

Building up a Tool Chain to support the Research Data Life Cycle

IASSIST 2017, Lawrence, Kansas (USA)

David Schiller (TBA21 Group), Dr. Ingo Barkow (HTW Chur)

TBA21 Group



- Offices:
 - Nürnberg (concepts and project management)
 - Szeged (programming)
- Software for:
 - assessment, questionnaires, survey management, data documentation, data analysis and data management
- Business plan (in short):
 - assist scientific institutes in running software projects
 - build open source tools
 - reuse software for further enhancement

HTW Chur



- University of Applied Sciences in South Eastern Switzerland
- Three departments: Living Environments, Applied Future Technologies, Entrepreneurial Management
- Swiss Institute for Information Science (SII)
- Team for Research Data Management
- Current projects: Swiss National School Monitoring (ÜGK), Program for the International Student Assessment (PISA), Swiss National Education Server (Edudoc), Multi Mode Big Data Repository (MMRepo)

The GLBPM (Barkow et al., 2013)

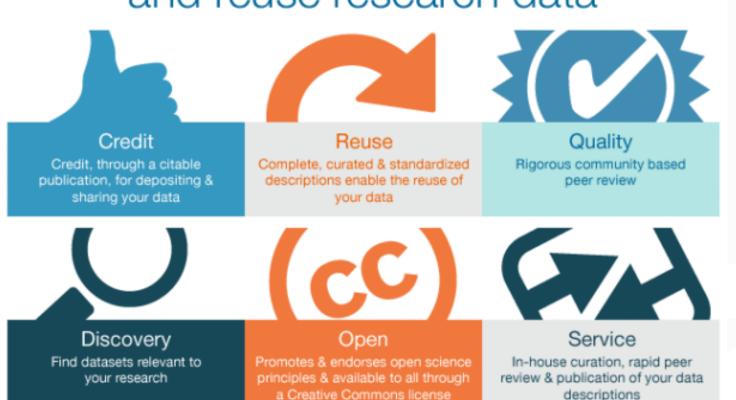


Generic Longitudinal Business Process Model: Overview 2 Archive / Data / Researc Evaluate / **Build / Rebuild** Collect Design / Process / Retrospective Preserve / Dissemination Publish Specify Needs Redesign Evaluation Analyse Curate Discovery 1.1 3.1 9.1 2.1 4.1 5.1 6.1 7.1 8.1 Define research Establish Identify sources Develop data Select sample Integrate data Ingest data & Deploy release Discovery of data questions, universe collection evaluation criteria and relevant mmastructure & high-level nationients research Design sampling 9.2 Set up collection Classify & code concepts 6.2 7.2 methods Gather evaluation Enhance metadata Preserve Create or enhance Access data inputs dissemination 1.2 infrastructure Run collection Explore, validate & 6.3 Design collection products Evaluate existing 8.3 9.3 clean data components Preserve data & instruments data & publications 4.4 7.3 Prepare data Conduct metadata 3.3 Finalize collection 5.4 Deploy access evaluation 1.3 8.4 Validate Impute missing control system / Specify data 6.4 Establish outputs instruments Analyze data data policies elements Ongoing Curation & needed Determine future 5.5 3.4 7.4 infrastructure actions Construct new Test production Promote Specify processing Prepare research variables and units dissemination systems / data cleaning papers Define specific products methods 5.6 concepts to be 3.5 2.6 Calculate weights 7.5 measured Manage disclosure Finalize Provide data Specify analysis risk production 5.7 plan citation support 1.5 systems Calculate 8.7 Planning, timetable aggregates 2.7 7.6 Publish research & needed Organize research Enhance data infrastructure 5.8 team discovery Anonymize data 1.6 7.7 2.8 Prepare proposal Design nage user Finalize data and get funding pport infrastructure outputs Project Management / Quality Management Metadata Management Use of External Standard Metadata (classifications, concepts, questions, variables)

The Data Life Cycle – reuse!







Nature Publishing Group

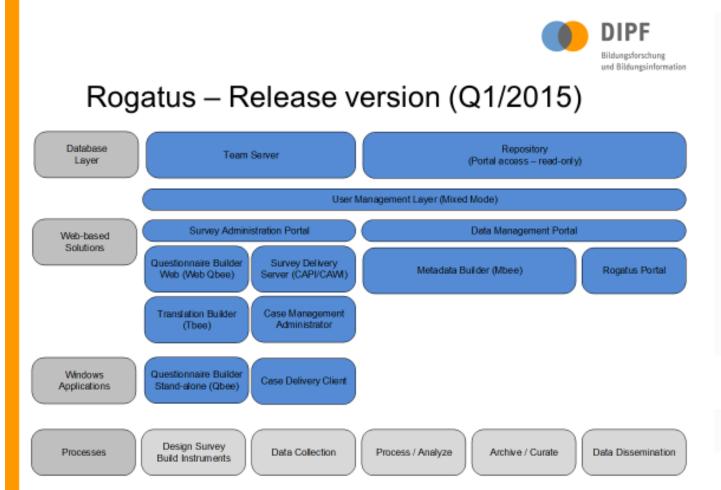
The Tool Chain – why?



- No tools, No standard (session C4, IASSIST 2017)
 - Standardised standards need to be in place
 - Standards need tools to support them and make them work in real life
- No tool chain, No Data Life Cycle (not yet a session)
 - Without chained tools information are not able to flow
 - Resulting in, e.g.:
 - Loss of data and quality of data
 - No, poor (or at least hard to create) data documentation
 - Wasted resources due to complicated data reuse

Tool Chain – first (nice) try





Barkow/Schiller NADDI 2014

Tool Chain - first (nice) try



 Why is there no "we completed the tool chain" talk at IASSIST 2017?

- different needs by clients
- not able to cover the complete tool chain
- lack of funding for project on meta level
- institutional changes (person)
- not feasible as a single organizational endeavour

How to build a tool chain...



- Support for standards and interfaces
- Sustainable long-term funding
- Work and develop in networks! Not institution specific
- Use workshops and conferences for exchange and to create understanding
- Commit yourself and your institution
- Believe in Open Science (because it is the knowledge you can trust)

How to build a tool chain...



Different perspective:

- Funding agencies: long-term funding for networks
- Data providers: coordinate data access with others
- Software developers: open source, working interfaces
- Research institutions: believe in credits for data reuse
- Researchers: cooperate with others and respect IT
- Standards: don't cover 100%, support tools

Summary



Open Science, The Data Life Cycle, Researchers, Institution, Standard Initiatives, Societies, You!

We all need the Tool Chain – lets build it!



David Schiller, schiller@tba21.net, www.tba21.net

Dr. Ingo Barkow, ingo.barkow@htwchur.ch, www.htwchur.ch









