



Researchers' Data Store and Requirements

Daan Broeder

The Language Archive, MPI for Psycholinguistics, Netherlands

EUDAT UF, Barcelona



March 2012



Goals

- Allow a researcher or project to upload data (-sets) to a save location.
 - Sharing (obligatory?)
 - Persistency
 - Elicit proper metadata description
- Should it be a Researcher's YouTube?
 - few metadata ☹
 - However: <http://www.youtube.com/watch?v=ZEIz5HInsmM> ☺
 - Individual resources only, no structure
 - Simple viewers
- Competition with existing and future 'commercial offerings'?
 - Google, MS, ...
 - What business model?
 - Who controls the data?



Requirement Inspiration

- CLARIN has been discussing the desirability of such a ‘deposition service’
- CLARIN NL
 - has created the CLARIN NL Digital Curation service; however too many interesting data sets waiting.
 - considering a CLARIN NL Deposition Service as an extension of CLARIN center offering
 - but CLARIN type of service too laborious for easy & fast use for many researchers



Technical issues to consider

- All uploaded data has to become part of the EUDAT domain
 - **Properly registered: PID, metadata, policies, ...**
 - But ..., no originating center
- research data sets have structure
 - Do we want to accommodate such structure?
 - (discipline specific) structural metadata needed
 - Support varying granularity: data-sets versus individual objects
- Accessibility & Sharing
 - Need sharing but do we need access restrictions?
 - who is in control and how?
- Visualization of data (sets)
 - Look at internal structure
 - YouTube data is easy, research data needs more complicated viewers



Usability & Uptake

- Usability
 - Who is it for? researchers + citizen scientist.
 - Should be easy to use, but we need to get proper metadata.
- Dilemma:
 - Too many questions -> too few depositions
 - Too few questions -> limited data usability
- How to encourage promotion of data from 'easy-store' to proper 'reviewed' repositories



Organizational Issues

- Access
 - Not all data can be completely open: IPR, privacy, ethics
 - We should guide the depositor to make his data maximally open
 - Do we support and promote one or more EULAs?
- Since there is no originating center
 - Who is responsible for the content?
 - Do we check content?
 - Idea to appoint 'volunteer' community centers, but this can still be a central service
- What promise or guarantee for persistency can we give the depositors?