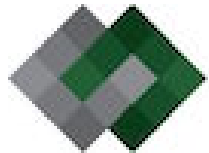


# WorldFair WP6 Social Surveys

Steve McEachern – Australian Data  
Archive

Hilde Orten – Sikt, Norway



ADA AUSTRALIAN  
DATA ARCHIVE



Sikt

# Who are we?

- Australian National University: <https://ada.edu.au> (WP6 lead)
  - The **Australian Data Archive (ADA)** provides a national service for the collection and preservation of digital research data. ADA disseminates this data for secondary analysis by academic researchers and other users.
  - The archive is based in the **ANU Centre for Social Research and Methods (CSRM)** at the Australian National University (ANU).
- Sikt - Norwegian Agency for Shared Services in Education and Research: <https://sikt.no/>
  - Sikt develops, acquires and delivers services for education and research. In collaboration with our users, we offer a common infrastructure for education and research. The aim is to free capacity for our customers, and to meet overarching goals of digitalisation, data sharing and open research.
- Both organisations:
  - are involved in collaborative international social survey projects, including the European Social Survey (ESS-ERIC) and the International Social Survey Program (ISSP)
  - Provide data archiving services
  - Conduct social survey research projects
  - Are members of the Data Documentation Alliance (DDI Alliance), and involved in the development of the DDI Cross Domain Integration standard (DDI-CDI)

# What will we do?

- Comparative study of the data management, harmonization and integration practices of one of the satellite countries – Australia, through the AUSSI-ESS – and the core ESS, an ERIC social science infrastructure.
- The project will examine both administrative procedures, data and metadata management, and technical environments.
- It will then leverage the DDI metadata standards to understand how such multi-national collections could be made increasingly interoperable and reusable through shared procedural and technical development, and
- Establish a set of guidelines and tools for the development of cross-national collections into the future

# In focus till now: FIPs

- Developed by GO FAIR and others to enable communities to describe their approach to FAIR in a common way
- FIPS are declarations of how each FAIR principle is implemented for a data source
- [Our FIPs at the current stage](#)

FAIR Principle name	Referring to Metadata/ Data	FIP question	FER Enabling Resource used in WP06 Social Surveys	Internal in organization	URL	Example
F1	MD	identifier service do you use for metadata records?	data describing the object and its location.	DDI URN		<a href="https://www.ietf.org/archive/id/draft-urn-66-00.pdf">https://www.ietf.org/archive/id/draft-urn-66-00.pdf</a>
F1	D	identifier service do you use for datasets?		DOI		<a href="https://www.doi.org/">https://www.doi.org/</a>
F1	D	identifier service do you use for datasets?		DDI URN		<a href="https://www.ietf.org/archive/id/draft-urn-66-00.pdf">https://www.ietf.org/archive/id/draft-urn-66-00.pdf</a>
F2	MD	What metadata schemas do you use for findability?		DDI-Lifecycle 3 - Study Unit		<a href="https://dataalliance.org/specification/DDI-Lifecycle-3.3.XM">https://dataalliance.org/specification/DDI-Lifecycle-3.3.XM</a>
F3	D	identifiers of your data to the metadata description?	data and the data they describe.	DDI-Lifecycle 3.3		<a href="https://dataalliance.org/specification/DDI-Lifecycle-3.3">https://dataalliance.org/specification/DDI-Lifecycle-3.3</a>
F4	MD	records?	data and provides search over that index.	GraphQL API	External	<a href="https://graphql.org/">https://graphql.org/</a>
F4	MD	records?	data and provides search over that index.	Published with Colectica web services.	Internal	<a href="https://docs.colectica.com/repositories/web-services/">https://docs.colectica.com/repositories/web-services/</a>
F4	D	Which service do you use to publish your datasets?	data and provides search over that index.	ESS Website landing page, API		<a href="https://www.european-socialsurvey.org/">https://www.european-socialsurvey.org/</a> , <a href="https://docs.nsd.no/">https://docs.nsd.no/</a>
F4	D	Which service do you use to publish your datasets?	data and provides search over that index.	EOSC Portal		<a href="https://marketplace.eosc-portal.eu/providers/en_enic">https://marketplace.eosc-portal.eu/providers/en_enic</a>
A1.1	MD	you use for metadata records?	pages are structured and exchanged.	HTTFS		<a href="https://en.wikipedia.org/wiki/HTTFS">https://en.wikipedia.org/wiki/HTTFS</a>
A1.1	D	you use for datasets?		HTTFS		<a href="https://en.wikipedia.org/wiki/HTTFS">https://en.wikipedia.org/wiki/HTTFS</a>
A1.2	MD	you use for metadata records?	al objects according to specified conditions	No auth. JSON in GraphQL	External	<a href="https://www.json.org/json-en.html">https://www.json.org/json-en.html</a> , <a href="https://graphql.org/">https://graphql.org/</a>
A1.2	D	you use for datasets?	al objects according to specified conditions	eduGAIN/OIDC, transport: GraphQL = data file format	External	<a href="https://edugain.org/">https://edugain.org/</a> , <a href="https://graphql.org/">https://graphql.org/</a>
A1.2	D	you use for datasets?	al objects according to specified conditions	Azure Active Directory, transport: Azure APIs = data file format	Internal	<a href="https://azure.microsoft.com/en-us/services/active-directory/">https://azure.microsoft.com/en-us/services/active-directory/</a>
A2	MD	What metadata preservation policy do you use?	ions under which metadata are to be part of a data management plan).	ERS Policy?		
I1	MD	records?		JSON in GraphQL		<a href="https://azure.microsoft.com/en-us/services/active-directory/">https://azure.microsoft.com/en-us/services/active-directory/</a>
I1	D	machine interoperability) do you use for datasets?		Parquet		
I2	MD	your metadata records?	ing web standards.	DDI-Lifecycle 3 structured codelists		
I2	D	your datasets?	ing web standards.	ISO3166-1 for country and ISO639-2 for language, NACE Rev 2 for Industry, ISCO08 for occupation, NUTS for External	External	
I3	D	your datasets?		DDI Controlled vocabularies, CESSDA vocabularies, ELIST	Internal	<a href="https://colectica.eu/processor/">https://colectica.eu/processor/</a>
I3	MD	records?		DDI-Lifecycle		
I3	D	What semantic model do you use for your datasets?		DDI-CDI		
R1.1	MD	records?		What to include here?		
R1.1	D	Which usage license do you use for your datasets?		CC BY-SA 4.0		CC BY-SA 4.0
R1.2	MD	the provenance of your metadata records?		CC BY-NC-SA 4.0		CC BY-NC-SA 4.0
R1.2	MD	the provenance of your metadata records?		DDI-Lifecycle 3		
R1.2	D	the provenance of your datasets?		DDI-CDI		
R1.2	D	the provenance of your datasets?		DDI-Lifecycle		
				DDI-CDI		
				PROV		

# In focus till now: FIPs

## Lessons Learned: Similarities and Differences

- We have both similarities and differences in our practices
- We use common standards (DDI), technologies (Nesstar, Colectica, ...) and resources (shared data dictionaries, data models and
- But which versions of which software/standards – creates differences in both practices and technical implementations
- But we can harmonise these
- We have been able to identify common resources – aligns well with the FAIR Enabling Resources
- And we are keen to build these out between us (Deliverable 2)

## Lessons Learned: Reusable FERs on a Detailed Level

- We have very good practices about enabling consistent resources
- We need to be able to point to key resources
- For social surveys we need specific questions, variables, response categories
- But also some domain-agnostic requirements: classifications/vocabularies, datums, harmonisations
- All of these are potential FERs
- And could be specified and reused as needed through ESS and similar international survey programs (ISSP, World Values Survey, ...)

# Lessons Learned: FER and Degree of Specificity

- We need to be very specific in what we can reference in FAIR enabling resources
- It's not clear that:
  - FERs can (currently) cope with the degree of specificity required
  - That we have resources defined with sufficient specificity to make the most of FERs if they can – i.e. how well do we identify specific resources that we use

# Lessons Learned: FAIR and Automation of Processes

- Social science has focused on FAIR for a long time (1960s onwards)
- Interoperability and reuse has been in focus
- But we have been very manual in our practices to implement this
- Automation is currently increasing and we want to upscale



# Ambitions

- We want to take FIPs and FERs out for a serious test drive
  - Can they really do what we need them to do?
- We need to explore suitable repositories and registries in order to be able to do this
- And then align with our standards at a machine-actionable level

- Questions?

# Deliverables

- D6.1 Cross-national Social Sciences survey FAIR implementation case studies (M8)
  - Conduct of a core set of case studies to examine existing practices, and the establishment of a policy and procedures library documenting best practice methods across the two
- D6.2
  - Establishment of, adoption of, or contribution to tools for assessing and comparing the FAIR status of machine-actionable content, with a view to
    - developing appropriate guidance and a checklist for making data FAIR; and
    - developing a discipline-appropriate means of assessing the FAIRness of social science survey data.
- D6.3 Pilot implementation of guidelines with ESS and AUSSI-ESS datasets (M21)
  - Piloting of proposed best practices from this case study in future round of ESS/AUSSI-ESS

