Research Data Management @ UGent

U4 Meeting: Research Data Lifecycle

Dries Moreels & Myriam Mertens, Ghent University Library
3 February, 2016
Overview

1. The beginnings
2. Where are we now?
3. Future plans
4. Challenges
The beginnings
Research Integrity as a starting point
Research Integrity


to prevent scientific misconduct & promote good research practices

**Research Integrity Policy Plan** (March 2015)

- identifies RDM as a way of enhancing the quality and credibility of research
- university committed to support faculties and researchers via training, tools, infrastructure
- faculties and researchers can develop own initiatives
Resulted in many RDM actions

- **pilot** project & faculty guidelines in Psychology & Educational Sciences
- mapping/awareness project in Arts & Humanities
- setup RDM Working Group (coordination)
- preparing university-wide RDM policy
- publish informational webpages (basic information, guidelines)
- raise awareness of good data collection & processing practices in libraries
- advertise support for data preservation & sharing by the library
- exploration of useful tools, e.g. for RDM planning
- encourage uploading of data alongside publications in institutional repository
Where are we now?
RDM Working Group (since end 2014)

- develop policy & strategy
- oversee implementations & pilots
- UGent Research Director invites:
  - ICT department
  - information security & data protection officer
  - research coordination office
  - technology transfer office
  - university library
  - faculty representatives
FACILITATING OPEN KNOWLEDGE CREATION
New role for University Libraries?

- it's in our **mission**: facilitating open knowledge creation

- tasked to preserve & unlock UGent scientific output, including research data
- operating the UGent Academic Repository
- commitment to open access & open science

  Immediate Deposit/OA Optional Mandate (since 2003)

- new: research data officer since June 2015
Research Data Management (RDM)

Proper handling of research data is an essential part of the research process, organisation and collaboration, but also from the viewpoint of scientific integrity (verifiability and reuse). Background information about the why of data management.

What is research data management?

All actions that result in research data that are safe, findable, accessible and should relate to the planning, collecting, organising, documenting, storing, preserving, and archiving of research data.

Data management tools

Since January, 1st of 2015 the faculty supports its researchers with regard to data management plan and second, there is support for documenting the research information. Follow the links below:

- Data Management Plan
- Data Storage Fact Sheets

Questions?

For all your questions with regard to research data management please contact the research office.

More information can also be found in the slides presented during the information sessions:

- Slides info session FPPW pilot
- Slides workshop Research Data Management 2015
Pilot by Faculty of Psychology & Educational Sciences

- Working Group Data Deposit (2013)
- Note to Faculty Council (2014)
  - collected research data should in principle be accessible for consultation by others (cf. APA guidelines)
  - research data should be preserved in a way that allows reproducing reported results & reuse
- DMP requirement for doctoral students since January 2015
- RDM practices included in evaluation criteria of (tenured) academic staff
- Faculty RDM coordinator provides researcher support
Digitisation and the internet have created a new dynamic within humanities research that concerns about reliability, scalability and reusability of digital data. Research plans in which they outline how their data will be stored, focusing on durability and philosophy have enough tools to support these developments.

Ghent University is currently looking into these issues on several levels. This interest from humanities scholars within the faculty.


A presentation on the importance of research data management and the role of Commissionis on April 24, 2014, was the start of this project. The faculty library and the Humanities Research Data project. In this project, an overview of all available information on the state of (digital) humanities research projects.

The project started in **September 2014** and ends in **April 2015**, with financial support from the University.

A survey will try to map out both the types of data that currently exist at the University. The survey was set up using international guidelines, the DANS project, and the Dutch DANS-project, which specialises in digital humanities.

The survey outcomes and the resulting report will form the basis for further research in collaboration with the University Research department, other faculties, the National Archives and the University.

**Project results will be presented to the Scientific Research Commission at the...**
Faculty of Arts & Humanities Research Data Project (2014-2015)

- faculty library initiative (digital humanities research support)
- chart digital data assets & current RDM practices
- via survey & in-depth interviews
- report to guide development of faculty policy
biblio.ugent.be

- academic bibliography & institutional repository of Ghent University
  - metadata of all publications (academic CV)
  - upload of fulltext (mandatory since 2007)
- new: data uploads to complement fulltext
  - max. 200MB per upload
  - open access or limited access (typically "UGent only")
  - embargo possible
  - no other requirements, deposit procedures
Upload document

- File*

  - Upload file:
    - Bladeren... Geen bestand geselecteerd.

  - Upload from internet location:

    Filename for new file:

- Kind of file*

  dataset

- Access Level*

  - Open access (the file is freely available, effective immediately)
  - Only in UGent Network
  - Only Author/Reviewer/Administrator
    - Switch automatically to open access on this day (YYYY-MM-DD):

  Submit  Cancel
Determinants of acceptance and subsequent uptake of the HPV vaccine in a cohort in Eldoret, Kenya


**abstract**
The development of Human Papillomavirus (HPV) vaccines provides new opportunities in the fight against cervical cancer. Many acceptability studies have revealed a high interest in these vaccines, but acceptance is only a precursor of behavior, and many factors, at the personal, community and provider level, may inhibit the translation of willingness to vaccinate into actual uptake. Through a longitudinal study in Eldoret, Kenya, HPV vaccine acceptability was measured before the implementation of a vaccination program (n=287) and the actual vaccine uptake, as reported by mothers, once the program was finished (n=256). In between baseline and follow-up, a pilot HPV vaccination program was rolled out via the GARDASIL Access Program, in which parents could have their daughter vaccinated free of cost at the referral hospital. The program was promoted at local schools: health staff informed teachers who were then asked to inform students and parents regarding cervical cancer prevention and the vaccination program. Even though baseline acceptance was very high (88.1%), only 31.1% of the women reported at follow-up that their daughter had been vaccinated. The vaccine was declined by 17.7%, while another 51.2% had wanted the vaccination but were obstructed by practical barriers such as a lack of information or time constraints. Being
- **online** RDM planning tool
- based on open source software by DCC
- available for Ghent University users since November 2015
- pilot for a fully localised version for Belgian research institutions
My plan (Horizon 2020 DMP)

6-month DMP

Data set reference and name (1 question, 0 answered)

Data set description (1 question, 0 answered)

Standards and metadata (1 question, 0 answered)

What standards will be used?

Guidance

UGent guidance on Data Format

UGent guidance on Data Capture Methods

UGent guidance on Documentation

UGent guidance on Metadata

Save

Not answered yet

Data sharing (1 question, 0 answered)

Archiving and preservation (including storage and backup) (1 question, 0 answered)
RDM training

- DMPPonline.be training sessions
  - in pilot faculties (December 2015)
  - others following: Economics, Bio engineering

- start of general training for PhD students via Doctoral Schools (February & March 2016)

- doubles as requirements gathering for further RDM infrastructure
  - 1 on 1 data interviews
  - based on DAF methodological framework
UGent RDM policy development

- started in 2013 within framework of Research Integrity
- various initiatives at central & faculty level
- growing involvement of university libraries

Growing awareness that RDM is more than integrity issue:

- also focus on benefits of data reuse and open science
- integrate issues of digital preservation
- consequences for ICT infrastructure
Key principles of UGent RDM policy

- RDM is an essential component of responsible research conduct
- research data are a valuable form of scientific output that should be made openly available
- (temporarily) restricting access may be necessary under certain circumstances

RDM is a shared responsibility

- university-wide services & infrastructure
e.g. mechanisms for storing, archiving, sharing data, general support & training
- faculties: domain-specific guidelines, procedures & support
- individual researchers: adopt RDM standards
Figure from S. Jones, G. Pryor & A. Whyte (2013), "How to Develop Research Data Management Services - a guide for HEIs". DCC How-to Guides. Edinburgh: Digital Curation Centre, p. 5.
Future Plans
Detailed institutional roadmap

- consolidate existing expertise
- complete requirements analysis
- establish systematic RDM policy framework
- coordinate actions accordingly
  - incremental but coherent approach
  - clarifying roles
  - planning for staff and material cost
Development of services

- roll out & further develop DMPonline.be
- expand RDM webpages & general training
- stimulate sound management of active data with online platforms for storage, processing, sharing of dynamic data, with version control & access rights management
  (e.g. cloud storage, github...)
WE FOSTER THE OPENNESS, INTEGRITY, AND REPRODUCIBILITY OF SCIENTIFIC RESEARCH

COS is a non-profit technology company providing free and open research tools to increase inclusivity and transparency of research. COS supports institutional incentives and practices to align more closely with scientific values.

osf.io
Everything your research needs to be a success

Manage your project
View all of your projects from one dashboard.

Quickly share files
Share key project information and allow others to use and cite it.

See project changes
See the latest project changes, who is contributing and historical file versions.

View project analytics
Access project data ranging from visits over time to top referring websites.

Archive your data
Computer or collaborator explode? You will never lose your projects.

Control access and collaboration
Add others to your projects to collaborate or control private access to view.

Supercharge your research
The OSF helps individuals, teams and institutions supercharge their research processes and outcomes.

Registration
Preserve the state of a project at any point in its lifecycle such as the onset of publication or any change in collaborating team members.
FAIR Data Principles

Preamble

One of the grand challenges of data-intensive science is to facilitate knowledge discovery by assisting humans and machines in their discovery of, access to, integration of, and analysis of task-appropriate scientific data and their associated algorithms and workflows. We describe FAIR - a set of guiding principles to make data Findable, Accessible, Interoperable, and Re-usable.

To be Findable:

F1. (meta)data are assigned a globally unique and eternally persistent identifier.
F2. data are described with rich metadata.
F3. (meta)data are registered or indexed in a searchable resource.
F4. metadata specify the data identifier.

To be Accessible:

A1. (meta)data are retrievable by their identifier using a standardized protocol.
A1.1 the protocol is open, free, and universally implementable.
A1.2 the protocol allows for an authentication and authorization process if necessary.
A2 metadata are accessible, even when the data are no longer available.
Adapting Biblio: more data-centric approach

- searchable register of institutional research data production (metadata)
- ensure access to the "long tail" of datasets & accompanying documentation
- link to related publications persistent identifiers for datasets

evolve into a full research data catalogue & repository
Introducing LibreCat

Our goal is to provide the open source set of programming components to build up digital libraries and research your local needs.

Unlock Research
As science evolves towards the principles of openness so digital library and research services will. LibreCat provides powerful tools to collect, distribute and present scientific output that matters.

Agile Development
Let's put agility on the menu and meet demands on time. We work to unify library workflows in accordance with research and administrative requirements. For this aim, we build on strong solutions.

Interope
We avoid multiple input respects digital library start with the fast growing lax research services to make..

librecat.org
Challenges
1. Specific national context
   - no funder requirements (yet)
   - no obvious centre of expertise

2. Specific institutional context
   - decentralised organisation model
   - new partnerships required for service delivery
   - challenging budgetary climate
3. Complexity of research data

- legal status not straightforward
- what about analogue data?
- what about data requiring restricted access regimes?

4. Great variety of research cultures

- long tradition of data sharing in some fields; hostility in other fields
- balancing domain-independent and domain-specific support
5. "Forever Cost" of Digital Preservation

- acceptable file formats?
- different versions of datasets?
- access regimes & data licensing?
- appropriate metadata?
- deposit procedures (e.g. copyright statements)?
- submission reviews (e.g. metadata accuracy, formats)?
- data validity?
- interoperability with research equipment such as GitHub, Taverna...?
6. Evolving role of the university libraries

• RDM as a "natural" extension of existing interests & activities
  
  e.g. preserving & sharing scientific output, open access movement

• library involvement in RDM also has "natural" limits
  
  e.g. good RDM practices change the research internal processes

• RDM presents opportunities & challenges to library staff

• new partnerships with researchers