REPORT

PLAN-E Workshop
Incentives and rewarding systems
For open data and sustainabile software
Lugano (CH), 9 September 2016

Ed. Patrick J.C. Aerts

On September 9th 2016, as part of the 5th plenary meeting of PLAN-E, a workshop was dedicated to the topic: Incentives and rewarding systems (by research councils and reviewing committees). The topic was introduced by:

- Patrick Aerts (NL)
- Ana Neves (Pt)
- Sverker Holmgren (SE).

1 Introduction

Background to the topic is, that in these modern times peer reviewed papers can no longer solely represent relevant and qualified scientific research output. However, implementing an extended alternative is easier said than done. The whole circuit around measuring performance, in this case impact analysis of reviewed papers, has a long standing -at least twenty years- advantage over that of sharing scientifically valuable data or academic quality software. Yet it is time to start addressing the issue. As it seems research councils in various countries have been thinking about the matter, but real steps have not been set yet. This workshop may help getting the topic on the agenda's where it belongs.

1.1 Patrick Aerts

From the presentation by Patrick Aerts the following elements are brought forward:

eScience research engineers @ NLeSC may:

- Want to join the industry
- Hop to other escience centers, in NL, EU or beyond
- Want to return to academia.

The first two may not pose a problem, but the last one will: eScience researchers work at academic research level, but their output are not reviewed scientific papers, at least not in the traditional

sense. So they will have an serious gap in their résumé once applying for an academic (research) position.

How to award scientists that have worked for years to compile complex data sets and decide to sharing it all with others, competitors included?

How to award scientists that compile software for future sustainability, meaning according to the state of art for low future maintenance and sustainable usage?

These are essential elements towards an Open Science world. Note that a recent Dutch national extensive survey with over 1000 respondents showed that only 25% of scientists are really prepared to share their data openly up front!

Generally, reviewing takes place by

- Research Councils at granting procedures
- Reviewing Committees at formal visitations of research groups
- Universities to measure its own research groups quality.

And reviewing often has a very formal format, including

- numbers of peer reviewed papers
- impact of the various publications (impact factors)
- other publications (books, chapters, proceedings, posters, public communications, etc.)

So, here something is missing: published open data and qualified software!

Some scientific mores, developed over the past twenty years, are perhaps difficult to change, such as the complex impact measurements of publications, which also form an essential part of Research Councils' accountability measures. But one can start by demanding to give proper reference to data (their collector/compiler) and to software (version, creator, ©-holder) so that in due course also their impact may be measured. Next step would be to allocate *value* (credits) to data sets made public for research and to software that has obtained a Software Seal of Approval (or any similar quality standard).

1.2 Ana Neves

From the presentation by Ana Neves, the following elements are referenced in the body text. She presented an all comprehensive overview of most aspects of data sharing and rewarding and encouragement views. Because any short-cutting through her contribution would do this overview short, that contribution is appended bullet-wise at the end of his report.

Barriers for data sharing are mainly in the domain of "business considerations": can I exploit my data further by not sharing or avoiding tensions between researchers. And there are practical issues, related to privacy and formal regulations regarding confidentiality.

There are quite some elements that can be brought into the discussion to stimulate and motivate data sharing, both by stick and by carrot. Examples are: helping with data quality processes, presenting ways to cover the cost, demonstrating scientific added values and carreer benefits. On the other hand: requiring publicly funded data to be open by principle, setting time constraint on opening data etc.

Funders are recommended to develop explicit policies for data sharing, making RDM part of project proposals, raise awareness, providing guidelines.

The ERAC opinion on Open Research Data adds to the arguments with aspects like education, monitoring, implementing FAIR principles, IPR issues and more.

The FCT strategy involves furthermore an Open Access repository network, stimulating Open Access publishing.

1.3 Sverker Holmgren

The contribution by Sverker Holmgren reflects the state of affairs and position of the Nordic research Councils from his point of view.

The complexities of keeping, publishing and sharing data are of organizational, legal, financial, ethical and technological nature and most of all, the escience researchers perspectives are missing. What is in it for them?

An initial attempt by the Norwegian Research Council to implement the OECD principles for data sharing in 2008 failed, because of the complexity of that problem at that time. In 2013 a renewed attempt started with a survey among 1500 Norwegian researchers. Barriers identified were lack of time and resources, lack of infrastructure and the lack of a proper rewarding system (up to even negative rewards by loss of future possibilities to publish articles). On the other hand the mindset was right: they want to share and generally agree on Open access to data principles.

In Sweden the Swedish Research Council was assigned to develop national guidelines for Open access to research publications and data. However, in the administrative system to store the scientific output there are only two types of publications: reviewed and non-reviewed publications (so no data or software). Instructions for reviewing scientist's merits, however, do allow for more than taking those papers into account, but do not mention open science, sharing of scientific results, leave alone experiences as escience researcher.

NordForsk does have "Ensuring Open Access" as one of its nine lead principles. It has a Nordic eScience Action Plan 2.0 and calls for Open Science efforts. An Open Science Impact Study and an Open Science and Research Award are scheduled for implementation.

2 Discussion

2.1 Issue 1: How did a particular community come to the data culture it now has?

- Some disciplines where open data works well e.g. molecular biology
    - All results stored in public databases
    - There is a culture that supports this, why did this happen?
        - Always comparing to existing corpus
        - Fighting a common enemy (commercialisation threat)
- Other disciplines have gone the other way, e.g. genealogy research where much data is now privately owned
    - Generators of data (the public) preferred to give data to commercial companies in return for easier functionality
- High energy physics
    - High cost of generating data
    - Not generally available however available to researchers in the area, and subsets made available for education
- Satellite data
    - Embargo period on data (1-2 years) then it is made available
        - Question: is this because they want to or have to?
- European Space Agency data
    - Need to preprocess data to make it interpretable by common tools
        - Processed data made available

        - Raw data available on request
      - Funded by European Space Agency
- Examples of where one discipline is required to access data from another discipline
    - E.g. Economics (using e.g. real time / social data)
    - Ecology - bird behaviour includes meteorology data, land use data
    - Transport - also want meterological, urban planning, other data
    - INSPIRE directive applying to environmental data

2.2 Issue 2: General issues/process

- EDUCATION is the first and foremost important contribution to getting where Open Science wants us to be. It is largely a change of mentality that needs to be achieved.
- Bottom up and Top down approaches are both required
- Carrots and sticks are also both required
- The party that pays, determines the conditions
- Let all disciplines and researchers enter Open Science at the same time, so all parties benefit without conflicts
- Use PLAN-E as a collaboration platform to enable members to profit from the good practice of other members (through site visits, etc)
    - E.g. how we work in our different e-Science centres to support researchers, share code, provide guidance
- Research Councils may benefit from attributing credits to data and software, because it helps them in turn to be easier accountable to their funders (usually ministries for science and education) in terms of scientific output.
- The NSF (US) started to use the term *research products* rather than *papers*, which implies that net to papers, also data sets, software, patents and copyrights are considered valuable research products. See https://www.nsf.gov/pubs/policydocs/pappguide/nsf16001/gpg_2.jsp#IIC2f.
- A more serious concern for data by Research Councils helps preventing data wrong doing
- Data infrastructure providers could make it easier to identify useful data, perform analysis
    - Figshare allows previewing of common data set types
    - Cloud providers make it easy to data mine data stored in their infrastructure
- Make it easier to write data papers
- Most publishers should as a rule require data to be deposited with paper before review
- Mutatis mutandis this should apply to software as well
- Research councils should consider funding open access publishing and for data storage
- For crediting software, see http://depsy.org/ (and the Nature article from there)
- For "Transitive Credit" as a Means to Address Social and Technological Concerns Stemming from Citation and Attribution of Digital Products, see http://openresearchsoftware.metajnl.com/articles/10.5334/jors.be/ .

2.3 Issue 3: Proper (data) referencing/Impact measuring

- By starting properly referencing data and software one can start measuring their impact
- Proper referencing data involves mentioning their collector/compiler
    - DOIs for data (through DataCite)
    - RRIDs in lifesciences
- Proper referencing software involves mentioning its version, creator and ©-holder
    - Software Citation guidelines from FORCE11
- Make proper referencing of data and software an academic proper conduct issue
- Start requiring -as a research council- proper referencing of data and software
- Idem for publishers
- Data journals are presently developing. This may help proper referencing
- Limits to the details:
    - People have no hesitation to properly reference, but how far does this go? One uses one another's data and then develops new insights. Is the original author of the data co-responsible for those new insights ?
    - For software the referencing type depends on the type of license.
- See also: http://www.digitalinfrastructures.eu/sites/default/files/Goodman_DI4R_09-30-2016_0.pptx
- Suggestions: to have Web of Science keep track of used data in published paper just like they keep track of funding agencies

- Use LICENSE file in software (where applicable)
- PLAN-E to consider starting:
    - an impact analysis endeavor through its members
    - to try and answer: who to award what (who contributed what to data or software)?
    - Volunteers mail their interest to Patrick Aerts.

2.4 Issue 4: Crediting software/Software Seal of Approval

- Research Councils will need criteria to grant credits to software
- How can data or software quality assessments be implemented (the higher the quality, the higher the reward)?
- Software engineering has a lot of prior work in this area
    - Verification tools
    - Code complexity and coverage analysis
    - ISO/IEC 25000 Software Product Quality
- Is FAIR a fair quality criterion?
    - Yes, but who defines when something is "FAIR enough"?
    - In other words: is there a global interpretation of different levels of FAIR?
    - Is FAIR implemented per community, per project?
    - Particularly for criteria which are more subjective (like interoperability, reusability)
    - Efforts are on their way to implement FAIR as properly as possible
    - Efforts are on their way to implement FAIR to software as well.
- Does a Software Seal of Approval (SSoA) help addressing the problem of rewarding?
    - Most likely it will, directly and indirectly.
    - Directly:
        - An international trustworthy organisation is responsible
        - Because there is an internationally recognized criterion for sustainable software
        - Easy to recognize label and easy to refer to and check
    - Indirectly:
        - User will use an SSoA to communicate a claim to be "better than the rest" which would encourage reward (although the SSoA itself might not do so)
        - SSA should encourage and support best practice as part of getting the seal, which will lead indirectly to other gains
        - Should not be(come) difficult or expensive to get an SSA
    - The SSoA should and will be about *the development process* and producers, rather than about the scientific quality of the code itself (similar to DSA)
    - One should be careful that the adoption of this process does not discourage people from publishing less good code in other words: the SSoA should not be the only lead for Research Councils
    - Software producers in room would put their software through the SSA process
    - Anyone interested in participating in the SSoA endeavor should mail to Patrick Aerts?
    - If it helps in deriving statistics from it, it may help for credits.

2.5 Issue 5: Open Science

- How can Open Science be practically used as a vehicle to stimulate thinking about rewarding open data?
- https://english.eu2016.nl/documents/reports/2016/04/04/amsterdam-call-for-action-on-open-science
    - Some insights about maintaining a software:
        - https://en.wikipedia.org/wiki/The_Cathedral_and_the_Bazaar
    - Lots of free online cloud tools available for open source software projects
    - Make it mandatory / force people to enter the Open Science world
        - UK funders use random sampling to check that data sets have been made available (but not enough effort to check every project)
        - Netherlands can withhold last 20% of grant if reports not completed
        - Could also restrict access to new funding
    - Need to ensure that funders are investing in infrastructure to make it easy to deposit and share data
    - Get people to make use of infrastructure part of their normal workflow (not just "deposit at end of project")
- Awareness at the site of the Open Science promotors (EC, governments, research councils) should also be created that openness of data is rather complicated matter. It's not only about IP versus openness of Privacy versus openness, it is also a technical matter, a cost matter and a philosophical matter. And the privacy topic does not make it easier, because the EC itself is issuing strong laws regarding privacy and security breaches.

3 Conclusions

Definitely work has to be done in this domain. And PLAN-E members can play an active role in it. This particular group is well positioned to know both from the urgent needs of escientists and of the technical matters involved.

There are various initiatives in progress, European-wide and national, but actual measures seem not taken so far. It is a spot-on topic at this moment in time.

Actions are to be taken from different angles:

- Research Councils
- European Commission
- Active research groups
- Librarians
- Data/Software service providers
- E-infrastructure providers
- Publishers[1]
- PLAN-E (or similar organisations), also on behalf of researchers at large

Suggestions done in the course of the discussions on the potential role of PLAN-E are:

- Look at what has already been done, and don't replicate but instead work with existing initiatives where possible
- Come up with a set of guidelines in this area that each member adopts and encourages amongst its own collaborators
- PLAN-E members could come up with information/tutorials for researchers to make it easier for them to share and cite data and software
- PLAN-E could compare how this is working in different countries and different disciplines and write down positive examples of where it does work and publish this
- PLAN-E could be used as a collaboration platform to enable members to profit from the good practice of other members (through site visits, etc)
    - E.g. how we work in our different e-Science centres to support researchers, share code, provide guidance

PLAN-E will discuss practical follow-on actions to be gtaken by PLANn-E or to be en courage by PLAN-E to be taken elsewhere.

This report will be distributed to European Research Councils, either directly or through kits members, depending on individual wishes of members.

4 Appendix: Contribution by Ana Neves, in bullet-form, adapted from her original slides.

Barriers for Data Sharing

- Concerns about adversarial science
- Tensions between researchers and opponents
- Business considerations related to data sharing, including the business value of data;
- Administrative issues
- How will data be made available and who will pay for data sharing?

[1] GigaScience sets a good example here. See http://www.digitalinfrastructures.eu/content/plenary-recordings for an extensive overview of reasons why early publishing is good for the author.

- Privacy and confidentiality issues
- Risks of participants being identified
- Informed consent issues.

Data Sharing complexities

- Huge diversity among disciplinary cultures and practices
- Generational differences in attitudes towards data sharing
- Conflict between researcher's personal benefits of withholding data vs. society's common benefits of data sharing

Everybody suggests that incentives towards data sharing should be provided for but few detail what concrete incentives should be given and how to provide them, *and* in such a way that outweighs researchers' personal benefits in keeping the data to themselves.

Incentives for Data Sharing

- Ensuring quality data sharing practices
- Developing common language and standards
- Planning and time limits for data availability
- Reducing tensions associated with data sharing
- Providing incentives to share data
- Objectively demonstrate (to funders) the benefits of data sharing
- Find ways to compensate (for the costs of data sharing).

Motivations for Data Sharing (intrinsic motivations for researchers)

- when data sharing is an essential part of the research process;
- direct career benefits derived from sharing through greater visibility of one's work, reciprocal data exchanges, and the reassurance of having one's data recognized as valuable by others;
- the norms that researchers are exposed to within their research circle or discipline;
- a framework of funder and publisher expectations, policies, infrastructure and data services as external drivers.

A Knowledge Exchange and Science Europe briefing paper on Funding research data management and related infrastructures yields the following:

- Half of the responding Research Funding Organisations (RFO's) implement measures to ensure that RDM related goals are adhered to by relevant stakeholders.
- Most of these are 'soft', but some are strong measures, where funding allocation is conditioned by the compliance with the RDM policies.
- Fewer than half of the responding RFOs provide incentives for data sharing, and none of these incentives are financial.
- Most of the RFOs stress that responsibility for RDM lies with the researchers and their institutions.

Recommendations for Funders (Project RECODE)

- Develop explicit policies for open access to research data with clear roles and responsibilities
- Adopt a comprehensive approach in funding the implementation of open access to and preservation of research data
- Reinforce the significance of the Data Management Plan to embed and promote data management as a distinct activity within the research process
- Raise awareness and promote open research data in view of leading an open science paradigm
- Foster collaboration with relevant stakeholders and networks.

Project RECODE Guidance on preparing and implementing a policy

- (…)
- Policy content. Policies should be developed with open access as the default.
- Data Management in grant applications. (…) DMPs should form an essential element of the grant proposal, be resourced adequately, be reviewed and their delivery monitored.
- (…) Grant agreement should include clauses on open access to research data, accompanied by (…) sanction mechanisms in cases of non-compliance, as well as clarification of related costs (…) eligible.

- Guidance to researchers. (…) develop appropriate tools such as templates for data management (…) and specify (…) eligible repositories/data centers for data deposit.
- Rewards to researchers. (…) such as the award of prizes for high-quality data (…)
- Policy monitoring mechanisms.

ERAC Opinion on Open Research Data

- Training of stakeholders and awareness raising
- Promote a better understanding of open research data through communication & awareness raising
- Establish training and education programs on Open Science
- Establish a reward system for data sharing activities
- Data dissemination should be considered when evaluating researcher performance
- Take the particular interests of individual researchers participating in a research project into account– including their "right of first use" of the data
- Appropriate author attribution of datasets and data citation standardization will help to fight against scooping and will ensure trust
- Ensure sound monitoring
- Incentives for Data Sharing
- Data quality and management
- Make data identifiable and citable
- Promote metadata standardization and production of metadata
- Promote innovative models for (open) peer-review and processes of quality assurance
- Strongly promote the use of data management plans
- Sustainability and funding
- Ensure the existence of FAIR (findable, accessible, interoperable and reusable) open research data infrastructures
- Ensure sufficient funding for open research data and for data sharing activities
- Legal issues
- Make IPR issues insightful

FCT Open Access to Research Publications Strategy

- To establish an Open Access repository network infrastructure: RCAAP since July 2008
- Mandate Open Access to Publications arising from publicly-funded research: FCT Open Access Policy
- Integrate compliance with the Open Access Policy into project reporting and grant management procedures of FCT-funded research

Implementation measures

Funder mandate for publicly-funded research to be made OA (2013)

- Infrastructural conditions
- Political conditions
- Research Institutions at national level
- European and International environment
- Generalised acceptance of OA principles by the research community

Portugal was deemed to be sufficiently mature for the adoption of a mandate towards availability in OA of research results obtained through public funding . Approved on May 5th 2014

It covers:

- Papers in scientific journals/conference proceedings
- Posters
- Books, book chapters and monographs
- PhD theses
- mandatory immediate deposit in one of the institutional repositories (RCAAP Portal)
- Embargo periods to full publication content allowed.

FCT Open Access policy

OA publishing through an Author-pays business model journal

- APC's eligible for refund as research costs subject to following conditions
- Immediate Open Access no embargo is allowed
- The definitive work must be allowed to be deposited in repositories other than the publisher's own repository
- Creative Commons CC-BY license (or equivalent) mandatory No restriction to access or re-use
- Prevision to set a cap for the APC amount, to be determined in accordance with international Open Access policies best practices

FCT Open Access policy on Research Data

Publications must reference its underlying research data and these should be made available to other researchers whenever requested, provided that all legal requirements are met

FCT Policy on management and sharing of research data – Recommendation

FCT Open Access Policy Implementation Overview

Benefits

Institutions

- Easier access to accurate, complete and updated information
- The same data can be used by the researcher's institution for career progression or other purposes

Funder

- Research Policy objectives
- both stimulate AND monitor compliance with the FCT OA Policy
- facilitate the adoption of OA practices by adding value to the researchers practices of depositing their outputs
- make those practices both easy to perform and a habit, to become as natural as the act of publishing
- Efficiency gains by streamlining the monitoring and research evaluation processes
- From manual procedures to semi-automatic
- Collection of precious data for research policy making that can itself be reused in multiple instances
- collect information on APC's so as to calculate expenditures with OA publications in Gold journals with APC's (new)
- track at an aggregated level the scientific outputs of FCT-funded research (new)
- easily check a projects' output

A Research Council's view – further thoughts

- Easier for a Research Council to use the "stick" than to reward with the "carrot"
- Benefits are already there most of the times but researchers are unware of them and also have strong personal motives to withhold the data
- Acknowledge the "right of first use" of the data
- it may be unrealistic to ask for sooner release (except for some disciplines)
- Rewarding procedures are dependent on the advancements of reliable data producers' identification and citation systems
- Funders can have an active role in stimulating progresses in this field
- Incentives for Data Sharing
- A Research Council's view – further thoughts
- Data Management Plans requirements
- If properly laid out with a view on ensuring maximum dissemination and reuse of data
- if adequately implemented as part of the research project selection process and monitored during the research project evaluation stages, then DMPs are crucial for an accurate mapping of the expected data collection, making it more difficult to "omit" or "hide" data or to keep it unjustifiably closed.

A Research Council can and should further encourage:

- that every publicly funded agency (such as weather and coastal services) gathering non-personal data must make such data openly available online; e.g. weather measurements, physical/chemical characterisation, biological records

- the provision of incentives for the research community to make the raw data underlying publications openly available as a default
- the provision of societal incentives (or penalties otherwise) for private companies to release as much data as possible on their products and methods – consumers should be fully aware of what they are buying
- the provision of computational and storage resources so that citizens and citizens' organisations may share freely reusable information and data
- investment on the universal access to the Internet, progressively, to the entire human population and the full geographical reach