

KonsortSWD

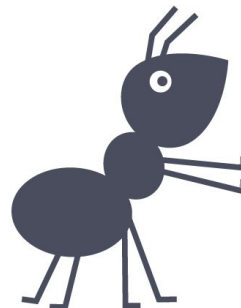


Consortium for the  
Social, Behavioural, Educational  
and Economic Sciences

CESSDA,  
Webinar on User Experience  
with FAIR Evaluation Tools and  
Services

Presentation  
11.10.2022

DOI  
[10.5281/zenodo.7409651](https://doi.org/10.5281/zenodo.7409651)



# Application of 'RDA FAIR Data Maturity Model' to assess the PID registration service in terms of FAIRness

Janete Saldanha Bach  
Claus-Peter Klas  
Peter Mutschke

*GESIS – Leibniz Institute for the Social Sciences*

## Janete Saldanha Bach



Janete Saldanha Bach is a Researcher at GESIS – Leibniz Institute for the Social Sciences, based in the Knowledge Technologies (KTS) Department, team FAIR Data and Human Information Interaction, working in the consortia KonsortSWD Project of the National Research Data Infrastructure (NFDI). She holds a Ph.D. and a Master's degree in Science and Technology Studies (STS) and a bachelor's degree in Information Science. Her research expertise is in Open Science, especially in research data management and data reuse in the Social Sciences. She is currently involved in consortium KonsortSWD, Task Area 5 Measure 1 – developing the conceptual framework for the PID registration service at a variable level and Task Area 5 Measure 2 Enhancing data findability.

## Claus-Peter Klas

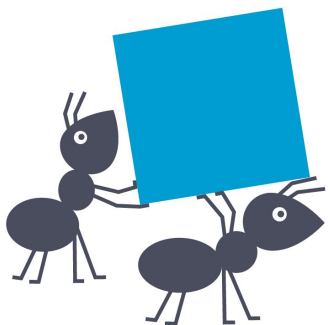


Claus-Peter Klas is lead of the Data & Service Engineering team in the department Knowledge Technologies for the Social Sciences of GESIS. He received his PhD in computer science at the University of Duisburg-Essen and was a postdoctoral researcher in the Department of Multimedia and Internet Applications, Faculty of Mathematics and Computer Science, University of Hagen, Germany. His research focuses on information retrieval, interactive information retrieval, information systems, databases, digital libraries, preservation and grid and cloud architectures. He developed the software Daffodil founded on a nation research project and worked in national and European research projects such as The European Film Gateway, SHAMAN (Sustaining Heritage Access through Multivalent ArchiviNg) and Smart Vortex (Scalable Semantic Product Data Stream Management for Collaboration and Decision Making in Engineering). He is currently responsible for several infrastructure projects within GESIS, such as daJra, SowiDataNet or Missy, all concerned with providing information and data for social scientists. In addition, he lead the measure PID Services in the national research infrastructure project NFDI. In his team, they are developing an open source DDI suite to support getting DDI into operation.

## Peter Mutschke



Peter Mutschke is deputy head of the department “Knowledge Technologies for the Social Sciences (KTS)” and leader of the team “FAIR Data and Human Information Interaction” of KTS. His research interests include Information Retrieval, Network Analysis and Open Science. He worked in a number of national and international research projects, such as the DFG projects DAFFODIL and IRM and the EU projects WeGov, SENSE4US, OpenMinTeD and MOVING. Peter served as a member of the management committee of the Leibniz research alliance “Science 2.0/Open Science” from 2013-2021. He founded and coordinates the GO FAIR Implementation Network “Cross-Domain Interoperability of Heterogeneous Research Data (Go Inter)”, and he is member of the steering committee of the FAIR Digital Objects Forum (fairdo.org) where he also co-chairs a working group on semantics. He is currently involved in consortia KonsortSWD, NFDI4DataScience and BERD@NFDI of the National Research Data Infrastructure (NFDI). ORCID: <https://orcid.org/0000-0003-3517-8071>.



# Agenda

- The PID Registration service
  - General goal and claim
  - The Research data granularity levels
  - Data citation using PIDs
- The PID Registration service: FAIR maturity level assessment
  - Criteria
  - Methodology
  - Results
  - Outcomes

- 
- General goal and claim

# Claim

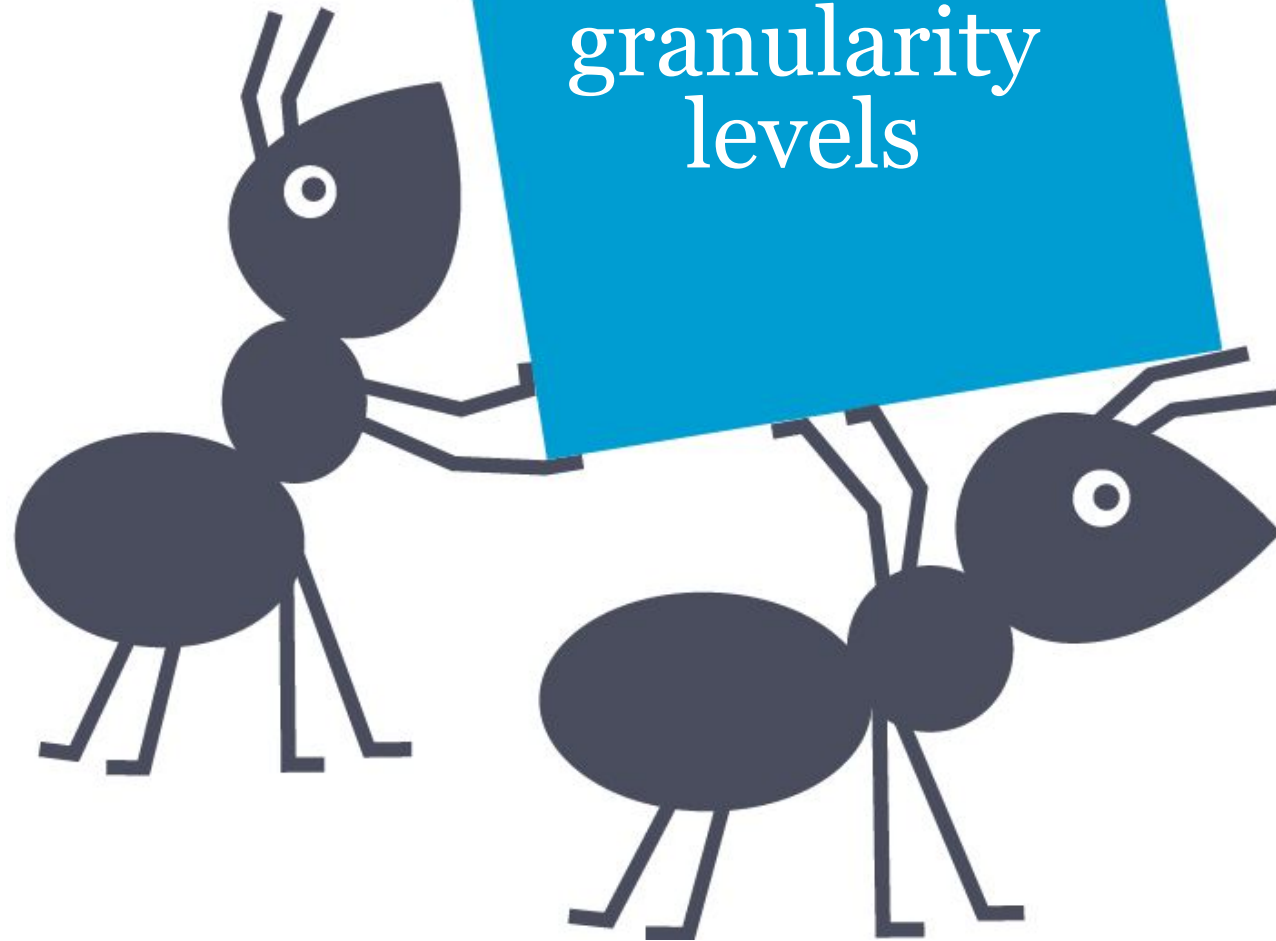
- Assigning a PID to a whole dataset is insufficient to unambiguously identify the information used and ensure an accurate data citation, thus, constraining the research results' trustworthiness.

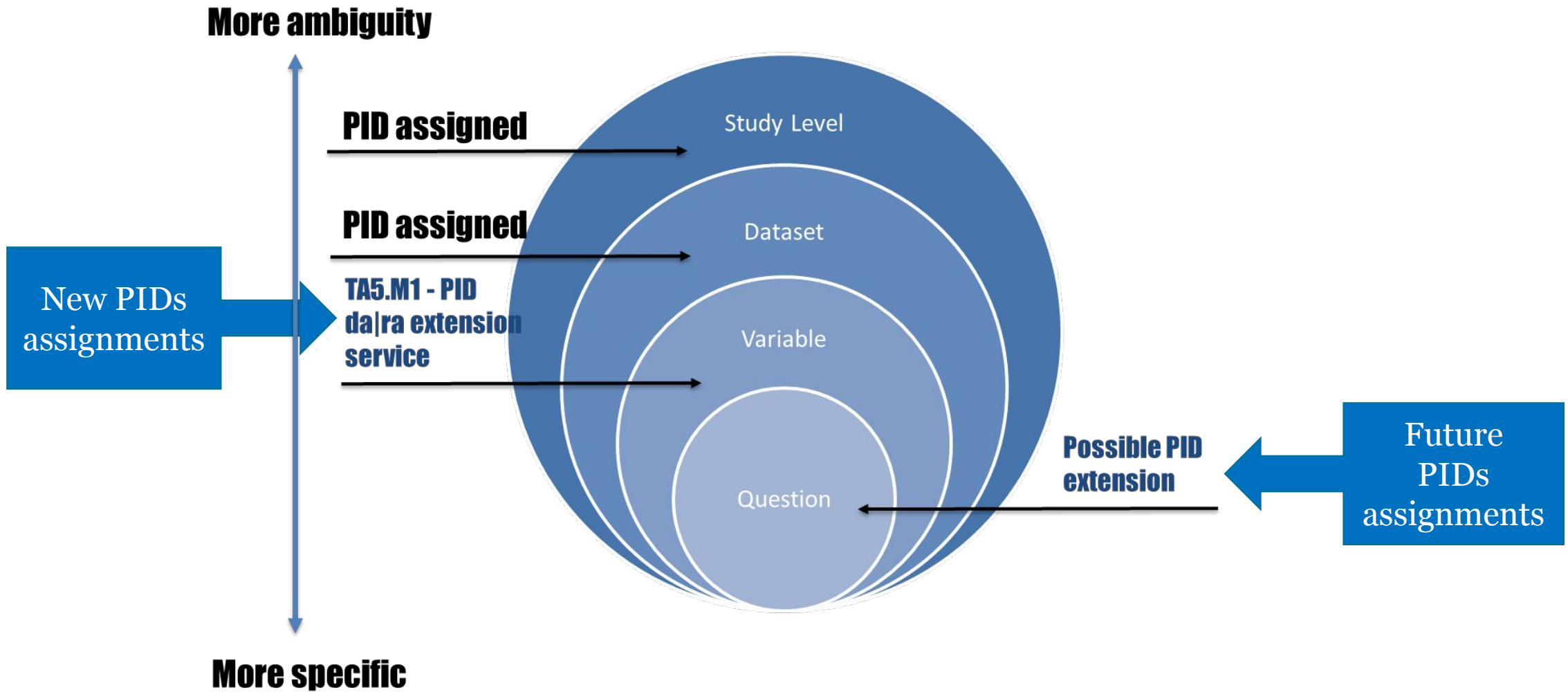


# General goals

- Identify survey variables, using one identifier – the PID – will simplify FAIR data management to boost subsequent citation, get direct (meta)-data access, and data reuse
- Since PIDs are machine-actionable, they are used as technical bridges to the FAIR principles that can increase traceability and foster reproducibility of research results in the Social and Economic Sciences

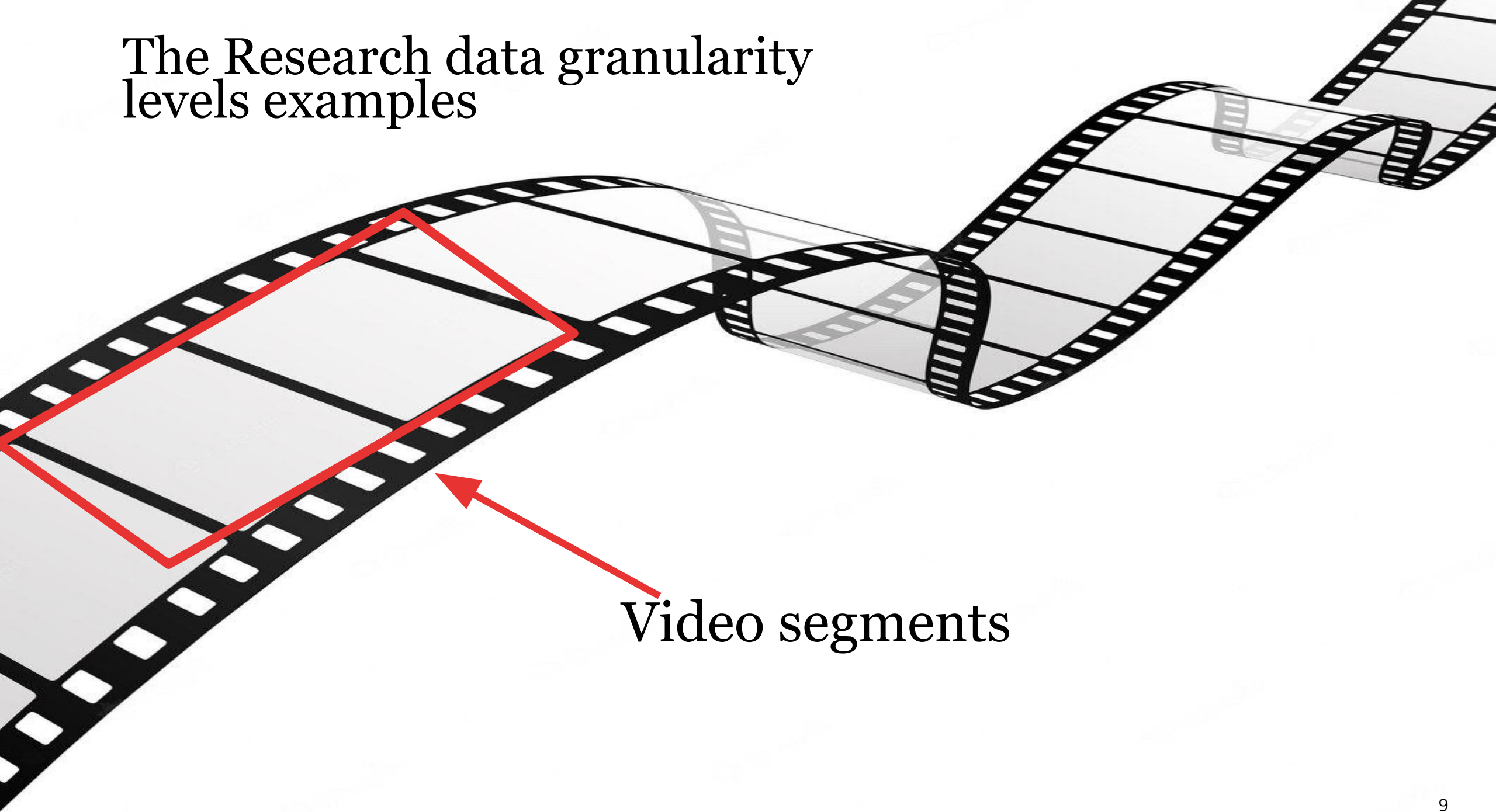
# The Research data granularity levels







# The Research data granularity levels examples

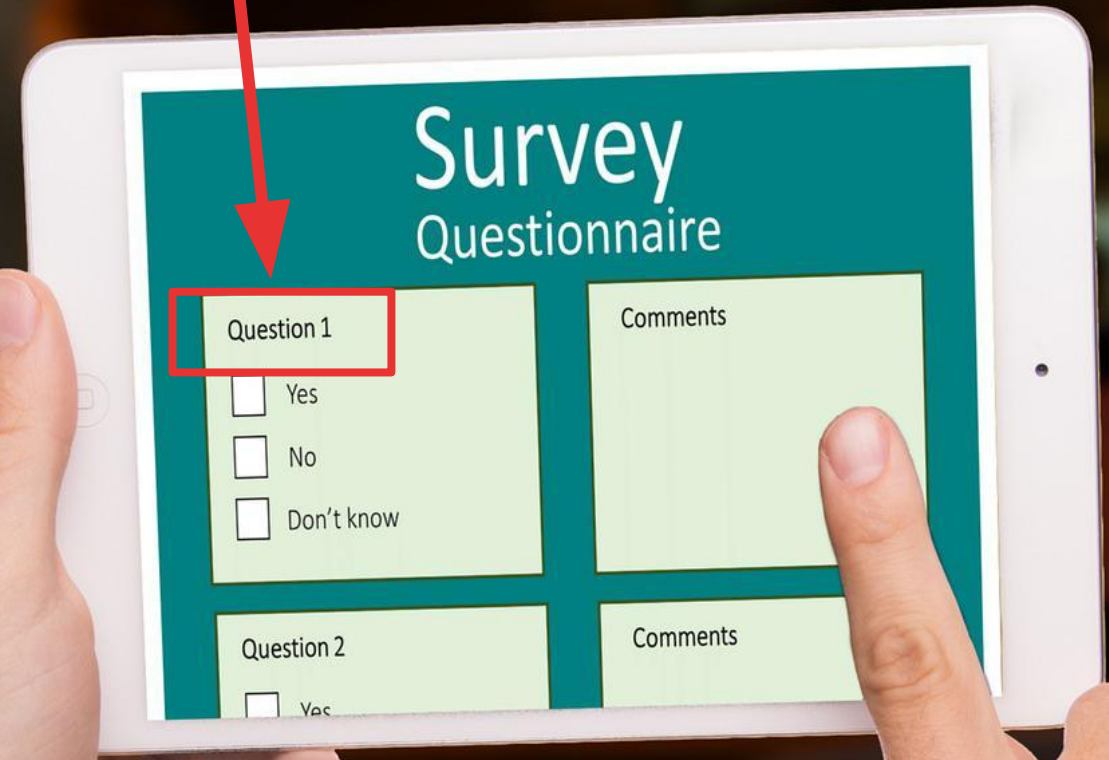


Video segments

# The Research data granularity levels examples

Use Cases

## Questions



The image shows a woman holding a tablet that displays a survey questionnaire. The survey is titled "Survey Questionnaire" and is divided into two columns. The left column contains two questions, "Question 1" and "Question 2". "Question 1" has three radio button options: "Yes", "No", and "Don't know". "Question 2" has a "Yes" option. The right column contains two "Comments" text input fields. A red arrow points from the word "Questions" above to the "Question 1" section, which is also highlighted with a red box. A finger is visible on the right side of the tablet, pointing towards the "Comments" field.

Question	Options	Comments
Question 1	<input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Don't know	
Question 2	<input type="checkbox"/> Yes	

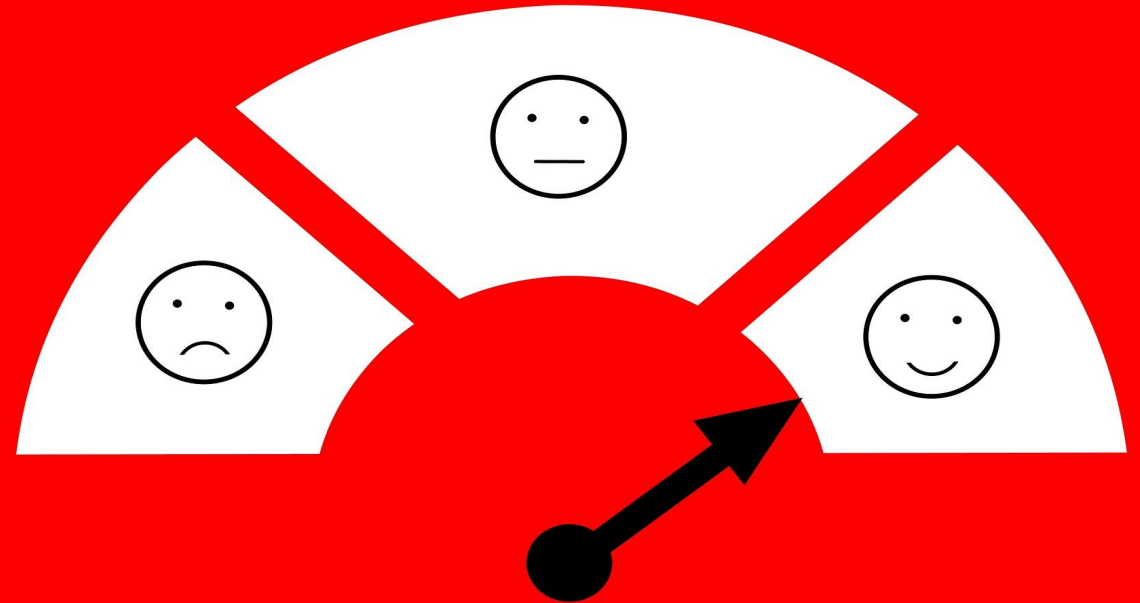
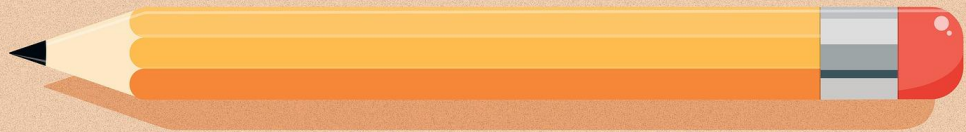


# Measures Scales



Yes

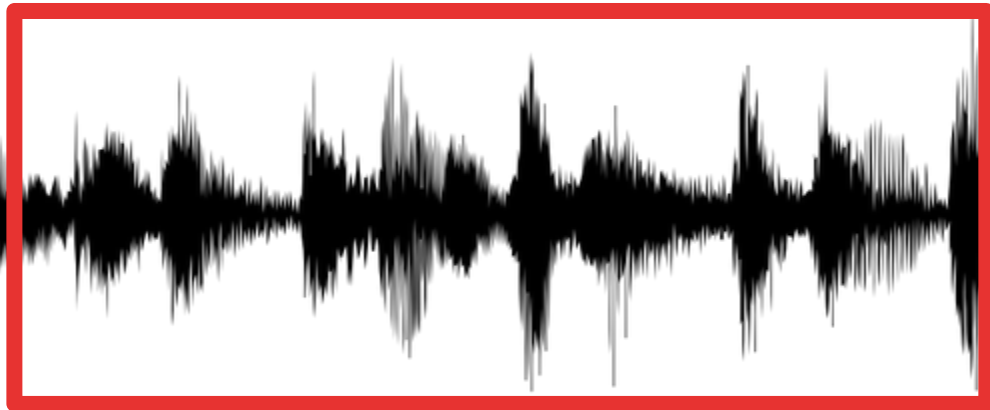
No



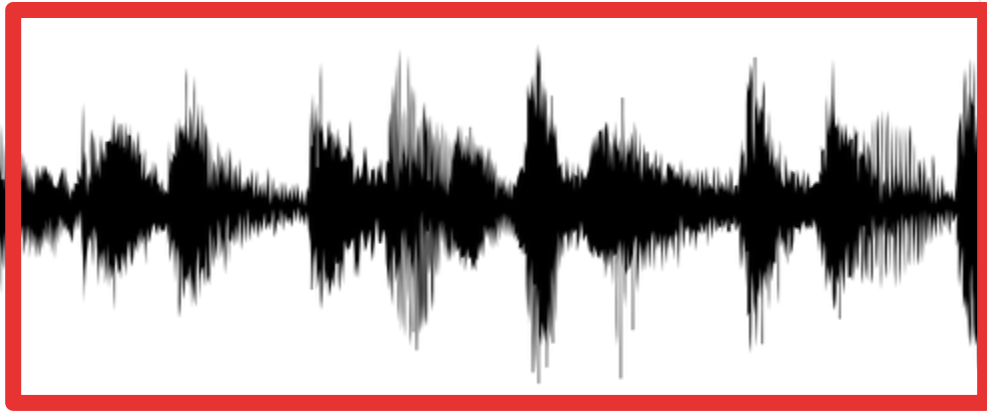


# The Research data granularity levels examples

**Audio segments**



# Audio files

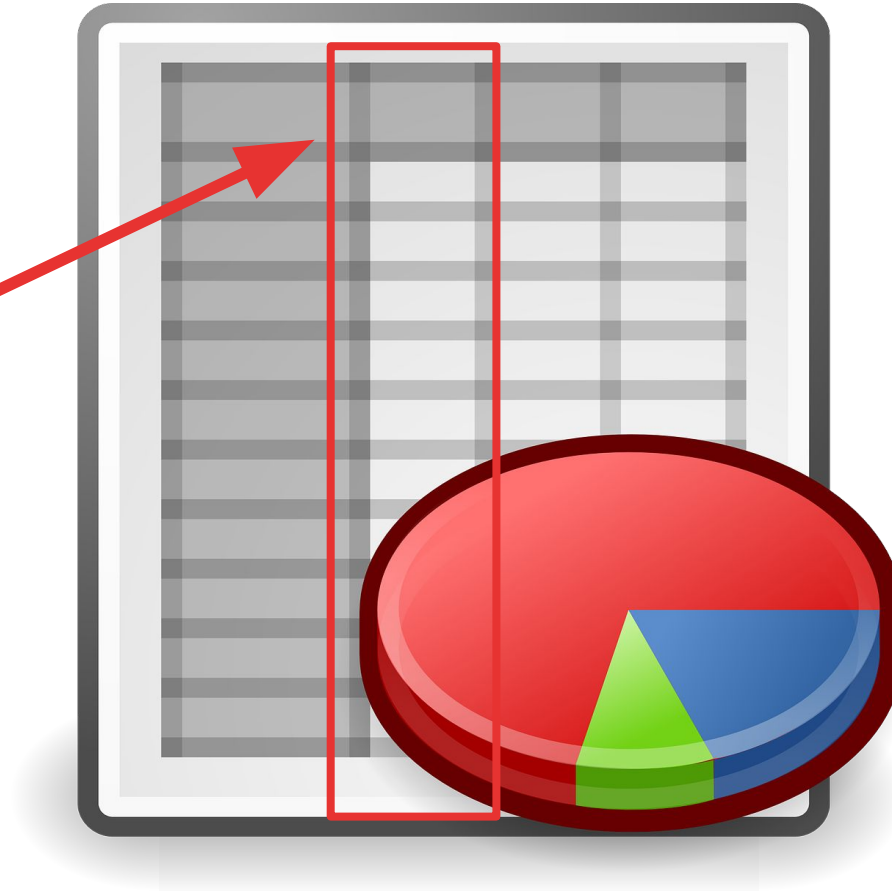


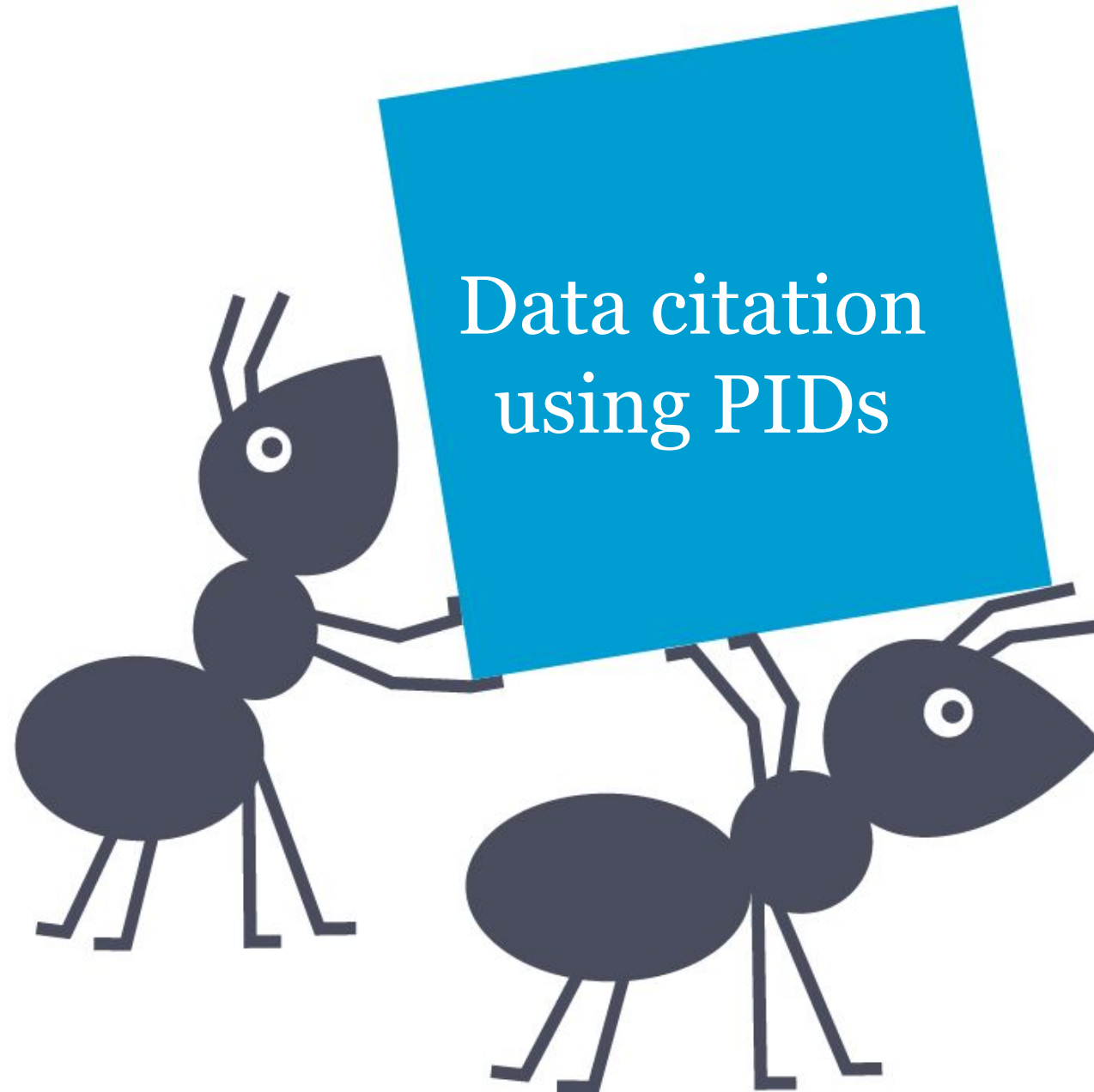
# Transcripts

great  
awesome you're amazing thumbs up  
brilliant excellent very good wonderful great  
bravo wow congratulations  
nice work super perfect good job  
congratulations Nice Work super awesome  
perfect wow good work brilliant  
great wonderful perfect superb nice work  
thumbs up wow Keep it up! excellent  
you're amazing

# The Research data granularity levels examples

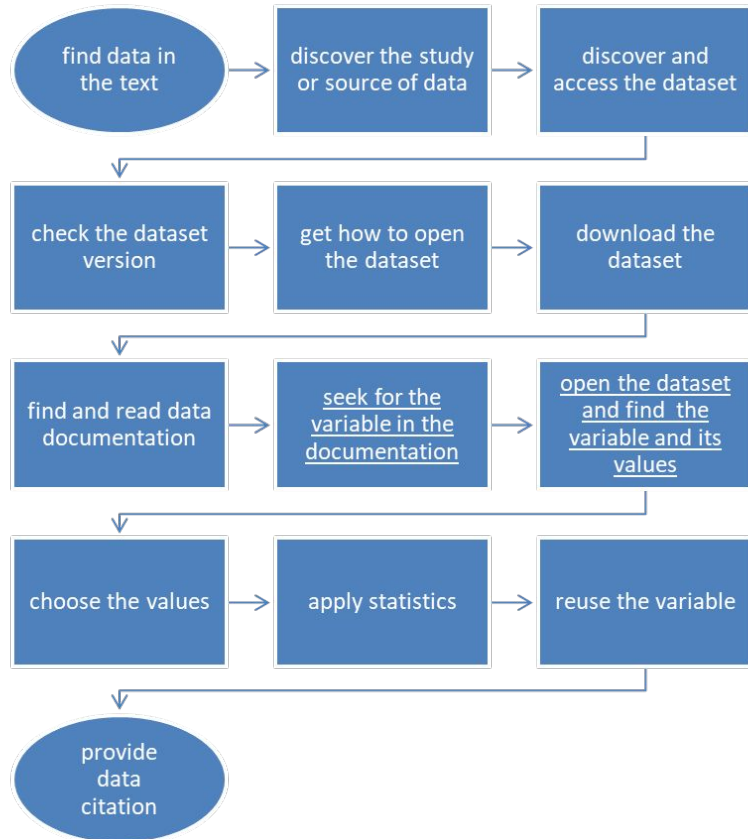
Variables





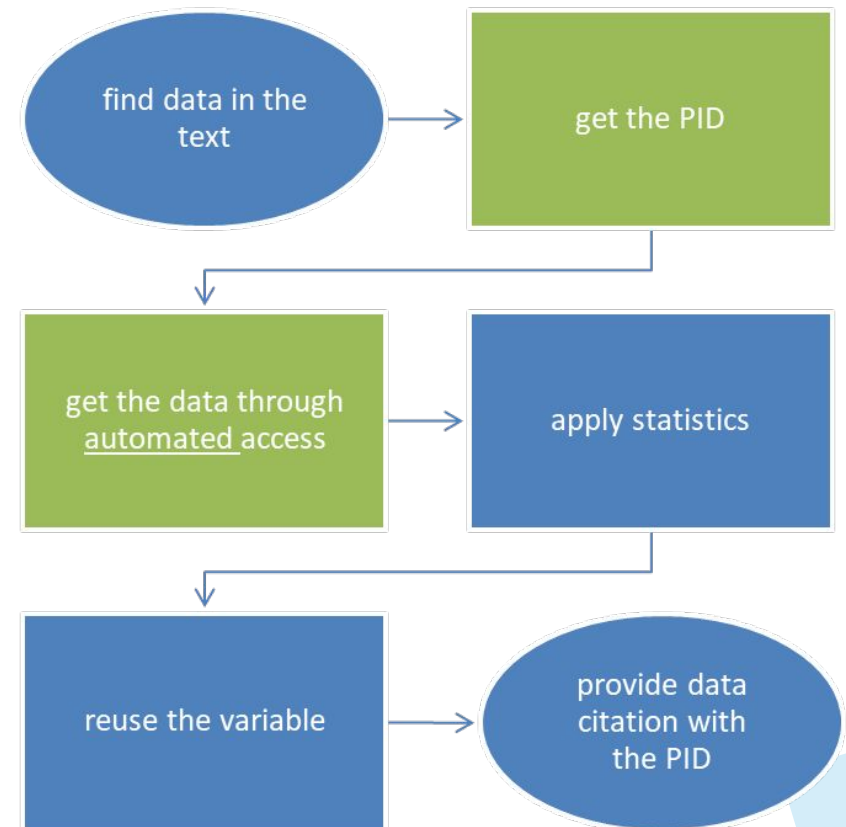
- Finding and getting the variable data using or not a PID:

### Process of accessing a variable without PID



X

### Process of accessing a variable with a PID





# Assigning PIDs for institutions such as:



# Research Data Centres (RDCs) potential users





The service  
FAIR maturity  
level assessment

# The service FAIR maturity level assessment: Criteria

- We assessed the service under the **FAIR Data Maturity Model** (RDA Working Group on FAIR Data Maturity Model, 2020, see doi: 10.15497/rda00050)

FAIR Data Maturity Model  
Specification and Guidelines



**DOI:** [10.15497/rda00050](https://doi.org/10.15497/rda00050)

**Co-chairs:** Edit Herczog, Keith Russell, Shelley Stall

**Published:** 25<sup>th</sup> June 2020

**Abstract:** Findability, Accessibility, Interoperability and Reusability – the FAIR principles – intend to define a minimal set of related but independent and separable guiding principles and practices that enable both machines and humans to find, access, interoperate and re-use data and metadata. The FAIR principles were defined in 2016 in an article by Mark Wilkinson et. al1. FORCE112 and GO FAIR3 provide further information on the principles. The principles have to be considered as

# The service FAIR maturity level assessment: Criteria

- The framework consists of **3 indicators classes**: Essential, Important, and Useful
- The sum of them is organized into **five levels**, according to the present indicator in each category
- When distributing the indicators per FAIR area, the principle of **Accessibility** and **interoperability** holds the majority of Essential and Important criteria for FAIRness

## 3 indicators classes in five levels

FAIR Data Maturity Model: evaluation framework		Level 1	Level 2	Level 3	Level 4	Level 5
Essential	20	20	20	20	20	20
Important	14		7	14	14	14
Useful	7				3	7
<b>Total</b>	<b>41</b>	<b>20</b>	<b>27</b>	<b>34</b>	<b>37</b>	<b>41</b>

## Indicators according to the FAIR Principles

Distribution of priorities per FAIR area					
Principle	Findable	Accessible	Interoperable	Reusable	Total
Essential	7	8	0	5	20
Important	0	3	7	4	14
Useful	0	1	5	1	7
<b>Grand Total</b>	<b>7</b>	<b>12</b>	<b>12</b>	<b>10</b>	<b>41</b>

# The service FAIR maturity level assessment: Methodology

- Applied the **stricter** evaluation method on each indicator, assessing them by passing or failing **binary answers**
- This approach was selected because the PID registration service is a widening solution to an established service through da|ra (da-ra.de)
- Link to assessment data:



Measure 5.1: PID Service for variables	Present	Not present		
FAIR Data Maturity Model: criteria framework	Pass	Fail	<b>Evidence</b>	<b>Comments</b>
RDA-F1-01M Metadata is identified by a persistent identifier	Pass		It has a variable PID assigned	Metadata and data is identified via DOI
RDA-F1-01D Data is identified by a persistent identifier	Pass		It has a variable PID assigned	Metadata and data is identified via DOI
RDA-F1-02M Metadata is identified by a globally unique identifier	Pass		Handle standard provides globally unique identifier	Metadata and data is identified via DOI
RDA-F1-02D Data is identified by a globally unique identifier	Pass		Handle standard provides globally unique identifier	Metadata and data is identified via DOI
RDA-F2-01M Rich metadata is provided to allow discovery	Pass		A metadata scheme is present to comply with the minimum metadata	Metadata is documented in DDI Lifecycle 3.2
RDA-F3-01M Metadata includes the identifier for the data	Pass		It includes the DOI of the study in which the variable to register appears	The DOI is part of the metadata
RDA-F4-01M Metadata is offered in such a way that it can be harvested and indexed	Pass		The metadata is in fact harvested and indexed at Gesis Search and/or other institutional repository as a service user	Metadata can be harvested via OAI-PMH



# The service FAIR maturity level assessment: Results

- The PID registration service **passed 33** indicators and failed 8
- The results for each level were in the range from **80% to 100%**

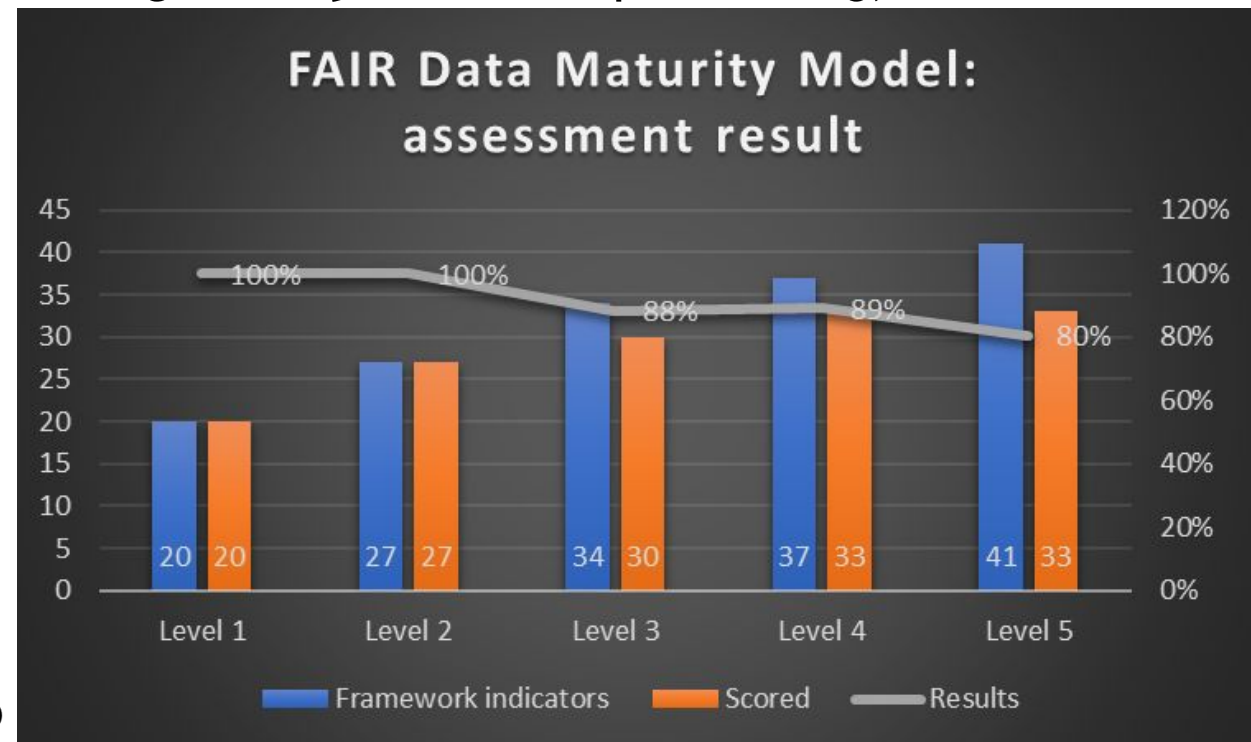


FAIR Data Maturity Model: assessment		
	Pass	Fail
Essential	20	0
Important	10	4
Useful	3	4
<b>Total</b>	<b>33</b>	<b>8</b>

Framework	Level 1	Level 2	Level 3	Level 4	Level 5
Essential	20 / 20	20 / 20	20 / 20	20 / 20	20 / 20
Important		7 / 7	10 / 14	10 / 14	10 / 14
Useful				3 / 3	3 / 7
Achieved indicators	20/20	27 / 27	30 / 34	33 / 37	33 / 41
Scored	20	27	30	33	33
Results	<b>100%</b>	<b>100%</b>	<b>88%</b>	<b>89%</b>	<b>80%</b>

# The service FAIR maturity level assessment: Results

- The results demonstrate **outstanding achievements at levels 1 and 2**, marking **100%** on the assessment measure
- The service achieves **88%** compliance at level 3 and **89%** at level 4. At level 5, the results show **80%** of passed indicators
- The service meets **all** indicators classified as **essential**
- The failed indicators concerned with **automatic features**, including references and/or qualified references to other data, and data is accessed automatically (i.e., by a computer program)





# The service FAIR maturity level assessment: outcomes

FAIR maturity level assessment of the PID service **confirm the initial assumption that:**

- PIDs on variable level improve/simplify FAIR data management because it:
  - Enables safe data **citation**;
  - Improves **findability**;
  - Fosters **reuse**;
  - Favors **reproducibility**;
- The **failed** indicators so far (**automatic features**) are **feasible to be implemented** in the future since it requires only the PID assigned to the variable and a code/do-file (i.e., by a computer program) designed to **get the data automatically**. It is a real potential advantage for the data provider and data users.

Janete Saldanha Bach, Claus-Peter Klas and Peter Mutschke. 2022. Application of ‘RDA FAIR Data Maturity Model’ to assess the PID registration service in terms of FAIRness. In *CESSDA, Webinar on User Experience with FAIR Evaluation Tools and Services*. Cologne, Germany, 11.10.2022, 27 slides.



# Thank you

PID Service report 2021

<https://doi.org/10.5281/zenodo.6397367>

The service is part of KonsortSWD project deliverable, NFDI funding number 442494171

