



RECODE

WP 1
Stakeholder Values
and Motivations



The
University
Of
Sheffield.

Thordis Sveinsdottir,
Bridgette Wessels and
Rod Smallwood

Overview

- Objectives of WP 1
- Methodologies
- Review – key points
- Case Studies findings
 - Archaeology
 - Bioengineering
 - Environment
 - Health
 - Physics
- Points for discussion

WP 1

Stakeholder Values and Motivations

- Objectives
 - To identify and map the diverse range of stakeholder values in open access and data dissemination and preservation
 - To map stakeholder values on to scientific ecosystems using case studies from different disciplinary perspectives
 - Conduct a workshop to evaluate and identify good practice in addressing conflicting value chains and stakeholder fragmentation

Methodology

- **Two stage document review**
 - 1st Stage – Broad Scoping of material and synthesis from stakeholder literature
 - 2nd Stage – Thematic analysis of a smaller sample of documents
- **Case study research**
 - Case studies within five fields
 - Archaeology
 - Bioengineering
 - Environmental research
 - Health and clinical research
 - Particle Physics and Particle Astrophysics
- **Stakeholder validation workshop**

Document Review – Key Findings

- **Consensus around the benefits of open data**
- **Scientific Values**
 - Self correction, secondary analysis, verification, replication, re-use, avoiding duplication
 - Open flow of knowledge, cumulative knowledge processes/dialogue, equal access to knowledge
- **Value of Data**
 - Data as currency, capital asset and raw material
 - Data as public investment (openness allows for a higher return on investment, re-use allows better value for money)
- **Motivations**
 - 21st Century Science (open, collaborative, interdisciplinary)
 - Meeting obligations
 - Cost effectiveness
- **Barriers**
 - Technological (vast datasets, lack of infrastructure, interoperability)
 - Cultural (competition within science, trust issues, lack of career related rewards)
 - Economic (sustainable business models and infrastructure)
- Overall, a high level discussion of openness and science, more clarity and focused strategy needed for the operationalisation of open data

Case Study 1 – Particle Physics and Particle Astrophysics

- Research ranges from large scale and, collaborative with tens or even hundreds of international partners to smaller scale experimental research
- Large scale research produces large quantities of data (data can be measured in hundreds of petabytes)
- Data from large scale experiments are analysed by a grid, consisting of thousands of interconnected machines around the world.
- Users need in-depth knowledge, specialised equipment, hardware and software in order to work with the data
- Not sufficient to share data only – context, methods, metadata are needed to give the data meaning.
- Particle physics are very competitive – e.g. The search for the Higgs Boson and Dark Matter Physics

Case Study 2 – Health and Clinical Research: The EVA project

- Health research is increasingly interdisciplinary- discipline specific research practices and traditions (e.g. medicine, genetics)
- Various stakeholders involved, e.g. pharmaceutical companies, patient organisations, ICT staff, research institutions, each of which may have a different take on the use and sharing of data.
- Health research benefits greatly from being able to link patient data from many sources, but complications arise when sharing genome and patient data due to legal and privacy issues
- Interoperability issues and work involved making datasets useable for other users from different disciplines. Coding, metadata, methods need to accompany datasets.
- Different levels of openness – data is open, but you need to fill out an application with institutional backing
- IPR issues arising from work with commercial partners, e.g. pharmaceutical industry

Case Study 3 – Bioengineering: Auckland Bioengineering Institute

- Openness and sharing of data and models are very important to bioengineering, as it is in essence an international and collaborative effort toward understanding physiological processes and the diagnosis and treatment of injury or disease.
- Sharing of models is advanced, but not of experimental data, due to competition between scientists and the work required to make the data open to other users
- Data derived from human subjects is subject to ethical and legal issues, problematic for openness
- High level knowledge and specific software is often needed to understand and work with the data
- Lack of common infrastructure to share open data

Case Study 4 – Environmental Research: JRC- Digital Earth and Reference Data Unit (EuroGEOSS and INSPIRE)

- Work with research data (satellite, spatial, numeric) from a variety of European stakeholders (public authorities, environmental research data)
- The aim is to integrate datasets from these different sources, to provide a single open access point to metadata information.
- Different limitations of use are attached with each data set, data can be in different formats and in different languages
- Results in complex licencing system for potential users
- Issues of long term data management – standards for storing and curating data
- How long do we keep data? Older data saved on cheaper storage solutions → access open, but not instant

Case Study 5 – Archaeology: Open Context.

- Data has a strong value within archaeology (documents sites that have since been disrupted by excavation)
- Archaeology is faced with the challenge of organising vast amounts of data gathered by excavations, some of which may date back a few decades. Much of this is not in digitised format.
- Data from any one excavation comes in different formats (spread sheets, relational databases, images: drawings and photos, and narrative text, e.g. excavation diaries)
- meaning of archaeological data is seen to arise from its context, therefore a dataset, where metadata including parameters, terminology, research questions, coding systems and methodologies are not clear, or missing, is likely to be un-usable by others.
- To establish context by linking up this different format, requires substantial work
- Some data may be sensitive and not amenable for open access (human remains, images from religious or cultural sites, location data)

Cross cutting issues for open data

- Opening access is not sufficient; open data needs to be meaningful and understandable; this requires work and funds
- Current climate of competition within science, could be a barrier.
- Reward systems for data work need to be considered
- Some data is sensitive (human and location data) and may need access control
- Sustainable infrastructure is needed to host data, current short term funding models are unable to ensure this
- Data Licencing standards need to be considered
- Peer review mechanism for data to ensure accuracy, validity and reliability