The iSchool View (Part 1):
Next Generation Data Roles & Pathways

Professor Liz Lyon
School of Information Sciences,
University of Pittsburgh, USA
RDMF17, London
Next Generation Data Roles & Pathways

1. Roles as Goals
   - Roles today
   - Roles beyond 2017

2. Capacity & capability challenges
   - Pathways and pain points
   - Skills

3. Take homes
Context

- iSchools educating prospective data scientists
- MLIS Program
- Curriculum development
- New Data Pathway(s)
A family of data science roles

(Lyon & Brenner IJDC 2015)
Job Analysis Part 1

(Lyon et al iPres Proc 2016)

Data Archivist

- Data documentation
- Data preparation
- Data integration

Data Librarian

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

Web authoring

Data Steward / Curator

- Data governance
- Data quality assurance
- Relational databases

Metadata
- Research
- Disciplinary data
- Statistical analysis

Data collections

Data visualization
Job Analysis Part 2

(Lyon & Mattern (2016) IJDC)

Data Journalist
- Experience as working journalist
- Data visualization
- Graphics creation
- Cartography tools
- Adobe Creative Suite
- CSS, D3, HTML5

Data Analyst
- Statistical analysis
- Data acquisition
- Python
- SQL
- R
- Excel
- Experience of data analysis / working as a data analyst
- Data management
- Domain knowledge
- Data mining

Data Engineer
- Quantified experience
- Experience in data engineering /processing/warehousing
- Experience with large IT /amounts of raw data
- Hadoop/MapReduce/Hive/Pig/ETL
- Data design & modelling
- SAS
Beyond 2017 ???

Some emerging themes....
1. Re-use and RoI

- Legacy data
- Provenance concerns
- Convert/upgrade to a usable resource

- Forensic Data Scientist
- Data Archaeologist
- Data Translator
2. Risk

- Data loss
- Data rescue
- Data triage

- Data Risk Assessor / Analyst
- Data Triage Scientist
- Orphan Data Curator / Archivist / Steward
3. Transparency, Truth & Trust

- Open science practices and workflows
- Retractions, fraud, fabrication
- Research integrity, record of science

- Reproducibility Librarian
- Research Data Auditor
- Data Verification/Validation Officer
- Transparency Agent
  - “Promote, demonstrate and action specific behaviours and practices for Open Science”

Liz Lyon (2016) LIBER Q
4. (Smart) prediction & decisions

- Data pool / data lake as analysis substrate
- Predictive models and simulations
- Driving education, environment, business
- Growth of Decision Science

- Cognitive Data Analyst
- Forecast Modeller
- Decision Engineer
Capacity & Capability Challenges

Pathways and pain points....
Workforce talent gap

- Short course or formal qualification?
- Carpentry or degree
- Agile / lightweight or deep dive
- We need *rapid capacity building*
Employer strategy

- *Recruitment choices: Doctorate or Masters*
- Desirable research methods and deep domain knowledge (eventually)     OR
- Quick to complete but may not address research skills / may be wholly service focused

*** BTW the Pitt MLIS *does* cover research methods and s/w….. ***
Likely graduate entry 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

*Domain disconnect? Data Science roles*
Immersive service delivery – Librarians in the Lab

- In laboratory or clinical setting
- Fully integrated
- Collaborative team science
- Data description & curation
- Data analysis & visualisation

Blend of math and social / power skills

Figure 8: The fastest-growing job areas require both analytical and social skills
US, change in employment skills by skills required, 1980 = 100

Take homes....

- **Scale up (fast)** – feasibility testing / pilot new approaches

- **Cultural change** - resistance to embracing new immersive models of service delivery

- **Competition** – for scarce funds and turf

- **Credibility & Credit** – ‘data scientists’ aspire to collaborate on equal terms with Faculty
“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.”

http://www.ischool.pitt.edu/lis/degrees/mlis-program.php

Thank you....

elyon@pitt.edu