Dataverse 'Installation Personas': A UX concept adopted to enable the navigation of Dataverse installations, for the purpose of sharing ideas and technical experiences by the international community – based on FAIR data principles

Janet McDougall[1], Philip Durbin[2], Tania Schlatter[3], Marina McGale[4], Alexander Jerabek[5]

[1]Austrailan Data Archive, Canberra, Australia, janet.mcdougall@anu.edu.au
[2]The Institute for Quantitative Social Science, Boston, USA, philip_durbin@harvard.edu
[3]The Institute for Quantitative Social Science, Boston, USA, tschlatter@fas.harvard.edu
[4]Austrailan Data Archive, Canberra, Australia, Marina McGale@anu.edu.au
[5]Library Services, Montreal, Canada, jerabek.alexander_j@uqam.ca

Dataverse is an open source repository to share, preserve, cite, explore, and analyse data. It is developed and maintained at Harvard's Institute for Quantitative Social Science (IQSS), along with many collaborators and contributors worldwide. As a community we aim to promote and learn about behavioural and technical solutions, and standards for curating, sharing, and preserving data that can be discovered and reused across disciplines, to reproduce and advance research. [1]

CONCEPT

By applying the UX concept of 'personas' to Dataverse installations, we can represent different Dataverse access models and technical configurations. Installation Personas can be used to describe and classify users' needs, experiences, behaviours and goals. Within the community we are discovering that we want to identify installations that may have already implemented a configuration, or are working towards a similar outcome, to share and build on experiences.

There are currently 46 Dataverse installations internationally, and there are countless configuration options available. There is a need to be able to easily find Dataverse implementations with specific installation models, versions, and configurations, whether to:
- check how a particular feature behaves
- compare or build on policies related to various archival models
- locate Dataverse communities within geographic locations
- identify installations by configuration (version, storage solution, access conditions...)

DESIGN AND DEVELOPMENT

The Installation Persona interface is still at the design and development stage, after collaborating at the Dataverse Community Meeting 2019 in June (https://projects.iq.harvard.edu/dcm2019). The aim is for Installation Personas to be ultimately maintained and remain current using installation metadata and 'non sensitive' configuration details extracted through Dataverse APIs. Standardised metadata is then usable by database, visualization, and mapping tools to build an interface based on FAIR data principles [2], to ensure the most benefits to the community.

Emphasis is for Installation Personas to become embedded in the Dataverse community as a tool for new and existing members to navigate within the community, and share materials and experiences.

Current plans are to build the interface on the Dataverse World View map (https://dataverse.org), with the search and visualisation functions based on metadata and vocabularies put forward since the 2019 community meeting:

- Configuration definitions - Out of the box; Standard; Public Only; Big Data; Cloud Compute; Sensitive Data
- Access Policy Models - Access Open; Access Conditional; Access Mediated; Access Restricted
- Data types - Qualitative; Quantitative; Derived; Curated; Self-deposit

It is anticipated that metadata will be harvested from community Dataverses using existing and new APIs, and stored for access by the Installation Persona interface.   Test scripts have been written to grab screenshots of the 46 community Dataverses and saved to a Dataverse dataset as part of the testing process to build 'Installation Personas' for each Dataverse.   The data files and scripts have been added to https://github.com/IQSS/dataverse-installations which is the new repo we created during the 2019 Dataverse Community Meeting.
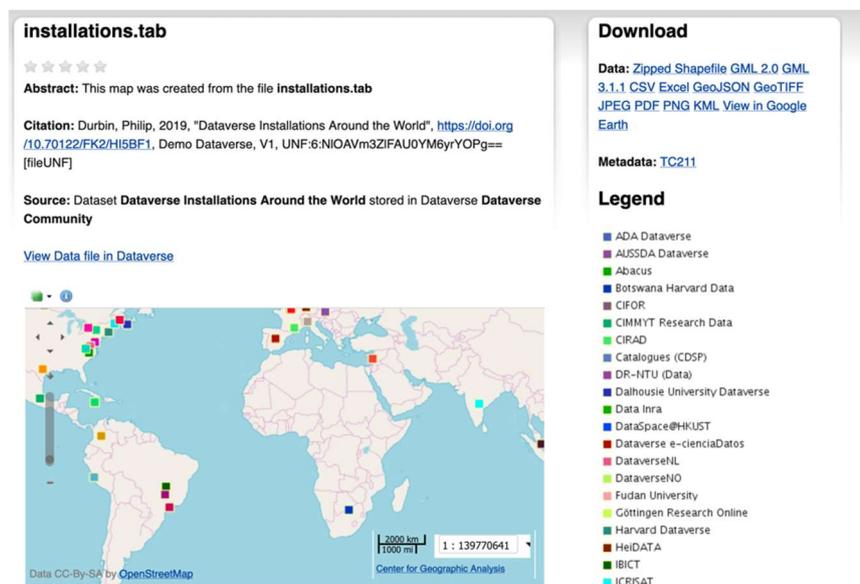

Figure 1: Dataverse World View Map of installations created using Dataverse, Geoconnect and Worldmap

FAIR DATA PRINCIPLES

Findable: Rich configuration, access policy, and data type metadata described by standardised vocabularies findable through the Installation Persona interface, and internet web searches.

Accessible: Configuration and policy implementation materials made available through the hosting institution after being located by the Installation Persona interface.

Interoperable: Community standardised vocabularies and metadata fields to describe 'Installation Personas' to enable them to be navigated, searched and visualized by the Installation Persona interface.

Reusable: Configurations and materials shared and made findable, accessible and interoperable through the Installation Persona interface.

REFERENCES

1. The Dataverse Project. Available from: https://dataverse.org/about accessed 30 August 2019.
2. ARDC FAIR Data. Available from: https://ardc.edu.au/resources/working-with-data/fair-data accessed 30 August 2019.