Strategies for Educating Digital Curation Professionals to Harness the Data Deluge -

Seamus Ross, iSchool University of Toronto

9th International Digital Curation Conference
"Commodity, catalyst or change-agent? Data-driven transformations in research, education, business & society"

24 to 27 February 2014
San Francisco, CA
Asked to talk about…

“…[h]ow data is transforming education and training for researchers and professional support staff. Drawing on digital curation and preservation curriculum development work from iSchools and other initiatives.”
Reflected on…

- Work of the DCC (Edinburgh, Bath, Glasgow)
- The DigCCurr family of projects (SILS-UNC)
- Data Curation Education in Research Centers (DCERC) (GSLIS at Illinois)
- DigCurV (EU Project)
- Crawling the websites of iSchool – iCaucus members
A Definition

- Digital Curation an umbrella term –
  - “Digital curation involves maintaining, preserving and adding value to digital research data throughout its lifecycle.” (DCC -- http://www.dcc.ac.uk/digital-curation/what-digital-curation)

- Here we will take a slightly broader view:
  - Digital curation involves maintaining, preserving and adding value to digital material throughout its lifecycle
    - That is it is Digital Curation applies more broadly than to purely research data
A Challenge

- Discussions of digital longevity have been driven by perspective of memory institutions which have not traditionally been in the data business.
- Perspective has shaped the ways we educate folks in digital curation.
- We have focused on preparing them for curating data and not for creating value from it.
Traditionally we have

- “[d]igital curation […] is about maintaining, and adding value to, a trusted body of digital information for current and future use” by adopting a lifecycle approach [22] and by foregrounding the need for “subject description and linkage to discipline-based ontologies […] descriptive information that allows re-analysis of datasets of scientific and scholarly significance”, as a prerequisite to ensuring future “fitness for purpose”.

- From Panos Constantopoulos and Costis Dallas, ”Aspects of a digital curation agenda for cultural heritage”  
  http://www.academia.edu/931035/Aspects_of_a_Digital_Curation_Agenda_for_Cultural_Heritage

Solid case examples of transformative power of data

“By 2018, the United States alone could face a shortage of 140,000 to 190,000 people with deep analytical skills as well as 1.5 million managers and analysts with the know-how to use the analysis of big data to make effective decisions.”

http://www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation
May 2011
The transformative book for me -- 2007
Big Data Market, but Digital Curation Market ????

http://wikibon.org/wiki/v/Big_Data/
Digital Materials larger than data--

Exhibit E2
How open data relates to other types of data

SOURCE: McKinsey Global Institute analysis
...Caught Canadian Sochi Fever...
It reminded me of an article in *Harvard Business Review*


- But I suspect that: Stijn Viaene in an essay about the data science ecosystem as a process of “modeling, discovery, operationalizing, and cultivation got it right when he wrote “It will take a lot of conversation to make data science work. Data scientists can't do it on their own. Success in data science requires a multiskilled project team with data scientists and domain experts working closely together.” *Data Scientists Aren't Domain Experts* (IT Professional, Dec 2013)
LIZ LYON, Data Scientist

- **data engineer** - focus on software development, coding, programming, tools
- **data analyst** – focus on business/scientific analytics and statistics e.g. R, SAS, Excel to support researchers and modellers, business
- **data librarian** – focus on advocacy, research data management / informatics in a university / institute
- **data steward** – focus on long term digital preservation, repositories, archives, data centres
- **data journalist** – focus on telling stories and news”
  
  © liz-lyon-microsoft-escience-chicago-october-2012-final.ppt (downloaded 17 Feb 2014)
An other definition

Gil Press writing in *Forbes* proposed: “A data scientist is an engineer who employs the scientific method and applies data-discovery tools to find new insights in data…. The application (and tweaking) of tools comes from their engineering, or more specifically, computer science and programming background. The best data scientists are product and process innovators and sometimes, developers of new data-discovery tools.'

Characteristics of a Data Scientist

- Domain Knowledge
- Rich Technical know-how
- Ability to apply it creatively
- Collaborative Spirit
- Intellectually Curious
- An imaginative storyteller

- HOW DO WE EDUCATE THESE KINDS OF INDIVIDUALS???
- And when we think we have how do we measure it
Digital Curation is Experimental

“The thing about big data right now is that it is experimental; you have to try putting different data sets together in different ways to see if you can get the intelligence out that you want.”

Virgin Atlantic IT director David Bulman in an interview with Computerworld UK (6 March 2013)
DigCCurr family of Projects

■ IMLS sponsored, SILS-UNC delivered – Lee and Tibbo over 7 years (1)

■ Matrix of Digital Curation Knowledge and Competencies (Overview),
  http://ils.unc.edu/digccurr/digccurr-matrix.htm -- focus is on

(1) Great summary of final meeting -- Alex H. Poole, "Curate Thyself" and the DigCCurr Experts' Meeting: Communication, Collaboration, and Strategy in Digital Curation Education” D-Lib Magazine, May/June 2013, Vol19, No. 5/6
The Landscape - iCaucus

- Data science/databases: 9 programs or specializations. See in particular Berkeley, UIUC, UMichigan, UMaryland, UWashington.
- Many programs offer courses in databases, but not necessarily data science.
- Digital Curation/Preservation: 8 programs with specializations/certificates/modules. See in particular Texas-Austin, UNC, UMichigan, UMaryland, Syracuse, UIUC, Humboldt
- 8 offer courses in digital librarianship or digital records management without offering a full program.

- Very preliminary and grateful to my colleague Andrew Drummond
See tomorrow’s presentation by Laura Molloy, one of the key architects of the DigCurV Framework

© DigCurV Project, http://www.digcurv.gla.ac.uk
Incoming Students

- Great bulk are from Arts and Humanities backgrounds
- Few with STEM domains
- Few come knowing statistics, CS
Knowledge base

- Curation – preservation sense
- Data Analytics
- Visualisation
- Statistics
- Machine Learning -- Automation
- Metadata & Annotation
- Information Architecture
- Data Quality
- Rights, Privacy, security
- Interoperability
- Collection Development
As cited in the Report of the *Blue Ribbon Task Force on Sustainability*
New kind of education experience:

- Empirical, Experimental, and exploratory
- Meeting of Action, method and theory
- Access to Case Studies and Senarios
- Access to diversity of tools and data resources
- Public and private partners supporting placements
- Apprenticeships
DCL at Simmons College

Digital Curriculum Lab

Workspace
- Interactive tutorials
- Course-specific modules
- Scenarios
- Discussion

Tools
- Software applications
- Tool-specific tutorials
- Standards

Content
- Audio
- Visual
- Documents
- E-records & data sets

Repository
- Processed content
- Evaluations
- Reports
- Samples

Next Steps

- Establish a Digital Curation and data science Canon
- Construct a plug and play shared Curriculum model in digital curation and data science
- Create digital curation and data science laboratories which foster exploration and experimentation
- Construct an educational resource data bank which provides resources to support relevant experimentation
- Educate our students in “collaborative laboratory work” as Jahnke and Asher – essential to ensure diversity of approaches (Automation, ML)
- Construct industrial collaborations – access to data, meaningful explorations,
- Educate our students to question established assumptions using exploratory and experimental data analysis techniques
Approaches

- Deliver students a diversity of Learning Opportunities
- Professionals must have a diversity of Abilities – some deeply technical, some social, some cultural.....
- Learning must be continuous – essential to attract students with an inquisitive minds who will see learning as a life long activity
- Pluri-disciplinarity of learning education, training, and experience
- Accreditation of programs and maybe certification of professionals
- Engender a desire in our graduates to engage in information entrepreneurship
Information Entrepreneurship

- Beyond traditional Employer relationships


October 2013
Accreditation & Certification

- Need mechanisms to assess the capabilities of digital curation professionals – whether they are data curators or data analysts
- Is accreditation of programs sufficient?
- Or should we certify professionals?
- What sorts of continuing and professional development should digital curation professionals need to take and how often?
- We need our credentials to be GLOBAL
In Conclusion

Future in Digital Curation Education is:

- Delivering meaningful and engaging exploratory empirical experimentation
- Attracting a broader range of students especially increasing the numbers from STEM
- Increasing the technical capacity of our graduates
- As an educational profession all our schools have a long way to go to deliver this educational capacity, but we have a tradition of creating intellectual change in Information domains
- To control the landscape we must drive the agenda
- We are way behind the curve --
© Canadian Hockey Men's Olympic Gold