



LEGACY4REUSE'23



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

Shuai Wang, Angelica Maineri





# ***AUTHORS***



**Shuai Wang**

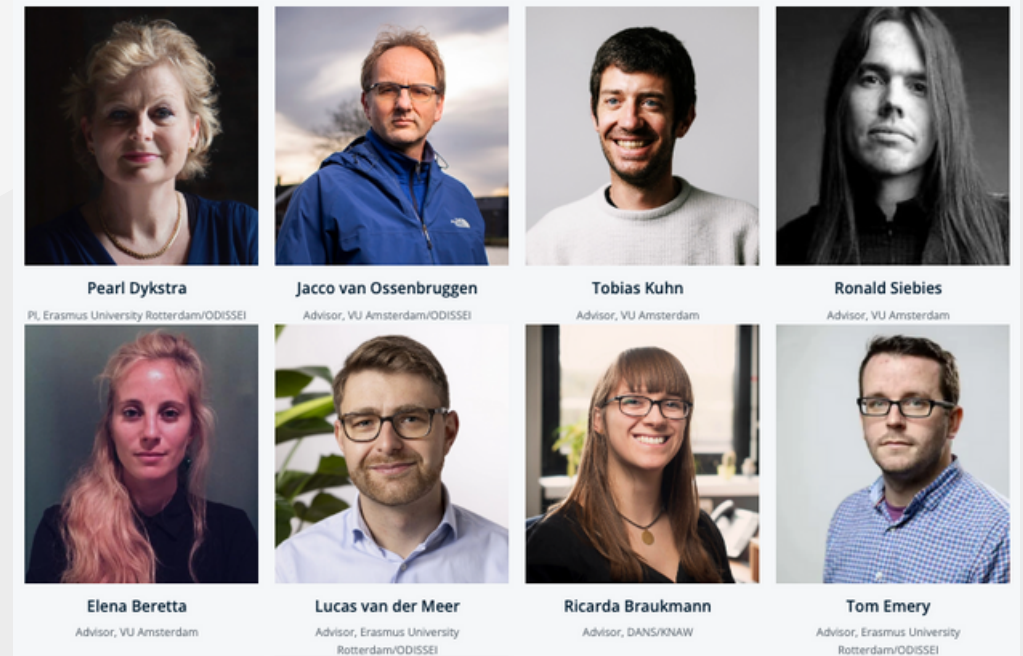
Scientific engiener  
VU Amsterdam



**Angelica Maineri**

Data manager  
ODISSEI &  
Erasmus University  
Rotterdam

# Special thanks to



**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 implementation

### **FER**

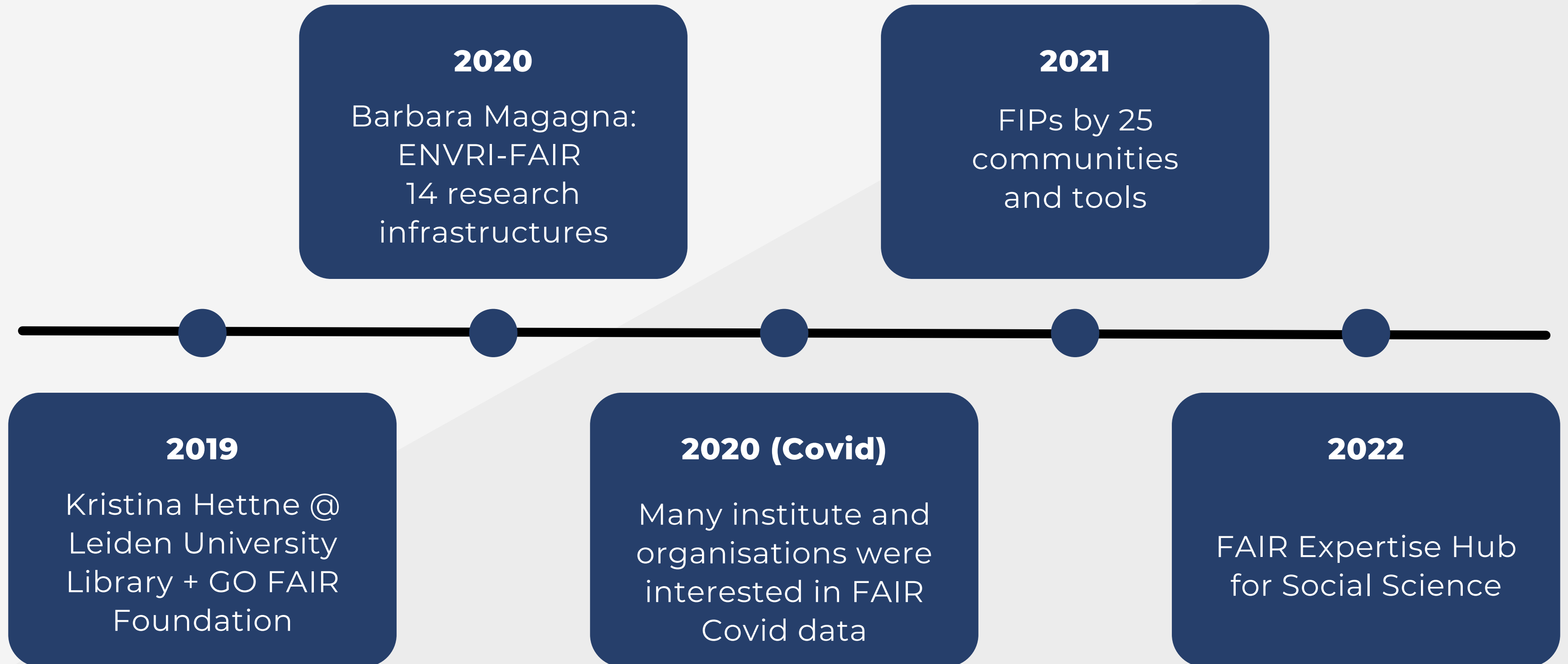
A resource that enables the implementation of FAIR behaviors

### **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**







# 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**

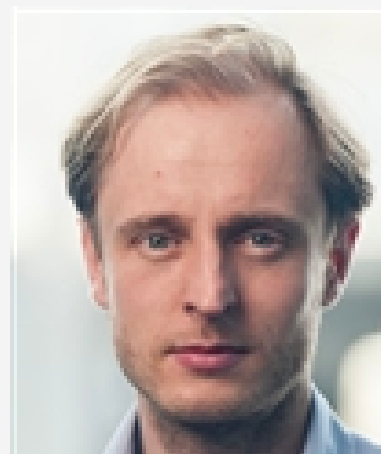
SSHOC-NL Socio-  
Economic History



**Richard Zijdeman**  
IISG



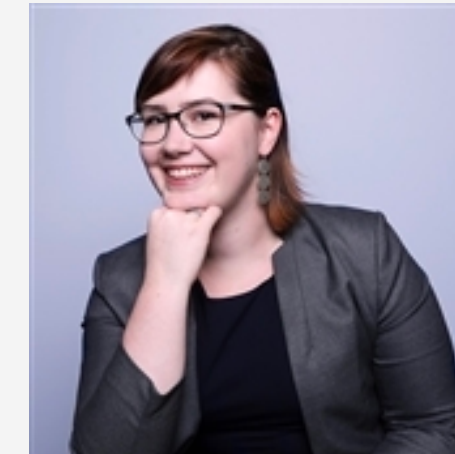
**Rick Mourits**  
IISG



**Auke Rijpma**  
Utrecht Universiteit



**Xander Wilcke**  
Vrije Universiteit  
Amsterdam



**Sytze Van Herck**  
Utrecht Universiteit







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

The screenshot shows the 'LGBTQ+ Linked Open Vocabulary FIP' page in the FIP Wizard. The interface includes a top navigation bar with 'Questionnaire', 'Metrics', 'Preview', 'Documents', and 'Settings'. Below this is a 'View' section with 'TODOs', 'Comments', and 'Version history'. The main content area is divided into two columns. The left column shows the 'Current Phase' as 'Defining FAIR Imple...' and a list of 'Chapters' from 'I. About' to 'VII. Register a new re...'. The right column contains the questionnaire text for 'II.1 Select your FAIR Implementation Community'. It includes instructions for selecting a community, a checkbox for 'Desirable: Defining FAIR Implementation Profile', and a text input field containing the URL 'http://purl.org/np/RAOGUK-HzDHdTUPVG9tJXEjnG\_p1mkqOyIWU10xf11\_8#LGBTQVoc'. Below the input field is a 'Clear answer' button and a note 'Answered 3 months ago by Shuai Wang.'. The bottom of the screenshot shows the start of 'II.2 Who is the Community Data Steward?' with instructions for providing an ORCID.

---

# ***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***

## **FIP4Upcycling**

Can FIPs provide  
suggestions for the  
upcycling of datasets?

**Research question**



---

# ***METHODOLOGY***

1. Manual examination of the upcycling of some datasets
2. Create a table with all the entries under assessment
  - **F**indability: Persistent and resolvable identifier, search engine and repository.
  - **I**nteroperability: metadata format, metadata knowledge representation language, the format for datasets, knowledge representation for datasets.
  - **R**eusability: 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. (volkstellingen.nl)
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. (www.licr.io)
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) 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 Zijdemans	Jul 2021
	4.0	Rick Mourits	Jul 2021
3	3.0	Richard Zijdemans	Feb 2021
	2.1	Richard Zijdemans, Ineke Kellij	Dec 2020
	2.0	Richard Zijdemans	Jan 2020
2	1.1	Richard Zijdemans	Feb 2019
	1.0	Richard Zijdemans	Mar 2016

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

Index	Page	URI	Comment
1	HISCO page	<a href="https://www.licr.io/sg.nl/">https://www.licr.io/sg.nl/</a>	No link to the datasets, nor any other page in the table.
2	IISG HISCO page	<a href="https://iisg.amsterdam/en/data/data-websites/history-of-work">https://iisg.amsterdam/en/data/data-websites/history-of-work</a>	Link to 1, 3, 4, 5
3	History Of Work Information System page on IISG Data-verse	<a href="https://datasets.iisg.amsterdam/dataverse/history-of-work">https://datasets.iisg.amsterdam/dataverse/history-of-work</a>	Link to 4
4	HISCO page on IISG Dataverse	<a href="https://datasets.iisg.amsterdam/dataverse/HISCO">https://datasets.iisg.amsterdam/dataverse/HISCO</a>	Link to 3.
5	The HISCO collaboration website	<a href="https://datasets.iisg.amsterdam/">https://datasets.iisg.amsterdam/</a>	Link to 3.
6	The HSNDB Occurrences page on IISG Dataverse	<a href="https://datasets.iisg.amsterdam/dataset.xhtml?persistentId=hdl:10622/88ZXD8">https://datasets.iisg.amsterdam/dataset.xhtml?persistentId=hdl:10622/88ZXD8</a>	The latest version of HISCO with the Persistent ID <a href="https://hdl.handle.net/10622/88ZXD8">https://hdl.handle.net/10622/88ZXD8</a> , 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

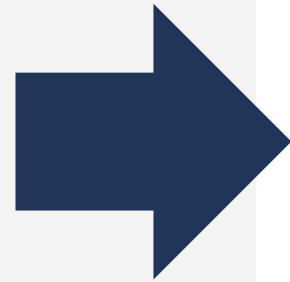
THE REFERENCE TO ALIGN  
YOUR OWN LEGACY DATA

**F** **A** **I** **R**

IMPLEMENTATION  
PROFILE

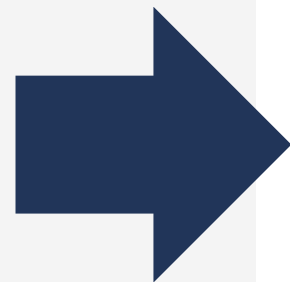


Repo



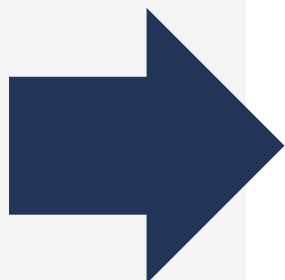
Aspects	LICR	Volkstellingen	HISCO	Gemeentegeschiedenis	HDNG	FIP
Year created	2017	1997	2002	2007	1970	-
Year last maintained	2017	2007	2020	2020	2021	-
Dataset PID	Handle	DOI, URN	Handle	DOI, Handle	Handle	DOI(to use), Handle
Repository	IISG Dataverse	EASY, DANS SSH Data Station	IISG Dataverse	DANS SSH Data Station, EASY, IISG Dataverse	Druid, IISG Dataverse	Dataverse, DANS SSH data station (future use)
Metadata format	Dublin Core, DDI, DataCite, DDI Codebook, JSON, OAI-ORE, OpenAIRE, Schema.org JSON-LD	XML, CSV	Dublin Core, DDI, DataCite, DDI Codebook, JSON, OAI-ORE, OpenAIRE, Schema.org JSON-LD	Dublin Core, DDI, DataCite, DDI Codebook, JSON, OAI-ORE, OpenAIRE, Schema.org JSON-LD	Dublin Core, DDI, DataCite, DDI Codebook, JSON, OAI-ORE, OpenAIRE, Schema.org JSON-LD	-
Knowledge representation language for metadata	XML, Dublin Core (DCMI 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	NO, TAB	Various	CSV, TAB	Shapefile (.shp)	RDF (TriX)	

KR

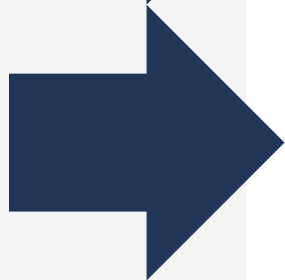




KR



Licence



	BOOK, JSON, OAI_ORE, OpenAIRE, Schema.or JSON-LD		BOOK, JSON, OAI_ORE, OpenAIRE, Schema.or JSON-LD	OpenAIRE, Schema.or JSON-LD	BOOK, JSON, OAI_ORE, OpenAIRE, Schema.or JSON-LD	
Knowledge representation language for meta-data	XML, Dublin Core (DCMI 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), GeoJSON	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

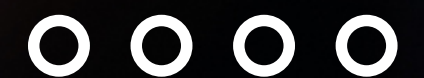
Discussed potentials of this approach and addressed problems and limits





# ***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](mailto:shuai.wang@vu.nl)

 [www.fairexpertisehub.org](http://www.fairexpertisehub.org)



# 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?](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=0CBIQjRxqFwoTCJDSjliu04IDFQAAAAAdAAAAABAD)

[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=0CBIQjRxqFwoTCJDSjliu04IDFQAAAAAdAAAAABAD](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=0CBIQjRxqFwoTCJDSjliu04IDFQAAAAAdAAAAABAD)

Image on the conclusion page :[https://www.google.com/url?](https://www.google.com/url?sa=i&url=https%3A%2F%2Fstock.adobe.com%2Fsearch%3Fk%3Dconclusion&psig=AOvVaw0H1eSQLGq4HIE485Br-Xxu&ust=1700644008920000&source=images&cd=vfe&opi=89978449&ved=0CBIQjRxqFwoTCMCu--Le1IIDFQAAAAAdAAAAABAD)

[sa=i&url=https%3A%2F%2Fstock.adobe.com%2Fsearch%3Fk%3Dconclusion&psig=AOvVaw0H1eSQLGq4HIE485Br-Xxu&ust=1700644008920000&source=images&cd=vfe&opi=89978449&ved=0CBIQjRxqFwoTCMCu--Le1IIDFQAAAAAdAAAAABAD](https://www.google.com/url?sa=i&url=https%3A%2F%2Fstock.adobe.com%2Fsearch%3Fk%3Dconclusion&psig=AOvVaw0H1eSQLGq4HIE485Br-Xxu&ust=1700644008920000&source=images&cd=vfe&opi=89978449&ved=0CBIQjRxqFwoTCMCu--Le1IIDFQAAAAAdAAAAABAD)