Bridging the Data Talent Gap: Positioning the iSchool as an Agent for Change

International Digital Curation Conference, February 2015

Liz Lyon, School of Information Sciences

Aaron Brenner ULS, University of Pittsburgh

Data science capability   2013-2014

Seizing the data opportunity

A strategy for UK data capability

Data as a career

As well as ensuring that we equip school leavers and graduates with the key skills, we also need to ensure that data analytics is considered an exciting and rewarding career to pursue – amongst schoolchildren and graduates, but also parents and the wider media.

Career pathways and progression routes

In 2011, the Harvard Business Review published an article referring to the data scientist as the sexiest job of the 21st century[22], yet as described earlier in this chapter, this is an area which is currently experiencing skills shortages.
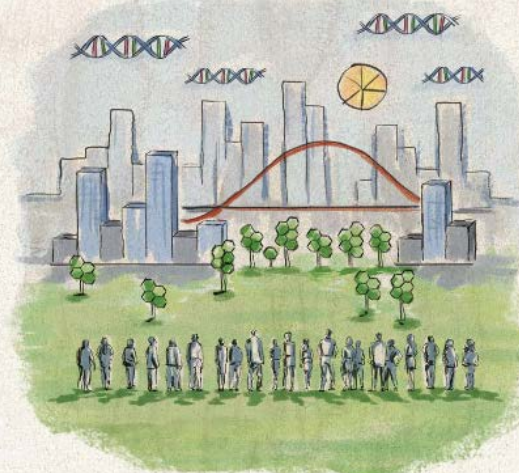
October 2013

Nesta...

creative skillset
Developing world class talent

MODEL WORKERS

How leading companies are recruiting and managing their data talent

5. Improve the supply of data talent with hybrid skill sets from education

Our research strongly supports the idea that UK universities are failing to produce graduates with the skills mix that data–driven businesses want. Perhaps this should not be a big surprise: there have long been concerns about how funding and organisational factors create hurdles to interdisciplinary teaching and research in UK universities.[37] While, say, the US major/minor system allows students to choose two fields of specialisation when they do a degree, UK students typically specialise in a single field. Anecdotally, one of our interviewees mentioned that PhD programmes in the US are in general broader and less specialised than in the UK too.

https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/254136/bis-13-1250-strategy-for-uk-data-capability-v4.pdf

http://www.nesta.org.uk/sites/default/files/model_workers_web_2.pdf

DATA SCIENCE ROLES

Where are they based?

Data Librarian | Data Journalist | Data Analyst | Data Engineer | Data Steward | Data Archivist

University / Research Institute

Newspaper Publisher

Corporate Sector

IT Company

Data Steward

Data Center

National Archive

DATA SCIENCE ROLES

What do they do?

Data Librarian	Data Journalist	Data Analyst	Data Engineer	Data Steward	Data Archivist

Advocacy, RDM, training

Telling stories, news vizualisations

Analytics, maths & stats models

Software development
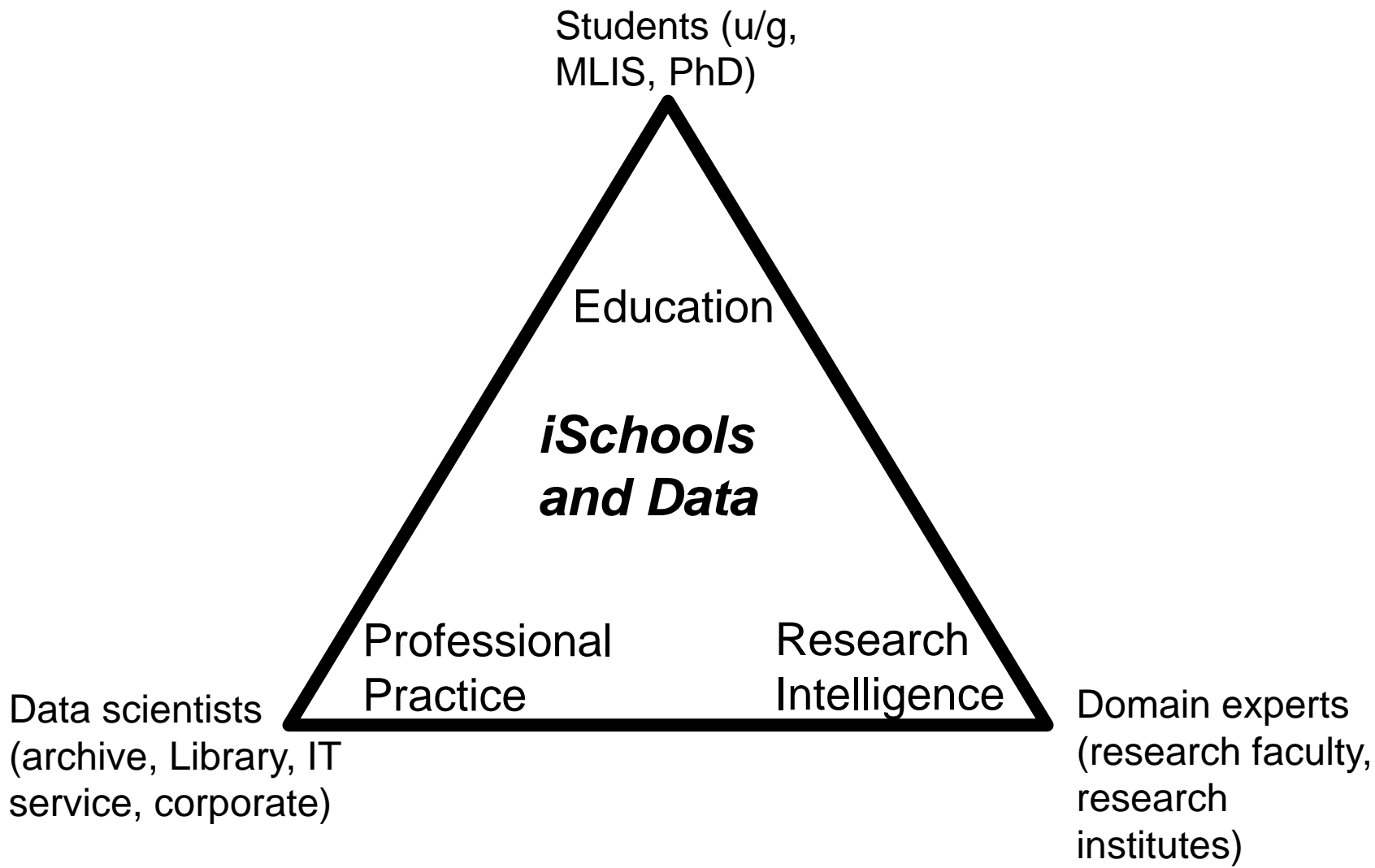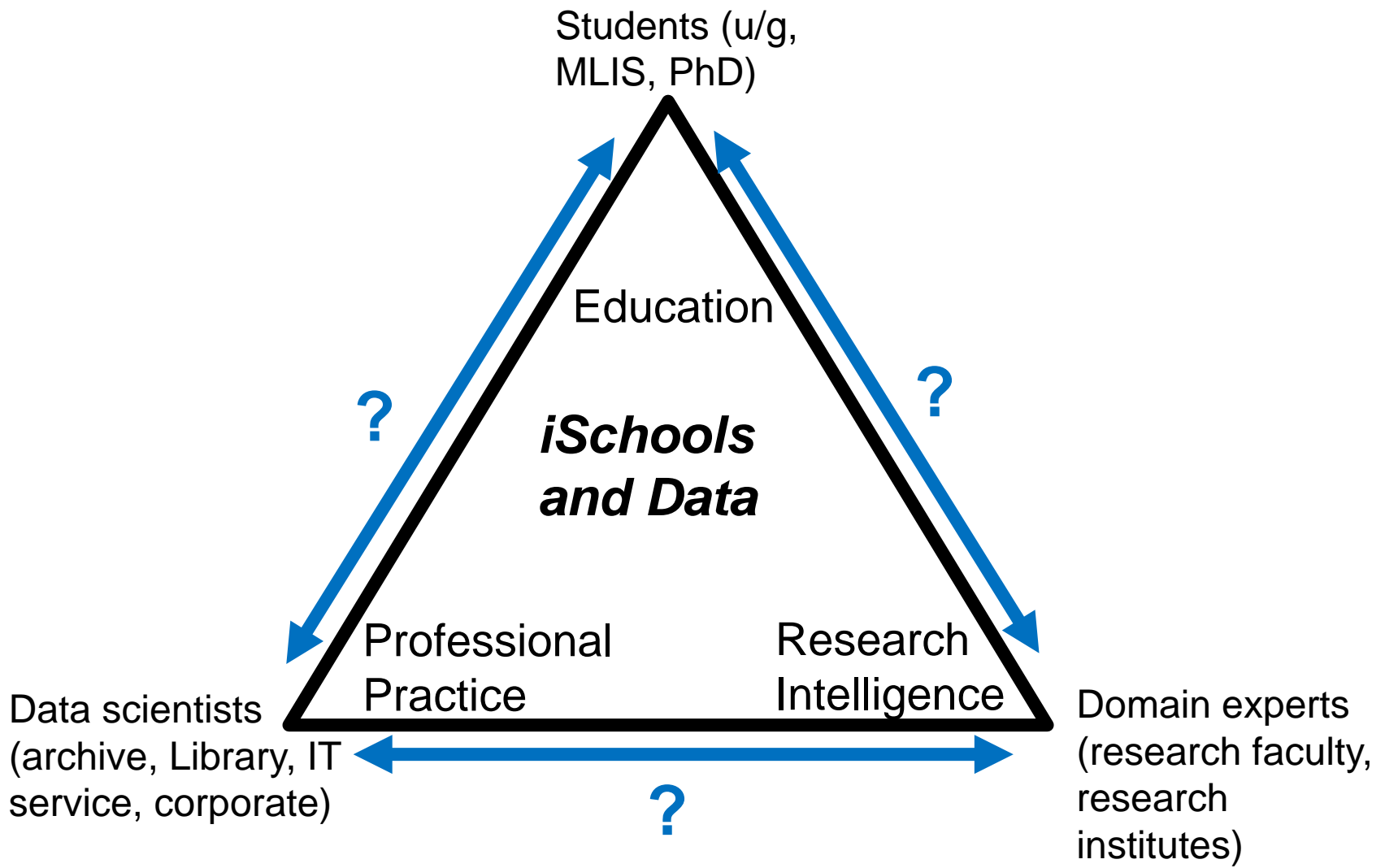
Curation

Long-term Preservation

iSchools Vision

- Educate a new breed of information professional, informed (but not constrained) by tradition
  - Technologically literate
  - Organizationally competent
  - Socially conscious
  - Ethically grounded
- Shape a new tradition
  - Pitt was one of 5 architects of the iSchool movement
    - Syracuse, Michigan, Washington, Drexel, Pitt
  - Now an international organization
    - 30 US universities
    - 3 Canadian universities
    - 18 European universities
    - 8 Asian universities
    - 3 Australian universities

Slide: Dean Ron Larsen, iSchool, University Pittsburgh

Students (u/g,
MLIS, PhD)

Education

*iSchools
and Data*

Professional
Practice

Research
Intelligence

Data scientists
(archive, Library, IT
service, corporate)

Domain experts
(research faculty,
research
institutes)

Students (u/g, MLIS, PhD)

Education

iSchools and Data

Professional Practice

Research Intelligence

Data scientists (archive, Library, IT service, corporate)

Domain experts (research faculty, research institutes)
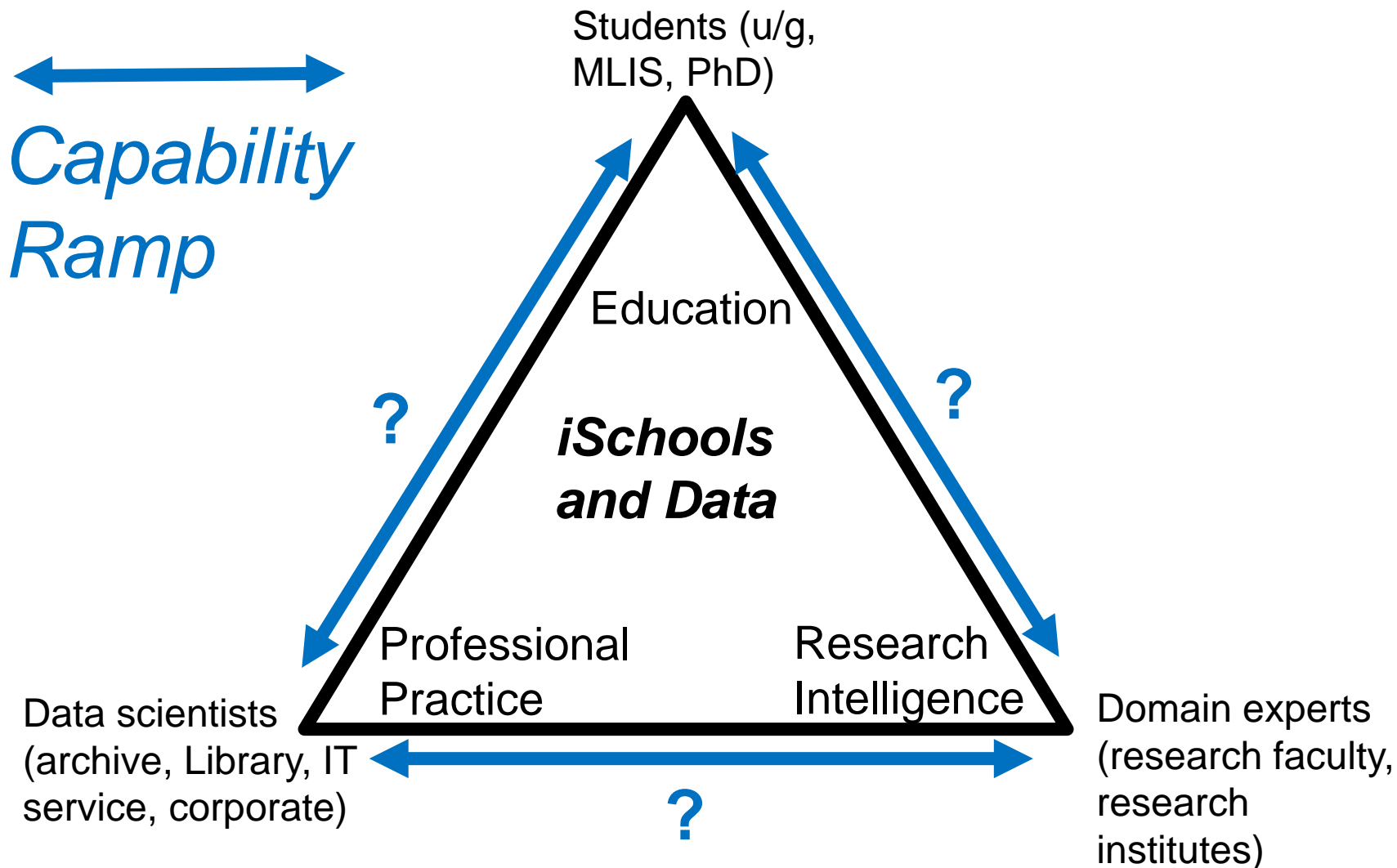
Concept of Ramps

- Shaping Ramps for data-intensive research (2010) Malcolm Atkinson & Dave De Roure
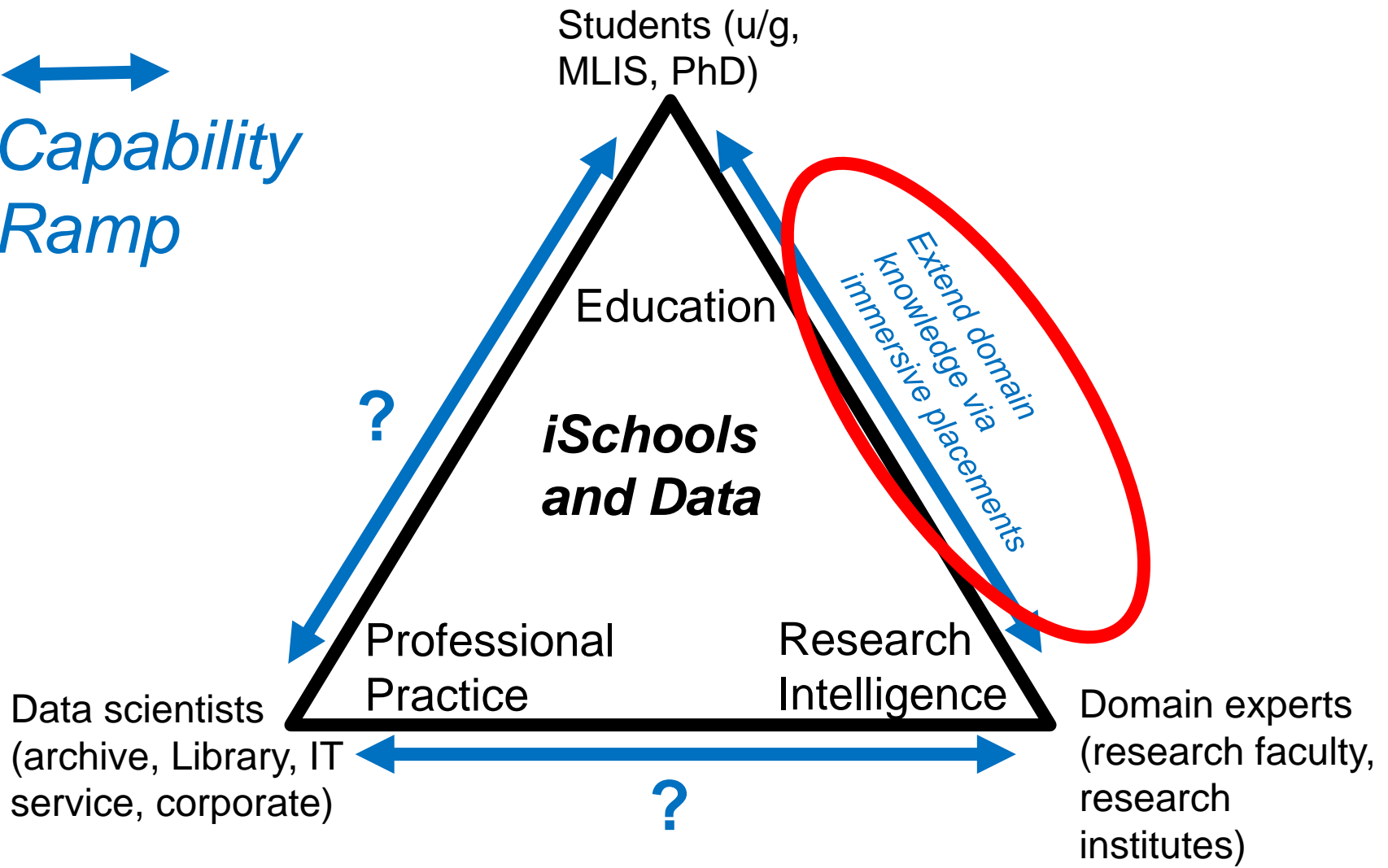 - Intellectual Ramp
  - Technology Ramp

- Participation Ramp (Lyon & Beaton 2015)

- *Capability Ramp?*

Capability Ramp

Students (u/g, MLIS, PhD)

Education

iSchools and Data

?

?

Professional Practice

Research Intelligence

Data scientists (archive, Library, IT service, corporate)

?

Domain experts (research faculty, research institutes)

1. *Immersive* MLIS graduate programs

Capability Ramp

Students (u/g, MLIS, PhD)

Education

iSchools and Data

Extend domain knowledge via immersive placements

?

Professional Practice

Research Intelligence

?

Data scientists (archive, Library, IT service, corporate)

Domain experts (research faculty, research institutes)

No connection detected...

Curation : domain disconnect

Research Data Management (MLIS 2014-2015)

1. Introductions & Overview
2. Data Landscape
3. Universities & Data
4. Data Requirements & Capability
5. RDM Roadmaps, Strategy, Services & Structures
6. Data Management Plans
7. No Class – Fall Break
8. **Immersive session with Researchers**

9. Disciplinary Data 1
10. Legal & Ethical Issues
11. Disciplinary Data 2
12. Data Centers
13. Data Advocacy, Skills, Training
14. Data Sustainability & Costs
15. Presentations

University of Pittsburgh

Research Data Infrastructures (MLIS 2014-2015)

1. Introductions, Syllabus Overview & Data Storage Part 1
2. Data Storage Part 2
3. Data Publication & Citation Part 1
4. Data Publication & Citation Part 2
5. Data Discovery
6. **Immersive session with Researchers**
7. Data Description & Data Standards
8. Data Repositories & Preservation Part 1
9. Data Repositories & Preservation Part 2
10. No class – Spring Recess
11. Data Sharing & Disciplinary Diversity
12. Citizen Science, Citizen Data
13. Data Science, Data Analytics & Data Visualization
14. Data, Society, Futures
15. Presentations & Summary Evaluation

Immersive sessions

- Librarians in the lab
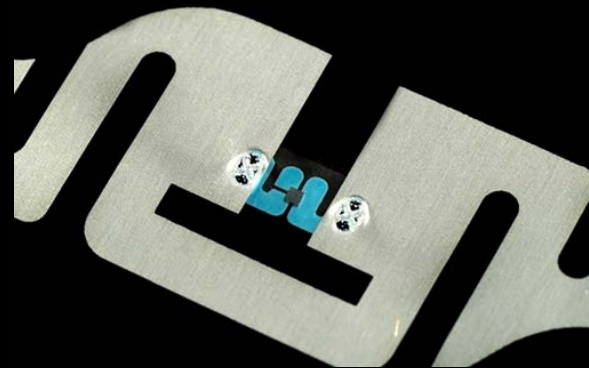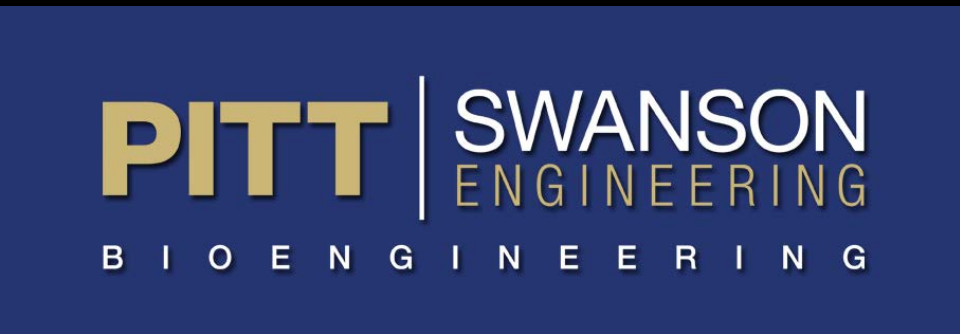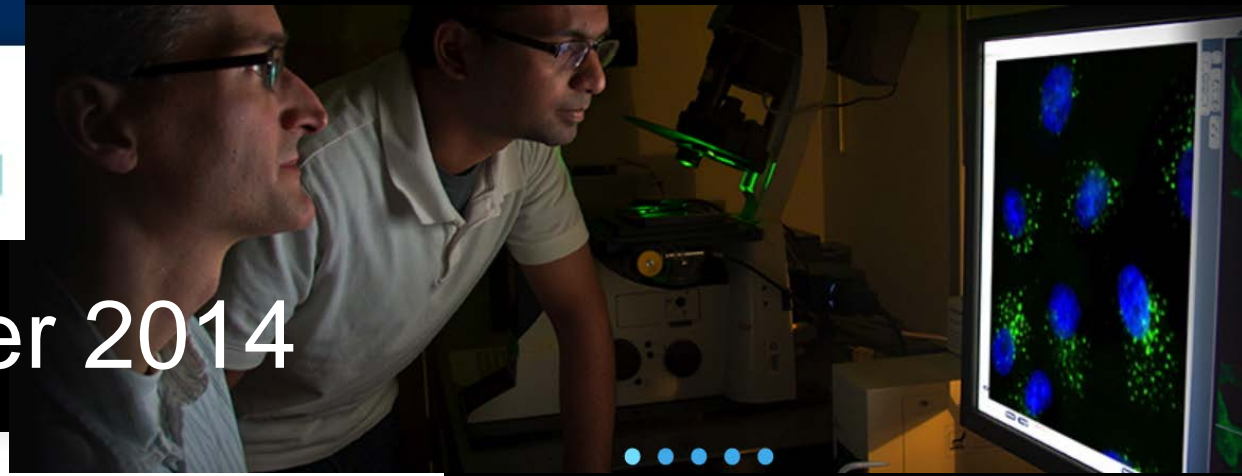- Data at scale : the lab may look like this

Photo Credits:Flickr NASA HQ

University of Pittsburgh

PITT PUBLIC HEALTH

Spring Semester 2014

The Renal-Electrolyte Division
University of Pittsburgh

PITT | SWANSON ENGINEERING
BIOENGINEERING

Center of Excellence
RFID
A University Center within the Swanson School of Engineering

Fall Semester 2014
Biomedical engineering

Department of Biological Sciences

Spring 2015 Immersive Unit

Biological Sciences
Evolution Lab

Evolution

Nothing in biology makes sense except in the light of evolution
--Theodosius Dobzhansky

Evolutionary biology is an exciting field with important implications for all other areas of the biological sciences. Research by faculty in the Evolution Group in the University of Pittsburgh Department of Biological Sciences spans a diversity of topics and organisms, ranging from the molecular evolution of bacterial genomes to the physiological ecology of mammals. A major area of faculty research is in microevolutionary processes and mechanisms, including the evolution of plant mating systems, the roles of selection and genetic constraint in the evolution of separate sexes in plants, and the role of gene interactions in the evolutionary process. The study of macroevolutionary patterns is also a major focus of the group, with emphasis placed on the phylogeny and systematics of birds, reptiles, amphibians, and fossil mammals.
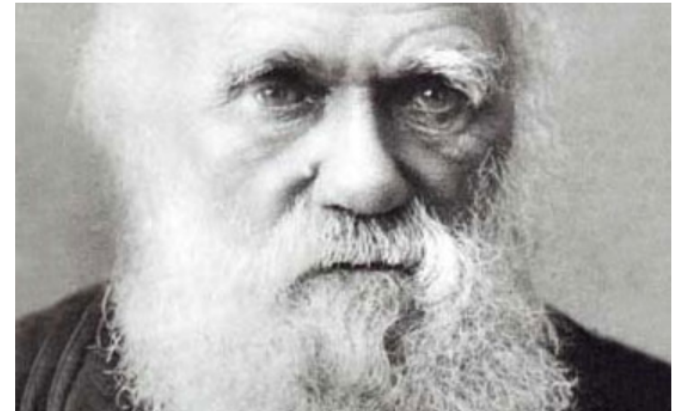
Student feedback

"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."

Faculty / Researcher feedback

"Explaining what one does to a new person is instructive, since it shows you what you do not understand and cannot explain. Discussion with the (LIS) student exposed some weaknesses in my own thinking"

"Learnt more intelligent way to organize the data…"

"Talked about issues with data storage and brought up suggestions …. about other things and most were things didn't know existed so helpful. "
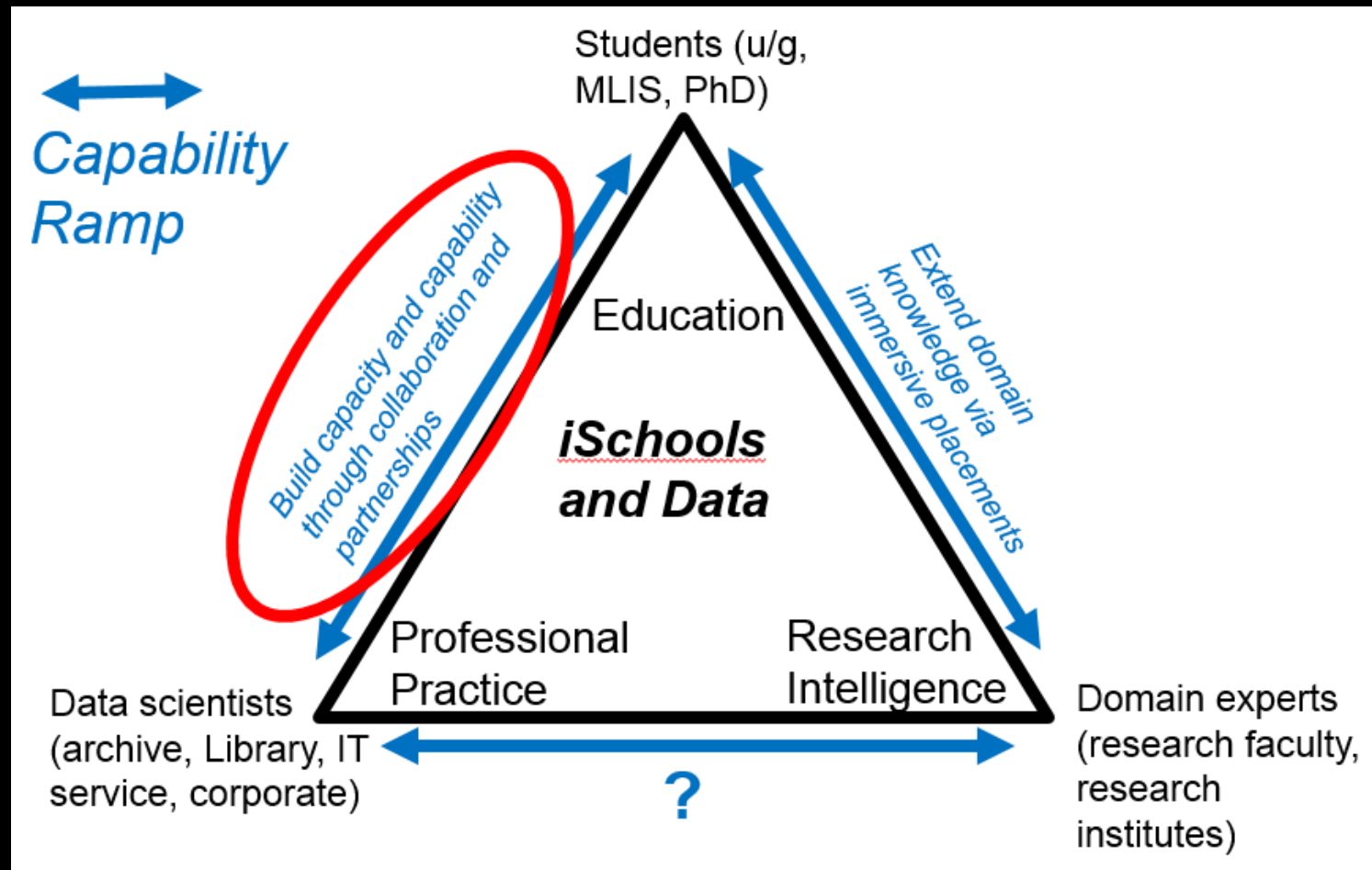
BILATERAL LEARNING

Immersive experience & Laboratory placement
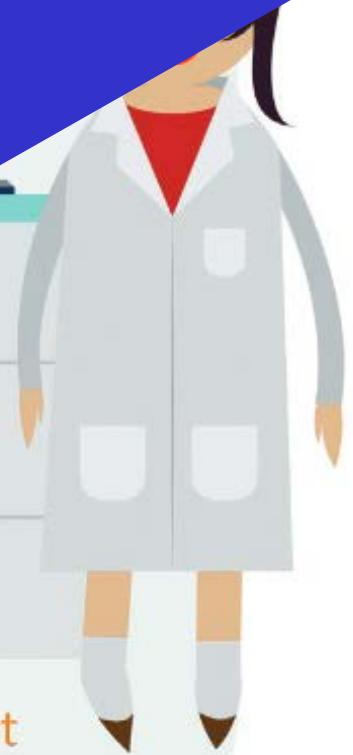
Data curation guidance & support

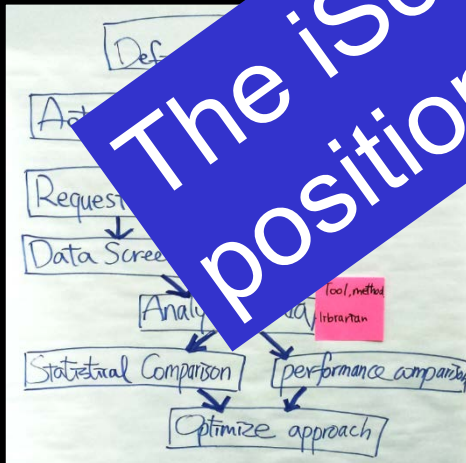The iSchool immersive experience is the Capability Ramp

2. Partnership with University Library System: Creating the Digital Scholarship (Data) Observatory

Digital Scholarship Observatory

- Physical space in Hillman Library
- Two joint faculty appointments between ULS + iSchool
- Developing suite of RDS
- Research to inform service development
  - *IDCC Poster Nora Mattern*
- Workshop

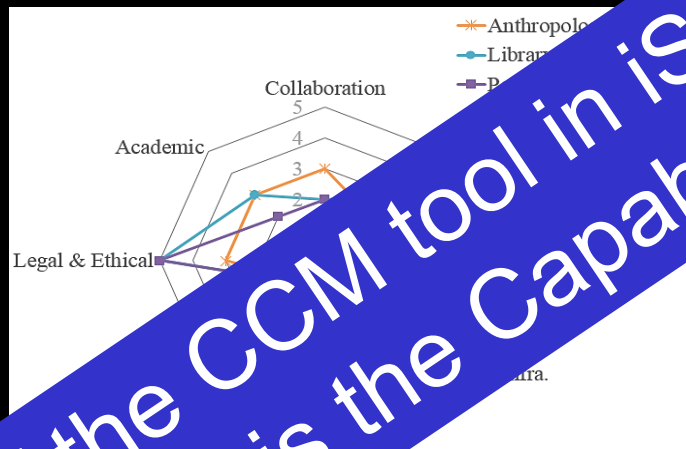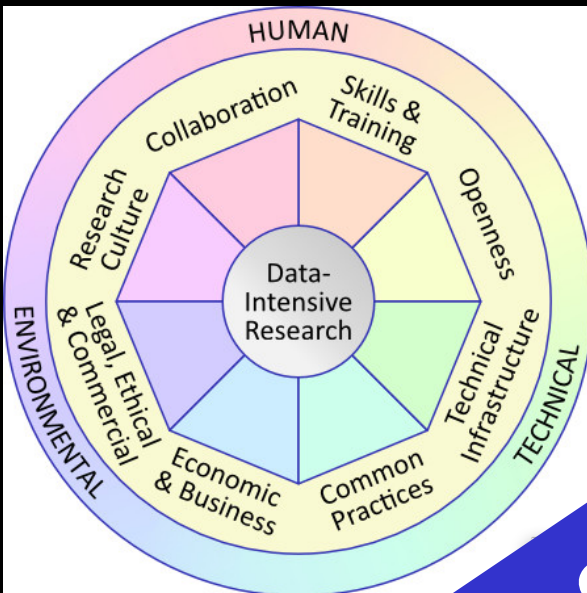**The iSchool joint appointment position is the Capability Ramp**

https://www.flickr.com/photos/kitpfish/4349591762/in/photolist-5RSTi1-58GdLP-7CmN6q

University Library System

Support for Research Data Management

"Manuscripts and records were, more or less, raw bits of data until a researcher 'mined' them to answer a research question. Therefore, libraries and archives had been partners in data management for a long time."
— James L. Mullins, Dean of Purdue Libraries

The ULS recognizes that researchers' datasets are -- like books, journal articles, and conference proceedings -- scholarly output that hold great potential for facilitating and informing future discoveries and knowledge creation. Among the hallmarks of all good research is good research data management practice. In alignment with its commitment to foster scholarly progress, the ULS is building support around research data management. The ULS can help you with the management of your data by meeting with you for a consultation or by connecting you to relevant resources.

We can assist you with:
- Creating a data management plan
- Understanding funder mandates
- Understanding publisher mandates
- Describing your data
- Choosing sustainable formats
- Locating repositories
- Citing datasets

Contact us about:

Creating a data management plan

Thinking about a new project proposal? Submitting an application to the NSF or another granting agency?

Data management begins with a plan. Increasingly funders are requiring that researchers submit formal data management plans with grant applications, but the creation of a data management plan is good practice even if it is not required by a funder.

The DMPTool is a free tool that guides you through the development of data management plans that meet specific funder mandates. Sample DMPs are also available through DMPTool. Researchers can find other useful starting points for thinking about data management planning at the following links:

- Digital Curation Centre (DCC) Checklist for a Data Management Plan
- The Inter-university Consortium for Political and Social Research (ICPSR) Guidelines for Effective Data Management Plans
- MIT Libraries, Write a Data Management Plan
- University of Oregon Libraries, Create a Data Management Plan

All University of Pittsburgh researchers should familiarize themselves with the Guidelines for Responsible Conduct of Research as a starting point for thinking about their data. Librarians at the University of Pittsburgh Health Sciences Library System have developed a useful LibGuide on data management that researchers should take a look at as well!

3. Mapping Disciplinary Data Practice: intelligence gathering to inform Shared Infrastructure and Services
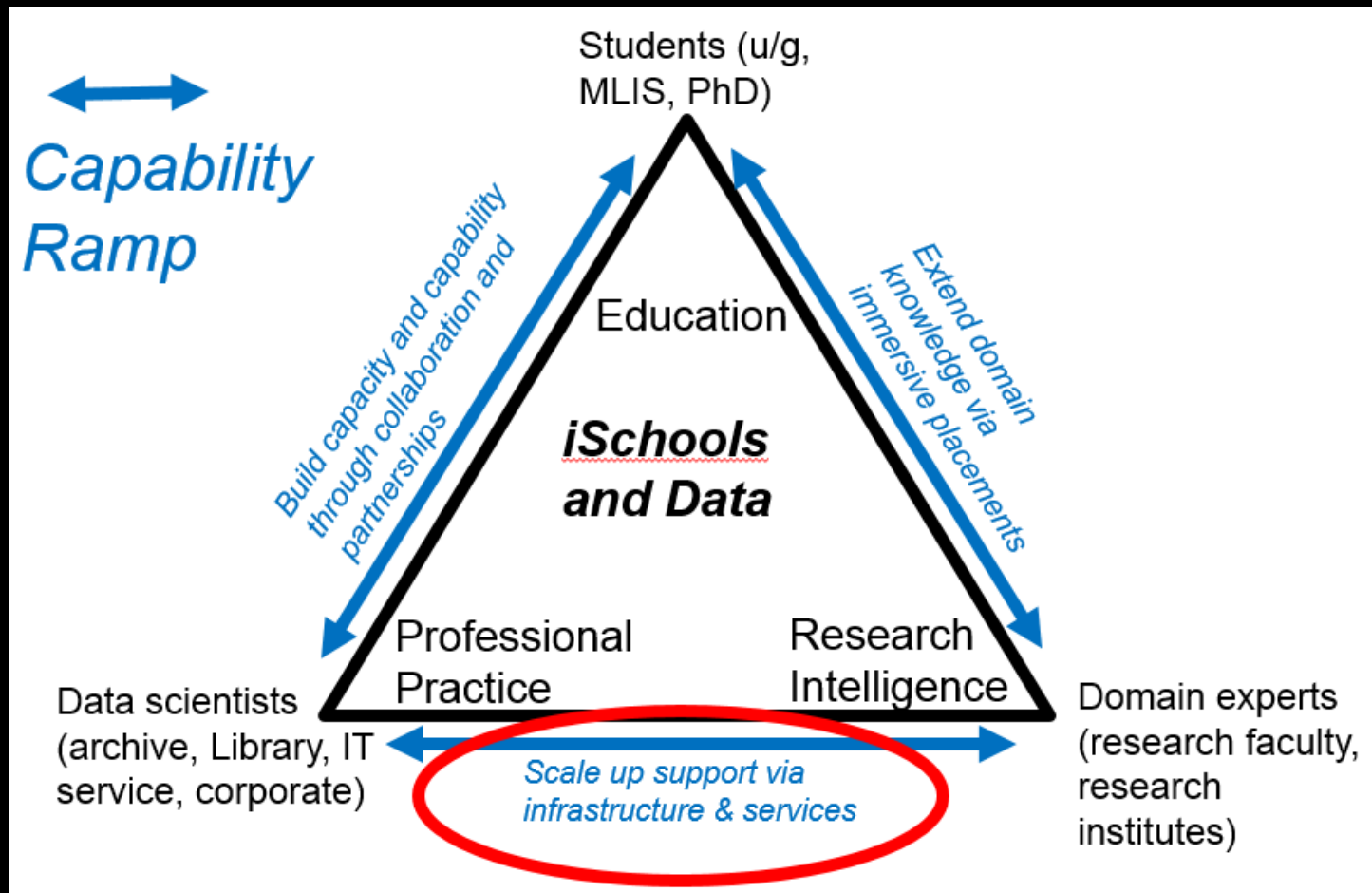
Collecting Community Capability Model Profiles

- Research work as part of a doctoral study
- Understanding disciplinary data practices
- Developing infrastructures
- Results in prep.

Application of the CCM tool in iSchool doctoral research is the Capability Ramp

Microsoft Research

iSchool Futures?

- Data is mission-critical for iSchools
- New data-centric programs emerging
- *"Translational data science: transition of skills, software tools and intelligence from the iSchool to the marketplace"*
- Engagement with the Research Data Alliance
- iSchools are key players in the data space….

Thank you….

IDCC2015

Professor Liz Lyon, School of Information Sciences, University of Pittsburgh   elyon@pitt.edu