

Towards Robust data Stewardship and Software Sustainability

Outlined ideas about a discipline-based approach

Patrick Aerts

## Present setting

- Data handling during and after project end urgently requires directives;
- There exists no general descriptive framework for the handling of data or the coding of software that sets minimum requirements for exchange, reusability, future access, re-traceability, etc. etc.
- Yet such requirements are increasingly more demanded by the community, funders, governments and the scientific community

## DS+SS: Definitions

- The following definitions:
  - Are primarily practical for the purpose of this discussion
  - Are preliminary and may require improvement
  - Sketch the issues at hand
- Data Stewardship: a careful and sustainability-ready handling of data, directed towards reusability and exchange during and after a project.
- Software Sustainability: coding practices (“ethics”) in support of reusability, verifyability and maintainability of software and the system of availability and maintenance of software.

## Examples of other definitions

### From: Science Europe Working Group on Research Data

- Data Stewardship is the management and oversight of an organization's data assets to provide business users with high quality data that is easily accessible in a consistent manner.  
<http://searchdatamanagement.techtarget.com/definition/data-stewardship>
- Data Management refers to the development, execution and supervision of plans, policies, programs and practices that control, protect, deliver and enhance the value of data and information assets.  
[http://en.wikipedia.org/wiki/Data\\_management](http://en.wikipedia.org/wiki/Data_management)

## DD+DS: coherence

- Data and Software are intimately connected
- Data can not be read, interpreted, handled without the proper software, unless it is printed matter
- Even for reading ascii-code software is required
- -> Software and data need to be treated in a coherent manner, to secure future use, re-use, retraceability, etc.

## DS, SS: differences

- Data Stewardship: data basically need to be kept as is whenever possible;
- Software Sustainability: Software needs to be kept up-to-date to remain useful;
- (Unless it is kept as an image of time: old versions of MS WORD, WORDPerfect, McWrite, Old games for old computers, etc. This is the concern of (national) libraries)

## Generic versus Specialized approach (1)

### Generic approach

- Many approaches towards data stewardship are generic in nature
- Providing data management systems and encouraging their use
- Universities/Institutes made into problem owners -> local solutions
- Strong focus on data like scientific publications, rather than as an abstraction of all objects that individually or collectively contain information or as searchable for information

## Generic versus Specialized approach (2)

### Specialized

- Specialized solutions address each (sub-)discipline
- May get better acceptance/adoption by the community
- May be much more suitable to serve the community needs
- But:
- Require a generalized framework on top, to ensure minimum requirements, such as mutual compatibilities, standards, exchangeability and other requirements not in the direct interest of a specific discipline



## The proposed new approach (step 1)

- A. General Framework
- General minimum requirements to be imposed on protocols to be developed;
- Guidelines for exchangeability and re-useability;
- Guidelines for the use of standards (think RDA);
- Manual on how to set up a protocol, including best practices and models;
- Links to laws and other regulations
- ... .

# PLAN-E Platform of National eScience/Data Research Centers in Europe

## The proposed new approach (step 2)

- Per discipline or subdiscipline have the scientific community define protocols for DS and SS;
- (take into account that the musicology will probably need other protocols than the archeology, both elements of the humanities)
- Set up expert groups per (sub-)discipline and make them responsible for setting up one or more protocols for their (sub-)discipline
- All protocols to adhere to the General Framework
- Publish those protocols as Scientific Publications

## Advantages

- Once the protocols are in place scientists
  - Can refer to these protocols when applying for grants, reviews, publications, etc.
  - The protocols are conceived and expressed in terms and language understandable by the discipline
  - Can add, change and keep the protocols up-to-date using the same route as the original protocol was
- In the medical sciences, scientific protocols are already in place for a long time, to describe their experiments, This may serve as an example

## Next steps

- Procedures as the sketched above can be put in place nationally
- Protocols can be defined without national directions by disciplines using to-become-available frameworks
- Important is that protocols established by disciplines gain the support from larger groups of scientists in that discipline across Europe and beyond
- PLAN-E can start a discussion on this topic, while, for example in The Netherlands, a demonstration will be set up

## Agenda

- Further Introduction by Peter Doorn - Director at DANS (Data Archiving and Networked Services)
- Notes by RDA, prepared by Herman Stehouwer - MPG
- Outlined ideas about a discipline- based approach by Patrick Aerts - Strategic Alliances at NLeSC
- Reflection on the presentations and discussion by all