
Jennifer Clark
Junior Data Scientist
Benefitfocus

Margaret Henderson
Director for Research Data Management
Virginia Commonwealth University

Jeroen Rombouts
Managing Director
3TU.Datacentrum

Library Connect
Webinar
Jennifer Clark
Junior Data Scientist
Benefitfocus
@jengerful

Girl, Interrupted:
From Librarian to Data Wrangler
The Science of Data

“an emerging area of work concerned with the collection, preparation, analysis, visualization, management, and preservation of large collections of information.”

- Jeffrey Stanton; An Introduction to Data Science, 2013
Data Curation == Data Governance?

Let’s not reinvent the wheel.

{Catching Up to Corporate: A Shift Towards Academic Data Governance}

DAMA International’s Data Governance Wheel
Data Curation != Data Science

Digital Curation Centre’s Curation Lifecycle Model
Evidence-based Discovery

<table>
<thead>
<tr>
<th>ID</th>
<th>HIRES_FAST</th>
<th>PREV_EXPERIENCE</th>
<th>TEAM_ENV</th>
<th>EXPERTISE</th>
<th>LONG_TERM</th>
<th>PERSONALITY_MATCH</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Corporate</td>
<td>1</td>
<td>1</td>
<td>Generalist</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>Academic</td>
<td>0</td>
<td>0</td>
<td>Specialist</td>
<td>NA</td>
<td>NA</td>
</tr>
</tbody>
</table>

Created with R - [http://www.r-project.org/](http://www.r-project.org/)
Going Rogue

Created with Processing - http://www.processing.org/
Data Science Lifecycle

Find it (Collection)

Show it (Visualization)

Wrangle it (Preparation)

Model it (Analysis)
Getting the Data Ready

Data Science Tasks

- Find it (Collection)
- Wrangle it (Preparation)

Librarian Skills

- Reference
- Attention to Detail
- Evaluation of Resources
- Coding Classes (Design of Databases and Python)
- Information Modeling (First order logic, Relational Algebra)
- Document Modeling (XML)
- Data Curation
Using the Data

Data Science Tasks

- Model it (Analysis)
- Show it (Visualization)

Librarian Skills

- Informetrics Class (WEKA, R)
- Coding Classes
- "Librarian Ethics" (ALA Bill of Rights)
- Communication of Information
Resources

• Coursera
  – Specialization in Data Science [https://www.coursera.org/specialization/jhudatascience/1?utm_medium=courseDescripLabel]
  – Calculus 1 [https://www.coursera.org/learn/calculus1]
  – Introduction to Data Science [https://www.coursera.org/course/datasci]
• Code Academy
• R
  – Jeffrey Stanton’s Introduction to Data Science [http://jsresearch.net/]
• O’Reilly Publishing [http://www.oreilly.com/]

Webinar
Library Connect
Thank You!

“A lot of people in our industry haven’t had very diverse experiences. So they don’t have enough dots to connect, and they end up with very linear solutions without a broad perspective on the problem. The broader one’s understanding of the human experience, the better design we will have.” - Steve Jobs

@jengerful     jenfanguy AT gmail DOT com
Margaret Henderson
Director, Research Data Management
VCU Libraries
Virginia Commonwealth University

Transitioning from Reference to Research Data: One Librarian’s Journey
“Our job is to advance discovery and support our faculty and their work, and insure student success. ... It is important to underscore the active role librarians play in advancing research and enabling success. And if we’re not, our institutions will not get the best out of their investment with us and there is the possibility that somebody else will come in and charge them more to do the same thing.”

-John Ulmschneider, University Librarian, VCU
http://animaldiversity.ummz.umich.edu/site/resources/Thomas_Belsk/clethrionomys_rutilus.jpg/view.html

Courtesy of http://www.anatomy.vcu.edu/microscopy/index.html
Second-mover advantages in the strategic adoption of new technology under uncertainty, Heidrun C. Hoppe
A Plan

1. Create a web presence and promote your service.
2. Conduct an environmental scan of VCU for data and data management resources that are available for researchers of all levels at the university.
3.

"Low-Hanging Fruit" by mookitty  [http://www.flickr.com/photos/mookitty/2375679549/](http://www.flickr.com/photos/mookitty/2375679549/)
Regulations and DMPs
4. Talk to researchers, students, and others involved with data.

https://creativecommons.org/ "Michael Carroll, Sarah Hinchliff Pearson and Diane Peters" / Joi / CC BY
5. Educate everyone.

National Archives [http://research.archives.gov/description/285702](http://research.archives.gov/description/285702)
What else?

altmetrics

FORCE11
The Future of Research Communications and e-Scholarship

DataONE
Data Observation Network for Earth

DRYAD

figshare

ORCID
Connecting Research and Researchers

Webinar
Library Connect

#LCwebinar
Margaret Henderson
Why Libraries?

1. Experience in Change Management
2. A Strong Set of Campus Relationships
3. A Physical Presence

Why Librarians?

Library and information professionals:
• need to become more involved with semantic web or users will reinvent wheel (i.e., ontologies)
• have the interpersonal and subject specialization for reference/consultation that IT doesn't have
• continue to help users find the information they need.

Select Data Education Sites

- DataONE [http://www.dataone.org/best-practices](http://www.dataone.org/best-practices)
- MANTRA [http://datalib.edina.ac.uk/mantra/](http://datalib.edina.ac.uk/mantra/)
- New England Collaborative Data Management Curriculum (NECDMC) [http://library.umassmed.edu/necdmc/index](http://library.umassmed.edu/necdmc/index)
- RDMRose [http://rdmrose.group.shef.ac.uk/](http://rdmrose.group.shef.ac.uk/)
- UK Data Archive [http://www.data-archive.ac.uk/home](http://www.data-archive.ac.uk/home)
- DCC [http://www.dcc.ac.uk/](http://www.dcc.ac.uk/)
  - How to Develop RDM Services – a guide for HEIs
Further Reading

• **Data Curation Profiles Directory** for completed profiles.
• **Data Curation Profiles Toolkit** to download guides to conduct data interviews.
• **E-Science Portal for New England Librarians: a librarian’s link to e-Science resources** blog and links are very useful and updated regularly.
• **Journal of eScience Librarianship** – data case studies and more
"There is no need for research libraries to start with all recommendations or to try to deliver a full spectrum of data services at once. Small steps will do."

LIBER working group on E-science. Ten recommendations for libraries to get started with research data management; 2012.
Jeroen Rombouts
Managing Director
3TU.Datacentrum

Lessons learned from developing 3TU.Datacentrum research data facility: staffing and more…
3TU.Datacentrum = ...

- Research data facility run by (currently) 3 Universities of Technology from the Netherlands.
- Offering products & services, before, during and after research, to make and keep valuable data accessible, discoverable and usable. Derived goals: provide tools & services and (re)skill researchers and support staff.

- TUD: 18.900 Students, 2.500 Research staff, 1.800 Auxiliary staff
  TU/e: 8.200 Students, 2.000 Research staff, 1.000 Auxiliary staff
  UT: 10.000 Students, 1.700 Research staff, 1.100 Auxiliary staff
- 3TU.DC = Approx. 15 fte in 3 locations (excl. ICT department, ...)
- Back-office based at TU Delft Library, front-offices at all partner institutions.
3TU.Datacentrum services

Data-labs

Collaboration platforms for research data (management) to enable exchange of data and other research material for collaboration and e.g. early review. ➔ improve standardization & documentation and lower archiving threshold.

Data-archive

Multi disciplinary, multi institutional data archive to ‘freeze’ research data and data descriptions for future use (with Data Seal of Approval and DataCite DOI’s)
➔ improve long-term accessibility, usability and discoverability.

Data-services

Training, dissemination and support on data management topics.
➔ improve data management and data-sharing planning and practice.

Data-R&D

Procedures, licensing, business models, training and technology for RDM.
➔ adopt, adopt, adopt & develop best practices.
A short history...

• 2007: Project initiated by libraries from 3 Universities of Technology shifted from aiming at Centres of Excellence to ready, able and willing research groups

• 2009: Lead time for project extended (within budget) difficult to get staff, data-acquisition = slow processes, founding DataCite

• > 2011: Project end transition to secure funding, setup shop

• 2013: Consortium agreement signed by 3 Universities Support from boards (expand to national, increase use), Research Data Netherlands founded, ...

• Recent highlights: National CODATA member with DANS, Research Data Netherlands expanded with SURFsara (Dutch HPC Center), co-organizer RDA 4th plenary, Dutch Data Prize, LC Webinar ;-)
Staffing – context & skills

**Library role** for research data?
responsible for university output and research input (a.o. data!)

Libraries changing
from subject librarians to product teams (TU Delft)

Valued ‘skills’

- Discovery & Delivery, Publication & Impact, Network (students and researchers), ...

Missing

- IT knowledge: tool developers, data formats & **infrastructure knowledge**, ...
- Soft skills: **understand** data producers and consumers to build relationship and provide effective support
Staff - training

1st: autodidact
In project phase through workshops, conferences, literature, etc.

2nd: train the trainer
Needed to reach a wider audience, found online material to teach researchers but not for support staff. Therefore developed: data-intelligence 4 librarians (RDM basics, 3TU.DC practical instructions, consultancy & acquisition skills). Later evolved to Essentials 4 Data Support.

3rd: train the users
Currently: experience with short module in information literacy, custom workshops. Developing blended learning module for ExtensionSchool, ...

3TU.Datacentrum
Current 3TU.Datacentrum team

5 groups: Development, Products, Technology, Communication, Acquisition.
0,8 Director (former product developer)
0,6 Data Steward (former product manager)
0,6 Data Engineer
2,2 Data librarians (former subject librarians, 2 levels)
2,6 Acquisition staff (2 at TUD, 1 at TU/e, 1 at UT)
0,4 Communication staff
3,2 IT staff (1 coordinator + 3 developers)

Other (library) staff occasionally involved, e.g., training, acquisition, events, projects, consultancy from education support team, research support team, IT department, Valorisation, Legal and Strategic Development departments.
What worked

1. Meet the needs
   - Sharing: large collection + cross institution collaboration, publications, Digital Object Identifiers (DOIs), ...
   - Credits: very rapid data collection, benchmark files, ...
   - Safe storage, (new) Funder requirements, ...

2. Find & work with ready, able and willing research communities
   - Offer support (students, refund APC, workshops, ...)
   - Co-develop tools

3. Build relation (on trust)
   - Send in people who understand what drives researchers
   - Solve other library ‘stuff’

4. Build and show good practices

5. Events: Dutch Data Prize, H2020, Roadshow OpenScience
What didn’t (yet)

1. Push (or pull) hard on the *not* ready and not willing
2. Use the wrong **language**: repository, metadata, archive, persistent identifier, open data, library, etc.

A. Establish proper data citation and references to data practices
B. Get journal editors to promote depositing data underlying articles in trusted archives
C. Get teachers to use freely available research data (large scale) in **education**
D. Major change of attitude/culture towards research data management with senior researchers
E. Policy support for better research data management (getting there) (this will only worked when data producers are ‘enabled’)

Webinar
Library Connect

3TU.Datacentrum
Challenges

- Create data sharing and data reuse incentives
- Create data ‘finding places’
- Provide clarity on data ‘ownership’
- Develop (inter)national funding model
- Change corporate image of library (and IT department)

Cartoons by Auke Herrema (from RDA 4th Plenary, Amsterdam).
Thank You! & Resources

3TU.Datacentrum
DataCite
Research Data Netherlands

Data Intelligence Training for Library Staff (art.)
Essentials 4 Data Support
Data-lab video (under construction)

http://datacentrum.3tu.nl/en/home/
http://www.datacite.org/
http://www.researchdata.nl/en/

http://dx.doi.org/10.2218/ijdc.v8i1.255
http://datasupport.researchdata.nl/
Library Connect Photo Contest
September to November 2014

Theme: Never underestimate the importance of a librarian

Grand prize: $1,500
Two runner-up prizes of $1,000 each

In 2003, Elsevier introduced a popular series of posters with the theme Never underestimate the importance of a librarian. Library Connect is reimagining those posters and related artifacts with photos crowdsourced from the global library community, and we are inviting librarians around the world to submit photos they have taken (or "made" for the real photo buffs) that exemplify the theme. The resulting pieces will be freely available under a Creative Commons license for download from the Library Connect website for use in your library.

The grand prize winner will receive US $1,500 and two runners up will receive US $1,000 to help offset travel and registration costs to a library conference of each recipient’s choice.

What Makes a Winning Photo?

Like those featured in the previous Never underestimate the importance of a librarian posters, your image should capture the essence of research, discovery and transformation within a physical, life, health or social science, or an arts and humanities setting. For example, one poster featured a crew conducting an experiment inside a space station, while another showed a surgical team in an operating theater. In both photos, a librarian is identified as a participant, but the accompanying text explains that while a librarian wasn’t actually there, the librarian played a vital role in enabling research breakthroughs and real world outcomes.

We are excited to see and share exciting visuals from the world and work that you help inform.

Upload your submissions today via the Submit an Entry tab. FAQs are provided below; please consult the Official Contest Rules on the Submit an Entry page for full details.

https://www.facebook.com/libraryconnect
Thank You!

Jennifer Clark
Jenfanguy AT gmail DOT com
@jengerful

Margaret Henderson
mehenderson AT vcu DOT edu
@mehlibrarian

Jeroen Rombouts
J.P.Rombouts AT tudelft DOT nl
@jprombouts