

Integrated Research Infrastructure for Social Science Project Report

ARDC HASS RDC and Indigenous Research Capability
Program

1.4 & 2.5 _WP1 & WP2_ IRISS Vocabulary Provenance Service Design _Milestone

ORGANISATION: Australian Data Archive (ADA)

Authors: Sameer Kamrani - Australian Data Archive (ADA)

30/11/2023

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

PARTNER ORGANISATIONS:

- 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)

PROJECT LEAD: Assoc. Prof. Steven McEachern (Project Lead)

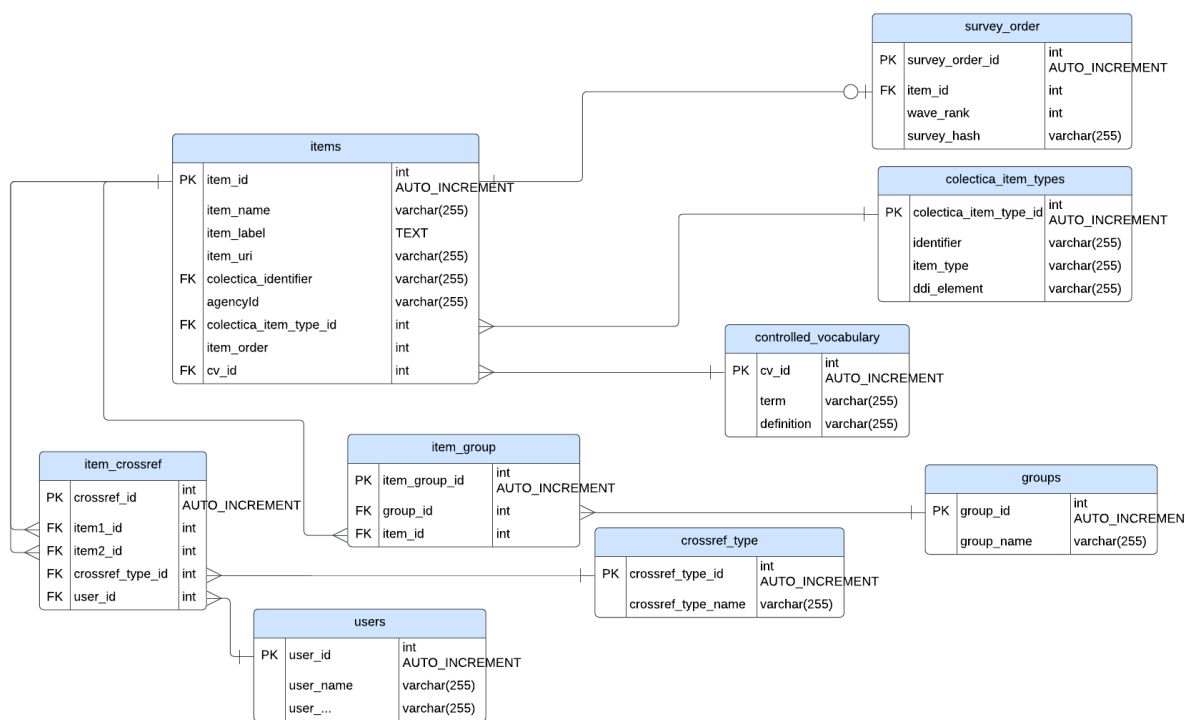
PROJECT COORDENATOR: Keila Silva

Schema Report: Harmonization of Survey Data

Introduction

This database is designed to harmonize and unify data collected from various surveys. It emphasizes creating a structured approach to consolidate diverse survey data, facilitating easier analysis and reporting. The schema also allows for the tracking of metadata associated with each survey item, as well as the relationships between them.

The Schema



Use-case Scenarios

- 1. Survey Data Consolidation:** A research institute conducts multiple surveys across different domains. They need a way to bring all survey data under one umbrella, ensuring consistency and allowing for cross-survey analysis.
- 2. Controlled Vocabulary Maintenance:** To maintain the integrity and consistency of survey data, a controlled vocabulary is used. This enables researchers to ensure that terms are used consistently across surveys. Additionally, this controlled vocabulary can be mapped to external vocabulary services. If specific terms do not exist in these external services, procedures can be put in place to integrate them into the vocabulary services as required.
- 3. Item Tracking in Multiple Surveys:** Researchers often reuse certain questions or items across different surveys. The database allows for tracking the same item across various surveys, noting its order and relevance in each.
- 4. Cross-referencing Survey Items:** Items in one survey can relate to items in another for various reasons, including but not limited to prerequisites, dependencies, conceptual equivalence, or as complementary information. The cross-reference feature allows for establishing these connections.

5. **Grouping Survey Items:** For better organization and categorization, survey items can be grouped under various categories, aiding in thematic analysis.
6. **User-driven Cross-referencing:** Given that different researchers might see different relationships between items, the database allows users to create their cross-references, capturing the user details for accountability.

The Synergy of Cross-Referencing: Unveiling Hidden Connections

Cross-referencing, while a technical act of linking data points, carries profound implications in the realm of research. Here are the benefits of such an intricate process.

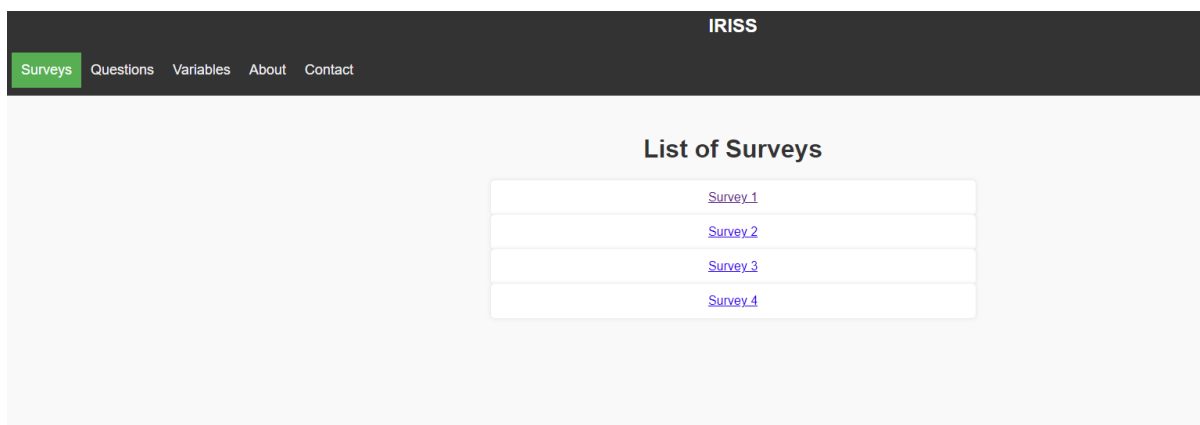
1. **Ontology and Knowledge Graph Construction from Cross-referenced Items:** The act of cross-referencing items inherently creates relationships between diverse survey data points, forming a rich web of interconnected information. These relationships can serve as the foundation for constructing ontologies and knowledge graphs.
2. **Collaborative Research Opportunities:** By allowing and tracking user-driven cross-references, the database encourages collaborative research endeavours. Over time, a collective wisdom emerges, reflecting the diverse perspectives of the research community.
3. **Integration with External Ontologies:** Apart from mapping to external vocabulary services, the controlled vocabulary can potentially be integrated with external ontologies or knowledge graphs. This can aid in semantic web research, enabling richer, context-aware querying and exploration of survey data.

Interface Overview

The interface, while currently presented as wireframes or a proof of concept, has been designed to mirror the intricacies of the underlying schema. It offers a visual representation of the harmonized survey data, making it more accessible and user-friendly. This integration between the schema and the interface ensures that the vast array of information stored in the database is presented in a structured and intuitive manner to the users.

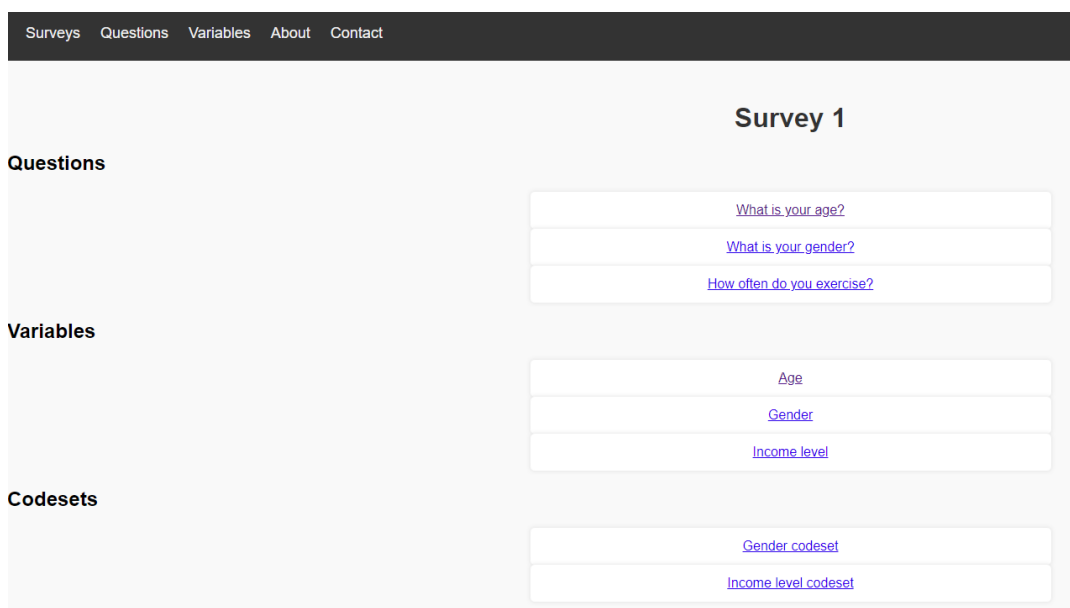
1. Surveys Overview

- This is the landing page that lists all the surveys available.
- It provides a navigation bar for quick access to other sections like questions, variables, about and contact.



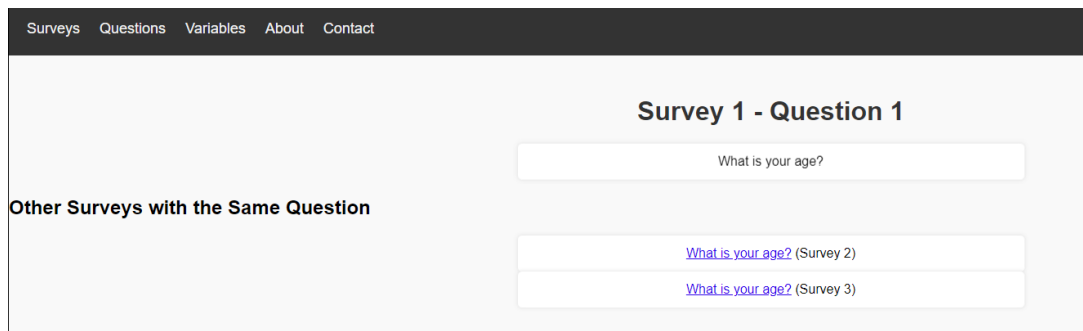
2. Individual Survey Page

- Each survey has its page that lists the questions, variables, and codesets associated with it.
- This page provides a deeper dive into each survey, allowing the users to explore its structure and contents.



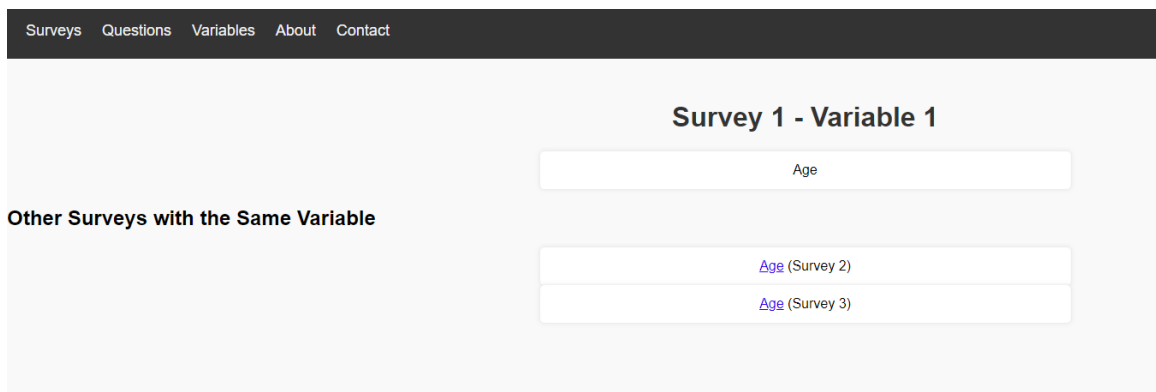
3. Individual Question Page

- This page focuses on a specific question from a survey.
- It lists other surveys that have the same question, facilitating cross-reference and highlighting the harmonization potential.



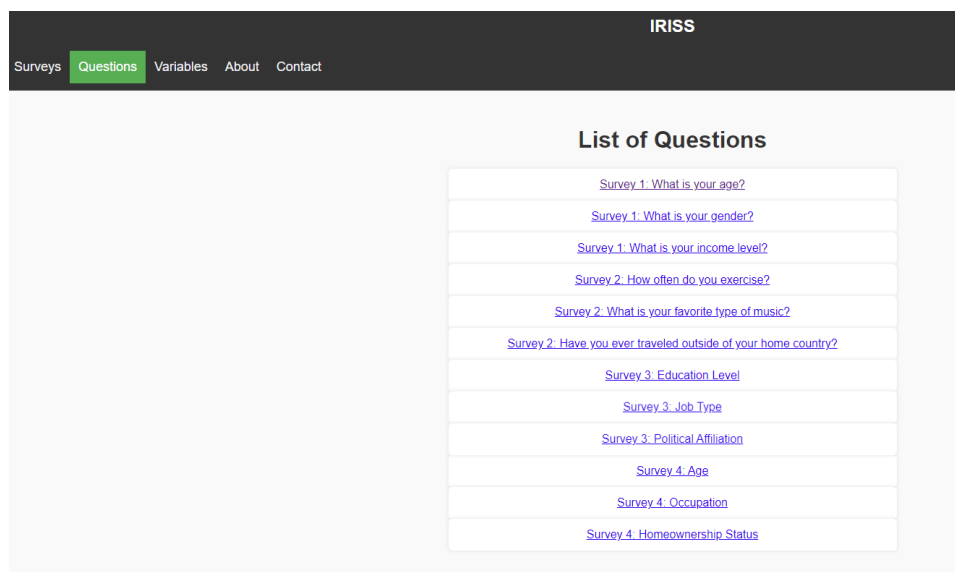
4. Individual Variable Page

- This page is dedicated to a specific variable from a survey.
- Like the individual question page, it also lists other surveys containing the same variable.



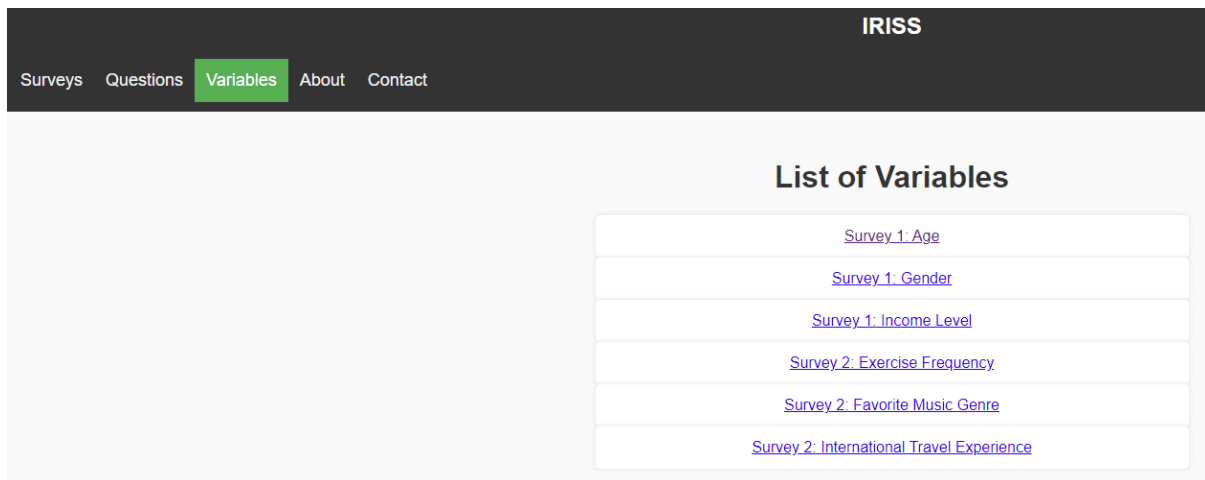
5. Questions Overview

- This page lists questions across all surveys, providing a holistic view of the kind of questions present in the system.



6. Variables Overview

- This page lists variables across all surveys, allowing users to understand the type of data collected and analysed.



Possible features in the Interface

- 1. Search Functionality:** Given the potential of large number of surveys, questions, and variables, adding a search bar would enhance the user experience. This allows users to quickly find specific items without scrolling.
- 2. Breadcrumb Navigation:** For deep pages like individual questions or variables, breadcrumb navigation can help in navigating back to higher-level pages.

Example of breadcrumb navigation

Survey1>Question1>Variable1

- 3. Metadata Display:** On individual pages, like questions or variables, displaying metadata (like associated groups, definitions from the controlled vocabulary, hyperlinks to external vocabulary service, and any other relevant cross-reference relationships) will enrich the information available to the users.
- 4. Interactive Visualization:** As the database grows, manual traversal becomes unmanageable. Integration of visualization tools like graphs or tree structures to visually represent relationships between items will enhance the user experience.
- 5. Pagination:** If the list of surveys, questions, or variables

The wireframes provide a clear and user-friendly structure for navigating and understanding the harmonized survey data. The platform can offer an enhanced, comprehensive, and interactive experience for users, maximizing the potential of the schema.

Conclusion

The harmonization of survey data schema is a comprehensive solution to manage, consolidate, and analyse survey data from diverse sources. Its structured approach not only ensures data integrity but also facilitates rich cross-survey insights. By capturing metadata, relationships, and user contributions, the database becomes a robust tool for any research institute or organization dealing with diverse survey data.