Break Out Session: Cultivating sharing, curating, and preserving data

Henk van den Hoogen
(Maastricht University)

Workshop Working on *'Data Stewardship, Meet your Peers!'*
1 DEC 2017

In collaboration with: Working Group Awareness/Engagement LCRDM

Your take home message

- What are the added values of RDM for a researcher
- How can you, as a data steward, motivate and convince
- Which positive messages do you have
- Which ‘enforcements'

Program for this session

Brainstorm between participants on:

- What do you consider carrots
- What do you consider sticks
- How can you use these in your work as a data steward
- Can these arguments cultivate RDM amongst researchers

Feedback groups and presentation on carrots & sticks

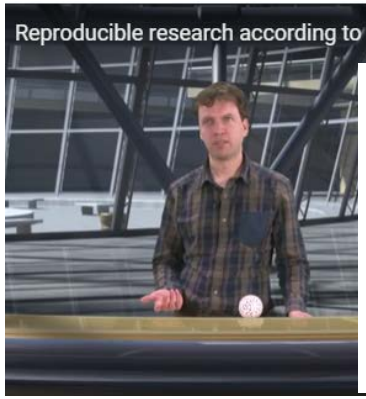
Discussion how to use and check

If time allows: Discussion statements

Carrots: RDM is rewarding

Reproducible research:

making data and programme code available



"other researchers **trust** your results much more";

"reproducible pubs with data and codes receive more **attention**, and **citations**";

"reproducible pubs facilitate more **collaborations**"

http://datasupport.researchdata.nl/start-de-cursus/ii-planfase/reproducible-research/

Carrots: RDM is rewarding

"*Citation Advantage* presently (at the least since 2009) amounts to papers with links to data receiving on the average 50% more citations per paper per year, than the papers without links to data."

Bertil Dorch: On the Citation Advantage of linking to data: Astrophysics, https://hal-hprints.archives-ouvertes.fr/hprints-00714715; 2012, last modification on : Wednesday, December 2, 2015 - 8:39:17 AM

"we find a robust citation benefit from open data, although a smaller one than previously reported. There is a *direct effect of third-party data reuse* that persists for years beyond the time when researchers have published most of the papers reusing their own data. A very *conservative estimate* found that 20% of the datasets deposited between 2003 and 2007 had been reused at least once by third parties."

Piwowar HA, Vision TJ. (2013) Data reuse and the open data citation advantage. PeerJ 1:e175 https://doi.org/10.7717/peerj.175

"Showing that articles with links to data get higher citation rates might increase the willingness of scientists to take the extra steps of linking data sources to their publications. In this presentation we will show this is indeed the case: articles with links to data result in higher citation rates than articles without such link

Edwin A. Henneken, Alberto Accomazzi (2011) Linking to Data – Effect on Citation Rates in Astronomy. http://arxiv.org/abs/1111.3618

Twin and family research

VU VRIJE UNIVERSITEIT AMSTERDAM Faculteit der Gedrags- en Bewegings-wetenschappen

JNNTR NIEUWS INFORMATIE ONDERZOEK PUBLICATIES VOOR STUDENTEN CONTACT

erzoek > Samenwerkingsverbanden REGISTER CHANGE OF ADDRESS

banden

Nationale en internationale samenwerking

Het NTR neemt via verschillende consortia deel aan grote onderzoeken waarbinnen wordt samengewerkt met andere tweelingenregisters, met andere onderzoekers die ook grote groepen kinderen volgen in hun ontwikkeling en met consortia waarbinnen gezocht wordt naar genetische varianten voor complexe aandoeningen en eigenschappen.

Binnen Nederland werkt het NTR onder meer samen met collega's binnen onderstaande groepen:

Twin VUmc (tijdelijk offline)
Amsterdam Public Health/EMGO+
Amsterdam Reproduction & Development (R&D)
Amsterdam Neuroscience
Twin 60++
BBMRI –NL (Biobanking and Biomolecular Resources Research Infrastructure)
BBMRI–EU (Biobanking and Biomolecular Resources Research Infrastructure)
GoNL (Genome of the Netherlands)
CID (Consortium Individual Development)
NBIC (Netherlands Bioinformatics Centre)

Samenwerkingen met andere tweelingregisters:

Internationale vereniging van artsen en onderzoekers
GenomEUtwin (collaboration European and Australian Twin Registers)
Avera (Institute for Human Genetics) & **Avera Twin Register**
Twinning Consortium

Internationale samenwerking in onderzoek naar kinderen en hun ontwikkeling:

ACTION Project
ENGAGE (European Network for Genetic and Genomic Epidemiology)
EAGLE (EArly Genetics and Lifecourse Epidemiology)
EGG (Early Growth Genetics)
PGC (Psychiatric Genomics Consortium)
CHARGE (Cohorts for Heart and Aging Research in Genomic Epidemiology)
GIANT (Genetic Investigation of ANthropometric Traits)
MAGIC (Meta-Analyses of Glucose and Insulin-related traits Consortium)
ENIGMA (Enhancing NeuroImaging Genetics through Meta-Analysis)
SSGAC (Social Science Genetics Association consortium)
IHGC (International Headache Genetics Consortium)
CARTA (Causal Analysis Research in Tobacco and Alcohol)
TAG (Tobacco and Genetics Consortium)
P3G (Public Population Project in Genomics and Society)

aring with
rticipants
& society

Sticks: RDM is a must

Code of Conduct for scientific practice:

vsnu
vereniging van universiteiten
association of universities
THE NETHERLANDS

Art 3.1 Replicability of research must be possible
Art 3.2 Quality of data collection, storage, manipulation
Art 3.3 Period of storage is min. 10 years
Art 3.4 Raw data are archived in such a way that they can be accessed at all times at minimal expense and effort

Other articles can also be used as RDM arguments

Sticks: RDM is a must

Institutional policies:

Utrecht, Groningen, UvA, VU: Research data ‘Open unless’
Nijmegen, Wageningen: RDM is a must (regulations for storage and archiving), control over data
Maastricht, Tilburg : code of conduct for RDM (verification)

=> Each University is working on RDM policies (control, archiving obligation, RDMP obligation etc.).

Sticks: RDM is a must

Funders requirements

website links ->	NWO	KNAW	H2020	FP7	ZonMW
RDM mandatory?	mandatory	mandatory	encouraged, opt-out possible	encouraged	mandatory
DMP requested?	after funding	after funding	after funding	not mandatory	after funding
Template available?	yes	no	yes	yes	yes
Centered around FAIR?	yes	yes	yes	yes	yes
Which should be made available?	relevant for reuse	all research data	validation for publications	as much as possible	all research data
Open access of data?	immediate open access, limited only when necessary				

Version 1.0 by Rosemary van Kempen & Jasper van Dijck

How to engage

- Lessons learned?

good practices, examples, ideas

- What is additionally needed?

support from others; embedding experiences

- First steps @ home?
plan of action

Discussion statement

Reliability of research is an incentive, but it is not a carrot for the individual researcher.

Reliability of research is mainly of importance for the university or research institute

It is not an argument that researchers will be sensitive for

Discussion statement

Credits for RDM should be made part of the regular research evaluation of researchers and research groups

RDM should be included in SEP
A RDM h-index should come to existence

Discussion statement

Researcher do not want to publish their data,
Since they are afraid that others will do nicer things with the data

"The Best Thing to do With (Your) Data will be Thought of by Someone Else"

Dr. Rufus Pollock, Co-Founder and
Director of the Open Knowledge Foundation