

*Data Archiving and Networked Services*

**Involving the Science Community  
in  
Data Stewardship and Software Sustainability**  
A Conceptual Approach

Patrick J.C. Aerts

ePLAN

netherlands  
eScience center  
by SURF & NWO

PLAN-E

DANS is an institute of KNAW and NWO

*Driven by data*

# How to involve the scientist?

- Data are produced by the scientists
- Software is created by the scientists
  
- But...
  
- RDM and Software Sustainability seem largely top-down directed
- What is in it for the scientists?
- Why choose for a one size fits all approach?

# Three take home messages

- Treat Software Sustainability and Data Stewardship on equal footing
  - *At least policy wise*
- Consider and treat Data and Software as *value objects*
  - *Then it starts making sense to spend some to keep the value or increase it*
- Make the stakeholder positions explicit, define their role and involve all
  - *Funders, scientists, executive organisations*

# Definitions

- The following definitions are used:
  - ✓ 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, verifiability and maintainability of software and the system for availability and maintenance of software.

# Coherence (of data and software)

- 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
- Ergo: Software and data need to be treated in a coherent manner, to secure future use, re-use, retraceability, etc.

# 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\*\*

\* We are aware of volatile data, websites, on the fly generated sheets, etc.

\*\* Old versions of MS WORD, WORDPerfect, McWrite, Old games for old computers, etc.

# 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 in the form of scientific publications, rather than data as an abstraction of all objects that individually or collectively contain information or are searchable for information

# Generic versus Specialized approach (2)

## *Specialized approach*

- 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

# Consider the following stakeholders categories

1. Governments, Research Organisations, Funding Organisations
2. Science Community, Society, incl. business, etc.
3. Other parties at the executive level (computing centres, data centres, libraries, policy organisations)

# Stakeholder roles and tasks

## Stakeholder

Governments, Funding parties, .
Science, Society, ...
Executive organisations

## Role, interest

• Accountability, (Cultural) Heritage, ...
• “Accelerating Discovery”->Protocols
• (Cost-)efficient, pro-active, goal focussed

## Task

➤ General Framework
➤ Design, publication, execution
➤ Infrastructure, portals, services

# 1. Governments, Research Organisations, Funding Organisations

Their interests are:

- Accountability for spending money in the first place
- Accountability for the way the allocated money is spend
  - giving credits to research output
  - verifiability of results
- Expediency (efficiency) of the money spend
- Act on behalf of society at large
  - Taking care of the social, economical and historical interests (cultural heritage)

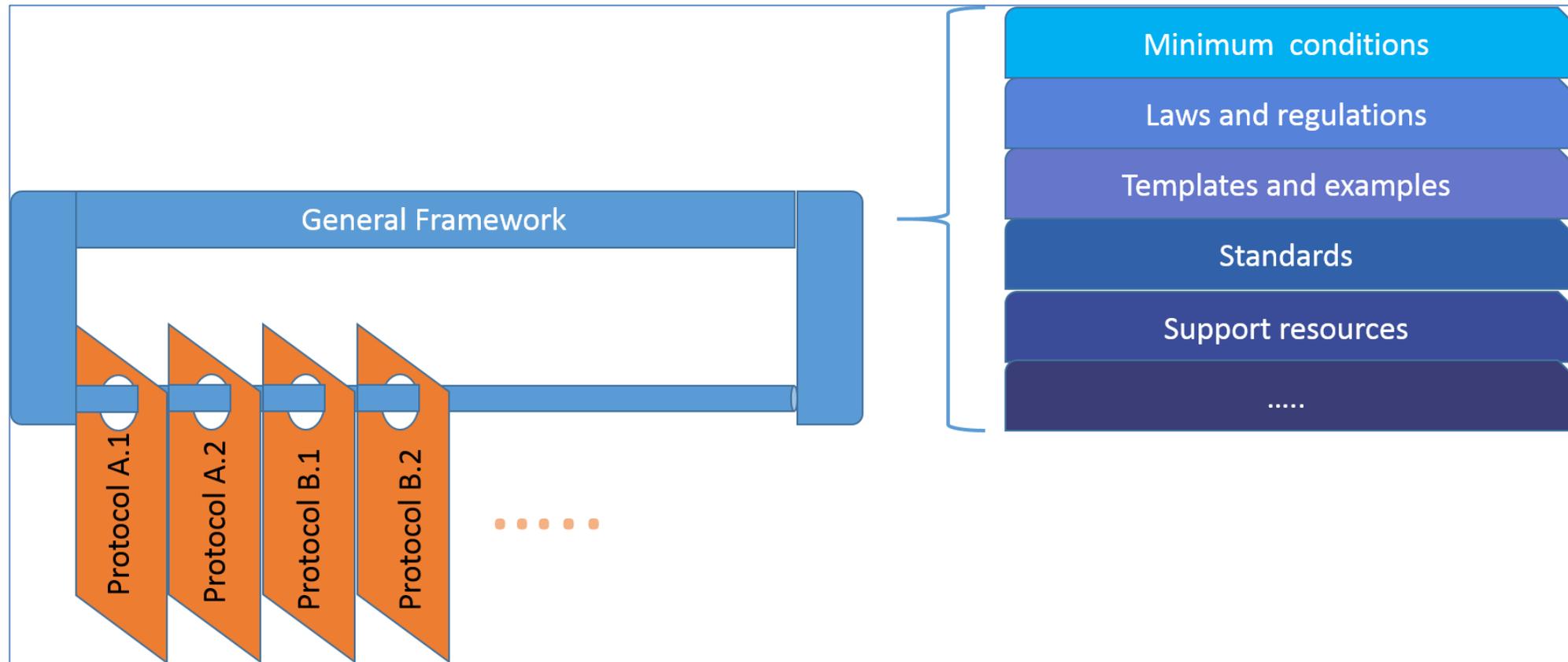
## 2. Science Community, Society, incl. business, etc.

- Contributing to social, individual well-being and prosperity
- Satisfying curiosity
- Accelerating research (results)
- Improving research in depth and scope
- Exchangeability of knowledge and basic data
- Open access , re-usability

### 3. Other parties at the executive level (computing centres, data centres, libraries, policy organisations)

- Providing the best and cost-effective services
- Providing infrastructures, management, planning
- Monitoring developments
- Inform the stakeholders
- Be aware and involved

# Effectively organize scientists' involvement



# The process

## 1. Set a general framework

- Minimum conditions;
- Legislation, national, European, international;
- Standards and references, best practices;
- Templates, check lists, FAIR principles,

## 2. Set up small expert groups in each (sub-)discipline

- Write protocols for SS+DS, as many as required, as few as possible;
- Fitting the requirements of the discipline, in their language, from their interest;

## 3. Publish the protocol(s) in discipline-related journals

- For later reference;
- For the sake of openness;
- Allowing for an international approach from the onset.

# What is in it for scientists?

- Once the protocols are in place scientists:
  1. No longer need to conceive their own data management/software sustainability plans;
  2. Can refer to these protocols when applying for grants, reviews, publications, etc.;
  3. The protocols are conceived and expressed in terms and language understandable by the discipline;
  4. Can add, change and keep the protocols up-to-date using the same route as the original protocol was;
  5. Quoting the protocols add to the credits of the expert group members
- In the medical sciences, scientific protocols are already in place to describe their experiments. This may serve as an example.

# For Software Sustainability:

- Consider setting up a Software Sustainability Initiative in each country;
- Consider forming a Software Sustainability Infrastructure built on these national initiatives
- This would add to the visibility and appreciation of the issue
- This would enable easy sharing of knowledge, insights, best practices

# Contact information

- DANS: [dans.knaw.nl](https://dans.knaw.nl) (Data Archiving and networked Services)
- NLeSC: [esciencecenter.nl](https://esciencecenter.nl) (Netherlands eScience Center)
- ePLAN: [eScience-platform.nl](https://eScience-platform.nl) (Platform of eScience/Data Research Centers in The Netherlands)
- PLAN-E: [Plan-europe.eu](https://Plan-europe.eu) (Platform of eScience Centers in Europe)
- p.aerts@esciencecenter.nl

# End-Of-Presentation

Thank you

# Responsibilities **cat. 1 stakeholders**

- Have a General Framework set up, involving
  - 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/Scenario” on how to set up a protocol
  - including best practices and models;
  - Links to laws and other regulations
  - ... .

# Responsibilities cat. 2 stakeholders

- Per discipline or sub (sub\*-)discipline have the scientific community define protocols for DS and SS:
  - 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
- \*take into account that musicology will probably need other protocols than archeology, both elements of the humanities

# Next steps

- Procedures as the sketched above can be put in place nationally, multilateral internationally, European, globally
  - Protocols can be defined without further national directions by disciplines using the General Framework
  - Important is that protocols established by disciplines gain the support from larger groups of scientists in that discipline across Europe and beyond
- 
- PLAN-E maintains the discussion on this topic, while, for example in The Netherlands, a demonstration will be set up
  - Estimated overall time needed: 3 years after establishing the framework