



Librarians and data

Andrew Cox, a.m.cox@sheffield.ac.uk,
Information School, University of Sheffield

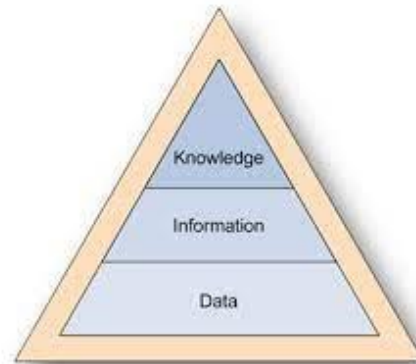


Image source: Adlib software
https://www.adlibsoftware.com/hubfs/Imported_Blog_Media/oilnewdata-blog



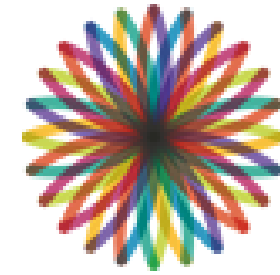


The
University
Of
Sheffield.

Information
School.

www.shef.ac.uk/is

A member of the
iSchools network





The Information School University of Sheffield, UK



Number 1 in the World 2021
for Library and Information Management.
QS World University Rankings by Subject 2021

Athena Swan Bronze 2021
Committed to Gender Equality.



Research data management



What is (research) data?

- Historical documents
- Interviews and focus groups and questionnaires
- Government survey data
- Field measurements or sensor data
- Results from experiments
- Simulation data
- Log data
- Images (e.g. brain scans)
- Moving images
- Etc ... etc
- Varied: quantitative and qualitative; print and digital; big and small.
- Complex and changing – not a static “thing”
- Not necessarily called “data”
- Ownership contested



2018 survey results:

Services Currently Provided by libraries

Ranking of services (providing any service – Basic, Well-developed, Extensive): Advisory rather than technical services predominate

1	Promote awareness of reusable data sources, such as data archives	83%
2	Offer advice on copyright and/or intellectual and/or licensing property rights relating to data and data management	81%
2	Data management training and/or data literacy instruction (e.g. to research students, early career researchers etc.)	81%
4	Maintaining a web resource/guide of local advice and useful resources for RDM	79%
5	Data Management Planning (DMP) advisory service	76%
5	Offer data citation advisory services	76%
7	Offer data publication advisory services	75%
8	Provide support for search and retrieval of external data sources	73%
9	Offer data storage advisory services	68%
10	Run a data repository/archive/store	67%
...		
24	Offer an advisory service on data mining	23%
25	Analyse and visualise datasets using Python scripts, SPSS, R and MS Excel software	21%
26	Rescue legacy data or perform data triage or forensic data recovery	16%

Source: Cox et al., 2019



The data role spectrum

Familiar <					> Unfamiliar					
Support for data search / access to data	Data literacy training and promoting awareness	Data collection management, including metadata	Gathering support requirements for services/tools	Data policy	Data Management planning advice	Data carpentry	Data curation	Data integrity	Embedded roles in a research team	Data analysis and visualisation

- *Close to existing roles*
- *Resources required*
- *Demand*

- Libraries are enterprising
- They do “easy” things first
- Importance of collaboration
- Development is quite slow
- Multiple models of “maturation”



Data and Artificial Intelligence



AI, premised on data

- AI as a set of functions: *Auto-suggest... auto-correct... grammar tools... recommendation...search ranking...captioning...auto-summarisation... translation tools...*
- AI as a set of technologies: “business analytics and data science; natural language processing, speech recognition and text to speech; machine learning, deep learning and neural networks; machine reasoning, decision making and algorithms; computer vision; and robots and sensors” (Gartner group)
 - Data could be speech, free text, images... training data
- AI as big Tech and associated industrial complex: Power of FAAMG (Crawford, 2021) – reliance on “cheap data”, cheap labour, material resources and energy
 - Our activity data extracted



Ethics of AI

- Bias – *biased historic data leads to biased outcomes however good the algorithm*
- Transparency, explainability and accountability
- Privacy – *risk of reidentification when data combined*
- Safety and security
- Human choice



Libraries and AI

Precursors: Text and data mining, digital humanities, data literacy, learning and library analytics, supporting data science

Limits: Cost, lack of scale of data

1. Knowledge discovery applications
 - Helping researchers use data – advice on copyright, help finding data/tools – creating a community
 - Procuring systems
 - Data and AI literacy within IL
 - Collections (special collections, print) as data
2. Chatbots – automation of Q&A – lack of scale of data, because questions too sparse/ varied?
3. Robotic Process Automation (RPA) – automation of routine admin
4. Managing users



4. Understanding or managing users

- Libraries are rich in data about users
 - Turnstile data
 - Circulation data
 - Usage of digital resources
 - Satisfaction surveys
 - Reference enquiries
 - Qualitative data, eg UX studies
- AI to analyse social media data or open text survey data
- AI to nudge users?
- Driver is metrification and demand for evidence of impact

Learning analytics debate

- Lack consent or student awareness of how their data is being used
- Lack ethics review
- Libraries few responsible use statements
- Benefits unclear or for institution not learner
- Privacy issues
- Chilling effect on free speech and expression

Jones et al. (2020)

- Issue of validity: “learning data” r learning



What librarians can bring to data societies

Skills and knowledge, attitudes, values and vision



AI's impact on jobs in libraries

- Replaced
 - Dominated
 - Divided
 - Complemented
 - Augmented
 - Rehumanized
(GPAI, 2020)
- As a profession premised on literacy (McKinsey, 2018)?
 - Already busy... do we need to write code?... or become data scientists?
 - Which direction are we heading in?



Skills and knowledge

- Collection management, including metadata, standards, IPR etc – *data not print collections*
 - *Importance of data provenance for validity*
- Procuring data and systems – *different types of system*
- Searching for data – *new landscapes of search*
- Teaching data & AI literacy – *new dimensions of literacy*
- Knowledge of users' need – *assembling data from wide range of sources*
- “Computational sense” (Twidale and Nichols, 2008)
- Contribution as researchers (Evidence base, 2021)

NOT (necessarily) data science (= computational thinking / statistics / domain knowledge)



Attitudes

- Service focus / balanced with sense of institutional agendas
- Collaborative skills / Influencing skills
- Commitment to professional development and learning
- Professional knowledge sharing



WEF (2020) – Ten skills to thrive in the 4th Industrial revolution

1. Complex Problem Solving
2. Critical Thinking
3. Creativity
4. People Management
5. Coordinating with Others
6. Emotional Intelligence
7. Judgement and Decision Making
8. Service Orientation
9. Negotiation
10. Cognitive Flexibility



Values

Example: CILIP's ethical statement

1. **Human rights, equalities and diversity**, and the equitable treatment of users and colleagues
2. The public benefit and the advancement of the wider good of our profession to society
3. Preservation and continuity of **access to knowledge**
4. **Intellectual freedom**, including freedom from censorship
5. Impartiality and the **avoidance of inappropriate bias**
6. The **confidentiality** of information provided by clients or users and the right of all individuals to privacy
7. The development of **information skills** and information literacy

- Understanding of how inequality reproduced?
- Relevance of sustainability, eg need for green AI
- Consideration of global South perspectives, eg issues around language and operating in low resource environments
- The ultimate drivers for datafication lie beyond our



The ethical framework for AI in education (2021)

1. AI should be used to achieve well-defined educational goals based on strong societal, educational or scientific evidence that this is for the **benefit of learner**
2. AI should be used to assess and recognise a broader range of learners' talents.
3. AI should increase the capacity of educational institutions whilst respecting human relationships
4. AI systems should be used in ways that **promote equity** between different groups of learners and not in ways that discriminate against any group of learners
5. AI should be used to increase the level of **control that learners have** over their learning and development
6. A balance should be struck between **privacy** and the legitimate use of data for achieving well-defined and desirable educational goals
7. Humans are ultimately responsible for educational outcomes and should therefore have an appropriate level of **oversight** of how AI systems operate
8. Learners and educators should have a reasonable **understanding** of artificial intelligence and its implications
9. AI resources should be designed by people who **understand the impacts** these resources w



An AI vision – the paradigm of the intelligent library

From searching to find a text to read

To interacting with the full text of the library collection

Cox, Pinfield and Rutter (2019)

The living systematic review





Your thoughts and questions!

Andrew Cox, a.m.cox@sheffield.ac.uk,
Information School, University of Sheffield



The impact of AI, machine learning, automation and robotics on the information professions: A report
<https://www.cilip.org.uk/general/custom.asp?page=researchreport>



References

- Cox A (2021) The impact of AI, machine learning, automation and robotics on the information professions: A report for CILIP <https://www.cilip.org.uk/general/custom.asp?page=researchreport>
- Cox A, Pinfield S, Kennan MA, Lyon L & Saffi L (2019) *Maturing research data services and the transformation of academic libraries*. *Journal of Documentation*.
- Cox, A. M., Pinfield, S., & Rutter, S. (2019). The intelligent library: Thought leaders' views on the likely impact of artificial intelligence on academic libraries. *Library Hi Tech*. 37 (3) 418-435.
- Cox A. and Verbaan E. (2018) *Exploring Research Data Management*. Facet.
- Crawford, K. (2021) *Atlas of AI*. Yale University Press.
- Evidence Base (2021). *The role of academic and research libraries as active participants and leaders in the production of scholarly research*. RLUK, <https://www.rluk.ac.uk/portfolio-items/rluk-scoping-study-report/>
- Global Partnership on Artificial Intelligence (2020). *Working group on the future of work*, <https://gpai.ai/projects/future-of-work/>
- IFLA (2020) *IFLA statement on libraries and artificial intelligence* https://www.ifla.org/wp-content/uploads/2019/05/assets/faife/ifla_statement_on_libraries_and_artificial_intelligence.pdf
- Institute for Ethical AI in Education (2021). The ethical framework for AI in education, <https://www.buckingham.ac.uk/research-the-institute-for-ethical-ai-in-education/>
- Jones, K. M., Briney, K. A., Gobin, A., Salo, D., Asher, A., & Perry, M. R. (2020). A comprehensive primer to library learning analytics practices, initiatives, and privacy issues. Jones, KML, Briney, KA, Gobin, A., Salo, D., Asher, A., & Perry, MR, a Comprehensive Primer to Library Learning Analytics Practices, Initiatives, and Privacy Issues. *College & Research Libraries*, 81(3), 570-591.
- McKinsey Global Institute. (2018). *Skill Shift: Automation and the Future of the Workforce* (Discussion Paper, May 2018). In McKinsey & Company (Issue May). [https://www.mckinsey.com/~/media/McKinsey/Featured Insights/Future of Organizations/Skill shift Automation and the future of the workforce/MGI-Skill-Shift-Automation-and-future-of-the-workforce-May-2018.ashx](https://www.mckinsey.com/~/media/McKinsey/Featured%20Insights/Future%20of%20Organizations/Skill%20shift%20Automation%20and%20the%20future%20of%20the%20workforce/MGI-Skill-Shift-Automation-and-future-of-the-workforce-May-2018.ashx)
- Twidale, M. B., & Nichols, D. M. (2006). *Computational sense: The role of technology in the education of digital librarians*. <https://researchcommons.waikato.ac.nz/handle/10289/40>
- World Economic Forum (WEF) (2020). The ten skills you need to thrive in the 4th Industrial revolution <https://www.weforum.org/agenda/2016/01/the-10-skills-you-need-to-thrive-in-the-fourth-industrial-revolution/>