# Integrated Research Infrastructure for Social Science Project Report

ARDC HASS RDC and Indigenous Research Capability Program

Milestone 2.1 **WP2 User requirements**_ARDC_VocabularyRequirements_IRISS.

**LEAD ORGANISATION:** Australian National University (ANU), Australian Data Archive (ADA), ANU Centre for Social Research and Methods (CSRM), Research School of Social Sciences (RSSS)

**Partners:**
- University of Queensland, Institute for Social Science Research (ISSR)
- University of Melbourne, Melbourne Institute
- Australian Urban Research Infrastructure Network (AURIN)
- Australian Consortium for Social and Political Research Inc (ACSPRI)

**Authors:**
30/13/2022

| Questions | Response/Decision | Next Steps |
|---|---|---|
| **Consent & scope – classifications** | | |
| Is there an ABS point of contact for ARDC? | ANU has Talei Parker, Jarrod Lange & Linda Skiller as contacts for: stakeholder engagement, geographic and Census services. ARDC will need to determine what point of contact is suitable for its purposes (as repository and broker). | AB and SM liaison. ARDC & ANU engagement with ABS on shared or delegated governance. |
| Have the target ABS classifications been identified? | Focusing on: Education, Age, Sex, State for Census demonstrator.<br>Top tier: Education (ASCED) and Geography (ASGC & ASGS) (as a priority for the demonstrators). Occupation (ANZSCO) and Industry (ANZSIC) as candidates for phase 2 of IRISS. With Geography – unclear here on optimum place for authoritative geographic (names) vs geometric (spaces) and the role of Loc-I (passing on ABS content).<br>Subsets: variables that appear in the 2021 Census data dictionaries relating to population and education and training. It may be important to focus on particular mnemonics e.g., HEAP, AGEP and SEXP for the demonstrator. | Plan for making Education top tier and population and education and training subsets across 2016 and 2021 Census data dictionaries machine accessible via RVA. |
| Does the project wish to express relationships between versions of classifications? | Yes – to classifications and variables. Unclear what the crossover is here with what's best to use RVA and Collectica for to make FAIR (classifications and variables respectively) to provide a national service (PID & referenceable). | Explore tech mix of Dataverse, Collectica and RVA (and Loc-I) as options for publishing and working with referenceable classifications and variables from Census data dictionaries for 2016 and 2021. |
| Has the project secured ABS consent for publishing ABS classifications via ARDC? | ABS content made publicly accessible is available under a CC-BY licence. If ABS wants to "own" the content on RVA (or elsewhere) this is idea. Alternatively, ADA can "guide" the development of the published output and maintain the procedural and technical documentation as a WP4 project output. | AB and SM liaison. ARDC & ANU engagement with ABS on shared or delegated governance. |
| Will ABS provide final review and acceptance of metadata and classifications published via ARDC? | As above. | As above. |
| **Encoding & IRI resolution – classifications** | | |
| What source formats are available from ABS? | Currently only .csv by download. Ideally access to a more expressive and accurate version of this content e.g., SDMX. | As above. |

| Questions | Response/Decision | Next Steps |
|---|---|---|
| Who will encode the classifications? (ABS, ARDC, other)? | ARDC with support from IRISS team on approach and relevance to use case/s. | IRISS team are to technically test the reuse of Census data dictionary classifications and variables for 2016 and 2021 in RVA. |
| Does ABS or the project have a preferred approach to encoding? (SKOS, XKOS, other)? | Project requires SKOS & XKOS – any alignment with DDI codebook or lifecycle models is beneficial. | ARDC & IRISS team review informatics as part of tech assessment of Dataverse, Collectica and RVA. |
| Will ABS define an IRI pattern? Does ABS have a redirection service for resolution of IRIs to the ARDC host? | Not ABS no on either question. ARDC is viewed as the partner/broker with ANU in publishing and making ABS content referenceable. Advice by experts in this space might be sought on the best approach. | |
| Machine-readable question libraries and survey response formats | | |
| Are there existing standardised approaches to modeling and encoding question libraries and survey response formats? | Exploring the potential for working with that idea of "cascading variables" and the role that Collectica might play as part of an informatics workflow for social science (along with Dataverse, RVA and LimeSurvey) | As above – Explore tech mix. |
| Are there existing examples of machine-readable registers of question libraries and survey response formats? | CLOSER (UK) use Collectica and they are an example of how DDI is used to encode survey and response information. | Explore the Collectica demo. |
| Integration | | |
| What is meant and whether an approach has been selected, or will be developed? VASSSAL will takes classifications and vocabularies from ACDC (Census Demonstrator project) and SPIRE, and integrate with international vocabulary services such as the CESSDA | IRISS will deliver proof-of-concept work through the demonstrators (WP4). We are in process of understanding where/how the different tech fits into an informatics workflow for social science (to make it possible to reuse ABS content pre and post data collection and as part of data integration. | Create a template for capturing the informatics workflow. |

| Questions | Response/Decision | Next Steps |
|-----------|-------------------|------------|
| vocabulary service and DDI Alliance vocabularies | | |

| Task Name | Priority | Assigned To | Due Date | Progress | Notes |
|---|---|---|---|---|---|
| **IRISS** | | | | **0 of 73 Done** | |
| Support services development (techdev) | Medium | Ingrid Mason | 08/31/23 | Quarter | |
| Framework for survey integration | Medium | Steven McEachern | 09/30/22 | Empty | |
| Export AUSSA from LimeSurvey to commence data integration with ABS Census data dictionary | Medium | Adam Zammit | 09/30/22 | Empty | Start with 2016, move onto 2021 |
| Socialise approach and results of reuse of ABS content with HILDA data and survey tool | Medium | Steven McEachern | 11/10/22 | Quarter | 15 Sept & 10 Nov RWG meeting |
| Engage with ABS on ASGS their use of Loc-I | Medium | Steven McEachern | 12/31/22 | Quarter | Jarrod Lange & Laurent Lefort (ABS) |
| Demonstrate use of demo RVA and CV with ABS Census artefacts | Medium | Steven McEachern | 12/31/22 | Quarter | CV reuse and fixity |
| Import ABS Census categories and classifications in demo RVA | Medium | Rowan Brownlee | 09/07/22 | Three Quarter | 2016 & 2021 Census dictionaries as reference-able |
| Import ABS Census concordances in demo RVA | Medium | Rowan Brownlee | 09/30/22 | Empty | 2016 & 2021 using XKOS |
| Transition ABS Census categories and classifications to prod RVA | Medium | Rowan Brownlee | 01/31/23 | Empty | Presumes MOU with ABS is in place |
| Import ABS Census questions in demo RVA | Low | Rowan Brownlee | 01/31/23 | Empty | Exploratory only |
| Oversee technical management | Medium | Marina McGale | 08/31/23 | Quarter | |
| Specifications for IRISS services | Medium | Steven McEachern | 09/30/22 | Empty | |
| Specification for Vocabulary service (solutions architecture) | Medium | Rowan Brownlee | 09/30/22 | Empty | |
| Engage with partners on uptake of CV | Medium | Steven McEachern | 08/31/23 | Quarter | |
| Requirements for RVA services (existing/new) | High | Ingrid Mason | 09/30/22 | Quarter | |
| Results of working with ABS content RVA demo | Medium | Rowan Brownlee | 08/31/22 | Quarter | Census 2016, 2021 dictionaries |
| Meeting to discuss if/how Census data dictionary Qs and variables are suitably encoded with ABS | Low | Steven McEachern | 12/15/22 | Empty | |
| Support CV workshop Q4 2022 | Low | Steven McEachern | 12/15/22 | Empty | |
| Update on internal ABS discussion on automation plans | High | Steven McEachern | 08/31/22 | Quarter | Linda Skiller & Talei Parker |

| Engage with ABS on MOU for governance | High | Adrian Burton | 12/31/22 | Quarter | Linda Skiller & Talei Parker |
|---|---|---|---|---|---|
| Manage partner requirements and engagement | Medium | Ingrid Mason | 08/31/23 | Quarter | |
| Workshop - documentation (as-is, to-be) DDI | Low | Ingrid Mason | 08/31/22 | Empty | Use AUSSA & ANU Poll |
| Workshop - harmonisation (as-is, to-be) R/Stata | Low | Ingrid Mason | 08/31/22 | Empty | Use AUSSA & ANU Poll |