

Social Sciences & Humanities Research Data Interest Group Proposal

Introduction

We are proposing a Social Sciences & Humanities Research Data (SSH RD) Interest Group under the auspices of the Research Data Alliance (RDA), to foster diverse professional exchange on issues particular to data originating from the social sciences and humanities.

Social Sciences and Humanities Research Data cover many disciplines, appear in many data types, deal with multiple objects and levels, and are very distributed – coming from various sources. It could be described as a patchwork quilt, lacking a grand design or focus. On the other hand, it is a way to cover the whole spectrum, to be flexible in collecting data.

There is a huge potential reuse of SSH RD – for researchers, but also for professionals outside universities, for companies, governments, and for citizens.

As a research data community, we are entering the **implementation phase** of the FAIR principles: we can see first results on Findability – with various catalogues coming available. But for Accessibility, it already complicates as many social data are **too sensitive** to share directly: access to social sciences research data is not dichotomous: open or closed but requires fine-tuning on making data accessible. For better Interoperability, we need alignment on controlled vocabularies and ontologies, as well as semantic techniques to relate data. A barrier for Reuse is **lack of clarity in data policies** – and their implications for researchers: what are conditions and requirements for providing access, and for using the data. For Reuse, new users want information about the **quality and provenance of the data**: where do they come from, how were they collected and curated, etc.

Social Science & Humanities Research Data Interest Group Focused Initiatives

This new SSH interest group will begin by helping to coordinate communications across the various current RDA groups of interest to our disciplines and to provide a place for our members to share solutions and concerns with others in our fields. These groups include the FAIRSharing and others noted by Braukmann in the “RDA Overview for the Social Sciences”¹. We will also seek input and coordinate with the external SSH community leaders and organizations such as CESSDA, DDI Alliance, IASSIST, ICPSR, IFDO and WDS. We will be open and inclusive seeking to use this group to connect the various organizations working to promote SSH data sharing. While this new interest group will

¹ <http://doi.org/10.5281/zenodo.1401105>

certainly be a coordinating group, we also aim to produce new RDA Working Groups to help provide solutions to challenges in our SSH domains. Given the complexity of SSHRD, we feel that the initial focus of our interest group should be very focused on defining working groups of immediate need to our communities. We foresee that to implement FAIR in our disciplines we need to engage all stakeholders: funders, producers, service providers, users; and as a new interest group we want to prioritise on three specific areas focused on the Social Science and Humanities research data communities.

- **quality of data**
find automated ways to investigate and provide information on quality
- **data policy**
align – and wherever simplify – data policies and their implications for making data available and for using data.
- **sensitive data**
estimates are that over 40% of the data in our community is too sensitive to make them openly available without any restrictions or measures.

This approach – to engage all relevant parties, and to focus on a limited number of topics – meets the Minimum Viable Ecosystem approach that is used for the EOSC, and has been used in building and expanding platforms, like Apple did (see frame)

In Apple's case, the MP3 player represented a limited feature set (music only) but established the necessary minimum ecosystem (MVE). Next there was iTunes to add new music. But the big leap happened when they added partners who increased the feature set, moving towards the full rollout: Adding AT&T as a partner expanded the iPod into the iPhone, a more complete feature set. The rest is history.
Source: <https://smartorg.com/innovating-to-create-an-ecosystem/>

Objectives

The initial objectives of the SSHRD interest group will be to bring together major community members seeking to coordinate international social science research data sharing. As noted above several existing groups within RDA are discussing topics of interest to the SSH disciplines but none focus specifically on the disciplines of social science and humanities with the key focus on data sharing. We will bring individuals and communities together to provide this forum and define future working groups to help solve the issues identified.

Participation

We want to include all stakeholders and reach as broad an audience as possible. We will initially begin working with existing RDA groups and further discussions between CESSDA, DDI Alliance and IFDO to start the conversations. But we will

immediately begin contacting the many other groups ranging from IASSIST to ICPSR that will be valuable in providing input to this interest group. Below is an initial matrix of the groups we feel would have interest in our efforts.

	Social Sciences	All Sciences
Policies (Open Science)	IFDO	OECD, CODATA, ORFG
Infrastructures Repositories Trust/Certification Archiving Sensitive Data	CESSDA, ICPSR, ODUM, ADA ERAN (EU), RAIRD (NO)	ESFRI, ARDC Figshare, Mendeley, Zenodo/OpenAIRE CoreTrustSeal Dataverse EGI (AAI)
Architecture Digital Objects PID Metadata Kernel Info	DDI	DOI / Handle, DataCite ISO RDA Data Type Registry
Data Production	ESS, SHARE, Wage- Indicator, GGP, EVS, WVS, ...	National Statistics, ILO, World Bank, ...
Researchers	Former SSRN?	YEAR (young researchers EU)

Outcomes

- Landscape analysis on sensitive data topics/issues, initiatives/activities/tools, best practices (and don'ts)
- Implementation guide on data policies building on ORFG Blueprint, Science Europe DMP
- Landscape analysis on tools for assessing data quality automated and integrated in existing data services tools (e.g. Dataverse, FigShare, Fedora, etc...)

Mechanism

We will begin by having biweekly video conference calls on a rotating time slot basis to help be inclusive for members in various time zones. We will propose an initial IG meeting at the RDA Plenary in Philadelphia Spring of 2019 to plan for the rest of the year. Initial calls will be coordinated by CESSDA, IFDO, and DDI Alliance but as the group grows, we will be adding more participants and organizational representatives from our matrix above.

First Six Months Timeline

Draft of first IG meeting agenda in Philadelphia
First call and then every two weeks following
Finalize IG meeting agenda for Plenary

Early January 2019
Late January 2019
Late February 2019

Contact list of organizations to invite to participate
RDA Plenary meeting
Suggested Working Group Topics
Final Draft of Working Group Suggestions

Early March 2019
April 2019
May 2019
July 2019

Initial and Potential Group Members

This group will initially be Co-Chaired by Jonathan Crabtree representing the Odum Institute, the International Federation of Data Organizations and the Global Dataverse Community Consortium, Ron Dekker representing the Consortium of European Social Science Data Archives (CESSDA) and Steve McEachern representing the DDI Alliance and the Australian Data Archive. Our initial Birds of a Feather meeting in Botswana was very well attended and we have thirty members from that meeting that are interested in joining this interest group. We will utilize our existing social science community network connections to increase this list significantly during the first six months once we begin advertising our intent to the larger RDA audience.