

# IRISS WP4 DEMONSTRATORS

SYNTHESIS WORKSHOP

# IRISS

IRISS is oriented around improving integration across the phases of the research lifecycle of the social sciences.

- Recap
- Synthesis
- Scope
- Approach
- Impact
- Feedback

# IRISS – RECAP

A key objective of the IRISS project is the enhancement of the capacity of Australian social science researchers to create, disseminate, integrate and use quantitative and qualitative social science data sources to generate new insights.

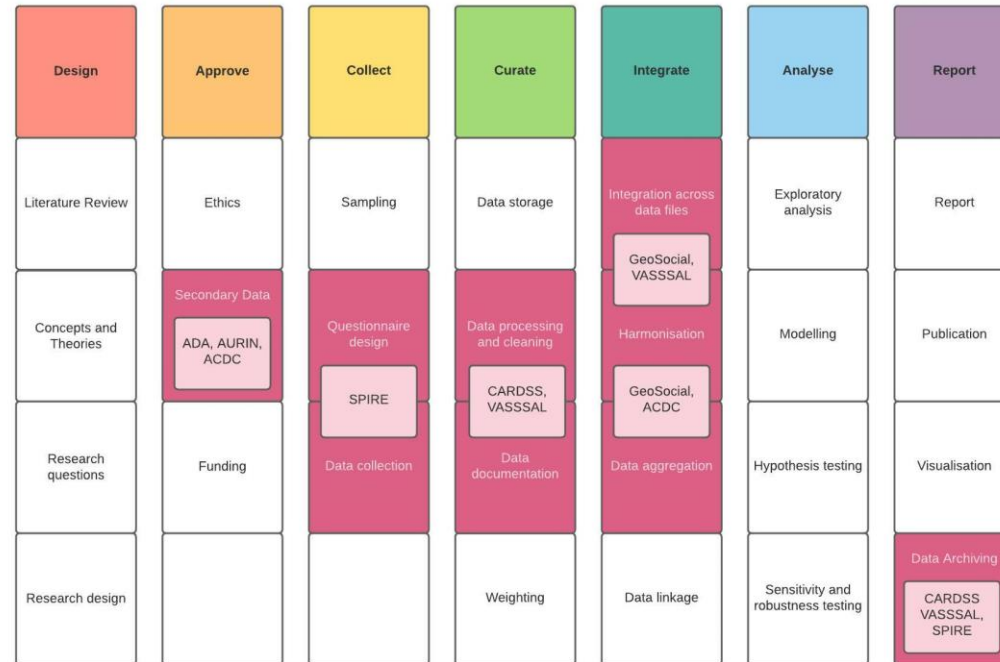
# IRISS – INTEGRATED INFRASTRUCTURE

A system to	By means of	In order to
<p>Coordinate data and research infrastructure components that support data-drive social science research.</p> <p>Enhance nationally significant datasets that underpin empirical social science research.</p>	<p>Using best practice data harmonisation techniques, domain informatics and interoperability standards.</p> <p>An integrated workflow between vocabularies, survey tools and data repository service</p>	<p>Establish a foundation for national research infrastructure for social science research.</p> <p>Create critical longitudinal, demographic and geospatial data infrastructure for social science researchers to build their evidence base on.</p>

# SOCIAL SCIENCE – RESEARCH LIFECYCLE

IRISS and the research process:  
New resources

Sources:  
Research Process - Bryman (2015)  
Framework - Western Australian Biodiversity  
Science Institute (2020)  
Technical Considerations - Gulliver, Famhi  
and Abramson (2021)



# WORK PACKAGES

WP2	VASSSAL – vocabulary access
WP3	GeoSocial – data integration
WP4	IRISS demonstrators
WP5	SPIRE – survey integration
WP6	CARDSS – data curation

# IRISS PERSONAS

Personas used to focus in on common requirements (data and system users)

Personas will need to be checked (reviewed) through interviews

Epics, use cases and user stories are built off each persona

1. Evan – the data scientist/analyst

---

2. Martin – the researcher (data collector)

---

3. Danielle – the researcher (data analyst)

---

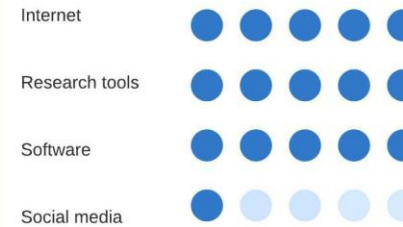
# EVAN – DATA SCIENTIST/ANALYST



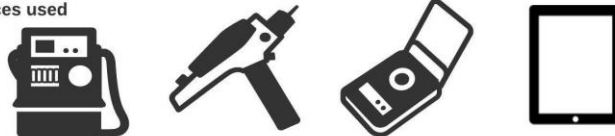
**Name:** Evan  
**Age:** 38  
**Occupation:** Data scientist/analyst

**Behavior/personality**  
Infrastructure-savvy; T-shape professional (specialist) or Alt-Academic; loves the outdoors

## Technology expertise level



## Devices used



## Favorite datasets and tools



## Goals


- Evan wants to establish new data governance and integration methodologies and provide new data products (enhanced with geo-spatial and vocabulary information) and question libraries
- He also wants to develop API protocols that enable interoperability between archive and research platforms and survey and analysis tools

## Challenges

- Evan has challenges with supporting social science researchers because data and system governance and interoperability models have not been developed that coordinate existing research data, vocabularies, platforms and tools from different sectors



# MARTIN – RESEARCHER (DATA COLLECTOR)



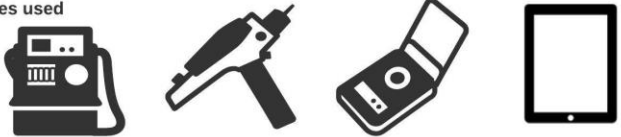
**Name:** Martin  
**Age:** 43  
**Occupation:** Social scientist

**Behavior/personality**  
Academic, tech-savvy, driven;  
passionate about social change  
through better policy; loves music


**Technology expertise level**

Internet	●	●	●	●	●
Research tools	●	●	●	●	●
Software	●	●	●	●	●
Social media	●	●	●	●	●

**Devices used**



**Favorite datasets and tools**



**Goal**

- Wants to set up his own data collection (survey design) with domain specific resources and advance his capacity to generate data outputs that are interoperable and reusable

**Challenges**

- Having to reinvent the wheel each time he sets up a survey and rekeying standardised census vocabularies and value data, geo-spatial information and question libraries into his data collection (survey tool)

# DANIELLE – RESEARCHER (DATA ANALYST)

**Name:** Daniella  
**Age:** 45  
**Occupation:** Social scientist

**Behavior/personality**  
Academic, tech-savvy, driven;  
passionate about economic change  
through better policy; loves history

**Technology expertise level**

Internet	● ● ● ● ●
Research tools	● ● ● ● ●
Software	● ● ● ● ●
Social media	● ● ● ● ●

**Devices used**

**Favorite datasets and tools**

**Goals**

- Daniella wants to request and then work on sensitive datasets (new data products) in analysis environments (safe settings) and use tools like R Studio, SAS or SPSS

**Challenges**

- Daniella has challenges with the current access and data transfer processes that do not support systematic monitoring of data movement and end-to-end security controls including encryption at rest and in transit

# IRISS PERSONAS

Evan	Data scientist/analyst
Martin	Researcher (data collector)
Danielle	Researcher (data analyst)
<i>Serena</i>	<i>Policy-maker</i>
<i>Yosef</i>	<i>Data Analyst</i>

# SERENA – POLICY MAKER

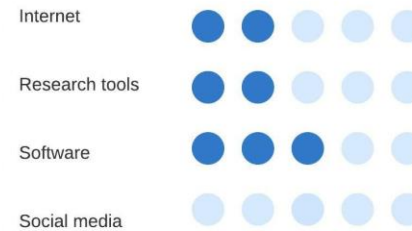


**Name:** Serena  
**Age:** 58  
**Occupation:** Social Policy Expert

#### Behavior/personality

Public servant, data-savvy, driven; passionate about social change through better policy; loves travel

#### Technology expertise level



#### Devices used



#### Favorite datasets and tools



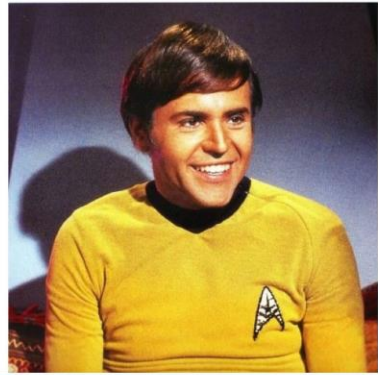
#### Goals

- Serena wants to know at a glance the judgement calls, curation decisions and the methods of data integration to understand the data dimensions and how fit-for-purpose the integrated data is for the research undertaken with it - akin to reading a nutrition or quality rating label

#### Challenges

- Serena has challenges when reading research publications knowing the quality of the underlying integrated data, to then be able to make an informed decision about incorporating the research findings in new policy papers

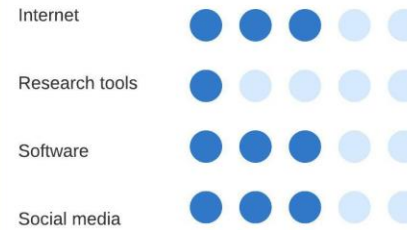
# YOSEF – DATA ANALYST



**Name:** Yosef  
**Age:** 32  
**Occupation:** Data analyst

**Behavior/personality**  
Software-savvy; T-shape professional (specialist) or Alt-Academic; loves cycling

## Technology expertise level



## Devices used



## Favorite datasets and tools



## Goals

- Yosef wants to use new data products (enhanced with geo-spatial and vocabulary information) and to learn to use API to make setting up his data collection faster and improve his analysis and learn to use R Studio

## Challenges

- Yosef has challenges with having to repeat the same procedures bringing different datasets together and integrating them with geo-spatial and vocabulary information - especially for comparative studies

# SYNTHESIS

- Practicalities

- Coordinate WP2 Vocabularies and WP3 GeoSocial
- Coordinate WP5 Survey and WP6 Curation
- Align CADRE and IRISS with sensitive demonstrator WP4
- Follow up workshops for deeper dives
  - AURIN demo of spatial integration work with ADA & PHRN
  - Linking survey setup (questions and variables) to data curation processes
  - Moving safe data from a survey tool into an analytical environment (UMELB)
  - Moving safe output into/out of an analytical environment (UMELB)
  - ARDC on engagement with ABS and reworking their content (ANU)

# SYNTHESIS

- Foundational work
  - Audience profile : context, user needs, domain expertise, technical skills
  - Governance structure for repurposing third party content e.g., ABS
  - Framework : data curation – harmonisation, recoding and provenance
  - Framework : data integration – different data sources and types
  - Framework : survey integration – different information sources and types
  - Definitions : data and survey integration, spatial and vocabulary info
  - 4D matrix : demonstrators, audience, software/tools, data/info sources

# SYNTHESIS

- Deliverables

- Longitudinal data integrated with vocabulary and spatial information and a survey tool (new data products delivered as the first exemplars that can be linked to the frameworks and matrix)
- Governance and best practice methods for data and survey integration
- Vocabulary and Transformation services

## Exploratory work

- Governance and best practice methods for reuse of questions and variables
- Novel data sources e.g., social media, satellite or private
- Data integration that has community impact e.g., disaster or disease
- Outline of research opportunities, requirements, and data integration challenges



# MATRIX

	Data Sources		Info Sources		Audience		Software/Tools	
	Longitudinal	Social media Satellite Private	Vocabularies	Spatial	Non-technical	Technical	Desktop	Scripting
<b>Spatial</b>				ASGC, ASGS	Data analyst, Policy maker		SPSS, Tableau	
<b>Sensitive</b>								
<b>Census</b>	ABS Census		ANZSRC, ANZCO					

# MATRIX

	Data Sources		Info Sources		Audience		Software/Tools	
	Longitudinal	Social media Satellite Private	Vocabularies	Spatial	Non-technical	Technical	Desktop	Scripting
<b>Spatial</b>				ASGC, ASGS		Data scientist/ analyst		R Studio
<b>Sensitive</b>	HILDA							
<b>Census</b>			ABS AGEF, ANZSOC					

# MATRIX

	Data Sources		Info Sources		Audience		Software/Tools	
	Longitudinal	Social media Satellite Private	Vocabularies	Spatial	Non-technical	Technical	Desktop	Scripting
<b>Spatial</b>				ASGC, ASGS		Data scientist/ analyst	Survey tool	
<b>Sensitive</b>								
<b>Census</b>	ABS Census		ANZSOC, ANZSIC					

# SCOPE

- Priorities for uptake and impact
- Datasets used as worked examples
- Info (vocabulary and spatial) used for integration

# SCOPE

- Priorities for uptake and impact
  - High utility longitudinal data for social science and allied domains
  - High utility vocabulary, value data and spatial info for social science
  - New data products are human readable and require low technical know-how

# SCOPE

- Datasets used as worked examples
  - Longitudinal and survey based datasets used as worked examples
    - ABS Census datasets in public domain (CC licence)
    - HILDA dissemination dataset (ADA, UQ, CADRE and UMELB Data Lab)
  - Show methods for and value of data and survey integration
  - Explore different data sources as future candidates for integration

# SCOPE

- Information used for integration in worked examples
  - ABS Census dictionaries, ANZSRC, ANZSIC, ANZSCO, ASGC, ASGS
  - Show methods for and value of reworking ABS content
  - Explore different information sources as future candidates for integration

# APPROACH

- Develop frameworks to *outline the extent* of the data curation and data and survey integration work that delivers value to social science research (scoping future work)
- Deliver end-to-end *worked examples* with longitudinal datasets and ABS content, make the methods for integration and curatorial decisions explicit (linked to areas in the frameworks)



# APPROACH

- Deliver easy to use data products and services targeted at the widest range and largest cohort of researchers (linked to matrix)
- Explore a range of novel information (vocabulary, questions, spatial) and data sources and their impact (scoping future work)

# IMPACT

Demonstrator	Delivers	Impact/s
Spatial	<p>Framework for enhancing longitudinal data with spatial info</p> <p>End-to-end example of data integration of longitudinal data with spatial info</p> <p>Exploration of different data and info sources</p>	<p>Research community</p> <ul style="list-style-type: none"> <li>• Social science</li> <li>• Allied domains</li> </ul>
Sensitive	<p>Assessment of project outputs for reuse in sensitive data environments</p> <p>Pilot study for future sensitive data support for technical audience and laypersons</p>	<p>Research community</p> <ul style="list-style-type: none"> <li>• Social science</li> <li>• Policy makers</li> </ul>
Census	<p>Framework for enhancing longitudinal data with vocabulary info</p> <p>End-to-end example of survey integration reusing longitudinal data with spatial and vocabulary info</p> <p>Exploration of different data and info sources</p>	<p>Research community</p> <ul style="list-style-type: none"> <li>• Social science</li> <li>• Allied domains</li> </ul> <p>Community</p> <ul style="list-style-type: none"> <li>• Disaster and</li> </ul>

# FEEDBACK & QUESTIONS

- Are the priorities set the right ones?
- Are the foundations and definitions set out the right ones?
- Is the balance right between:
  - illustrating social science as a domain and an enabler of allied domains?
  - delivering useful outputs (for research community) and exploring options for novel research (for research impact)?