University of Pittsburgh

SCHOOL OF
Information Sciences

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
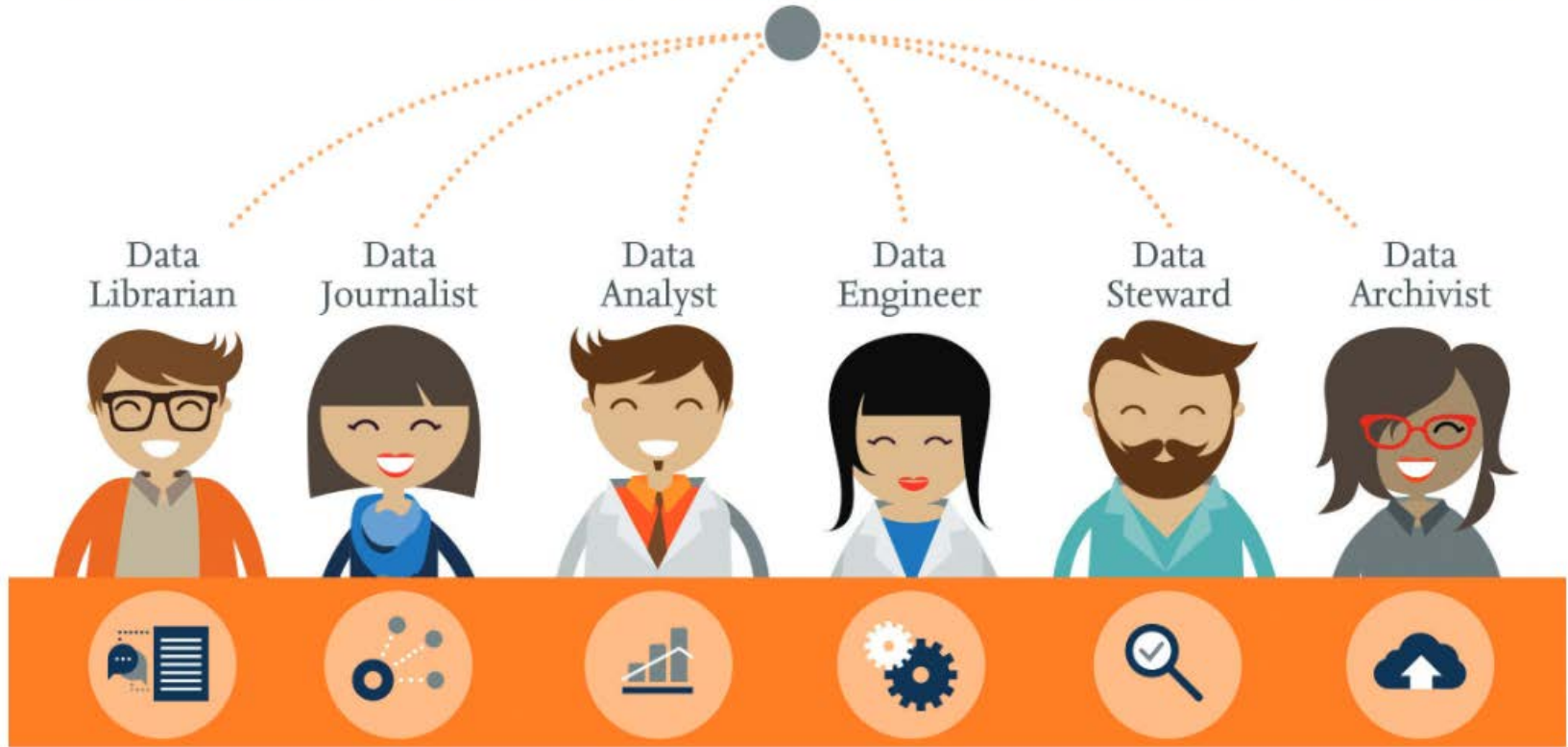
iSchool Context

- 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….

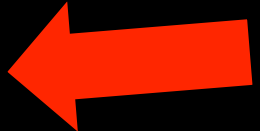A family of new data science roles

Lyon & Brenner (2015) IJDC

Q1
What are the real-world
requirements for these roles?

Linking data roles, skills & curriculum
Lyon et al (2016) iPres Proc, Lyon & Mattern (2016) IJDC In Press

- 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

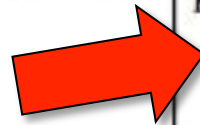- Informing development of a Data Stewardship Pathway

Data Librarian

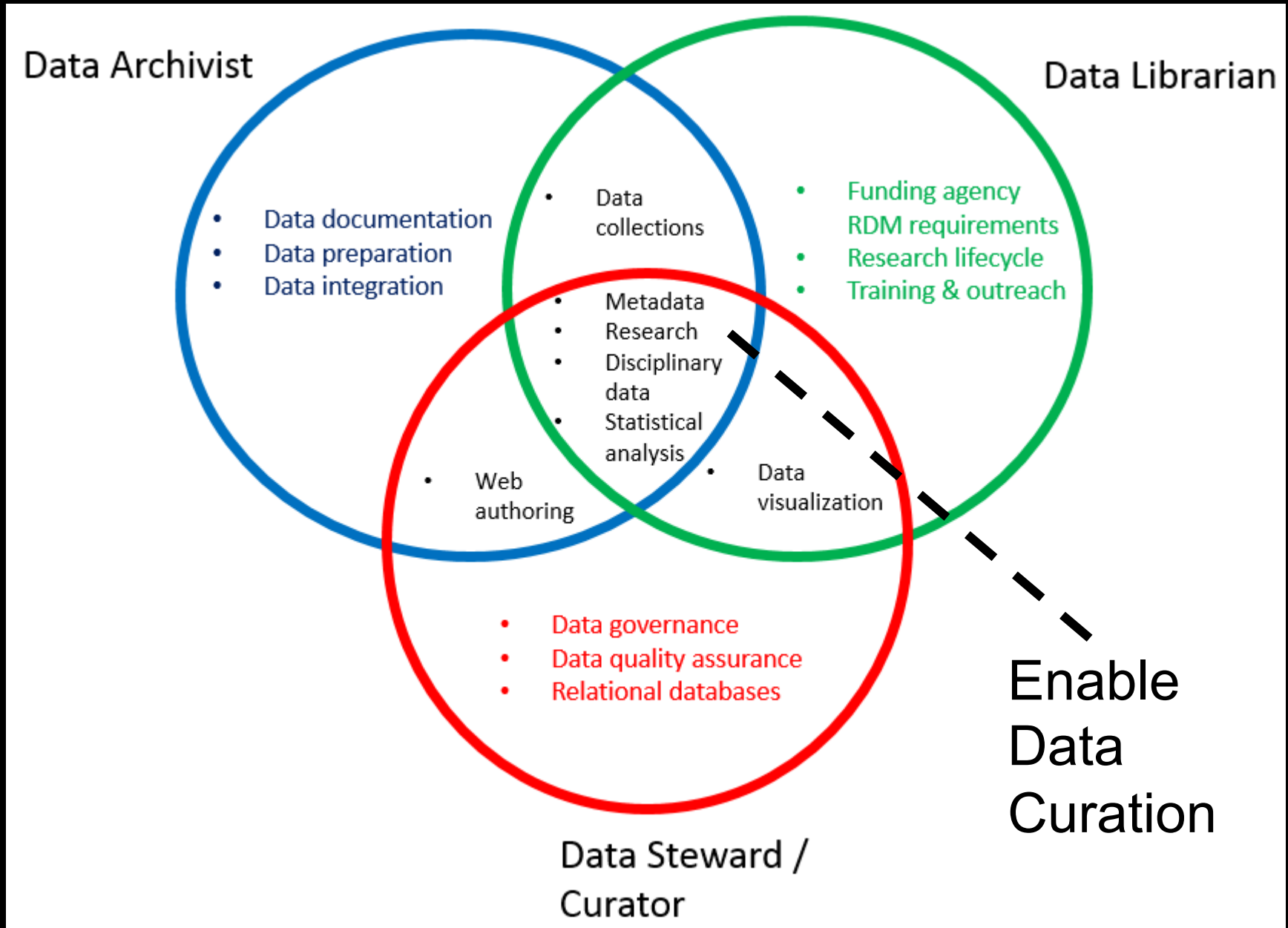| Data Librarian				
Education	**Experience**	**Knowledge**	**Skills**	**Competencies**
ALA-accredited degree in library and /or information science	Experience conducting qualitative and/or quantitative research	Knowledge of RDM activities and roles throughout research lifecycle	Ability to work well in collaborative teams	Competence with qualitative and quantitative analysis software packages (e.g. Atlas.ti, NVivo, SPSS, R)
ALA-accredited degree in library and/or information science **or** advanced degree in relevant discipline	Experience designing and delivering RDM training and outreach	Knowledge of RDM trends/current research, particularly in academic setting	Strong oral, written, and interpersonal communication skills	Competence with programming languages, (e.g. JavaScript, Python, and PHP)
ALA-accredited accredited degree in library and/or information science **and** a graduate degree in relevant discipline	Experience delivering RDM consultation support	Knowledge of metadata standards for data discovery and preservation	Project management effectiveness	Competence with GIS software
	Experience assessing user data needs and designing RDM services in response	Knowledge of sources for locating and depositing disciplinary data	Analytical and organizational skills	Competence with visualization tools
	Experience acquiring data resources for a library collection	Knowledge of funders' data management requirements		

Lyon et al (2016) iPres Proc

Data Steward / Curator

Data Steward/Curator				
Education	**Experience**	**Knowledge**	**Skills**	**Competencies**
Unspecified Bachelor's degree Bachelor's degree in discipline relevant to data that is at the focus of work (i.e. health sciences and biological) Bachelor's degree in an "analytical" major such as math, business, computer science Advanced degree in informatics-related field	Experience analyzing and understanding data as a researcher Experience with metadata schemas, structures, and standards Experience with data governance	Knowledge of data management and quality assurance practices Knowledge of metadata schemas and ontologies Knowledge of data governance Knowledge in discipline relevant to data Knowledge of database structure and development	Ability to work effectively in collaborative teams Oral, written, and interpersonal communication skills Ability to communicate effectively with researchers from a variety of disciplines and backgrounds Ability to learn new technologies quickly and to adapt to change Analytical and organizational skills	Competence with relational database systems (e.g. Microsoft Access; MySQL) Competence with Microsoft Excel Competence with data visualization tools Competence with web authoring tools, Drupal

Lyon et al (2016) iPres Proc

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

"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*
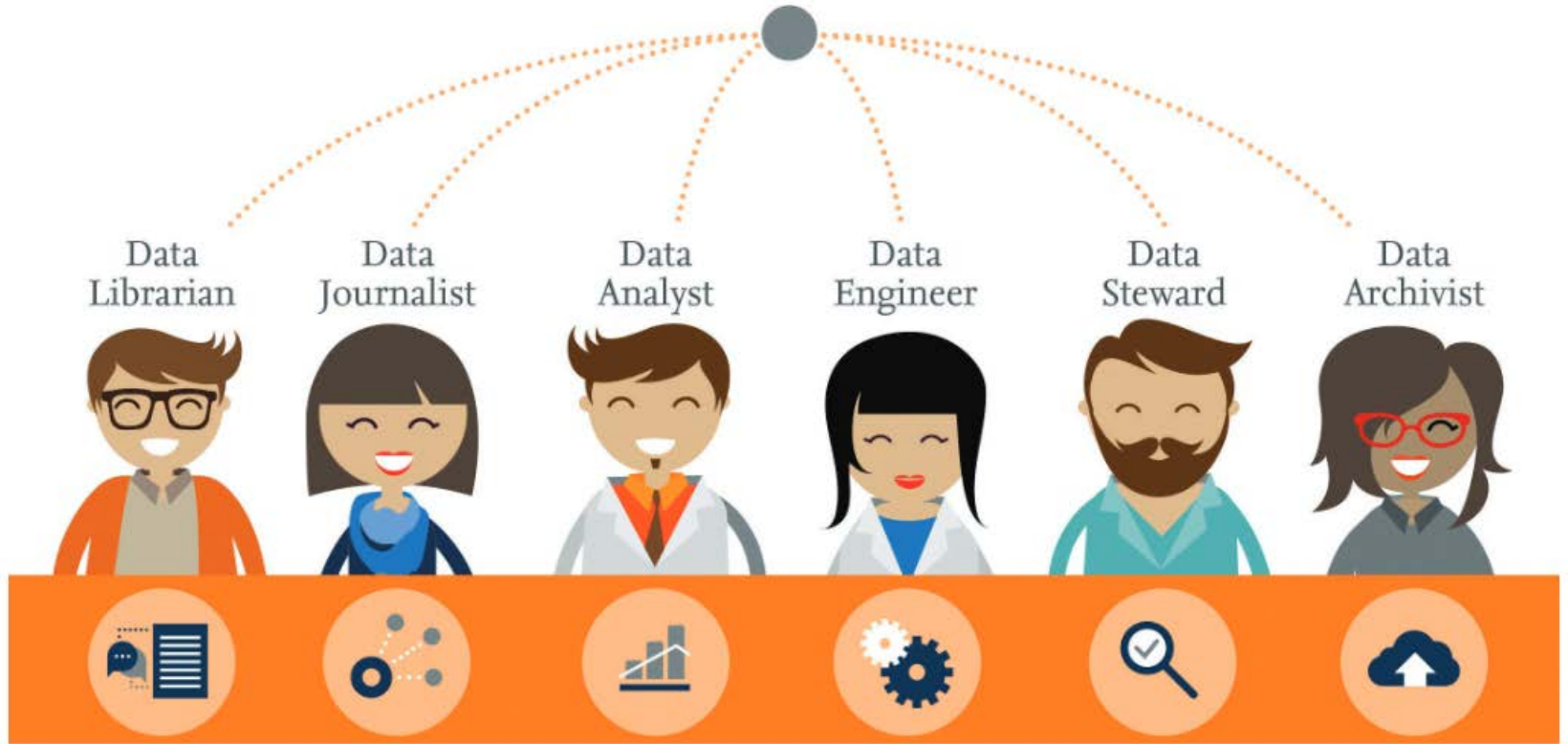
Recruitment pool? Backgrounds?

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

- Archival informatics
- Bioinformatics
  - Bioimage informatics
- Biodiversity informatics
- Business informatics
- Cheminformatics
- Community informatics
- Computational informatics
- Development informatics
- Disease informatics
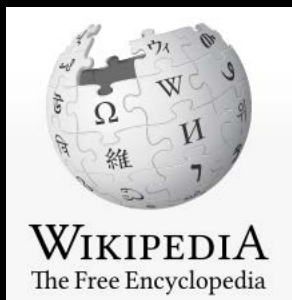
- Ecoinformatics
- Education informatics
- Engineering Informatics, an interd
- Environmental informatics
- Evolutionary informatics
- Forest informatics
- Geoinformatics, an interdisciplinar
- Health informatics
  - Consumer health informatics
  - Imaging informatics
  - Public health informatics
- Hydroinformatics
- Irrigation informatics
- Laboratory informatics
- Legal informatics

- Materials informatics
- Medical informatics
- Music informatics
- Neuroinformatics
- Pervasive Informatics
- Social informatics
- Technical informatics
- Translational research informatics

WIKIPEDIA
The Free Encyclopedia

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

No connection detected...

Curation : domain disconnect ?

*Q2
How can prospective data scientists learn about diverse disciplinary practices?*

RDM Class

1. Data: What? Where? When? Who? Why? How?
2. No Class (Labor Day)
3. Universities & Data Part 1: Understanding Data Requirements & Capability
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

Lyon (2016) New Review Academic Libraries

Librarians in the Lab?

Photo Credits:Flickr NASA HQ

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."

Lyon (2016) New Review Academic Libraries

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

Engage with research projects/ project participation and data analysis *"uniformly considered to be low priority services for future development"*.

*Cox, Kennan, Lyon & Pinfield (2016) JASIST In press*

UK

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

Australia

Future top priority	
Run a **data repository/archive/store**	21/34 (62%)
Offer a research data management **advisory service** to researchers	21/34 (62%)
Offer a service creating or transforming **metadata** for data or datasets	20/34 (59%)
Offer advice on **copyright and/or intellectual and/or licensing** property rights relating to data and data management	20/34 (59%)
Offer research data management **training** and/or data literacy instruction	20/34 (59%)
Maintaining a **web resource/guide** of local advice and useful resources for RDM	20/34 (59%)
Provide a **data catalogue** including your institution's research data	16/33 (48%)
Offer **data citation** advisory services	16/34 (47%)
Offer **data publication** advisory services	13/34 (38%)
Provide access to **tools** to support research data management	13/34 (38%)
Provide advisory services on the **curation** of active data	12/33 (36%)
Selecting, accessioning and/or deselecting and de-accessioning data/data sets for deposit in a repository	12/34 (35%)
Preparing data/data sets for deposit in a repository	12/34 (35%)
Offer **data storage** advisory services	11/33 (33%)
Promote awareness of **reusable data sources**, such as data archives	10/34 (29%)
Carrying out long term **preservation** of research data	9/34 (26%)
Carrying out the **curation** of active data	8/33 (24%)
Provide advisory services on the technical aspects of long term data **preservation**	8/33 (24%)
Directly participate with researchers on a research **project** (as a team member)	7/34 (21%)
Provide support for **search and retrieval** of external data sources	6/34 (18%)
Offer an advisory service on **data analysis/mining/visualization**	4/34 (12%)

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

Lyon (2016) New Review Academic Libraries
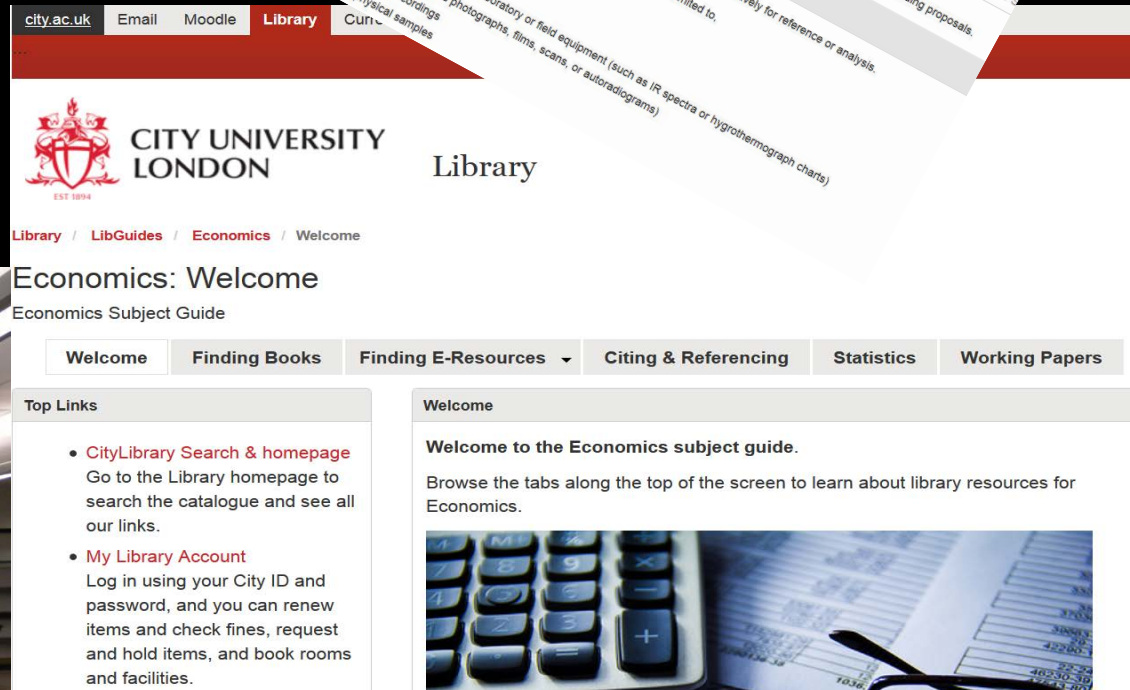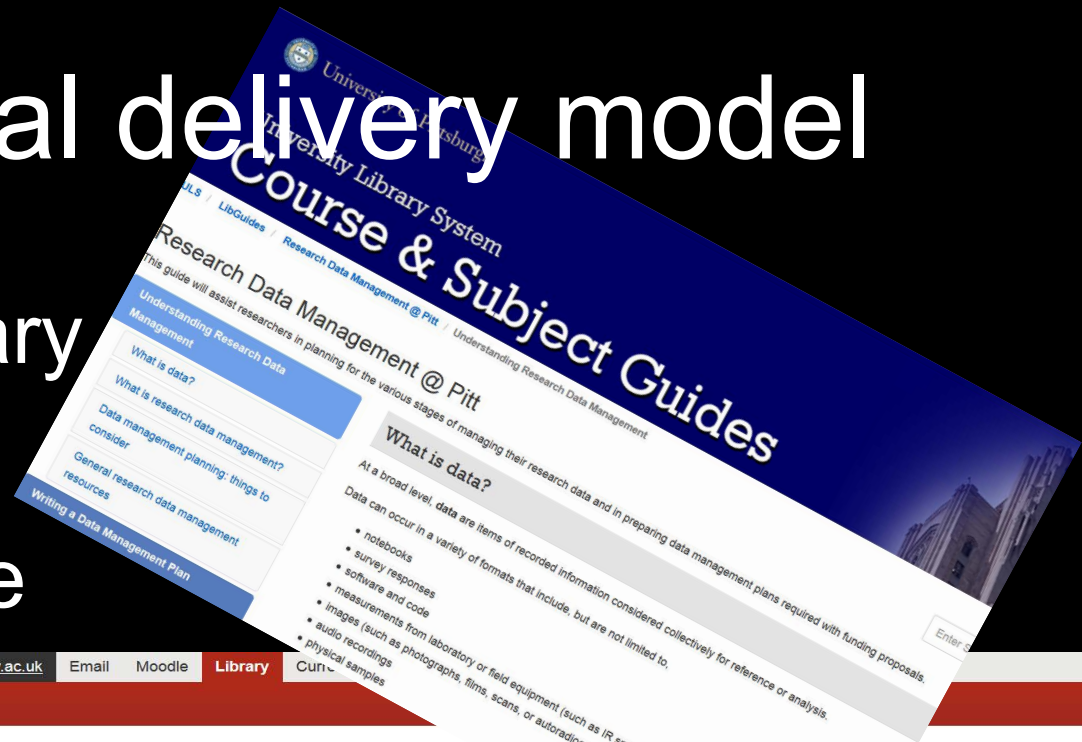
2. Hybrid delivery model

- Assigned to Faculty / Department
- Liaison
- Consultancy
- DMP
- RDM training

Need Help?
Ask a Librarian!

Your liaison librarians can help you find the information and resources you need.

www.mcgill.ca/library

McGill Library. Everything you need.

UNIVERSITY OF CALIFORNIA
DMP Tool
Build Your Data Management Plan

https://www.flickr.com/photos/brownlessbiomedicallibrary

Lyon (2016) New Review Academic Libraries

3. Immersive delivery model – Librarians in the Lab

- Laboratory or clinical setting
- Fully integrated
- Collaborati... science
- Data descri... curation
- Data analysis & visualisation

Re-engineered ☺

Lyon (2016) New Review Academic Libraries

Librarians on rounds?

Clinical Informationist?

Guise (1997)
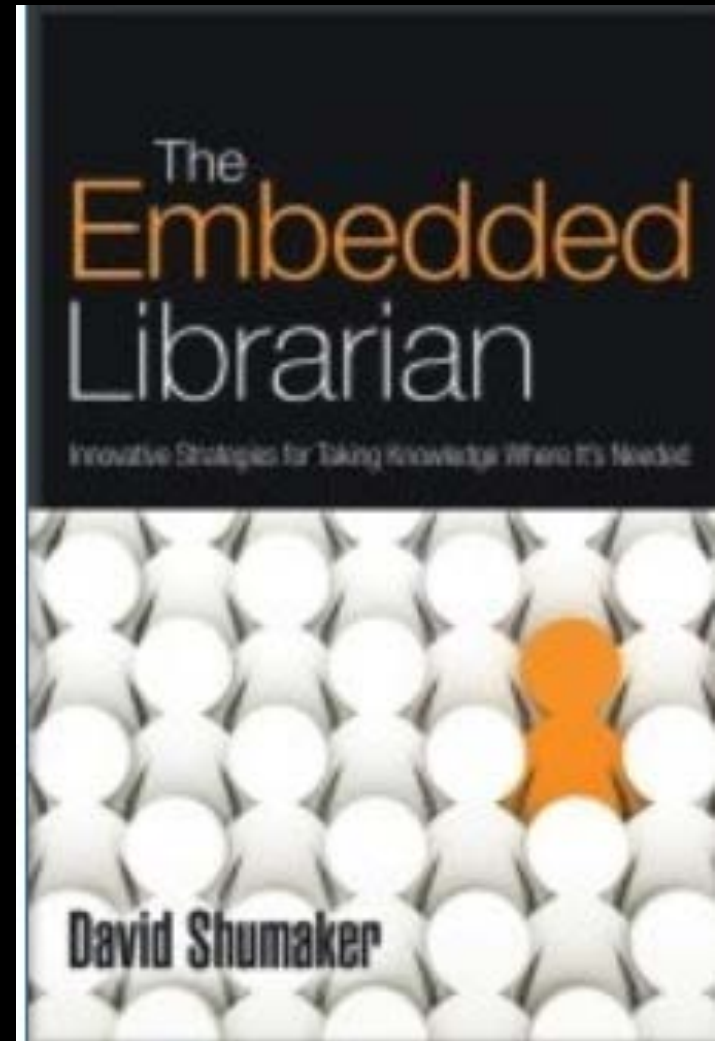*Davidoff & Florance (2000)*
Detlefson (2002)

A variation on the theme

- 'Embedded librarian'
- Embedded vs immersive

- T-shaped professional?
- I-shaped model?
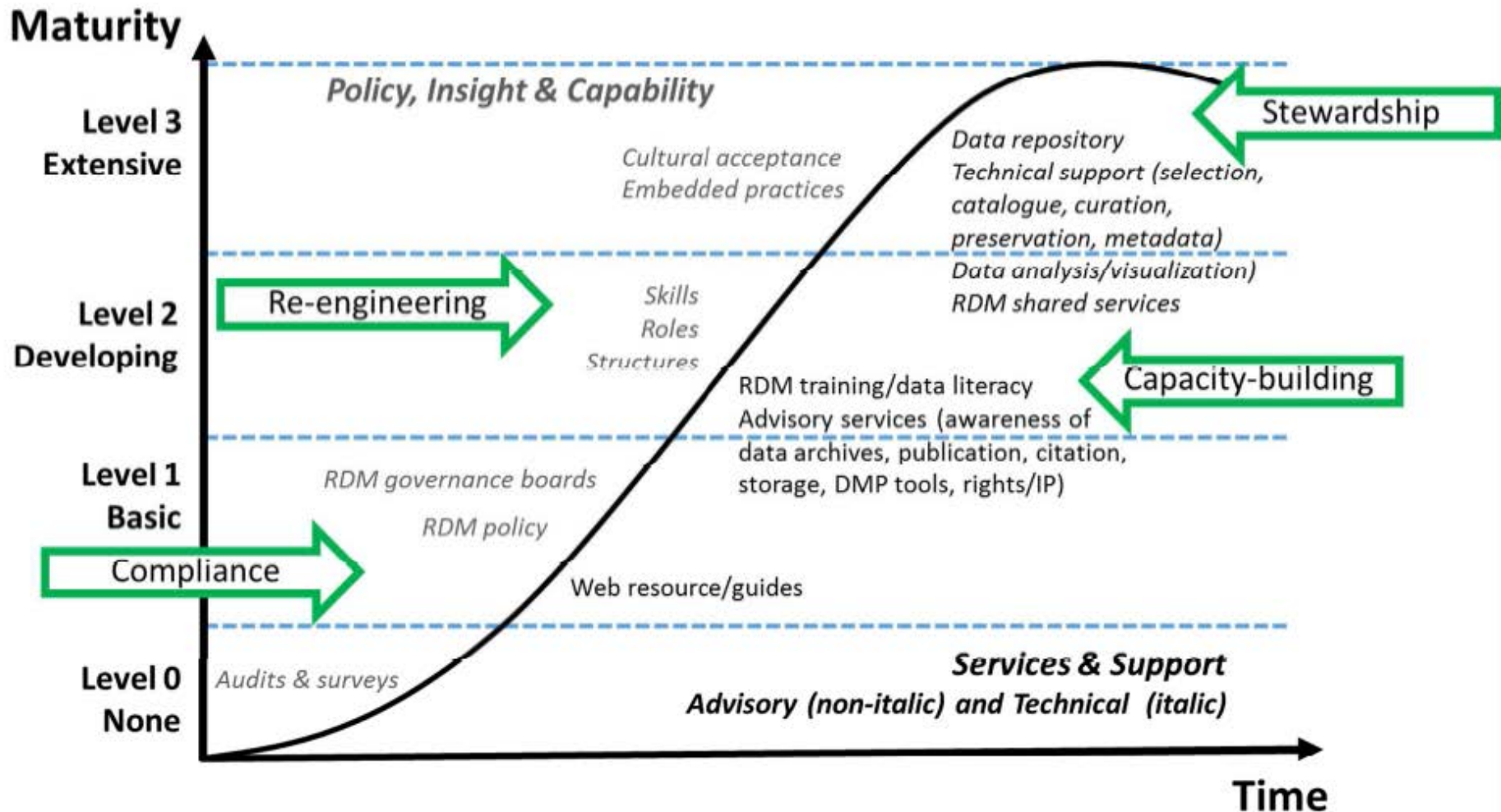
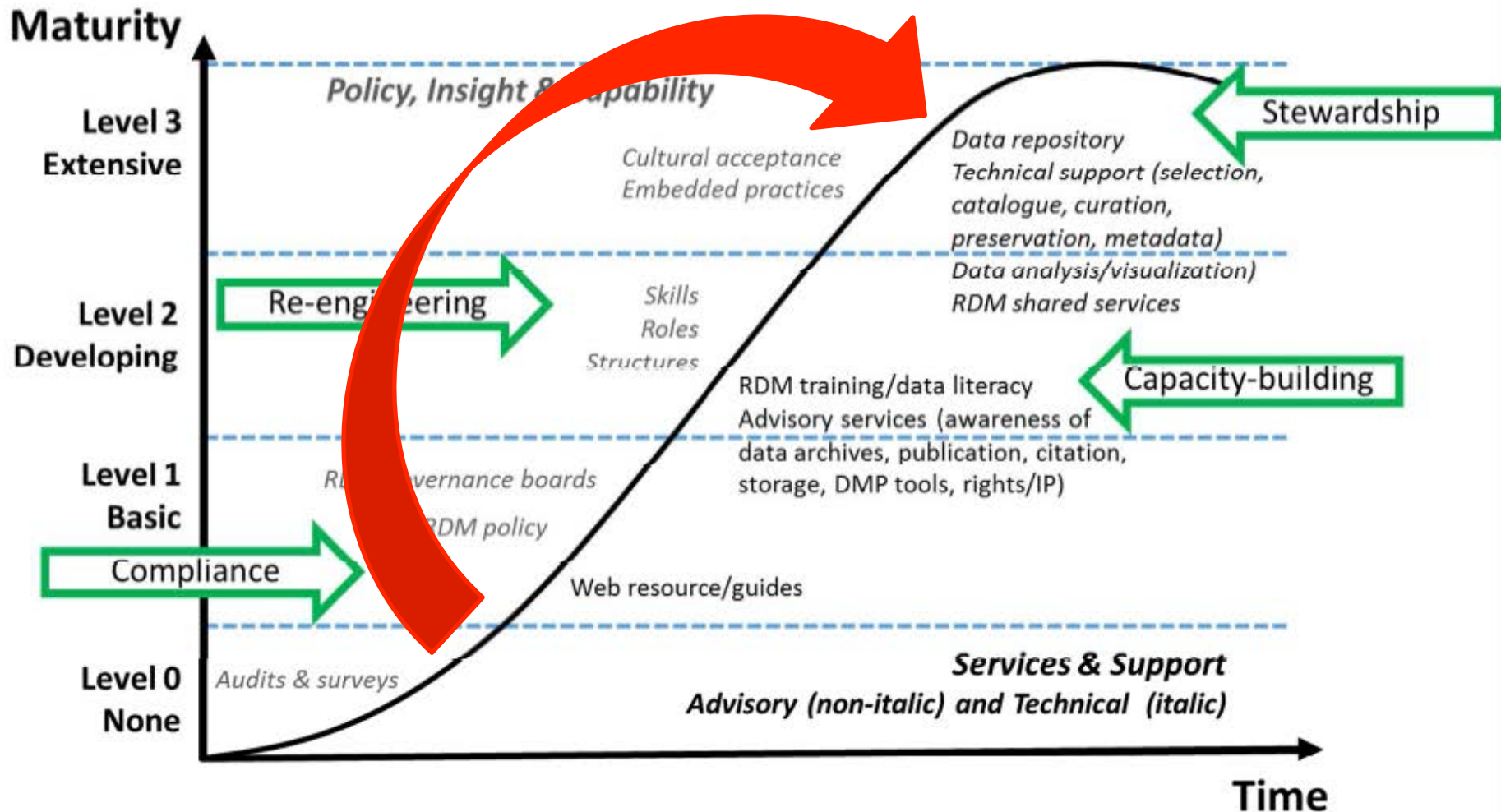*Shumaker (2012)*
Martin (2013)
Stanton et al (2012)

The Embedded Librarian

Innovative Strategies for Taking Knowledge Where It's Needed

David Shumaker

Q5
How do we view the maturity of Research Data Services?

Research Data Service maturity

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

Forward trajectory toward 2020

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

Lyon (2016) New Review Academic Libraries

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