Solid Scenarios For Software Sustainability

Patrick Aerts
NWO/eScienceCenter/Data Archiving & NS

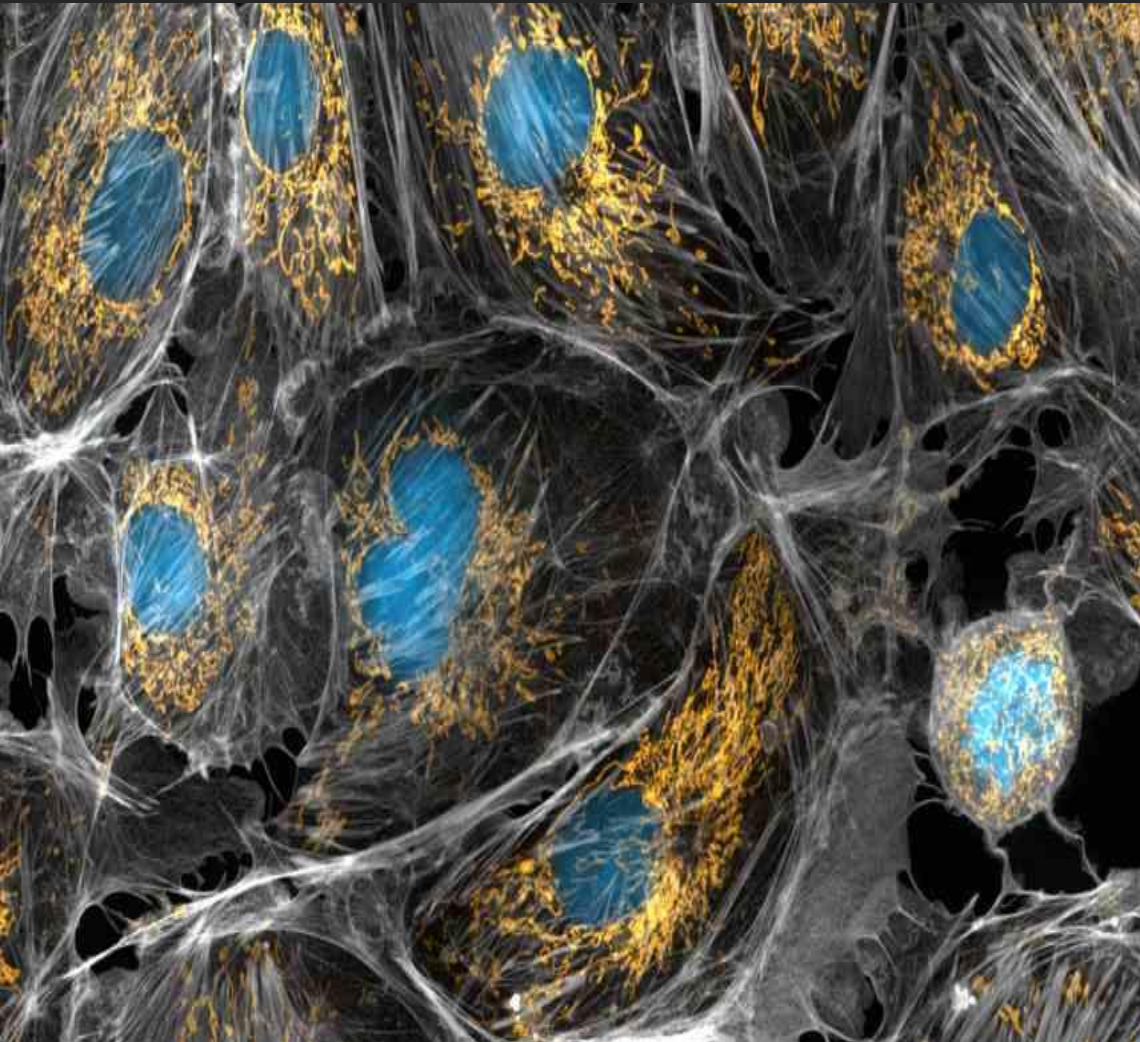# Three take-home messages:
(while you are still with me)

1. Treat Software Sustainability and Data Stewardship on <u>equal footing</u>

   At least policy wise

   Seriously consider linking up with RDA, under a separate chapter

2. Consider Software (and data) as <u>value objects</u>

   Then it starts making sense to spend some to keep the value or increase it

3. Make the <u>stakeholder positions</u> explicit, define their role and involve all

   Funders, scientists, executive organisations

Data Archiving and Networked Services
DANS
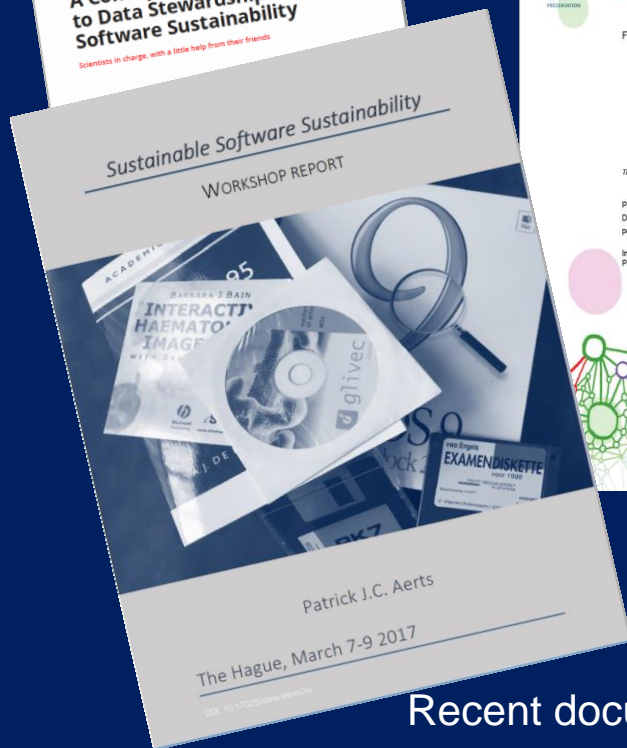
Recent documents on software sustainability

Communities
Concerned with Software Sustainability

- *Science*
- *Archives and Libraries*
- *Sound, Vision, Motion pictures*
- *Arts*
- *Gaming*
  - ➤ *Probably communities are the most important vehicle to keep software products alive*

eScience center

Cross domain issues on software sustainability

- Legislation
  - Contradictory law
  - Copyright violation

- Obsolete/unavailable hardware
  - Inaccessible data
  - Unrecoverable software

- Bit rot/Link rot/Reference rot
  - Just too late...

- How to format reference to software products

http://www.nature.com/news/

the-trouble-with-reference-rot-1.17465

Past

- So much legacy code
- So much to be restored
- So much to put into the Software Heritage Archive
- Selection: what to keep?
- But sometimes restauration is worth while

Present

Education

Advice

Education

Improved coding

Education

Future

- Most codes yet to be compiled
- Rules to keep
- Rules to discard
- FAIR rules for software
- Easy to maintain
- Problem solves itself?
- (And think beyond containers)

Back to the solid scenario

- A solid Framework model and protocols

- A solid software heritage archive

- A solid International Software Sustainability Infrastructure (Under the RDA flag?)

eScience center

There's much more

But it would take another 5-10 minutes…

Sorry

You're missing:

- Who are the stakeholders
- FAIR for Software
- Software Heritage Archive
- Domain protocols
- Harmonisation efforts (among Research Councils)
- Rewarding systems
- Software Seal of Approval

eScience center

The End

(For the Panel discussions that is)

Domain defined SoSu Protocols

Science Europe M.O.'s (and others) to set Data Protocols Framework (Terms of Reference for Domain Protocols)

General Framework

Minimum conditions

Laws and regulations

Templates and examples

Standards

Support resources

.....

Protocol A.1

Protocol A.2

Protocol B.1

Protocol B.2

. . . . .

Domain Data Protocols (DDPs) to be openly published

DANS
Data Archiving and Networked Services
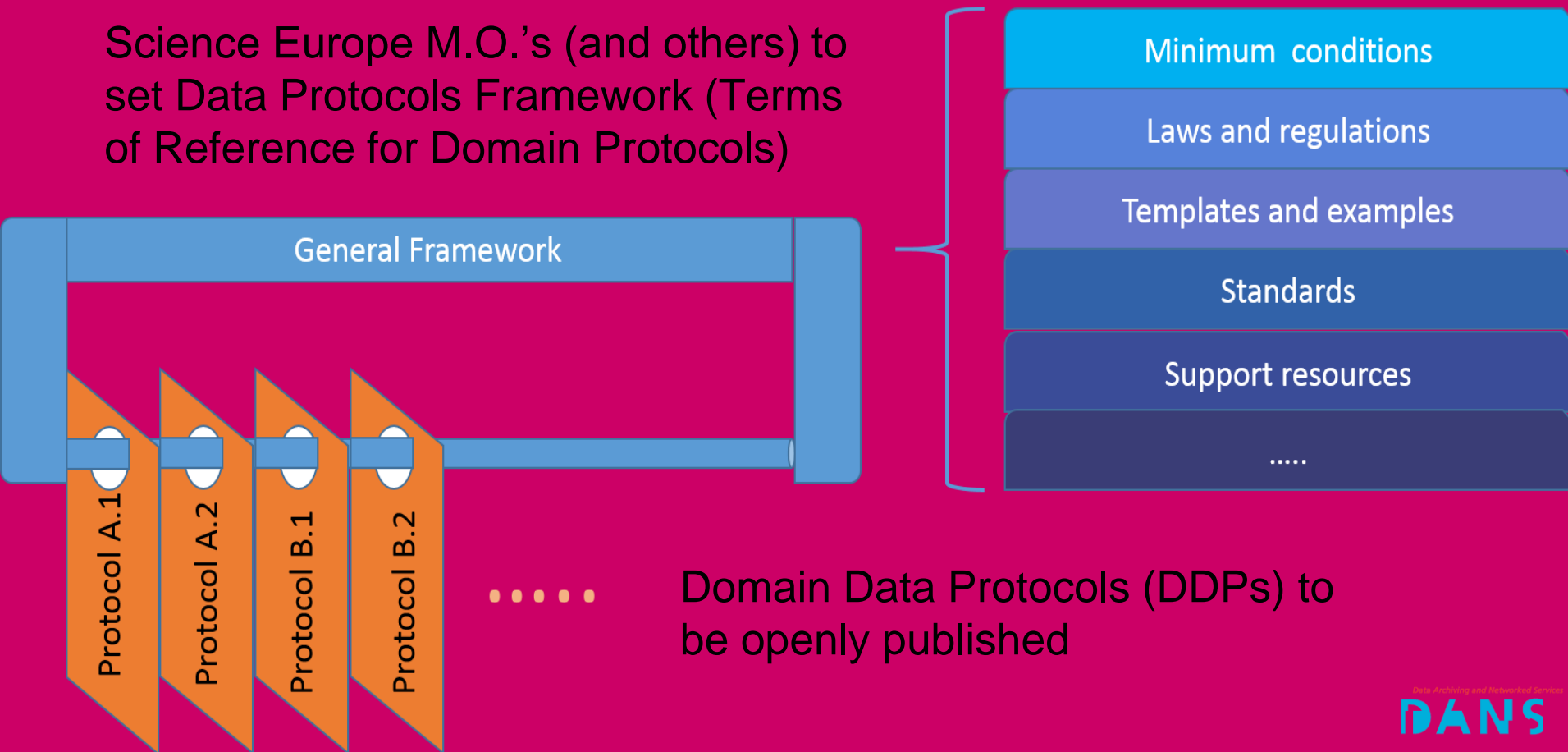
> Let funders set minimum conditions for Research Data Management Protocols and Software Sustainability Protocols (Domain-defined protocols)

> Let domains (disciplines) define their own suitable RDMP's and SoSuP's

Software Sustainability Protocols

- Protocols as such are much used in the medical domain and in archeology

- For software: protocols as yet under-developed

- Rules, best practices and guideliness largely to be established

- SSI (UK) leads the trends

DANS

Software Heritage by *Inria* — inventors for the digital world — 50 YEARS

The Software Heritage archive

Home / The Software Heritage archive

Our long term goal is to **collect** all publicly available software in source code form together with its development history, replicate it massively to ensure its **preservation**, and **share** it with everyone who needs it.

The **Software Heritage archive** is growing over time as we crawl new source code from software projects and development forges. We will incrementally release archive **search and browse** functionalities — as of now you can check whether source code you care about is already present in the archive or not.

Serving the world heritage

Home / Mission / Serving the world heritage

Cultural heritage is the legacy of physical artifacts and intangible attributes of a group or society that are inherited from past generations, maintained in the present and bestowed for the benefit of future generations.

Software in source code form is produced by humans and is understandable by them; as such it is an important part of our heritage that we should not lose. Software is furthermore a key enabler for preserving other parts of our cultural heritage that we would *de facto* lose if we lose the software needed to access them. **Preserving software is essential for preserving our cultural heritage.**

Software is fragile
unlike words carved in stone
it can be deleted or get
corrupted

A major, global effort in software sustainability

Softwareheritage.org

NOKIA Bell Labs   HUAWEI   (intel)   Microsoft   DANS

Rewarding Scientific Research output (other than reviewed papers)

Rewarding scientific reviewed papers is a solid, but a limited form of crediting scientists for their work

Modern forms of scientific output involve

- Data, preferably FAIR and open

- Software, preferably written at the state-of-the-art
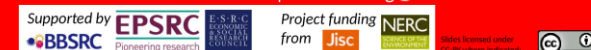
But how to involve these?

- Led by DANS
  - Workshops to define concept and approach
- Who/what should "receive" the seal of approval?
- Definition of key criteria

Software Sustainability Institute

Data Archiving and Networked Services
DANS

Software Sustainability Institute

Welcome!

Sustainable Software Sustainability Workshop

Tuesday, March 7

Wifi password: wifinoi825

Data Archiving and Networked Services
DANS

www.software.ac.uk

Setting up a
Software Seal of Approval

9th March 2017, DANS/SSI Software Sustainability Workshop
Neil Chue Hong (@npch), Software Sustainability Institute
ORCID: 0000-0002-8876-7606 | N.ChueHong@software.ac.uk

- DANS-SSI Workshop Sustainable Software Sustainability discussed the aspects of a **Software Seal of Approval**
- Should we have one?
- On what criteria?
- For what purpose and goals?
- Automatic evaluation?
- Self evaluation?
- Would it help Research Councils to reward software efforts?

SSoA
Concluding:
we are going to work on it!

FAIR Principles

The FAIR Guiding Principles for scientific data management and stewardship

Mark D. Wilkinson, Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, Jan-Willem Boiten, Luiz Bonino da Silva Santos, Philip E Bourne, Jildau Bouwman, Anthony J Brookes, Tim Clark, Mercè Crosas, Ingrid Dillo, Olivier Dumon, Scott Edmunds, Chris T Evelo, Richard Finkers, Alejandra Gonzalez-Beltran, Alasdair J G Gray, Paul Groth, Carole Goble, Jeffrey S. Grethe, Jaap Heringa, Peter A.C. 't Hoen, Rob Hooft, Tobias Kuhn, Ruben Kok, Joost Kok, Scott J. Lusher, Maryann E. Martone, Albert Mons, Abel L. Packer, Bengt Persson, Philippe Rocca-Serra, Marco Roos, Rene van Schaik, Susanna-Assunta Sansone, Erik Schultes, Thierry Sengstag, Ted Slater, George Strawn, Morris A. Swertz, Mark Thompson, Johan van der Lei, Erik van Mulligen, Jan Velterop, Andra Waagmeester, Peter Wittenburg, Katherine Wolstencroft, Jun Zhao, and Barend Mons

Open data is about MORE THAN DISCLOSURE it must be "Fair"

- Findable
- Accessible
- Interoparable
- Reusable

http://www.nature.com/sdata/   nature publishing group

How FAIR are you?

- However complicated to apply –in general-, the concept is broadly praised
- The concept could even be applied to software (with some changes)
- A colorful implementation is available shortly
- Impact analysis should follow
- Then Research Councils will have instruments to reward data and software, objectively

Badges for assessing aspects of data quality and "openness"

OPEN DATA   OPEN MATERIALS   PREREGISTERED

These badges do not define good practice, they certify that a particular practice was followed.

BRONZE: data is openly licensed, available with no restrictions, accessible and legally reusable.

SILVER: satisfies the Bronze requirements, the data is documented in a machine readable format, reliable and offers ongoing support from the publisher via a dedicated communication channel.

GOLD: satisfies the Silver requirements, is published in an open standard machine readable format, has guaranteed regular updates, offers greater support, documentation, and includes a machine readable rights statement.

PLATINUM: satisfies the Gold requirements, has machine readable provenance documentation, uses unique identifiers in the data, the publisher has a communications team offering support. This is an exceptional example of an information infrastructure.

LINKED OPEN DATA
★ On the web OPEN LICENSE
★★ Machine-readable data
★★★ Non-proprietary format
★★★★ RDF standards
★★★★★ Linked RDF
IS YOUR DATA 5 ★?

★	make your stuff available on the Web (whatever format) under an open license[1]
★★	make it available as structured data (e.g., Excel instead of image scan of a table)[2]
★★★	make it available in a non-proprietary open format (e.g., CSV as well as of Excel)[3]
★★★★	use URIs to denote things, so that people can point at your stuff[4]
★★★★★	link your data to other data to provide context[5]

5-star deployment scheme for Open Data

Sources: Open data institute (UK), Centre for open science (US), Tim-Berners Lee

Thanks for your attention

Still remember the take-home-messages?

Patrick.Aerts@dans.knaw.nl
www.dans.knaw.nl

Treat Software Sustainability and Data

Stewardship on equal footing

At least policy wise

Consider Software (and data) as value

objects

Then it starts making sense to spend some to keep

the value or increase it

Make the stakeholder positions explicit,
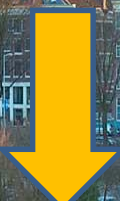
define their role and involve all

Funders, scientists, executive organisations

Stakeholders?

Who   **Why**   **What**

Who	Why	What
Governments, Funding parties, .	Accountability, (Cultural) Heritage, …	General Framework
Science, Society, …	"Accelerating Discovery"->Protocols	Design, publication, execution
Executive organisations	(Cost-)efficient, pro-active, goal focussed	Infrastructure, portals, services