

Information landscapes, technological ecosystems, enterprise discoverability, humanities open data conundrums and web search analysis for the public good:

Reflections on a pandemic

Gary Horrocks, editor – eLucidate

Little did I know as I completed the editorial for volume 16, issue 2 of eLucidate back in spring 2020 that the world was about to experience the turbulent descent into the maelstrom of the devastating Covid pandemic. Face-to-face courses, meetings and workshops in London were hastily abandoned as UKeiG explored the potential of online delivery by sponsoring three successful CILIP ‘Webinar Wednesdays’ on the topics of open access monographs, natural language processing and advanced search commands.

At this point in the international health crisis new-fangled Zoom was merely an onomatopoeic expression

UKeiG colleagues battened down the hatches and began to formulate a CPD programme that would work in a virtual environment.

Our annual Members’ Day in June 2020 also transitioned to online and continues in this format. Lucy Mckenna from the ADAPT Centre, Trinity College Dublin, spoke to her eLucidate paper on [‘NAISC-L: a linked data interlinking tool for information professionals.’](#) The tool is designed to meet the requirements of the library, archive and museum domain improving linked data accessibility. ‘In the context of NAISC-L, interlinking describes the task of connecting entities, such as people, places, objects or works, held in a dataset to related or identical entities held in another. Such interlinks have the potential to expose collections to larger audiences and to allow for richer user searches.’

Open data practices in the humanities

This segued beautifully into the topic of creating an open data science ecosystem. Mark J. Hill, representing the Computational History Research Group at the University of Helsinki, presented on the conundrum of ‘Open Data Practices with Closed Data Sources’, a personal reflection on the emergence of open science in the humanities.

Applying the principles of open data to the paradigm of subjects like history, literature, theology and philosophy is a difficult and complex process. Semantics alone is an issue.

What is humanities 'data' in a discipline that generates photographs, art, geospatial entities, musical scores, poetry, as well as journal articles, long form publications like monographs and other forms of scholarly communication? Access is also limited. Vast amounts of information have yet to be digitised; stored in yet to be excavated archives. Humanities data is derived from the spiritual as well as temporal. with significant qualitative, as well as quantitative relevance and value. It requires considerable intellectual analysis and value addition to render it discoverable and to be able to cross reference it with complementary sources of information from other disciplines and domains.

Hill noted that the drive to apply open science to the humanities has catalysed the convergence of significant international expertise across a spectrum of disciplines including history, linguistics, geography, philosophy, computer and data science. There are numerous initiatives underway to collaborate with colleagues across the world to overcome the reluctance, even apprehension amongst some researchers, to liberate their datasets and create 'a virtuous cycle of better data and better science through open practices.' He focused on two specific library catalogue projects, the English Short Title Catalogue and Eighteenth-Century Collections Online, highlighting the work that goes into bibliographic database projects from data cleansing to data interpretation, understanding and visualisation, contextual harmonisation and enrichment.

The [English Short Title Catalogue](#) (ESTC) has an offline version not publicly available where the raw data is scrutinised and transformed. Considerable efforts are made to disambiguate scholarly output that has multiple editions, editors and locations of publication. A typical workflow entails the conversion of raw data from XML to CSV, the extraction of relevant information depending upon the context of the research (the physical details of document; the authors, collaborators, publication history - print location and year of publication - for example.) The data is harmonised and enriched with data from other open sources, generating new research outputs and datasets.

The [Eighteenth Century Collections Online](#) (ECCO) project's goal is to 'combine open data with closed data in a reproducible manner to provide answers to researchers cautious or sceptical of the accuracy of the data they are using.' The harmonised metadata is then combined with [full-text sources](#) (ESTC and ECCO) to enable text analysis and mining in powerful and innovative new ways. 'Using full text to enrich metadata feeds back into the loop by providing better quality data and detecting subjects, topics and even new forms of data.'

Open access data in the humanities is very much a goal, through 'grey areas remain'. Extensive intellectual efforts have been invested in creating an infrastructure to encourage the dissemination of datasets. Where researchers are reluctant to release their closed data, they are sharing their code, research methods and data analysis tools. 'Data manipulation processes should be open and reproducible. The aim is to make all this data available in one

form or another at some point which would create a rich source of historical material and information available to researchers.'

Analysing web searches for public good

In November 2020, UKeIG hosted the sixth annual Tony Kent Strix Award memorial lecture - 'Analysing Web searches for public good: inferring the health of populations and individuals' - presented by 2019's worthy winner Professor Ingemar J. Cox, Department of Computer Science at University College London. The prestigious international award is given in recognition of an outstanding practical innovation or achievement in the field of information retrieval and was inaugurated in 1998 by the Institute of Information Scientists. It is now presented by UKeIG in partnership with the International Society for Knowledge Organisation UK (ISKO UK), the Royal Society of Chemistry Chemical Information and Computer Applications Group (RSC CICAG) and the British Computer Society Information Retrieval Specialist Group (BCS IRSG).

In a General Data Protection Regulation (GDPR)/data privacy context we're all cynically aware of cookies and how our search histories and website behaviour can be utilised to 'develop and improve' personalised services, determine search results and target advertising, but retracing digital footprints can also provide a wealth of invaluable health information at both the population and individual level. The data presented in Google and Twitter searches has immense epidemiological value and can surface signs of a range of issues including adverse drug reactions and the prevalence, spread and virulence of a disease. It can also be utilised to assess the effectiveness of national public health interventions like vaccination programmes.

Professor Cox presented on research efforts to harness the power of search in a public health context, giving a celebratory and collaborative nod to team colleagues at University College London (UCL), Public Health England (PHE), the Royal College of General Practitioners (RCGP), Microsoft Research and Google. PHE define syndromic surveillance as: 'The process of collecting, analysing and interpreting health-related data to provide an early warning of human or veterinary public health threats, which require public health action.' Researchers can infer significant demographic information from search histories, including gender, age, sexual orientation, religion, politics and economic indicators (a universal credit claim as opposed to a house price search, for example), but what of health information?

It's common knowledge that Doctor Google is the source of billions of inquisitive and hypochondriacal health-related enquiries, but analysing and attempting to interpret the motivations and behaviours that initiate these digital trails can be perilous. There is immense potential for the early detection of a disease outbreak, an influenza or Covid hotspot, if there is a correlation between an increase in web searches, their geographical

location and a rise in confirmed cases. A cluster of searches for non-prescription cold remedies, for example, may forecast the prevalence of illness in the community.

Disambiguating search queries and inferring conclusions from them is a fallible process that carries its own health warning. Researchers need to account for interest rather than infection. A news story or documentary on BBC Radio 4 may spike curiosity. Are you sick, or just intrigued? Timing is key. An adverse reaction to a drug may be short term problem for an individual, but how do you track a wider population crisis six months down the line?

The potential for identifying people at risk of experiencing a 'health event' based on their search histories is huge.

'Online behaviours don't cause diseases; but online behaviours are indicative of offline behaviours.'

Could frequent searches for fast food restaurants be associated with an increase in the risk of obesity, cancer or myocardial infarction, or searches for pornography be a risk marker for predicting the likelihood of sexually transmitted infections?

Professor Cox posed a fundamental question: what is reasonable inference? The anonymisation of search engine query data is crucial for obvious legal, ethical and privacy reasons. Yet the public health benefits that the data provides are invaluable and lifesaving.

Defining and developing the enterprise search experience

Martin White, Principal Analyst at [SearchResearch Online](#) opened the 2020 Strix Memorial Lecture with a presentation on 'Defining and developing the enterprise search experience', where he articulated the challenges of corporate information retrieval based on projects he had undertaken over the last two decades. 'There is a great deal of interest at present around the topic of the digital employee experience (DEX). Enterprise search is a very important discovery and integration platform, yet little attention (and no research) has been paid to the user experience in this area.'

Martin has written extensively on [enterprise search management](#), including his definitive July 2020 report [Managing Enterprise Information](#), a guide to good practice that embraces the value of information, assessing information risk, governance, information behaviours, working in teams, information discovery, knowledge management and records management.

International organisations generate massive amounts of structured and unstructured multilingual content, 'very little of which is curated for quality.' Value-added 'discoverability' interventions like indexing, metadata and abstracting are scarce, alongside the development

of metrics to measure the impact of information on business-critical objectives. 'Enterprise search is different to web search', asserts White. Employees must negotiate a complex organisational terrain of internal information that is often pushed to them by a multiplicity of enterprise applications including resource planning and customer relationship management systems, human resources applications and corporate email alongside external web content and social media. Version control of business documents is a key issue. There are committee minutes and actions, products and project management information, policy briefs, working papers, technical notes, annual reports and relentless news and updates.

White argues that technology is not the issue, that 'IT is neither user nor ornament' in this context. Staff are not concerned with the enterprise search technology but with a successful interaction and query resolution. Trust and overall search satisfaction is fundamental. 'It doesn't matter how sophisticated the system is in terms of the features and functionality. What matters is that the user can make an informed judgement based on retrieved content that best serves their information need, reinforces their trust in the application, maintains the highest possible level of overall search satisfaction and supports the achievement of business and personal objectives.' The development of 'user cordial interfaces' that are quick, intuitive, flexible and task oriented, presenting relevant and current information, should be the key priority.

White articulated his vision up to 2025. 'Search technology has matured over the last forty years to the point where innovation in text indexing will not result in a significant increase in relevance in managing enterprise content. As repositories of text, data, image and video increase in depth and complexity, federated search applications delivering measurable impacts on organisational performance will be in demand and will be much easier to justify. This demand will be satisfied by search suites and by search-based applications. To maximise the adoption and value of these applications, novel approaches to task specific interaction through the user interface will develop rapidly. Delivering the best possible enterprise search experience will be a key differentiating factor between competing vendors.'

Managing the information landscape

The UKeIG annual Members' Day in July 2021 included a fascinating presentation from Dr Andrew Whitworth, Senior Lecturer at the Manchester Institute of Education, University of Manchester, UK. 'Managing the information landscape' was based on his book [Mapping Information Landscapes - New Methods for Exploring the Development and Teaching of Information Literacy](#) (2020, Facet Publishing). His research interests lie with professional and staff development, particularly around the areas of information management, information literacy, information and communications technology and distance education.

He acknowledged that he had been inspired by Professor Annemaree Lloyd's 2010 book 'Information Literacy Landscapes'.

In an homage to his love of maps he deployed them as an ingenious geospatial metaphor to emphasise their role in the development of digital literacy. You really need to be grounded, to have a sense of place and context, to be information literate.

What makes a map a map? 'They aren't confined to physical geography. We can use them to situate ourselves within any landscape. Maps allow learners to develop a collective understanding of what information is important to them, to represent this understanding graphically or discursively, and to communicate it to others.' An organisational chart, for example, maps roles and responsibilities. Mind and concept mapping enables the diagrammatic representation of the association between topics and sub-topics. Whitworth introduced the concept of 'psychogeographics'; mental mapping that stimulates exploration and reflection.

In the context of digital information landscapes maps provide global overviews and communicative spaces that offer perspective and insights that enabling you to engage with assets and resources.

'A map is an aesthetic device that enables exploration, triggers directions and ideas. Why can't the search process be the same?'

He identified the subject domain as a landscape inhabited by resources, solutions and 'clusters of knowledge'. We are obliged to judge their quality and relevance and make decisions to rectify an information deficit, in much the way we would with an A-to-Z street atlas or satellite navigation application.

The library technology ecosystem

At the UKeig Members' Day in June 2022, Ken Chad, of [Ken Chad Consulting](#) presented on 'The library technology ecosystem. Where are we now. What does the future hold?' He has worked on [trends in the library technology market](#) for many years and his focus on this occasion was on the higher education (HE) sector and the two core revenue generating activities that drive change in HE: teaching/learning and research. 'Libraries are judged primarily by their ability to support those activities. In the UK, an institution's performance in the Research Excellence Framework (REF) and Teaching Excellence Framework (TEF) are critical to funding.' He cited SCOUNL's 2017 report '[Mapping the Future of Academic Libraries](#)' (Stephen Pinfield, Andrew Cox and Sophie Rutter) and referenced the theme of landscapes, confusing and complicated in the HE context, and in a constant state of transition and flux. Institutional repositories of open access research outputs were morphing into Research Information Systems (RIS); aligned to wider institutional goals

supporting the research information lifecycle including research funding, research impact (bibliometrics/altmetrics), open science, AI-based resource discovery, open and linked data and the emergence of university presses and the library as publisher. 'There is a growing awareness of how RIS workflows are increasingly intersecting with those in the library.' Chad argues that the opportunities are endless for the library and information science community, enabling us to 'draw a new map of support and services for researchers', working closely with professional services and research support teams across the organisation.

The 'marketised' teaching and learning HE landscape is equally complex. 'Students as consumers are impatient if the university cannot deliver the resources they need, and there is a dichotomy between the provision of library resources and learning resources.' It's increasingly difficult to identify what constitutes a resource that is managed (and paid for) by the library and what is an educational resource and paid for from a learning budget. Within this context, what future does a traditional library management system (LMS) have in delivering an increasingly diverse and eclectic portfolio of learning resources? Google also continues to dwarf conventional library systems.

Like research, the library and information services role in supporting teaching excellence and the student experience is integrating into a much wider learning paradigm. AI, data and learning analytics offer huge potential to map student behaviour, tracing their digital footprints to determine support needs but also to evidence the value and impact of library and information services on student achievement and success. Privacy issues loom if a student is unaware that their data is being processed, regardless of the purpose or outcome.

Chad poses a challenging question as all of these research and learning systems and services assimilate. Who pays for the ecosystem?

'LibTech' becomes 'EdTech' or 'ResTech'

The traditional provision of scholarly publications to support research and teaching becomes one facet of much more diverse research and education information systems. He argues that some current library vendors are still mired in supporting legacy systems preventing the visionary transformative approach needed. 'Only those that can deliver new higher value platform solutions will thrive in the longer term.'

AI - utopia or dystopia?

Andrew Cox, senior lecturer at the Information School, University of Sheffield, presented at the 2022 UKeig Members' Day on 'Artificial Intelligence: What you need to know.' He spoke to his independently researched report - [The Impact of AI, machine learning, automation and robotics on the information profession](#) - published by [CILIP](#) in 2020 with the support

of [Health Education England](#). It called for a 'joined-up and coherent response from information professionals, enabling us to maximise the benefits of AI, machine learning, automation and robotics for information users while mitigating the emerging risks.'

AI is suddenly terribly important and everybody's talking algorithms, rushing to create strategies and mourning the demise of boring, unintelligent mankind. It is often viewed as a disruptive piece of science fiction imposed on a soon to be redundant, disintermediated library and information services profession, but Cox argued that so much is out there already, immersed in our daily work and leisure routines.

- Autocorrection, predictive text, spelling and grammar checks
- Transcription/captioning of meetings
- Translation tools
- Search personalisation and recommendations.

We're all aware that tools like autocorrect often can be exasperatingly inaccurate. (Well and we'll/ill and I'll immediately come to mind. Our smart phones are not so smart.) 'They're not magic,' says Cox, and arising new technologies including machine learning, natural language processing, sentiment analysis and text and data mining will require our knowledge and critical skills to evaluate them, inform their development and explore their potential and benefits. AI requires human input to be of any relevance or value. (UKeIG explored some approaches at a members' forum way back in June 2019: ['The impact of Artificial Intelligence: brave new world?'](#))

AI is also built on data, so data literacy, management, validity and integrity are fundamental, and we are well positioned to offer these skills as a data-rich profession that regularly collates:

- Turnstile data and circulation data in libraries
- Usage of digital resources
- Satisfaction surveys
- Reference enquiries
- Qualitative data, including user experience (UX) studies.

Cox's message is not to obsess on the technology that underpins AI, but to prioritise creative thinking and developing a strategic roadmap and vision for the future. IT departments often dominate decision making when it comes to the procurement and acquisition of enterprise systems. It's crucial to collaborate with them, cross-fertilise skills sets and converge in some areas. While we don't all need to morph into code writers *per se*, a root and branch review of our competency frameworks is essential, perhaps existential.

Undoubtedly AI will render some routine tasks redundant, but the opportunities are endless. We are not staring disintermediation down the barrel of an AI gun. In an uplifting rally cry, Cox argued that the library and information services community has always valued user requirements and has endless imagination and vision, with a:

- Proactive, can-do attitude
- Service focus that is informed with a knowledge of institutional agendas
- Collaborative and influencing skills set
- Commitment to professional development, learning, knowledge sharing and addressing our weaknesses.

So, let's keep calm, carry on and get on with it.