in only [14 out of 67 projects](#) (about 21%).
- The project, PsychFileDrawer, dedicated to replication of published articles in experimental psychology, shows a [replication rate 3 out of 9](#) (33%) so far.

### CHALLENGES IN IRREPRODUCIBLE RESEARCH

Science moves forward by corroboration – when research advances faster when people waste less time pursuing false leads. However, it is often considered to be the final word, but there are too many things that can go wrong.

There is growing alarm about results that cannot be reproduced. This is worrying: replicating an experiment is a main foundation of the scientific method. Levels of scrutiny, complexity of experiments and statistics are increasing. Scientists, institutions and funders all have a part in tackling this problem. We need substantive steps to improve the transparency and robustness in what we publish, and to promote awareness within the scientific community. We hope that the articles contained in this collection will

Recent reports in the **Washington Post** and the **Economist**, among others, raise the concern that relatively few scientists' experimental findings can be replicated. This is worrying: replicating an experiment is a main foundation of the scientific method.

# Fair Scientific Knowledge: Benefits All

Transparent



Accessible



*Science and Society:  
Knowledge is exchanged, tested and reflected between  
the various interest groups.*

# Fair Scientific Knowledge: Benefits All

Transparent



Accessible

*open to the reuse & judgement of others*

Can I **reproduce** the same analysis with same materials ?  
Are my results consistent with the other study



Researchers

# Fair Scientific Knowledge: Benefits All

Transparent



Accessible

*open to the reuse & judgement of others*



Innovators

Can I **generalize** the outcomes of the scientific study to solve another problem?  
Does results valid with different data and different analyses?



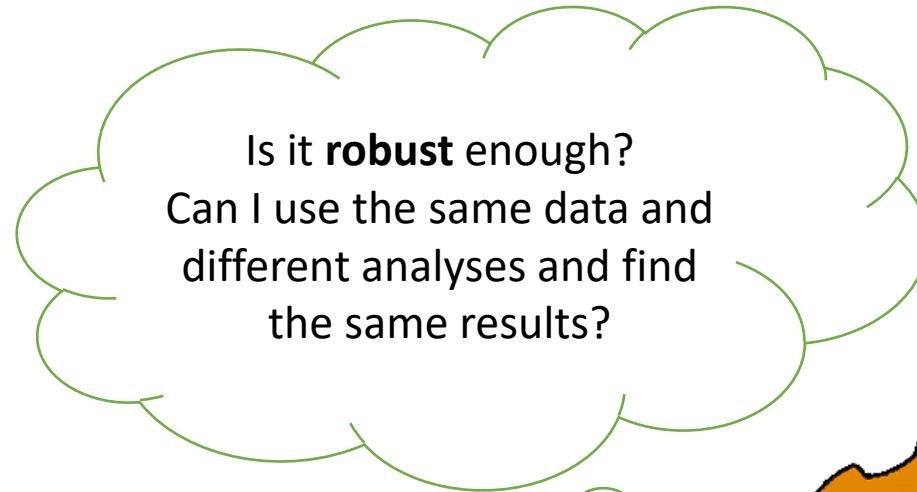
# Fair Scientific Knowledge: Benefits All

Transparent



Accessible

*open to the reuse & judgement of others*



Citizens

# Fair Scientific Knowledge: Benefits All

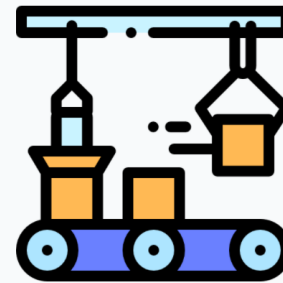
Transparent



Accessible

*open to the reuse & judgement of others*

Can I **replicate** the results in my setting with different data? So that I can develop products.



Industry

# Fair Scientific Knowledge: Benefits All

Transparent



Accessible

*open to the reuse & judgement of others*



Data Scientist

Can I reuse the data and apply data-driven analytics methods to have better solutions?  
Will AI and machine learning perform better?



# Fair Scientific Knowledge: Benefits All but YOU?

Transparent



Accessible

*open to the reuse & judgement of others*



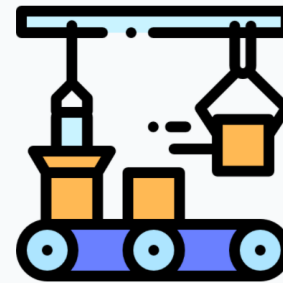
Data Scientist



Researchers



Innovators



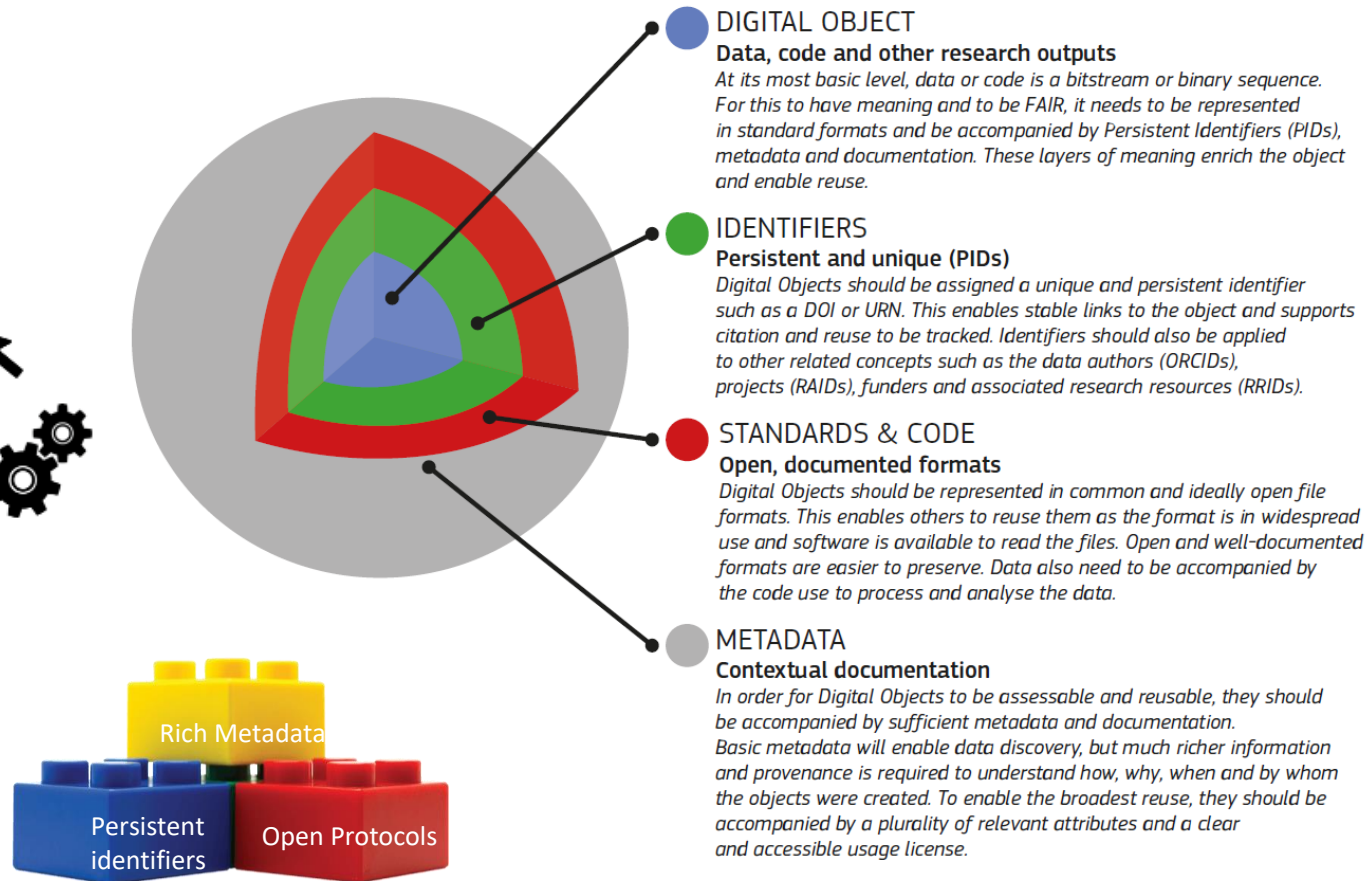
Industry

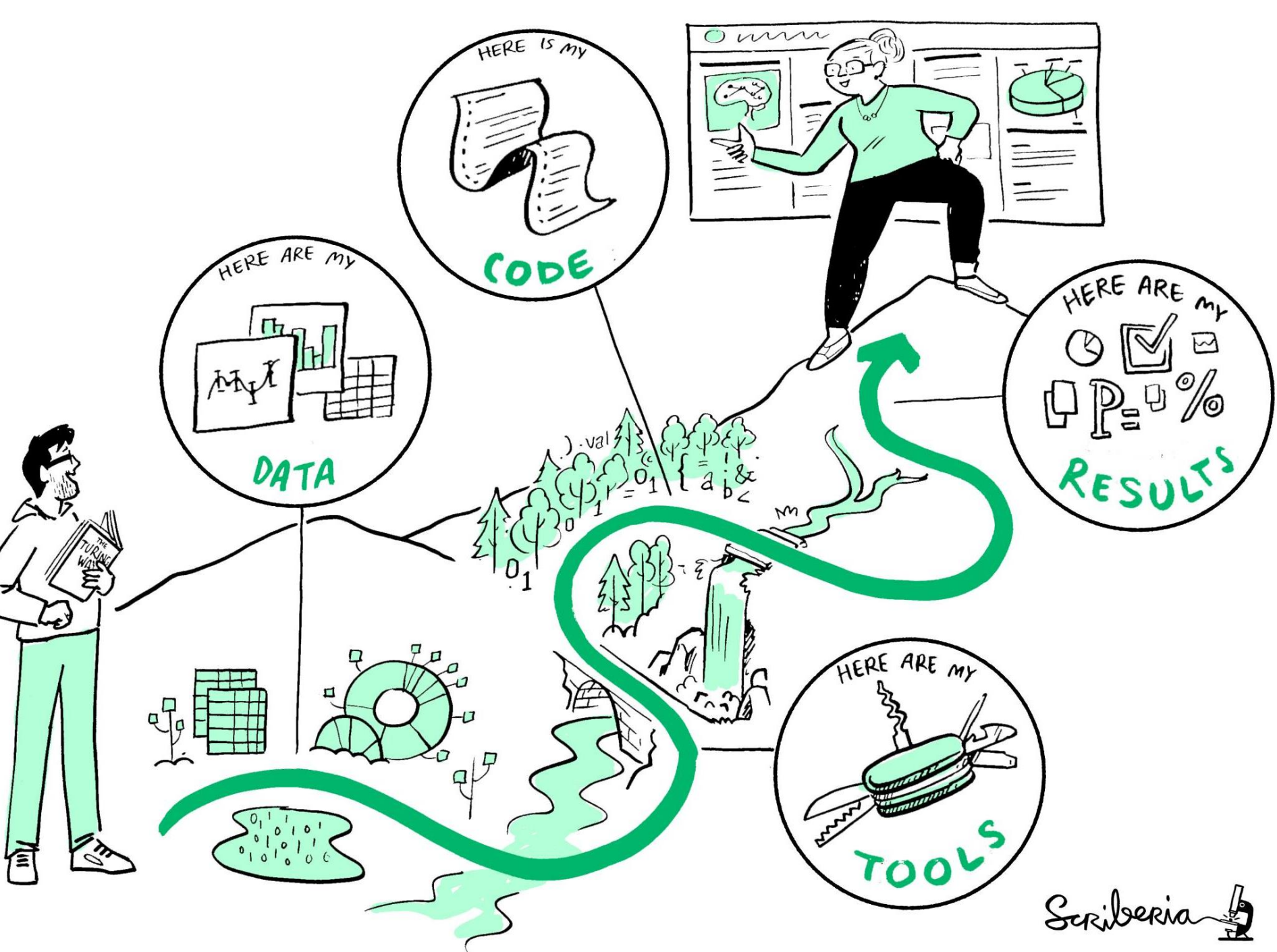


Citizens

# FAIR for Data / Algorithms / Services...

**F**indable   
**A**ccessible   
**I**nteroperable   
**R**eusable 





# Making your data open

- Which data should I share ? Where?
  - raw data / cleaned & integrated data / processed data / analyzed data / published data
- When I share it, can someone reuse it?
  - Human readable: Your colleague asks about the meaning of excel columns.
  - Machine readable: Search engines find the relevant data
- Can I really share it ?
  - Privacy / consent / ethics / licence
- Do we still have the data ?
  - It was stored in a computer at the lab 6 years ago.
  - Which version is the final ?





# Making your methods open

- Can you identify materials
  - Which cell lines did we used? What was the test kid?
- Did you document your study protocol?
  - I have everything at lab books: hypotheses, experiments and initial analysis or interpretation
  - Can someone read it?
- Did you register your study?
  - I got ethical approval does it enough?
  - Why do I need to deposit read only protocols to domain registries ?



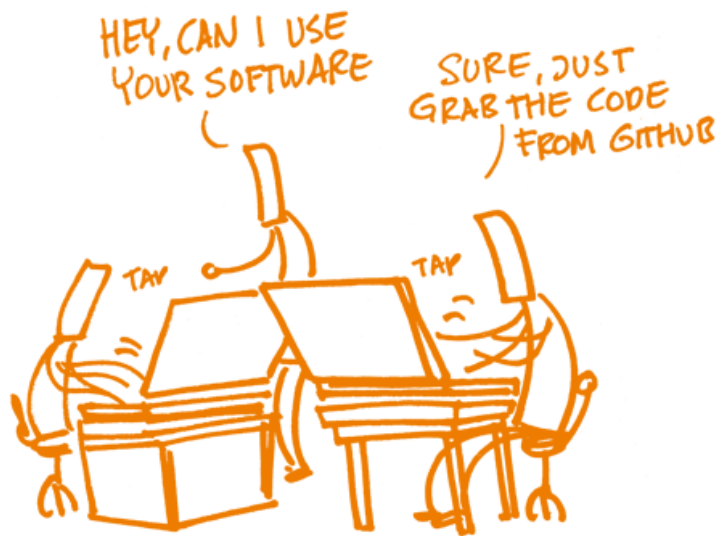
# Making your analysis open

- Do you make your research software open?
  - The software you developed (or used) for analysis, simulation, visualization
  - The models / algorithms that you developed / used

OPEN SOURCE

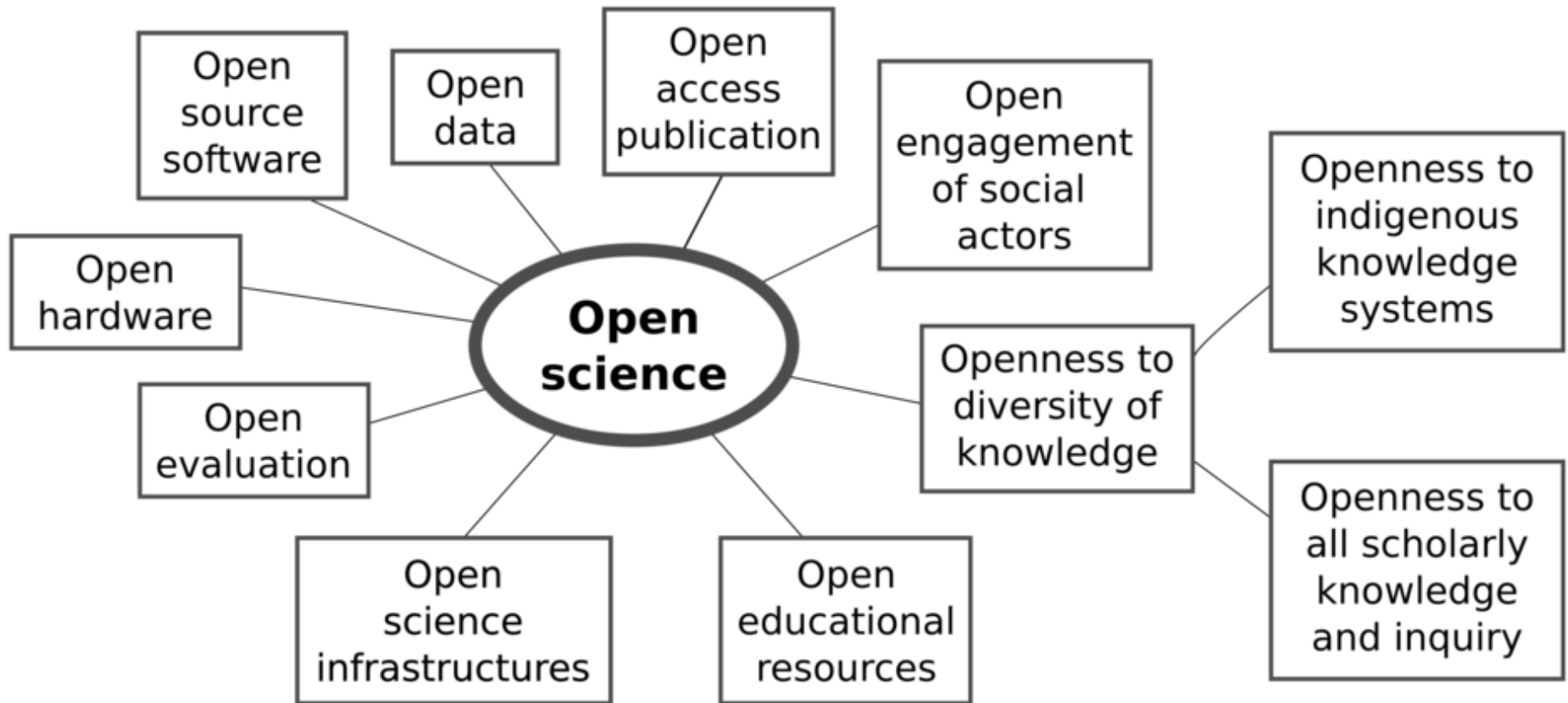
vs

CLOSED SOURCE

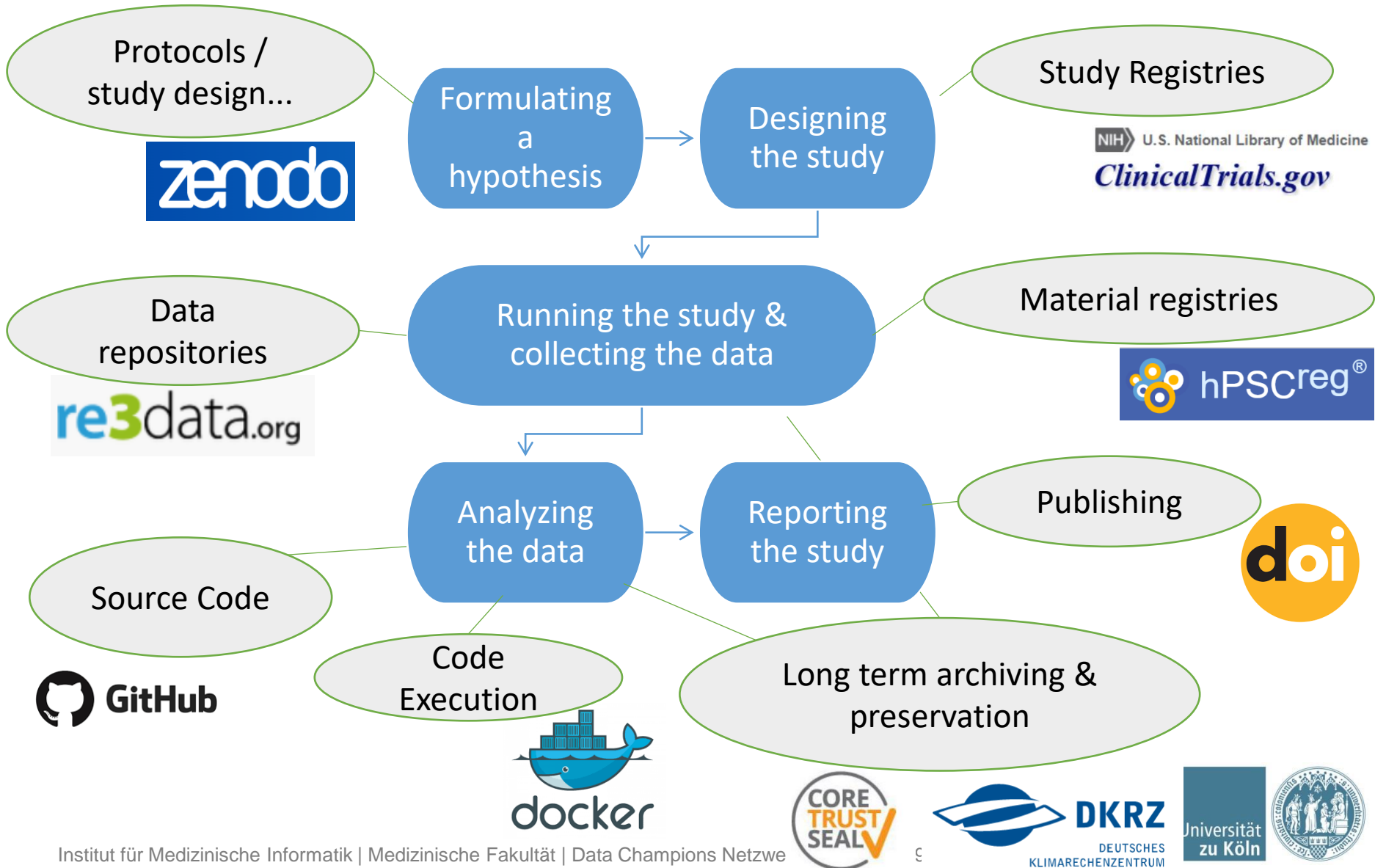




# The Complex Nature of the Open Science



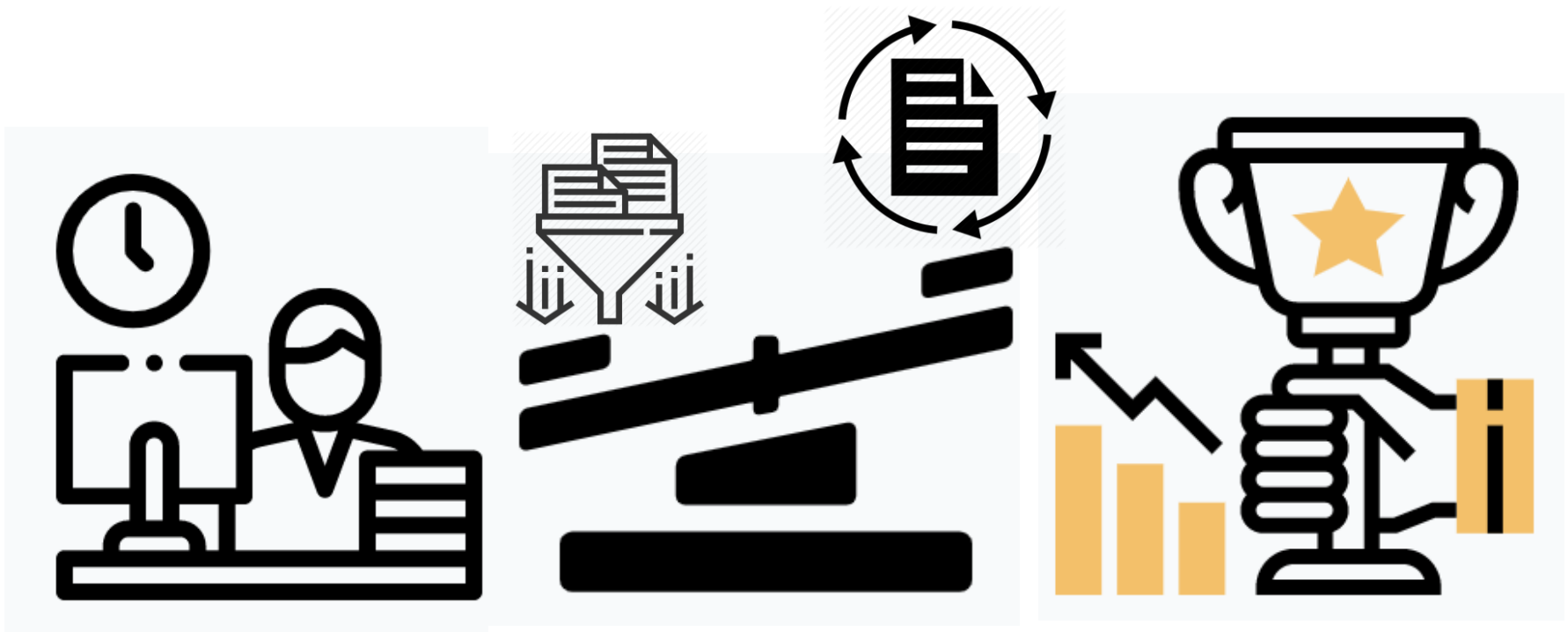
# Research as a Network of FAIR Digital Objects



# Open Science is an investment

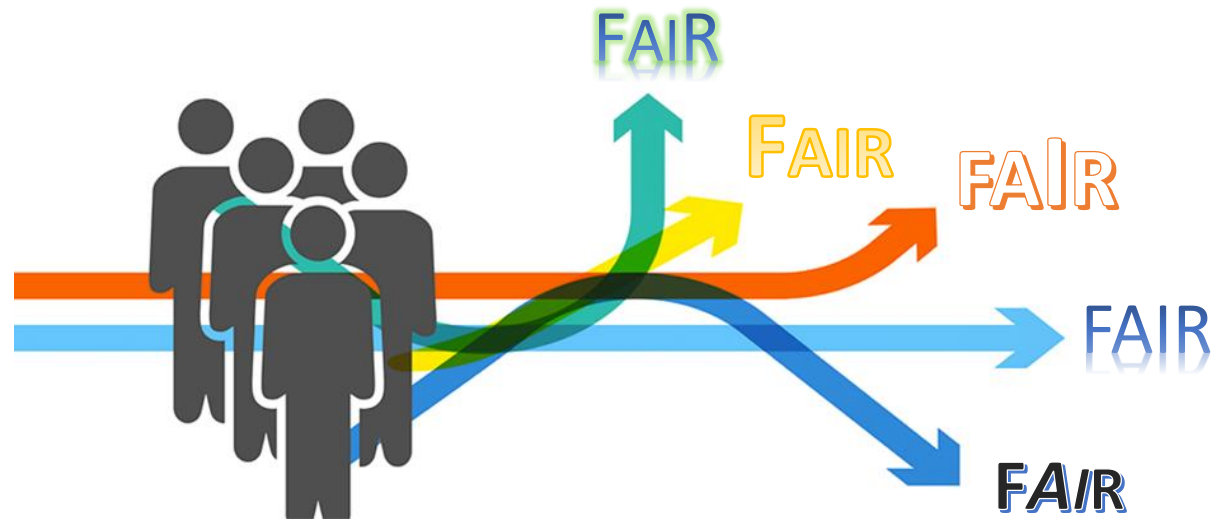
FAIR Data

Producers  $\neq$  Beneficiaries



# Many ways to FAIR

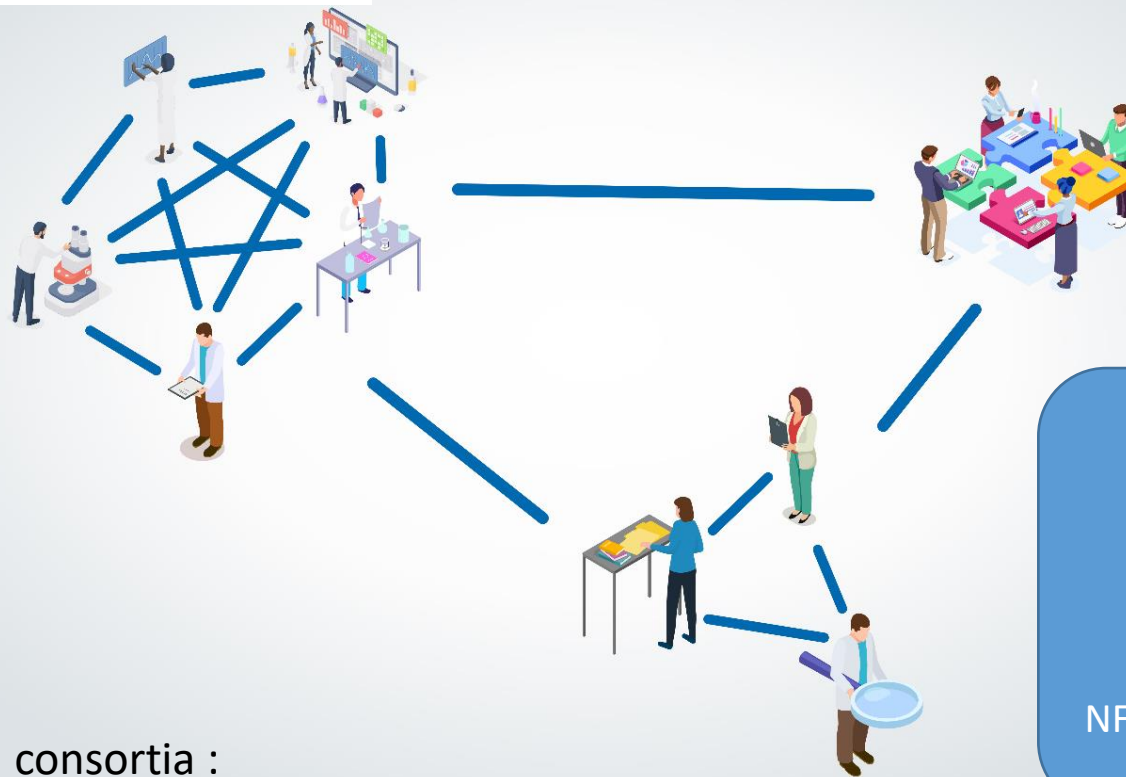
- Each research community has different needs
- Communities may be at very different stage: e.g. different practices, data collection methods, nature of the data/research
- Different data sharing systems / standards



# Scientist should not be alone: Research Data Infrastructures

**DFG** Deutsche  
Forschungsgemeinschaft

**NFDI**



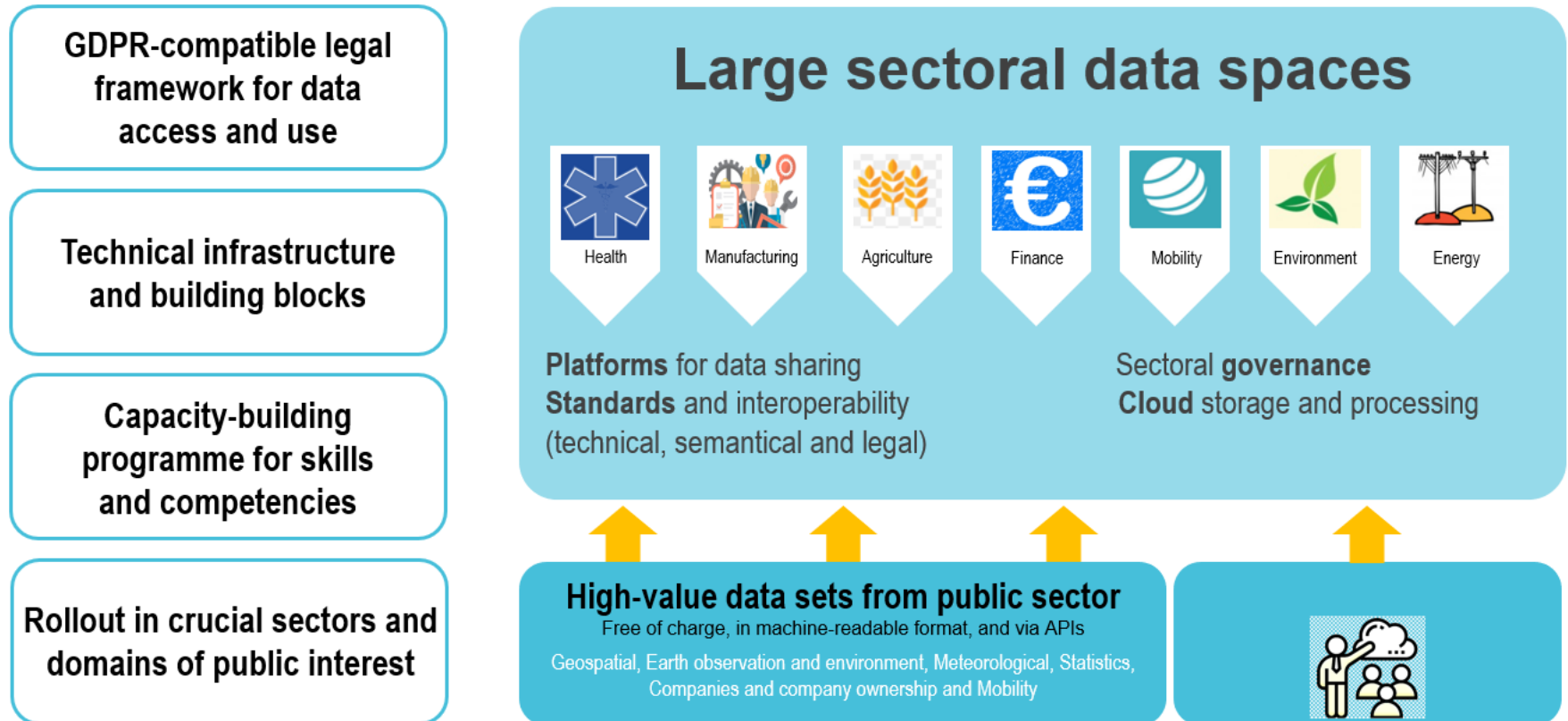
Uni Köln is  
a partner of:  
GHGA  
NFDI4Culture  
NFDI4Chem  
NFDI4Health  
NFDI4DataScience  
....

List of funded consortia :

[https://www.dfg.de/en/research\\_funding/programmes/nfdi/funded\\_consortia/index.html](https://www.dfg.de/en/research_funding/programmes/nfdi/funded_consortia/index.html)

# Common European Data Spaces

<https://digital-strategy.ec.europa.eu/en/policies/strategy-data>

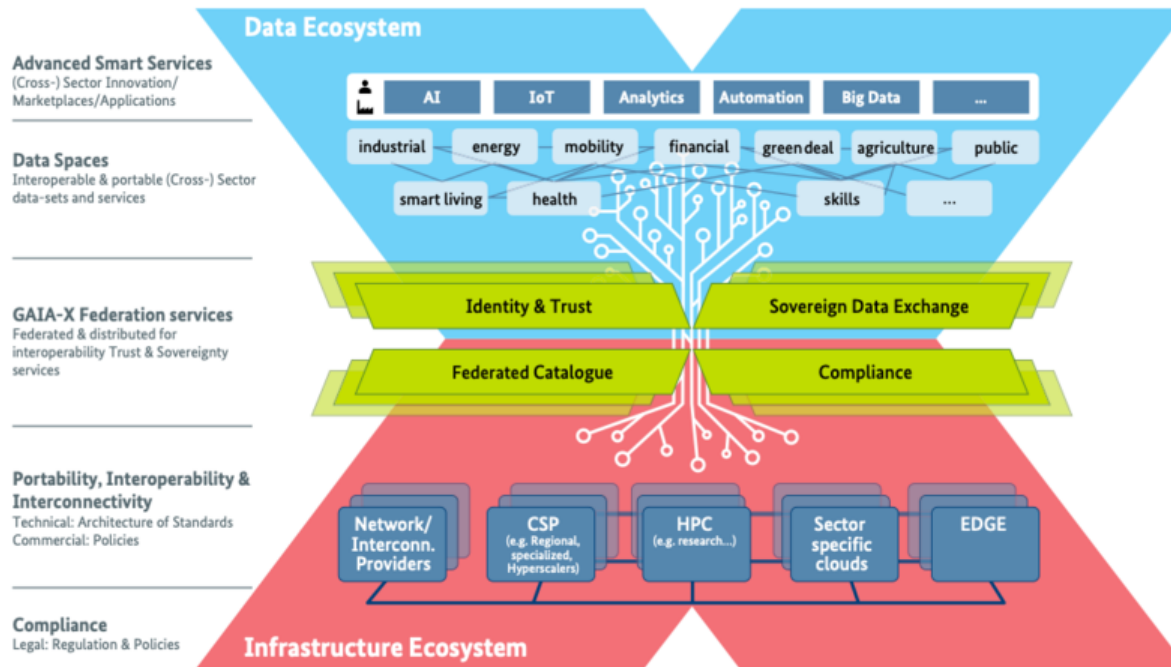


The figure is adapted by EHTEL from a slide from the European Commission (Yvo Volman, Head of Unit, DG CNET, Data Policy and Innovation)



# Industrial Data Spaces : GAIA-X

<https://www.data-infrastructure.eu/GAIA-X/Navigation/EN/Home/home.html>



Source: BMWi

## Gaia-X and NFDI: FAIR Data Spaces for Research & Industry

<https://www.nfdi.de/start-of-the-fair-data-spaces-project/?lang=en>

Köln MII is a partner in with Distributed Analytics & Data Science Ethics

# Your Support at Home

- Cologne Competence Center for Research Data Management
- For the Medical Faculty:
- MeDIC Data Integration Center:
  - MII Initiative: HIGHmed Project
  - Integrates real world patient data for research
  - Newly established Medical Informatics Institute: supports clinical use cases



*making the data intensive research  
and innovation possible ...*

Prof. Dr. Oya Beyan  
Institut für Medizinische Informatik  
Medizinische Fakultät  
Universität zu Köln  
Zülpicher Str. 58, 50674 Köln  
+49 221 478-96238  
[oya.beyan@uni-koeln.de](mailto:oya.beyan@uni-koeln.de)

