

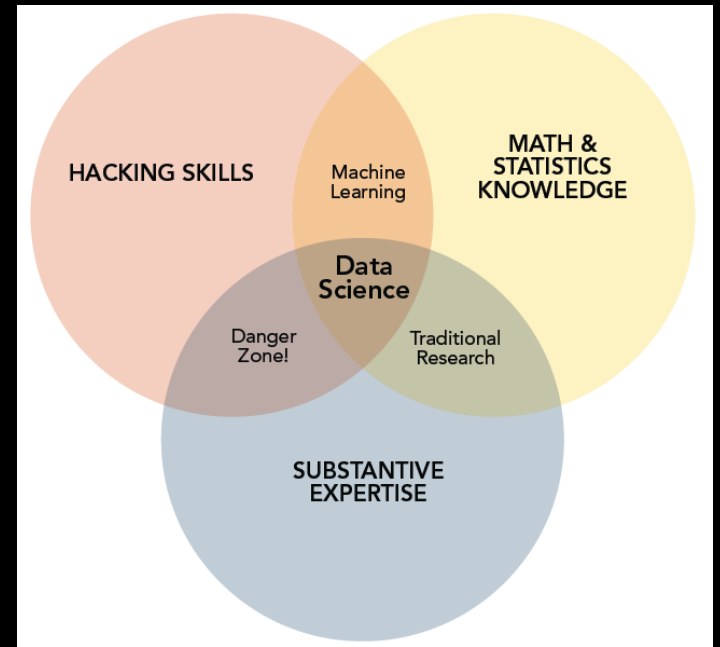
Shifting to Data Savvy: The Future of Data Science in Libraries

Liz Lyon and Chris Erdman

LIBER Conference July 2018, Lille France

Agenda

- Data Savvy Report
- The Management Gap
- Data Science Roles
- The Skills Gap



- Supporting Data Savvy Communities
- Training Strategies
- Library Carpentry
- Help needed!



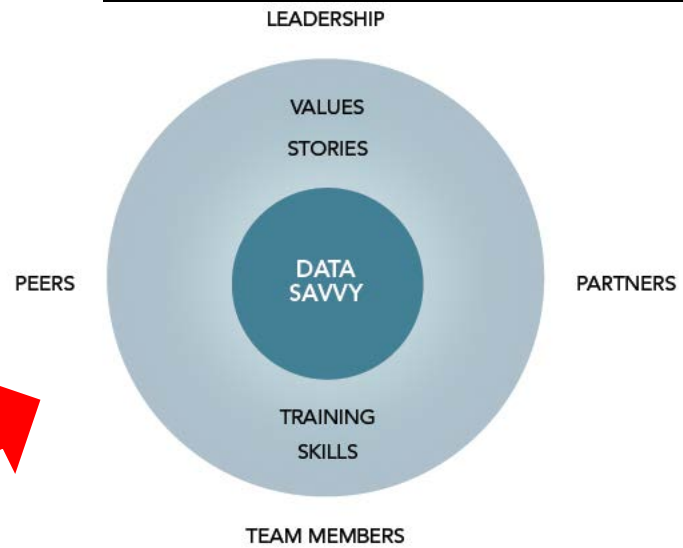
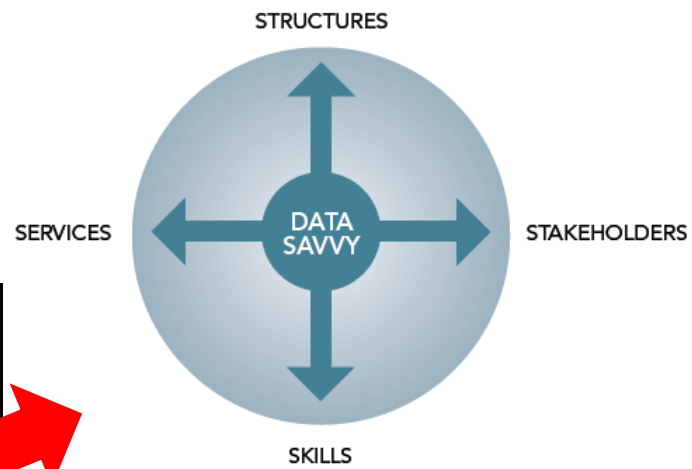
Data Science in Libraries White Paper

<http://d-scholarship.pitt.edu/33891>

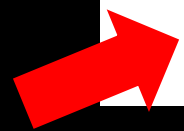
- IMLS-funded study
- Data-savvy concept

Shifting to Data Savvy: The Future of Data Science In Libraries

Matt Burton
Liz Lyon
Chris Erdmann
Bonnie Tijerina



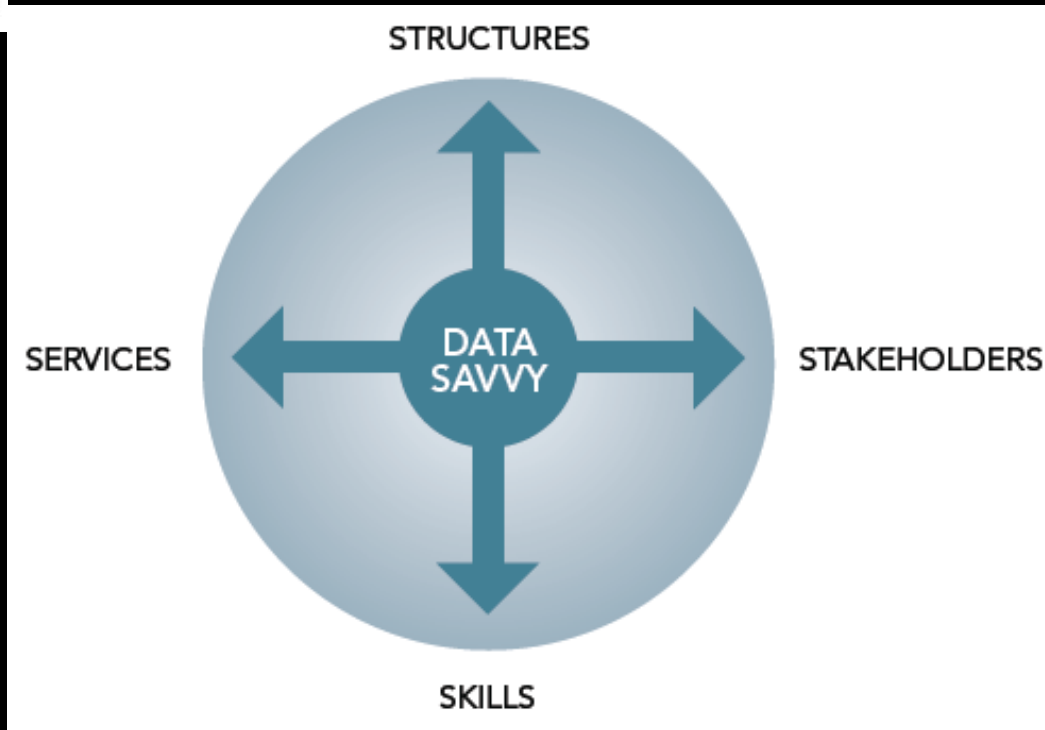
- Multi-faceted framework
- Re-engineering tension
- Credentialing tension
- Data-savvy 360 model organization



Data Science in Libraries White Paper 1

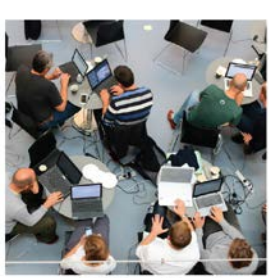


- Skills gap
- Management gap



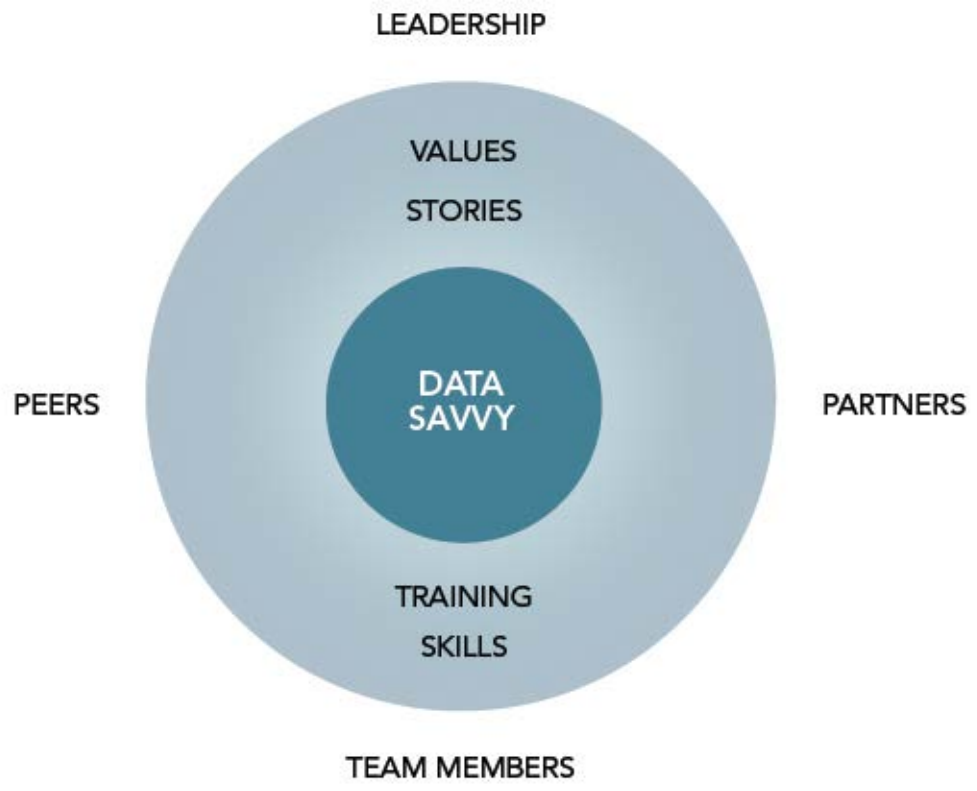
- **Re-engineering tension:** data stewardship vs data science
- **Credentialing tension:** PhD + deep disciplinary knowledge vs MLIS + broad information skills

Data Science in Libraries White Paper 2



Shifting to Data Savvy:
The Future of Data Science
In Libraries

Mike Burson
Liz Lyon
Chris C. Spitzer
Sara S. Tipton



- Skills gap
- Management gap

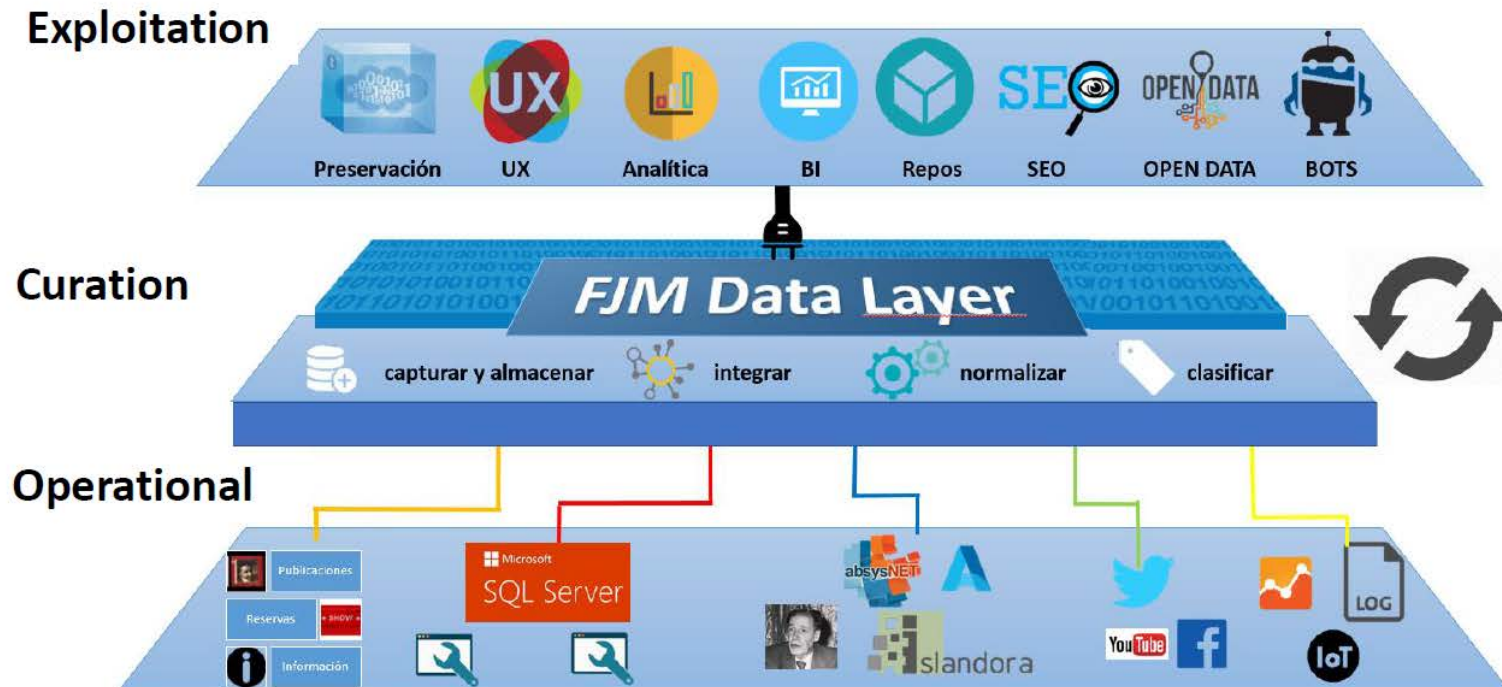
- Service opportunities, internal data, external (research) data
- Leadership, data-savvy teams, collaboration, partners
- Embedded data roles, scaling challenges, credibility, skills

Exemplar

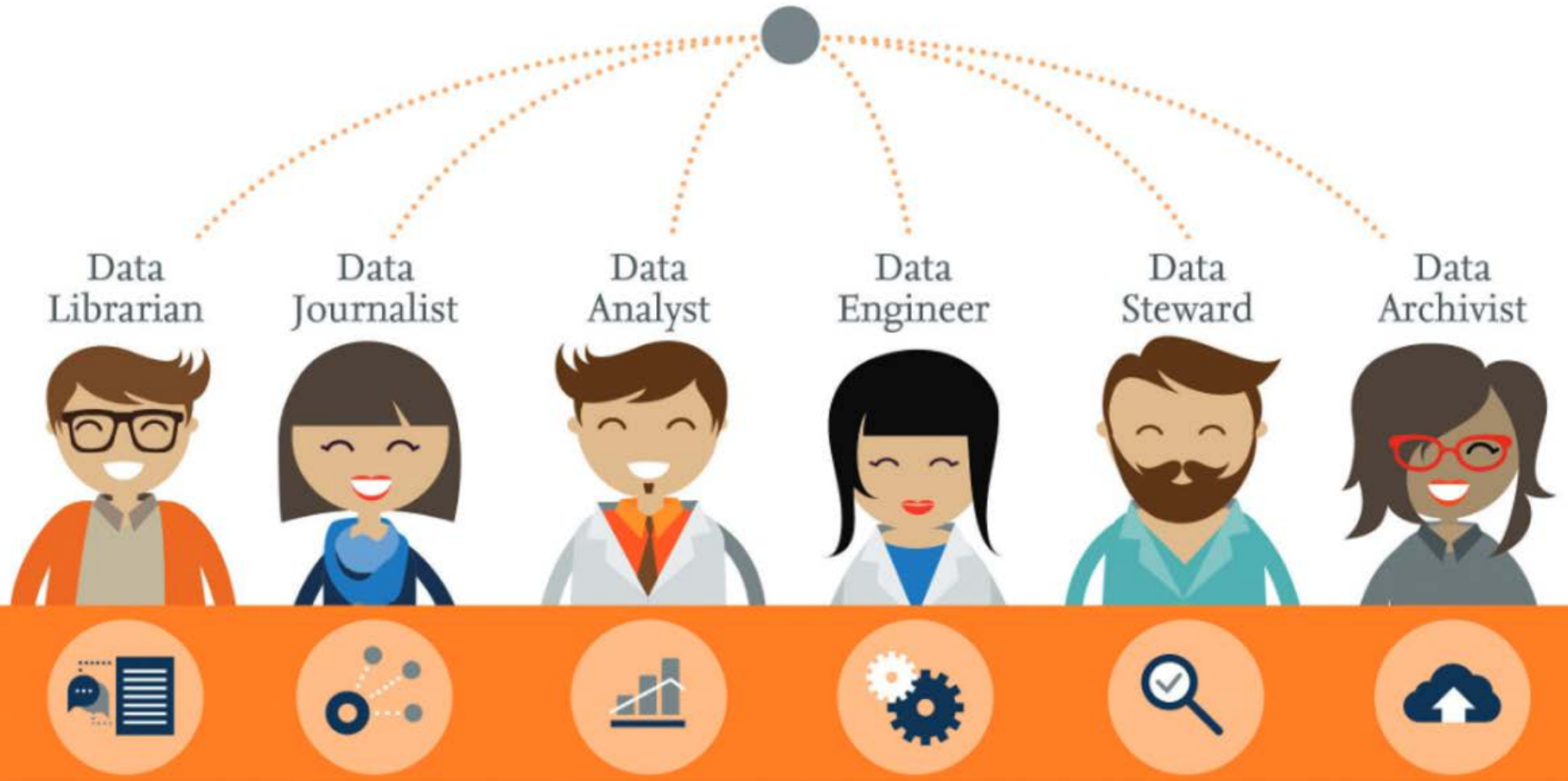
IDCC – Barcelona 21 February 2018

Luis Martinez-Urbe
Data Scientist
DataLab, Library
Fundación Juan March

The DataLab promotes a data centric vision to enable data exploitation via curation



Data-savvy roles

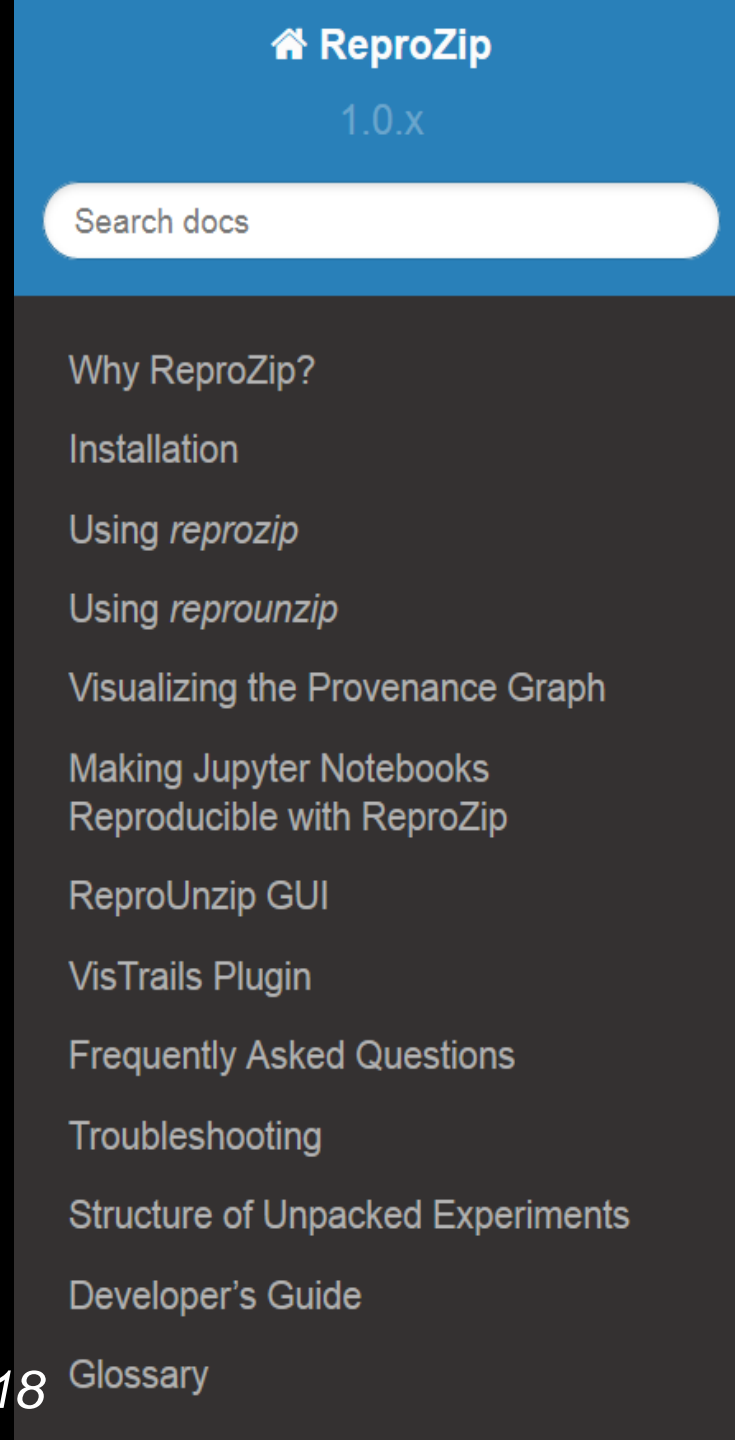


Lyon, L. Learning about research data in the lab at the Pitt iSchool.

<https://libraryconnect.elsevier.com/articles/learning-about-research-data-lab-pitt-ischool>

Next Gen data stewardship: FAIRer data roles?

- *Reproducibility Librarian*
- *Research Data Auditor*
- *Data Verification/Validation Officer*



RDM & Reproducibility Librarian @ NYU (Vicky Steeves!)



NYU Libraries / Research Guides / Elmer Holmes Bobst Library / Data Management Planning / Reproducibility

Data Management Planning: Reproducibility

Information on best practices and standards for data management planning.

- Home
- Data Creation
- Metadata
- Data Sharing and Publishing
- Reproducibility
- Data Management Plans for Grants
- Resources

General Information



NYU LIBRARIES & IT

Location

Staffed Hours: Summer 2017

Contact Us

Reproducibility RSS Feed

Reproducibility

The ultimate goal of data management is the reproducibility of an experiment and the reuse of its results.

Reproducibility of scientific research is imperative: it helps researchers verify results, and it allows others to build on them, advancing the global body of scientific knowledge.

However, with experiments becoming increasingly complex and digital, researchers have to rely on data described in papers, or secondary data if it is supplied. This leaves out data critical to understanding the composition of an experiment: descriptions of column names in tabular data, libraries used in scripting or computational experiments, algorithms used in machine learning, even software used to view files.

Nobel Winner Retracts Research Paper

By KENNETH CHANG
Published: March 7, 2008



A team [Unravels Nobel for 2004](#) ***Nobel Laureate Retracts Two Papers Unrelated to Her Prize***

By KENNETH CHANG SEPT. 23, 2010

Related

- Unravelir Nobel for 2004) Email
- Share
- Web Lir Tweet
- Retraction Reveals a Map in th (Nature) Save
- Original I Reveals a Map in th (Nature) More

[Linda B. Buck](#), who shared the 2004 [Nobel Prize](#) in Physiology or Medicine for deciphering the workings of the sense of smell, has retracted two scientific papers after she and her colleagues were unable to repeat the findings.

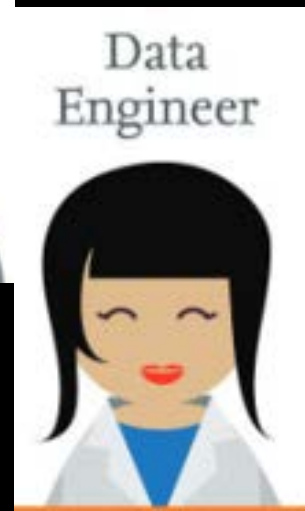
The retractions, which did not concern the work for which Dr. Buck won the Nobel, were published Thursday on the Web sites of the journals where the papers appeared. [One had been published in the Proceedings of the National Academy of Sciences](#) in 2005, the other in the journal [Science](#) in 2006.

"I sincerely apologize for any confusion that its publication may have caused," Dr. Buck wrote in the retraction of the Science paper.

The retractions follow [a separate one, two years ago](#), of a paper by Dr. Buck that was [published in the journal Nature in 2001](#).

Next Gen data science roles?

- Data pool / data lake as analysis substrate
 - Predictive models and simulations
 - Cognitive computing applications
 - Growth of Decision Science
-
- *Cognitive Data Analyst*
 - *Forecast Modeller*
 - *Decision Engineer*



Library Data Analyst (faculty)

- Advertised March 2017
- Virginia Tech
- Promote value of library data
- Collects, analyzes, reports, documents....
- Expertise in SPSS, Stata, R, NVivo etc
- Business Intelligence Systems
- Liaison role with new Office of Academic Decision Support



Repositioning LIS education curriculum

MLIS Redesign: Stewardship and Science

“we need research-heavy librarians”

“we are looking for librarians with skills to support not only information discovery, but also analysis and creation -- in other words those who can provide expertise to students and faculty across the full research lifecycle.”



University of Pittsburgh

School of Computing
and Information

LIS Training opportunities.....

...over to Chris