

0000



LEGACY4REUSE'23



# TOWARDS RIGOROUS DATA UPCYCLING USING THE FAIR IMPLEMENTATION PROFILE FOR THE SSHOC-NL SOCIO-ECONOMIC HISTORY COMMUNITY

Shuai Wang, Angelica Maineri

0000

# Special thanks to

## AUTHORS



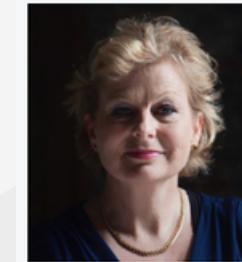
**Shuai Wang**

Scientific engineer  
VU Amsterdam



**Angelica Maineri**

Data manager  
ODISSEI &  
Erasmus University  
Rotterdam



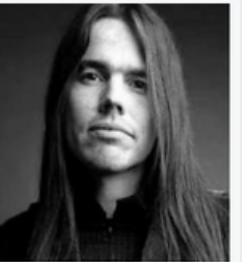
Pearl Dykstra  
PI, Erasmus University Rotterdam/ODISSEI



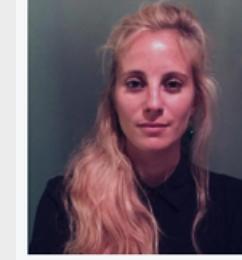
Jacco van Ossenbruggen  
Advisor, VU Amsterdam/ODISSEI



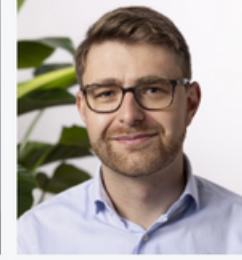
Tobias Kuhn  
Advisor, VU Amsterdam



Ronald Siebies  
Advisor, VU Amsterdam



Elena Beretta  
Advisor, VU Amsterdam



Lucas van der Meer  
Advisor, Erasmus University Rotterdam/ODISSEI



Ricarda Braukmann  
Advisor, DANS/KNAW



Tom Emery  
Advisor, Erasmus University Rotterdam/ODISSEI



**Rick Mourits**  
IISG



**Tycho Hofstra**  
**University  
Library, VU  
Amsterdam**

<https://fairexpertisehub.org/>

# TABLE OF CONTENTS

- Introduction to FIP
- SSHOC-NL-SEH FIP
- RQ & Methodology
- Legacy Datasets
- Upcycling using FIP
- Discussion
- Conclusion and future work



# **FAIR IMPLEMENTATION PROFILE**

## **Definition**

A list of decisions made by a community to FAIRify their data

The FIP is a collection of FAIR implementation choices made by a community of practice for each of the FAIR Principles.

### **FIC**

A group of people who are using/publishing data on the same topics with some intention to conduct FAIR implementaiton

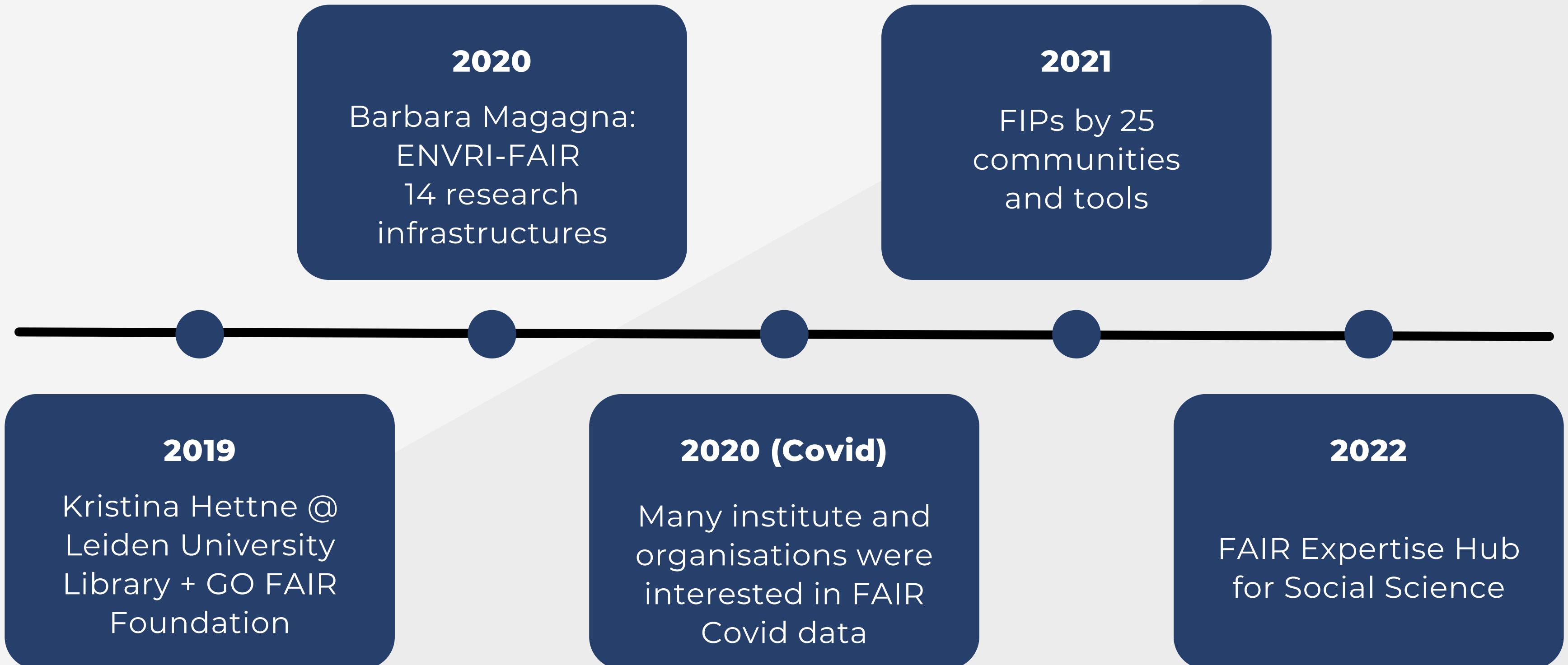
### **FER**

A resource that enables the implementation of FAIR behaviors

### **Their relations**

A FIC used a FER  
A FIC is using a FER  
A FIC will use a FER

# **A MINI HISTORY OF FIP**



**VODAN  
AFRICA**

**HEALTH-RI**

**ZONMW  
COVID  
PROGRAM**

**DUTCH  
CANCER  
FOUNDATION**

**DANISH  
E-INFRASTRUCTURE  
CORPORATION**

**ENVRI-FAIR**

**WORLD  
FAIR**

**EOSC  
NORDIC**

**NASEM  
ROUNDTABLE ON  
OPEN SCIENCE /  
ORFG**

**NIAID / NIH  
OFFICE OF  
DATA  
SCIENCE  
STRATEGY**



○ ○ ○ ○

# **FIPS FOR SOCIAL SCIENCE**

- Two FIPs in WorldFAIR
  - One FIP from past attempts
  - Data in social science can be very different (e.g. more qualitative data)
  - Diversity in data (bio, geospatial, images)
  - Different data management behaviors
  - Less FAIR

# COMMUNITIES

## **ESS**

European Social  
Survey (ESS)

## **AUSSI-ESS**

Australian Social  
Survey International –  
European Social Survey

## **GESIS SSSR**

GESIS Social  
Science Survey  
Research

## **SEH**

SSHOC-NL Socio-  
Economic History

## **MCAL**

Media Content  
Analysis Lab  
scholars

## **LGBTQVoc**

LGBTQ+ Linked  
Open Vocabulary

**Nov 2022 - Apr 2023**

**SEH**

SSHOC-NL Socio-  
Economic History



Radboud Universiteit Nijmegen



**Richard Zijdeman**  
**IISG**



**Rick Mourits**  
**IISG**



**Auke Rijpma**  
**Utrecht Universiteit**



**Xander Wilcke**  
**Vrije Universiteit**  
**Amsterdam**



**Sytze Van Herck**  
**Utrecht Universiteit**



**Universiteit Utrecht**

0 0 0

Meetings

Processing

Review

Publish

# CREATING FIPS

- Meetings/emails with members of the communities.
- All FIPs were created on the FIP Wizard
- Download in PDF and \*.trig

LGBTQ+ Linked Open Vocabulary FIP Share

Questionnaire Metrics Preview Documents Settings

View

Current Phase Defining FAIR Imple...

Chapters

I. About ✓

II. Declare your FAI... ✓

Select your FAIR Imple...

Who is the Community ...

Specify the start date f...

Specify the end date fo...

III. Declarations for ... ✓

IV. Declarations for ... ✓

V. Declarations for I... ✓

VI. Declarations for ... ✓

VII. Register a new re... ✓

In this section of the FIP Wizard, you, the Community Data Steward, are requested to answer questions that will declare, in a machine-readable way, the FAIR Implementation Community.

✓ II.1 Select your FAIR Implementation Community

In the drop-down list, select your FAIR Implementation Community. If your community is not yet listed, you will need to first create a nanopublication referencing your FAIR Implementation Community. Once you create this nanopublication, you will then see it in the drop-down list. A Wizard for creating nanopublications referencing a FAIR Implementation Community can be accessed below (in Chapter VII. Register a new resource as a nanopublication).

Desirable: Defining FAIR Implementation Profile

**LGBTQ+ Linked Open Vocabulary Community**

The LGBTQ+ Linked Open Vocabulary Community consists of researcher, librarians, linguistics experts that are contributing to the development and use of vocabularies such as Homosaurus, QLIT, etc. The community is centered around the Homosaurus project.

[See more here](#)

 [http://purl.org/np/RAOGUK-HzDHdTUPVG9tJXEjnG\\_p1mkqOylWUi10xf11\\_8#LGBTQVoc](http://purl.org/np/RAOGUK-HzDHdTUPVG9tJXEjnG_p1mkqOylWUi10xf11_8#LGBTQVoc)

Clear answer

Answered 3 months ago by Shuai Wang.

✓ II.2 Who is the Community Data Steward?

As the Community Data Steward, you take the responsibility to represent your FAIR Implementation Community, and for reporting the decisions your community has made on how to implement the FAIR Principles. Provide your ORCID here (like this: 0000-0003-2 without <https://orcid.org/>).

---

# **THE SSHOC-NL SOCIO-ECONOMIC HISTORY FIP**

- The identifier captured by the FIP are Handle and DOI (to be used in the future) for datasets.
- Dataverse and DANS SSH Data Station (to be used in the future) are the chosen search engines.
- XML, HTML, RDF are the knowledge representation languages captured in this FIP for its datasets.
- That for metadata includes RDF (OWL, RDFS), and SKOS.
- Finally, CC-BY-SA is being used by the community with CC-BY-NC to be used in the future.

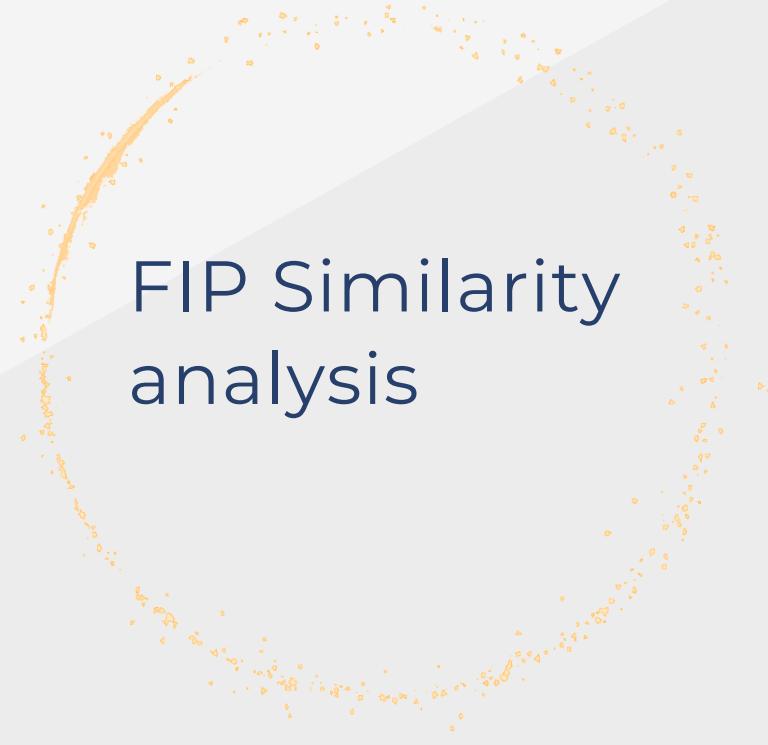


---

# ANALYSIS USING FIPS



Use of FAIR  
Enabling  
Resources



FIP Similarity  
analysis



FAIR  
Convergence  
Analysis

## **FIP2DMP**

Aligning DMPs with  
community standards  
using FIPs

# ***FAIR IMPLEMENTATION PROFILE***

**Research question**

## **FIP4Upcycling**

Can FIPs provide  
suggestions for the  
upcycling of datasets?

---

# **METHODOLOGY**

1. Manual examination of the upcycling of some datasets
2. Create a table with all the entries under assessment
  - **Findability:** Persistent and resolvable identifier, search engine and repository.
  - **Interoperability:** metadata format, metadata knowledge representation language, the format for datasets, knowledge representation for datasets.
  - **Reusability:** the licence of datasets.
3. Compare against the FIP

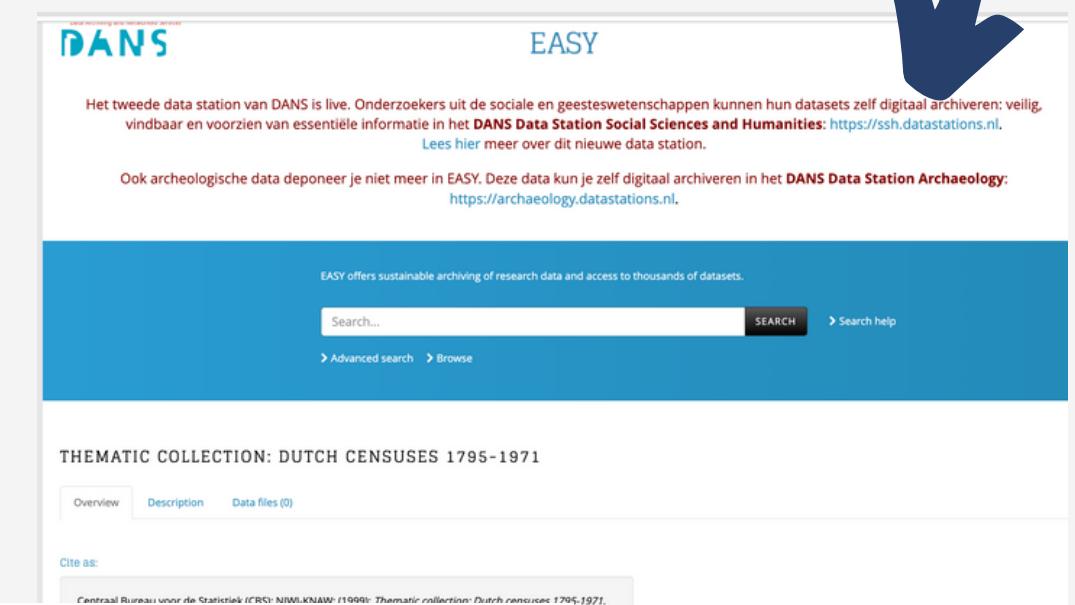
# LEGACY DATA AND DARK DATA

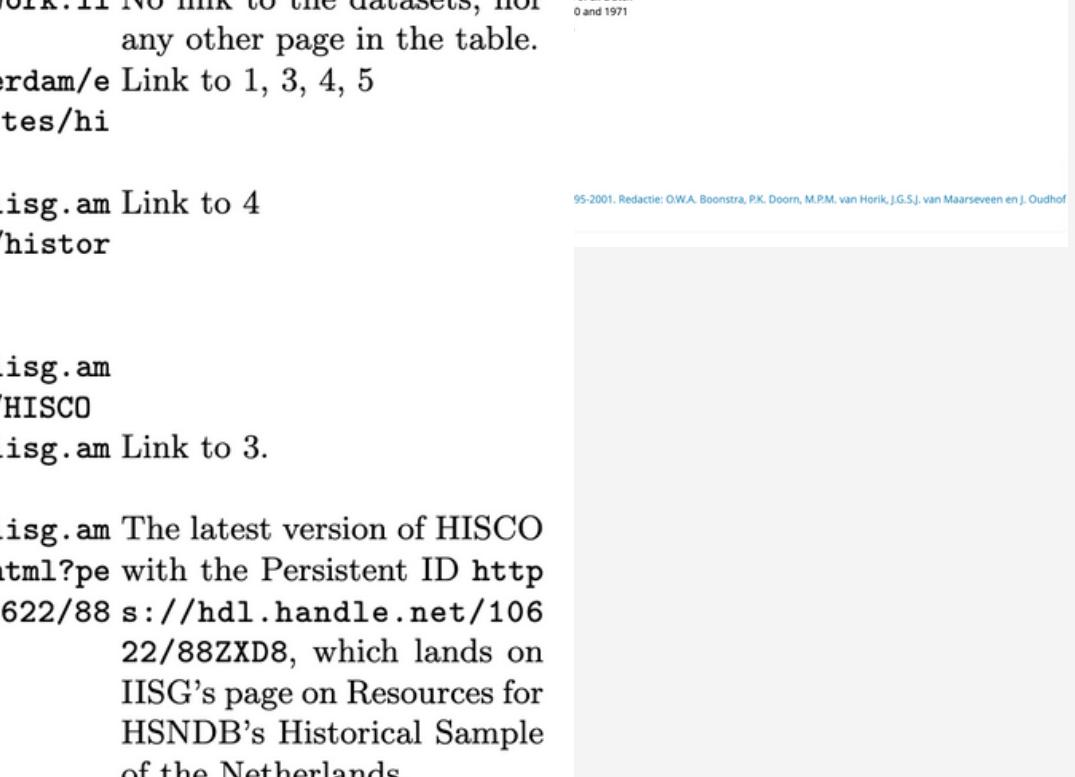
Dataset	Year created	Year last maintained	Comment
Volkstellingen	1997	2007	Some census tables published between 1795 and 1971. ( <a href="http://volkstellingen.nl">volkstellingen.nl</a> )
Linked International Classification for Religions (LICR)	2017	2017	A classification system that provides mappings to various other well-known systems such as IPUMS, NAPP, HL7. ( <a href="http://www.licr.io">www.licr.io</a> )
History Of Work Information System (HISCO)	2002	2020	A multilingual class scheme that provides historical occupational codes and classification of occupational activities
Gemeentegeschiedenis (Municipal History)	2007	2020	The boundaries of Dutch municipalities since 1812.
Historical Database of Dutch Municipalities (HDNG)	1970	2001	Historical Dutch Municipalities

- The official website ([www.licr.io](http://www.licr.io)) is no longer accessible.
- Volkstellingen was published on EASY. But EASY is not maintained anymore
- Gemeentegeschiedenis has 1542 datasets published in the EASY repository but are being migrated to the DANS SSH Data Station
- The datasets of Gemeentegeschiedenis were converted to GeoJSON and maintained by IISG.
- HDNG: the dataset can pass through multiple people
- HISCO was published at different locations but no link redirecting to the latest place

Version	Version History on Dataverse	Contributors	Date Published
4	5.1	Ineke Kellij	Jul 2023
	5.0	Rick Mourits, Richard Zijdeman	Jul 2021
	4.0	Rick Mourits	Jul 2021
3	3.0	Richard Zijdeman	Feb 2021
	2.1	Richard Zijdeman, Ineke Kellij	Dec 2020
	2.0	Richard Zijdeman	Jan 2020
2	1.1	Richard Zijdeman	Feb 2019
	1.0	Richard Zijdeman	Mar 2016

**Table 1.** The version history of the HDNG dataset on the Dataverse of IISG





Index	Page	URL	Comment
1	HISCO page	<a href="https://iisg.amsterdam/e/data/data-websites/history-of-work.ii">https://iisg.amsterdam/e/data/data-websites/history-of-work.ii</a>	No link to the datasets, nor any other page in the table.
2	IISG HISCO page	<a href="https://iisg.amsterdam/e/data/data-websites/history-of-work">https://iisg.amsterdam/e/data/data-websites/history-of-work</a>	Link to 1, 3, 4, 5
3	History Of Work Information System	<a href="https://datasets.iisg.am/stardam/dataverse/history-of-work">https://datasets.iisg.am/stardam/dataverse/history-of-work</a>	Link to 4
4	HISCO page on IISG Data- verse	<a href="https://datasets.iisg.am/stardam/dataverse/HISCO">https://datasets.iisg.am/stardam/dataverse/HISCO</a>	
5	The HISCO collaboration website	<a href="https://datasets.iisg.am/stardam/">https://datasets.iisg.am/stardam/</a>	Link to 3.
6	The HSNDB Occupations page on IISG Data- verse	<a href="https://datasets.iisg.am/stardam/dataset.xhtml?persistentId=hdl:10622/882288ZXD8">https://datasets.iisg.am/stardam/dataset.xhtml?persistentId=hdl:10622/882288ZXD8</a>	The latest version of HISCO with the Persistent ID http://hdl.handle.net/10622/882288ZXD8, which lands on IISG's page on Resources for HSNDB's Historical Sample of the Netherlands.

**Table 3.** Attempts for the upcycling of HISCO

SOMEONE TO TALK TO?



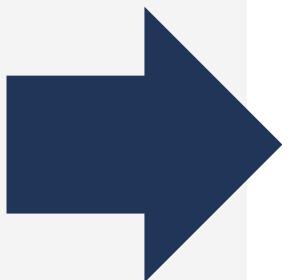
**FAIR**  
**IMPLEMENTATION**  
**PROFILE**

FOR OTHER COMMUNITIES TO  
ALIGN TO

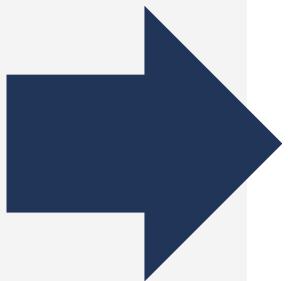
THE REFERENCE TO ALIGN  
YOUR OWN LEGACY DATA

**FAIR**  
IMPLEMENTATION  
PROFILE

**Repo**



**KR**



Aspects	LICR	Volkstellingen	HISCO	Gemeentege-schiedenis	HDNG	FIP
Year created	2017	1997	2002	2007	1970	-
Year last main-tained	2017	2007	2020	2020	2021	-
Dataset PID	Handle	DOI, URN	Handle	DOI, Handle	Handle	DOI(to use), Handle
Repository	IISG Data-verse	EASY, DANS SSH Data Sta-tion	IISG Data-verse	DANS Data Station, EASY, IISG Dataverse	Druid, IISG Dataverse	Dataverse, DANS SSH data station (future use)
Metadata format	Dublin Core, DDI, DataCite, DDI Code-book, JSON, OAI_ORE, OpenAIRE, Schema.or JSON-LD	XML, CSV	Dublin Core, DDI, DataCite, DDI Code-book, JSON, OAI_ORE, OpenAIRE, Schema.or JSON-LD	Dublin Core, DDI, DataCite, DDI Code-book, JSON, OAI_ORE, OpenAIRE, Schema.or JSON-LD	Dublin Core, DDI, DataCite, DDI Code-book, JSON, OAI_ORE, OpenAIRE, Schema.or JSON-LD	-
Knowledge representation language for metadata	XML, Dublin Core (DCMI Metadata Terms), EASY Metadata Terms, DataCite, HTML, JSON-LD	Dublin Core (DCMI Metadata Terms), EASY Metadata Terms	XML, Dublin Core (DCMI Metadata Terms), DataCite, HTML, JSON-LD	XML, Dublin Core (DCMI Metadata Terms), DataCite, HTML, JSON-LD	XML, Dublin Core (DCMI Metadata Terms), DataCite, HTML, JSON-LD	RDF (OWL, RDFS), SKOS
Data for	NO TAB	Various	CSV TAB	Shapefile (dbf)	PDF (Trig)	

**Licence**



	book, JSON, OAI_ORE, OpenAIRE, Schema.or JSON-LD		book, JSON, OAI_ORE, OpenAIRE, Schema.or JSON-LD	OAI_ORE, OpenAIRE, Schema.or JSON-LD	book, JSON, OAI_ORE, OpenAIRE, Schema.or JSON-LD	
Knowledge representation language for metadata	XML, Dublin Core (DCMI Metadata Terms), EASY Metadata Terms, DataCite, HTML, JSON-LD	Dublin Core (DCMI Metadata Terms), EASY Metadata Terms	XML, Dublin Core (DCMI Metadata Terms), DataCite, HTML, JSON-LD	XML, Dublin Core (DCMI Metadata Terms), DataCite, HTML, JSON-LD	XML, Dublin Core (DCMI Metadata Terms), DataCite, HTML, JSON-LD	RDF (OWL, RDFS), SKOS
Data format	NQ, TAB	Various	CSV, TAB, PDF, TXT	Shapefile (.dbf, .shp, .shx), Geo- JSON	RDF (Trig), CSV	-
Knowledge representation language for datasets	SKOS, RDF, RDFS	Not found by manual examination	None	None	RDF, SKOS	XML, HTML, RDF (RDFS, OWL)
Licence	CC0 1.0	Not found	CC0 1.0	CC BY-SA 4.0	CC0 1.0	CC-BY-SA, CC-BY-NC (new to the community)



## DISCUSSION

- Upcycling is a continuous progress
- FIPs are a live document
- The FIP could have missed some entries
- FIP can help handle differences due to the dynamics of HR
- Not good at capturing resources beyond FER (yet)
- No automatic way to handle the upcycling process
- A dataset could be useful for multiple communities

# Conclusion



Close examination of the upcycling of datasets  
in the SSHOC-NL Socio-Economic History  
community

Address problems in the upcycling of datasets

By mapping to questions of FIP, we explore the  
use of FIP as a community standard for the  
upcycling of datasets

Provide a list of suggestions for datasets

Discussed potentials of this approach and  
addressed problems and limits



## FUTURE WORK

- The FIP could be updated
- Take more datasets for analysis
- Other communities (excluding social survey communities?)
- Accessibility could be included in some future work
- Aligning with other communities when upcycling



○ ○ ○ ○

Do you think it worth further development?

# THANK YOU

## CONTACT



shuai.wang@vu.nl



www.fairexpertisehub.org



Platform Digitale Infrastructuur  
Social Sciences & Humanities



• • • •



○ ○ ○ ○

# References

- Erik Schultes et al. “Reusable FAIR Implementation Profiles as Accelerators of FAIR Convergence”.
- Mark D Wilkinson et al. “The FAIR Guiding Principles for scientific data management and stewardship”. In: *Scientific data* 3.1 (2016), pp. 1–9
- Barbara Magagna et al. “FIPs and Practice”. In: *Research Ideas and Outcomes* 8 (2022). doi: 10.3897/rio.8.e94451.
- Hana Pergl Sustkova et al. “FAIR Convergence Matrix: Optimizing the Reuse of Existing FAIR-Related Resources”. In: *Data Intelligence* 2.1-2 (Jan. 2020), pp. 158–170. issn: 2641-435X
- Kristina Hettne et al. “FIP2DMP: Linking data management plans with FAIR implementation profiles”.

# References (useful links)

FAIR Implementation Profile: 3-part video series:

<https://www.go-fair.org/2021/08/02/fair-implementation-profile-3-part-video-series/>

WorldFAIR: <https://docs.google.com/presentation/d/1vDgMCWW-RM806RnDul5ataGh4md5VX49u5NjcolVMU8/edit?usp=sharing>

FIP questionnaire and FER types: <https://bit.ly/WF-FIPs-Questions>

The FIP Wizard: <https://fip-wizard.ds-wizard.org/>

Image on slide 14: <https://www.google.com/url?sa=i&url=https%3A%2F%2Fwww.verywellmind.com%2Fwhat-to-do-when-you-need-someone-to-talk-to-5089236&psig=AOvVaw1wTGhIERzSkO7mAnJ7HlvT&ust=1700596573423000&source=images&cd=vfe&opi=89978449&ved=OCBIQjRxqFwoTCJDSjliu04IDFQAAAAAdAAAAABAD>

Image on the conclusion page :<https://www.google.com/url?sa=i&url=https%3A%2F%2Fstock.adobe.com%2Fsearch%3Fk%3Dconclusion&psig=AOvVaw0H1eSQLGq4HI E485Br-Xxu&ust=1700644008920000&source=images&cd=vfe&opi=89978449&ved=OCBIQjRxqFwoTCMCu--Le1IDFQAAAAAdAAAAABAD>