Librarians in the Lab: Radical re-engineering or research reality check?

National Data Service Workshop, October 2016
Professor Liz Lyon, Doreen E. Boyce Chair, School of Information Sciences, University of Pittsburgh
Agenda

1. Data Science roles & requirements
2. Skilling up for data
   - MLIS Data Stewardship Pathway
   - Student experiences in the Lab
3. Research Data Services reality
4. Radical Re-engineering
   - Research Data Service Models
   - Challenges and Benefits
iSchool Context

- Educating prospective data scientists
- Curriculum review of MLIS Program
- Focus on a family of data science roles
- Some roles may be located in a library....
A family of new data science roles

Lyon & Brenner (2015) IJDC
Q1
What are the real-world requirements for these roles?
Linking data roles, skills & curriculum


- Analysis of real-world positions for six data roles
- Part 1: data librarian, data archivist, data steward
- Part 2: data analyst, data engineer, data journalist
- Map to current iSchool courses
- Informing development of a Data Stewardship Pathway
<table>
<thead>
<tr>
<th>Data Librarian</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Education</strong></td>
</tr>
<tr>
<td>ALA-accredited degree in library and/or information science</td>
</tr>
<tr>
<td>ALA-accredited degree in library and/or information science or advanced degree in relevant discipline</td>
</tr>
<tr>
<td>ALA-accredited accredited degree in library and/or information science and a graduate degree in relevant discipline</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Education</td>
</tr>
<tr>
<td>-----------</td>
</tr>
<tr>
<td>Unspecified Bachelor’s degree</td>
</tr>
<tr>
<td>Bachelor’s degree in discipline relevant to data that is at the focus of work (i.e. health sciences and biological)</td>
</tr>
<tr>
<td>Bachelor’s degree in an “analytical” major such as math, business, computer science</td>
</tr>
<tr>
<td>Advanced degree in informatics-related field</td>
</tr>
</tbody>
</table>

Lyon et al (2016) iPres Proc
Real World Job analysis Part 1

Lyon et al (2016) iPres Proc

Data Archivist
- Data documentation
- Data preparation
- Data integration

Data Librarian
- Funding agency
- RDM requirements
- Research lifecycle
- Training & outreach

Data Steward / Curator
- Metadata
- Research
- Disciplinary data
- Statistical analysis

Enable Data Curation

Web authoring
- Data visualization

Data governance
- Data quality assurance
- Relational databases
“The Data Stewardship Pathway will provide an introduction to data curation, digital preservation and data science, and will frame these topics within the broader context of data informatics, digital scholarship, research integrity, disciplinary diversity and cultural change.”

Launching a new MLIS Pathway…..
• Describe policy trends in open data, open science and open scholarship

• Assess disciplinary practices in creating, using and sharing research data

• Develop an understanding of legal, ethical and sensitive data

• Learn good practice in describing and documenting data

• Recognise the benefits of effective storage and curation of active data

• Demonstrate a knowledge of long-term data archiving and preservation

• Show an awareness of the principles of data publication & citation

• Meet federal funding agency data policy requirements for DMPs

• Use a range of tools for data cleaning, data analytics and data visualization

• Evaluate the impact of big data on society
Recruitment pool? Backgrounds?

Lyon & Brenner (2015) IJDC
Likely graduate degrees

- **data engineer** - computer science, engineering
- **data analyst** – mathematics, statistics, business studies
- **data librarian** - arts & humanities
- **data steward** - arts & humanities
- **data journalist** – journalism, media studies, communications studies
Flavours of disciplinary/domain (X-) informatics

STEM: strongly represented...
“Very few librarians are likely to have specialist scientific or medical knowledge - if you train as a research scientist or a medic, you probably won’t become a librarian.”

RLUK/Mary Auckland (2012): Reskilling for Research
Curation : domain disconnect ?
Q2
How can prospective data scientists learn about diverse disciplinary practices?
RDM Class

2. No Class (Labor Day)
4. Universities & Data Part 2: Policy, Strategy & Services
5. Data Management Plans
6. Open Data & Data Sharing
7. Data from a Legal Perspective
8. Immersive Session with Faculty Researchers Pt 1
9. Immersive Session with Faculty Researchers Pt 2
10. Data from an Ethical Perspective
11. Data Centers: From Google to National to Local
12. Disciplinary Data Diversity: Part 1
13. Disciplinary Data Diversity: Part 2
14. Making a Career in Data
15. Student Presentations & Course Evaluation
Research Data Infrastructures

1. Data Storage Part 1
2. Data Storage Part 2
3. Data Publication & Citation Part 1
4. Data Publication & Citation Part 2
5. Data Description & Data Standards
6. Data Discovery
7. Immersive session with Researchers
8. Data Repositories & Preservation Part 1
9. No class – Spring Recess
10. Data Repositories & Preservation Part 2
11. Data Sustainability & Costs
12. Citizen Science, Citizen Data
13. Data Science, Data Analytics & Data Visualization
14. Data, Society, Futures
15. Student Presentations & Course Evaluation
• Innovative *immersive sessions in the lab*
  e.g.
  – Evolution Lab (Biol)
  – Materials Chemistry
  – Pharmacy
  – Public Health
  – Engineering
  – Renal Medicine

Data at scale: the lab may look like this!
Student feedback

“It was great to see a real-life example of how a lab generates and uses data.”

“We learned not only about the specifics of their research but about the lifecycle of data.”

“This was a valuable experience. It was very practical and illuminated some of the struggles that one may encounter in discussing data as its own area of research.”

Researcher feedback

“Conversations with students and their feedback were helpful to the project. We never have these conversations-how to back up data, back up strategies, how the University could support it…”

“Main thing we learned was about the DMPTool - very helpful.”

“One issue was file naming – knowing to do that initially would have helped….Students with own naming conventions for simulation files and couldn’t go back to them and understand the data”.

BILATERAL LEARNING

Immersive experience & Laboratory placement

Data curation guidance & support
Q3
What Research Data Services are Libraries providing to researchers?
Reality check - RDS provision
(International survey: Australia, UK, Ireland, New Zealand, Canada, Germany, Netherlands)

• Advisory services
  – Web resource/guide “most common service, well-developed / extensive”
  – Training / data literacy “positioned as a growing service ie basic / well-developed”

• Technical services
  – Data repository “best considered as basic”
  – Curation of active data, create/transform metadata, prepare data for deposit, long term preservation of research data “predominantly no service”

Cox, Kennan, Lyon & Pinfield (2016) JASIST In press
Future top priority

<table>
<thead>
<tr>
<th>Service</th>
<th>UK</th>
<th>Percentage</th>
<th>Australia</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Offer research data management training and/or data literacy instruction</td>
<td>51/81 (63%)</td>
<td>21/34 (62%)</td>
<td>7/81 (9%)</td>
<td>4/34 (12%)</td>
</tr>
<tr>
<td>Maintaining a web resource/guide of local advice and useful resources for RDM</td>
<td>51/81 (63%)</td>
<td>21/34 (62%)</td>
<td>7/81 (9%)</td>
<td>4/34 (12%)</td>
</tr>
<tr>
<td>Offer a research data management advisory service to researchers</td>
<td>48/81 (59%)</td>
<td>20/34 (59%)</td>
<td>6/34 (18%)</td>
<td>4/34 (12%)</td>
</tr>
<tr>
<td>Offer advice on copyright and/or intellectual and/or licensing property rights relating to data and data management</td>
<td>41/78 (53%)</td>
<td>20/34 (59%)</td>
<td>6/34 (18%)</td>
<td>4/34 (12%)</td>
</tr>
<tr>
<td>Provide a data catalogue including your institution’s research data</td>
<td>40/77 (52%)</td>
<td>16/33 (48%)</td>
<td>6/34 (18%)</td>
<td>4/34 (12%)</td>
</tr>
<tr>
<td>Run a data repository/archive/store</td>
<td>41/79 (52%)</td>
<td>16/34 (47%)</td>
<td>6/34 (18%)</td>
<td>4/34 (12%)</td>
</tr>
<tr>
<td>Provide access to tools to support research data management</td>
<td>38/80 (48%)</td>
<td>13/34 (38%)</td>
<td>6/34 (18%)</td>
<td>4/34 (12%)</td>
</tr>
<tr>
<td>Offer data publication advisory services</td>
<td>35/78 (45%)</td>
<td>13/34 (38%)</td>
<td>6/34 (18%)</td>
<td>4/34 (12%)</td>
</tr>
<tr>
<td>Provide advisory services on the curation of active data</td>
<td>34/79 (43%)</td>
<td>13/34 (38%)</td>
<td>6/34 (18%)</td>
<td>4/34 (12%)</td>
</tr>
<tr>
<td>Offer data citation advisory services</td>
<td>33/81 (41%)</td>
<td>12/33 (36%)</td>
<td>6/34 (18%)</td>
<td>4/34 (12%)</td>
</tr>
<tr>
<td>Promote awareness of reusable data sources, such as data archives</td>
<td>30/80 (38%)</td>
<td>11/33 (33%)</td>
<td>6/34 (18%)</td>
<td>4/34 (12%)</td>
</tr>
<tr>
<td>Offer data storage advisory services</td>
<td>27/79 (34%)</td>
<td>10/34 (29%)</td>
<td>6/34 (18%)</td>
<td>4/34 (12%)</td>
</tr>
<tr>
<td>Carrying out long term preservation of research data</td>
<td>25/78 (32%)</td>
<td>9/34 (26%)</td>
<td>6/34 (18%)</td>
<td>4/34 (12%)</td>
</tr>
<tr>
<td>Offer a service creating or transforming metadata for data or datasets</td>
<td>25/79 (32%)</td>
<td>8/33 (24%)</td>
<td>6/34 (18%)</td>
<td>4/34 (12%)</td>
</tr>
<tr>
<td>Provide advisory services on the technical aspects of long term data preservation</td>
<td>22/79 (28%)</td>
<td>8/33 (24%)</td>
<td>6/34 (18%)</td>
<td>4/34 (12%)</td>
</tr>
<tr>
<td>Selecting, accessioning and/or deselecting and de-accessioning data/data sets for deposit in a repository</td>
<td>21/79 (27%)</td>
<td>8/33 (24%)</td>
<td>6/34 (18%)</td>
<td>4/34 (12%)</td>
</tr>
<tr>
<td>Preparing data/data sets for deposit in a repository</td>
<td>21/79 (27%)</td>
<td>7/81 (9%)</td>
<td>7/81 (9%)</td>
<td>7/81 (9%)</td>
</tr>
<tr>
<td>Provide support for search and retrieval of external data sources</td>
<td>17/81 (21%)</td>
<td>6/34 (18%)</td>
<td>7/81 (9%)</td>
<td>7/81 (9%)</td>
</tr>
<tr>
<td>Carrying out the curation of active data</td>
<td>12/79 (15%)</td>
<td>5/34 (15%)</td>
<td>7/81 (9%)</td>
<td>7/81 (9%)</td>
</tr>
<tr>
<td>Offer an advisory service on data analysis/mining/visualization</td>
<td>8/80 (10%)</td>
<td>4/34 (12%)</td>
<td>7/81 (9%)</td>
<td>7/81 (9%)</td>
</tr>
<tr>
<td>Directly participate with researchers on a research project (as a team member)</td>
<td>7/81 (9%)</td>
<td>4/34 (12%)</td>
<td>7/81 (9%)</td>
<td>7/81 (9%)</td>
</tr>
</tbody>
</table>

Cox, Kennan, Lyon & Pinfield (2016)
JASIST In press
Q4
How should Library Research Data Service models be re-engineered?
1. Transactional delivery model

- In the physical Library
- Remote
- Access & Reference
- RDM Advocacy
- RDM LibGuides


https://www.flickr.com/photos/smiling-gardener
2. Hybrid delivery model

- Assigned to Faculty / Department
- Liaison
- Consultancy
- DMP
- RDM training
3. Immersive delivery model – Librarians in the Lab

- Laboratory or clinical setting
- Fully integrated
- Collaborative science
- Data description & curation
- Data analysis & visualisation

Re-engineered

Clinical Informationist?

Guise (1997)
Davidoff & Florance (2000)
Detlefson (2002)
A variation on the theme

- ‘Embedded librarian’
- Embedded vs immersive

- T-shaped professional?
- I-shaped model?

Shumaker (2012)
Martin (2013)
Stanton et al (2012)
Q5
How do we view the maturity of Research Data Services?
Research Data Service maturity

Cox, Kennan, Lyon & Pinfield (2016) JASIST In press
Q6
What are the challenges and the benefits of re-engineering RDS?
Challenges to overcome

- **Workforce capacity** - to scale up provision
- **Workforce capability** – recruit new staff or up-skill
- **Trust in Library data service** – to be perceived as a trusted source of expertise
- **Credibility** – aspire to collaborate on equal terms with Faculty researchers
- **Leadership** – demonstrate vision, insight in articulating a future data stewardship strategy
- **Cultural change amongst library professionals** – resistance to embracing new immersive models of service delivery
Benefits to advocate and sell

- Data support at the researchers’ point-of-need *(here and now)*

- LIS professionals fully integrated at the coalface *(in the field, in the business, in the lab...)*

- Default listings in citations with attribution + credit *(LIS “co-authors”)*

- LIS data science roles act as “transparency agents” *(enhance research integrity & open science)*

*Lyon (2016) Liber Q
Lyon et al (2016) ASIST*
Thank you....
elyon@pitt.edu

Professor Liz Lyon, Doreen E. Boyce Chair, 
School of Information Sciences, University of Pittsburgh