

# Using Persistent Identifiers at the GESIS Data Archive

Wolfgang Zenk-Möltgen, GESIS - Leibniz Institute for the Social Sciences



This work is licensed under [Creative Commons Namensnennung 4.0 International Lizenz](https://creativecommons.org/licenses/by/4.0/).

CESSDA

Persistent Identification within CESSDA, CESSDA Seminar in Cologne, Germany, Feb 18-19, 2016

# Questions

- Specific Requirements
- Decision on PID System and Provider
- PID System and Workflow
- Challenges for Introduction
- Satisfaction with current system
- Future plans concerning PIDs
- Role of CESSDA in this

# Specific Requirements at the GESIS Data Archive

- Integration into existing systems, e.g. the Data Catalogue (DBK)
- Citation requirement for data users
- Clear understanding of what to identify
- Guarantee for unique names
- Expectation of stability
- Usage of important metadata
- Support for internal workflows

# Decision on PID System and Provider

- GESIS joined the DataCite consortium
  - Using DOIs (Handle) as PI system with existing infrastructure
- Supporting citation of research data
  - By establishing da|ra to enable other agencies to assign DOIs
  - To register DOI names for the Data Archive and other GESIS products
- Introduction of a versioning policy at the Data Archive: We identify each published version of a dataset

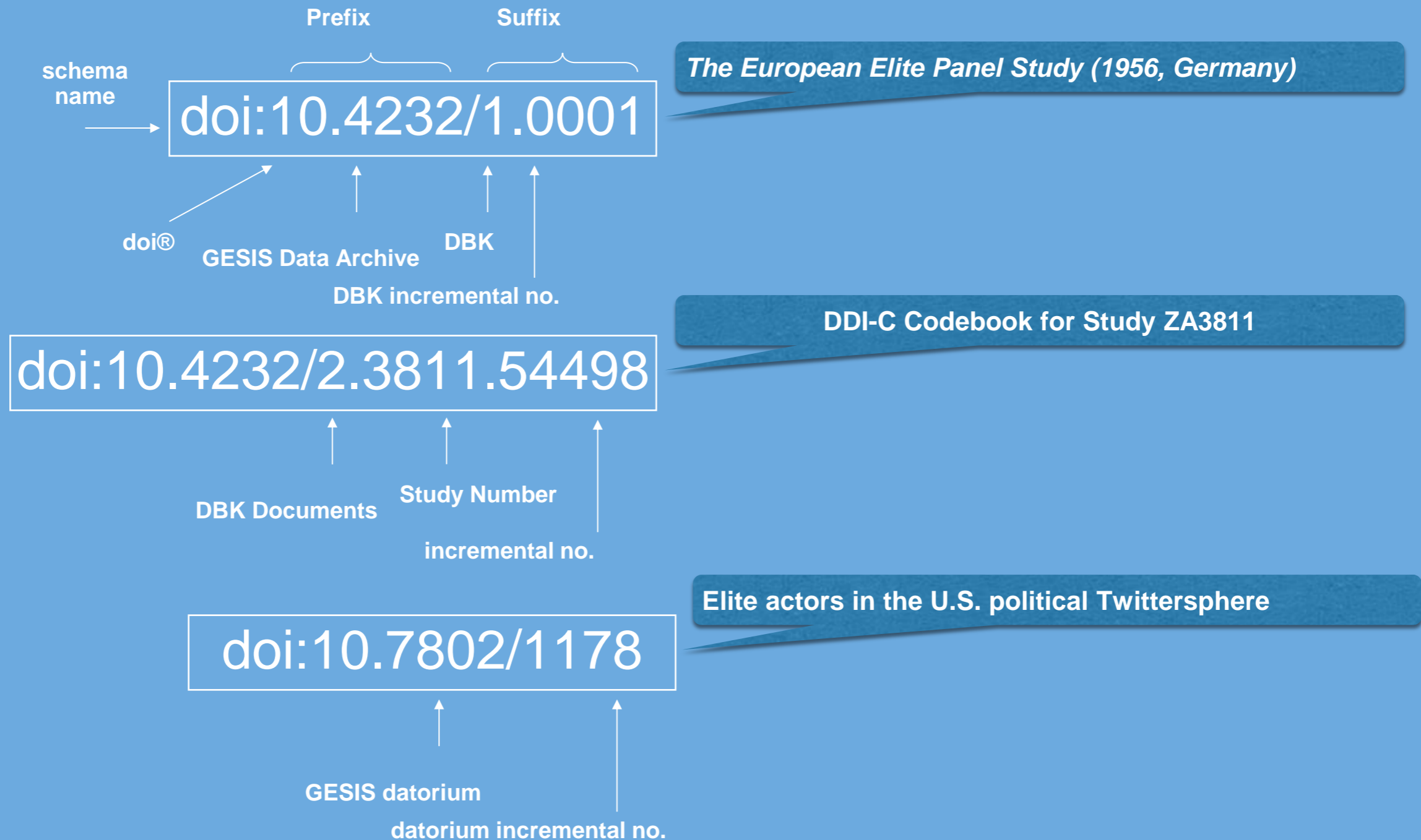
# PID System and Workflow

- The GESIS Data Archive uses da|ra
- Integration into the Data Catalogue DBK is done via the da|ra webservice
  - da|ra is the service provider for DOI allocation:
    - A DOI proposal is transmitted from DBKEdit for each published version of a dataset
    - Together with the URL and some metadata
    - DOI allocation takes place here
    - Changes of URL and/or metadata have to be transmitted from DBKEdit again
    - DBKEdit user may send the registration immediately or wait for a script that registers once a day

# Workflow Example


- DBKEdit is used as management system at the Archive:
  - Introduction of mandatory version numbers in 2010
  - Format: N.N.N (in accordance to DDI standard) starting with 1.0.0
  - Rules for increasing major, minor, revision number
  - Documenting also the version date, author, title and description
  - Errata of the current version are documented (optional)
  - Corrected errata are kept for documenting a new version
  - Version History and Errata are published
  - Current version of a dataset is registered with the DOI

# DOI Name Structure



# Data Catalogue DBK Example

gesis 

 | [Contact](#) | [Email to Data Service](#)

## GESIS Data Catalogue 2.1

[Home](#) | [Search](#) | [Browse](#) | [Overview](#) | [News](#) | [About](#)


[Login](#) | [Shopping Cart](#)

[Groups](#)  
[Products](#)  
[Studies](#)  
[Geographic Codes](#)



### ZA5234: Eurobarometer 73.4 (May 2010)

[Bibliographic Citation](#) | [Content](#) | [Methodology](#) | [Data & Documents](#) | [Errata & Versions](#) | [Groups](#)

[Citation](#) 

European Commission (2012): Eurobarometer 73.4 (May 2010). TNS OPINION & SOCIAL, Brussels [Producer]. GESIS Data Archive, Cologne. ZA5234 Data file Version 2.0.1, doi:10.4232/1.11479

Citation

Study No.

ZA5234

Title


Eurobarometer 73.4 (May 2010)

Other Titles

• Financial and Economic Crisis, the Future of the European Union, Globalization, and European Citizenship (Subtitle)

Current Version

2.0.1, 2012-10-15, doi:10.4232/1.11479

[Alternate IDs](#) 


• doi:10.3886/ICPSR34384.v1 (DOI), 2.0.1

Date of Collection

05.05.2010 - 28.05.2010

Principal Investigator/  
Authoring Entity,  
Institution

• Papacostas, Antonis - European Commission, Brussels DG Communication, Public Opinion Analysis Sector

[Categories](#) 

• International Institutions, Relations, Conditions

Member of the



da|ra Registration agency for social and economic data



DBKSearch 2.1 © GESIS



# Versioning Example

gesis

 | [Contact](#) | [Email to Data Service](#)

**GESIS Data Catalogue 2.1**

Home
Search
Browse
Overview
News
About

Login
 
Shopping Cart

Groups

 - -

Products

Studies

Geographic Codes

## ZA5234: Eurobarometer 73.4 (May 2010)

Bibliographic Citation
Content
Methodology
Data & Documents
Errata & Versions
Groups

Errata in current version
Version 2.0.1
Version 2.0.0
Version 1.0.0

## Versions

Version 2.0.1 (current version) 2012-10-15 Archive edition update [doi:10.4232/1.11479](https://doi.org/10.4232/1.11479)

Changes in this version

Date	Code	Description	Date	Changes
2012-10-12	V675 / C1	Derivation description in value labels for codes 1 to 3 is not correct.	2012-10-15	(1) -- LOW (coded 4 for V95 and 1,2,3, or 0 for V676; or coded 5 for V95 and 1, 2, 3, or 4 for V676) -- (2) - (coded 3 for V95 and 1,2,3, or 0 for V676; or coded 4 for V95 and 4 for V676) -- (3) + (coded 1 for V95 and 1,2 or 0 for V676; or coded 2 for V95 and 1,2,3, or 0 for V676; or coded 3 for V95 and 4 for V676)

## Changes

Member of the

Registration agency for social and economic data

FIND, ACCESS, AND REUSE DATA

DBKSearch 2.1 © GESIS

# Challenges for Introduction

- Introduction of the versioning policy:
  - We want to identify datasets
    - Not study series, studies, data collection events
    - Not variables or questions
    - Independent from specific format (SPSS, STATA, SAS)
- We keep track of dataset versions, but not for metadata versions
  - Documentation inside the dataset
  - DBK only holds metadata for current version, but da|ra for each version (=DOI)
- Update of URLs (due to server migration) was not possible via the da|ra webservice (not it is)
  - All updates required all metadata, which we did not have for „older“ versions

# Satisfaction with current system

- Very good, requirements are met:
  - Integration into DBK
  - Citation requirement (minor differences)
  - What to identify
  - Unique names
  - Stability
  - Important metadata
  - Support for internal workflows

# Future plans concerning PIDs

- Usage for other types than datasets
  - DDI metadata
- Review different metadata expressions for the same object
  - DDI-Codebook
  - DDI-Lifecycle
  - DataCite
  - Dublin Core
- Review metadata „completeness“
- Implement single source of metadata for all workflows

# Role of CESSDA in this

- Coordination of policies
- Providing best practises
- Exchange of metadata

Thank you!

Wolfgang Zenk-Möltgen, GESIS - Leibniz Institute for the Social Sciences

[wolfgang.zenk-moeltgen@gesis.org](mailto:wolfgang.zenk-moeltgen@gesis.org)

CESSDA