SPECIAL ARTIFICIAL INTELLIGENCE ISSUE

JUNE 2022

Trends and Issues in Library Technology

International Federation of Library Associations

IFLA IT Section

TRENDS & ISSUES IN LIBRARY TECHNOLOGY

IFLA Information Technology Section — Special Issue on Artificial Intelligence

Feature Articles

IFLA Summer Dublin Conference

Departments

Editor's Notes

Special Artificial Intelligence Issue

Dear Colleagues,

I am very happy to introduce this special AI themed *Trends and Issues in Library Technology* with a great cross-section of AI focused articles and important IFLA IT section conference information. For the conference, there is information on IFLA IT-sponsored sessions, co-sponsored IT panels with other sections, SIGS and satellite conferences also co-sponsored with the group. In keeping with this issue's AI theme, I want to highlight an exciting AI satellite conference preceding the WLIC which will take place in Galway, Ireland and encourage you to attend. See https://ucg.ie/ifla/ and the Satellite program in this issue (p. 37) for more information.

IFLA WLIC 2022 DUBLIN

Inspire, Engage, Enable, Connect

87th World Library and Information Congress
26-29 July 2022, Dublin, Ireland

For this issue, we have a wide spectrum of great articles on AI topics for libraries. Juja Chakarova, of the International Atomic Energy Agency, Vienna, gives a historical overview of AI in the context of libraries. Lynn Kleinveldt, Cape Peninsula University of Technology, South Africa, overviews new AI learning environments and libraries. Elena Sánchez Nogales, Alicia Pastrana García, José Carlos Cerdán Medina, of the National Library of Spain, present a project regarding Digital Transformation, Data Reuse and AI at the National Library of Spain. Marcelo Lorca González, from the Library at the National Congress of Chile, details a project on the conversion of MARC to BIBFRAME and implications towards semantic web-linking. From Greece's International Hellenic and Aristotle University of Thessaloniki and Datascouting, Afrodite Malliari, Ilias Nitsos, Sofia Zapounidou, Stavros Doropoulos overview an AI-based Open Audio-Visual Archives (OAVA) project for enhanced search and retrieval. And from the University of Sheffield, UK, Andrew Cox forwards a compelling set of ethics scenarios regarding artificial intelligence and the importance of examining ethical implications of AI for information professionals.

Our chair, Edmund Balnaves, also gives a quick update on the IT section's 2022 progress and WLIC 2022 conference information. To better kick-off our AI theme, I have written an editorial on our AI articles, drawing together connections and introducing themes and authors.

I would also like to thank our wider Communications team: first, IT Section Communications Director, Francois-Xavier Boffy for help with layout; and Wouter Klapwijk, South Africa, for his valuable editing skills. Finally, a special note of appreciation to all our section authors. They have all done amazing work contributing articles to this special AI focused issue. I would also mention that if you are looking to volunteer or help with TILT, we are searching for a skilled volunteer graphic designer/layout artist who can speed up our publication processes. We are also always looking for upcoming high caliber contributions and projects. Our submissions door is now open again, so feel free to submit a proposal or query for upcoming issues. Also, if you have not joined online already, I encourage all of you to visit, join and contribute to our new IFLA IT Social Media Facebook Group available at https://www.facebook.com/groups/iflaitsection

I hope everyone will be attending the upcoming conference in Dublin. I had recently read that the top five things to do and see in Dublin are: 1) Trinity College, one of the oldest universities in the world; 2) Grafton Street where you can shop till you drop; 3) St. Stephen's Green, Dublin's Central park where writers James Joyce and W.B. Yeats have memorials commemorating contributions to Irish poetry and literature; 4) The National Gallery of Ireland; 5) Guinness Brewery and Storehouse.

To all our readers, I look forward to meeting you at our upcoming annual conference in Dublin and I extend my best wishes for safe travels.

Sincerely,

Ray

Ray Uzwyshyn, Ph.D. MLIS ruzwyshyn@txstate.edu
Editor, Trends and Issues in Library Technology

IT Section Chair's Corner

Update from the Chair

The WLIC congress is fast approaching: a chance at last for the library community to gather together in person.

To celebrate that event, the Information Technology Section has challenged IFLA with an interesting array of session offerings. The IFLA professional council has come to the party, accepting all our proposals for the congress.

As a result, we are now actively organising:

* A satellite conference on Artificial Intelligence hosted by National University of Ireland, Galway. An exciting programme has been prepared, as well as an opportunity experience the cultural and technological offerings of Galway. See https://library.nuigalway.ie/ifla/

* A main session on Artificial Intelligence. Taking the detailed experience of the satellite, we will bring a Town Hall styles session to the main conference to open the floor to discussion of all aspects of AI: library operations, ethical considerations, new innovations.

* A further session in conjunction with the Subject Analysis and Access (SAA) Section with on the Agile methodology for project management. The Agile methodology is established practice in IT and offers libraries a new and practical way to engage in projects of all sorts.

The IT section proposal for a new IFLA Special Interest Group on Artificial Intelligence has been formally approved by the IFLA Governing Board. This group will provide an opportunity to explore this fast developing area within the Association and more widely. The satellite conference on AI will feature workshops in practical application of AI as well as case studies in ethical issues. The Section is also planning further Webinars in this territory.

I look forward to meeting our existing and new colleagues in the Information Technology Section, Big Data SIG and new Artificial Intelligence SIG in Galway and Dublin. Make sure you register soon.

Edmund

Edmund Balnaves, Ph.D., ejb@prosentient.com.au
Chair, IFLA IT Section

IT Section Conference Notes: A Solid WLIC IT Agenda

Francois-Xavier Boffy, francois-xavier.boffy@univ-lyon1.fr
IFLA IT Communications Coordinator

The River Liffey and Conference Bridge, Dublin

The IT Section and sponsored SIGs will be involved in many different sessions and events surrounding the International Federation of Library Associations World Library and Information Congress 2022. Here is a current overview offered to IFLA IT members and other parties interested in IT related topics:

The IT Section will sponsor two satellite meetings:

- New Horizons in Artificial Intelligence in Libraries - in line with a newly approved AI Special Interest Group. 21-22 July 2022, Galway, Ireland https://ucg.ie/ifla/
- Library Carpentry Workshop - organized with the Science and Technology Libraries Section (STL), Continuing Professional Development Workplace Learning Section (CPDWL) and Special Interest Group in Big Data . 22-23 July 2022, Waterford, Ireland

The IT Section will also sponsor three open sessions during World Library and Information Congress 2022, 26-28 July:

- Perspectives on data access and use at scale: Lessons from the Field - organized by Special Interest Groups in Big Data and Digital Humanities / Digital Scholarship 26 July at 3pm.
- Artificial intelligence: new horizons and implications for libraries - in collaboration with the Advisory Committee on Freedom of Access to Information and Freedom of Expression FAIFE; invited speakers will reformulate highlights and perspectives from the Galway satellite meeting and spark continuing discussions with the World Library Congress audience. 27 July at 2:45 pm
- Agile in the Library: methods and tools for project management, collaboration and innovation - organized with the Subject Analysis and Access (SAA) Section. 26 July at 1:15 pm.

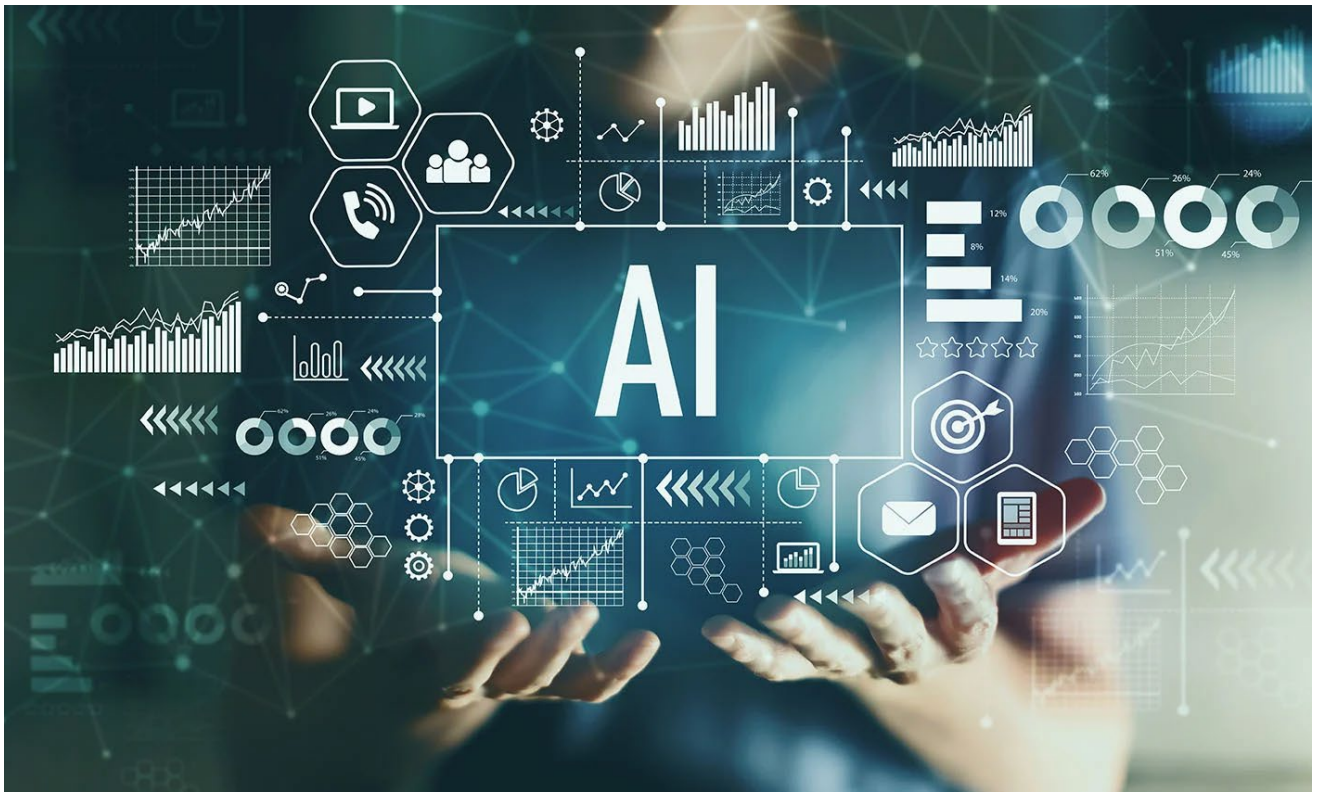
Other events and open sessions may also raise interest of tech-friendly librarians: sessions about combatting misinformation / fake news Health and Biosciences Section (**HBL**) Social Science Libraries Section (**SSL**) and News Media Section (**NM**); Open Science & infrastructures issues (Science and Technology Libraries Section, (**STL**), Health and Biosciences Section (**HBL**) and Knowledge Management Sections (**KM**); or new virtualized reference services, Reference and Information Services Section (**RISS**)

See the conference website for current WLIC program updates. And looking forward to seeing everyone in Ireland.

Editorial Overview

Introduction: Artificial Intelligence in Libraries

Ray Uzwyshyn, ruzwyshyn@txstate.edu
Texas State University Libraries



AI in Libraries and Education, Tierney, Courtesy Adobe Stock

Introduction

The world is changing, and technological paradigms of AI are quickly being adopted in the world of libraries and information management. With a newly approved 2022 IFLA Special Interest Group in AI, this issue introduces libraries and information professionals around the globe engaging with these new leading-edge AI IT paradigms. This issue of *Trends and Issues in Library Technology's* AI brings together authors, researchers, and practitioners. Articles range from general to high level introductions to algorithmic overviews, AI projects and AI histories. All articles focus through a lens of libraries and information management. The issue also overviews intersectional similarities among learning, libraries and education environments. There are new possibilities for educating next generations of information professionals into AI paradigms. Thematic AI topics range from the role of data and metadata for AI model training to machine learning. Larger roles are also being played by national libraries in opening possibilities through natural language processing, harnessing large datasets for AI and new collaboration possibilities with associated supercomputing centers.

Conversion to BIBFRAME triples is also contextualized and detailed. National library perspectives can act as a gateway towards helping semantic web-linking and future AI harnessing possibilities. Complex AI -related projects surrounding online audio-visual archives are also overviewed. These explore the potential of AI, speech recognition and natural language processing. Global linguistic online archives can now improve search and retrieval mechanisms to create higher information quality audio-visual archives.

Important, but often overlooked topics of AI, information professionals and ethics are also examined from perspectives of AI ethics scenarios. Scenarios may be both teaching tools for information professionals but also developmental contexts for system designers. AI ethics scenarios help organizations think more widely about social implications of the currently occurring AI revolution.

All of these articles together create a vibrant picture of AI's future possibilities and challenges for libraries and

information centers. From these perspectives, our special issue's authors need to be lauded. They have written insightful historic overviews, but also taken up the pragmatic new challenge of AI. This ranges from the algorithmic and programmatic to larger project management challenges and current new ethical questions arising from AI technologies' sea change in technological paradigms affecting information professions.

Articles, Authors, and Global AI Perspectives

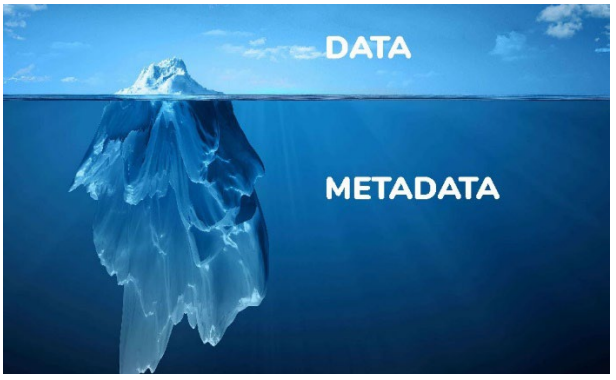
Juja Chakarova begins this special issue of TILT with a broad historical overview of AI beginning with Aristotle's logic and then refocusing Arnold Toynbee's original definition of 'Industrial Revolution'. Toynbee first employed the term 'Industrial Revolution' in the 19th Century and Chakarova frames Toynbee's characterization within historical phases of industrial revolutions (IR, 1-4) She traces these through the last two centuries to our present era and AI.

Chakarova's overview then sets the stage for the twentieth century AI luminaries from Alan Turing to John McCarthy and Douglas Hofstadter. Current AI definitions, subfields, ethical considerations, and futures for libraries are also explored. Chakarova speculates that the fourth Industrial Revolution will combine physical, digital, and biological worlds enabling quantum computing, robotics, cyber physical systems and artificial intelligence. According to Chakarova, we have somewhat unwittingly already entered this fourth IR AI phase.

Following Chakarova, Lynn Kleinveldt continues focusing this historical optic on AI within present necessary parameters of teaching, learning and libraries. Kleinveldt advocates that library must now add 'Algorithmic (AI) Literacy' to more well-known library literacy areas of 'information' and 'digital literacies'. This will empower 'Generation AI' with these new and necessary interdisciplinary competencies.

In her article, Kleinveldt offers a compelling list of rationales for the necessity of AI literacy in larger digital and information literacy curricula in academic libraries, higher education and learning environments. Kleinveldt argues for the library as a natural third place for a previously lost interdisciplinarity in traditional compartmentalized curricula. 'Algorithmic' Literacy should be organically added to the 'information' and 'digital' literacies already taken up by libraries to fill other gaps. Kleinveldt eloquently forwards imperatives of adding algorithmic AI literacy both to Gen AI library schools (i.e., 'i-Schools') and Information Science curriculums. Higher educational institutions need to be taking up these new cyber-physical (robotics), network and artificial intelligence paradigms through academic library interdisciplinary third space curricular perspectives.

The next two articles shift angles and move us to pragmatically realized projects. They approach libraries and AI from fundamentally important but different angles of data and metadata.



Data, Metadata and AI Ecosystems

Providing data and metadata in suitable formats is a primary need for AI. This ranges from the training samples needed for 'Machine Learning and Deep Learning' to large scale 'Natural Language Processing'.

In their article on 'Digital Transformation, Data Reuse and Heritage Collections at the National Library of Spain', Elena Sánchez Nogales, Alicia Pastrana García and José Carlos Cerdán Medina provide excellent overviews of the National Library of Spain's' (BNE) work in these areas. Nogales, Garcia, and Medina focus on novel forms of institutional cooperation and the new BNElab's work to promote the use and reuse of the national library's (BNE's) resources. This ranges from data, metadata and digital resources hosting to open licenses for Spanish public sector data. These are all needed for areas such as machine learning and artificial intelligence.

Nogales, Garcia, and Medina's article also focuses on the Spanish Web Archive as a training ground for AI-based Natural Language Processing models. The National Library of Spain has partnered with the Barcelona Supercomputing Center to train Spanish language models using a supercomputer. The most powerful supercomputer in Spain (*MareNostrum*) is utilized to clean massive WARC data sets. Nogales, Garcia, and Medina detail their fascinating project journey and its use of neural network technology based on Transformer (previously used for the English language natural language processing). Transformer is now trained and refocused towards Spanish datasets to create an AI model that can understand Spanish vocabulary, expression and writing rules on expert levels. The model called *MarIA* can also understand abstract concepts and infer word meaning from usage contexts. The project is a milestone in the application of artificial intelligence for the Spanish language and a great example of library/supercomputing centers collaboration and use and reuse of large datasets.

Equally foundational, Marcelo Lorca, of the National Congress of Chile, presents an important and detailed set of programmatic steps of the National Library of Chile's

transition from MARC records to BIBFRAME. Converting to BIBFRAME fundamentally improves bibliographic data by integrating it with and enabling it for other data resources on the semantic web. Specifically, BIBFRAME opens future AI possibilities for the data now recategorized within triples. The BIBFRAME format allows conversion of traditional library catalogue records towards compatible semantic web formats which can further be linked to each other and other web resources. Lorca details the National Chilean Library's path through these challenges and possibilities on deep programmatic levels. Both projects will be interesting to follow as other national libraries follow and/or remix and continue to develop these AI/Metadata and new data formatting-centered paths.

The potential of AI Natural Language Processing, data and Machine Learning is also explored further in Malliari's, Nitsos, Zapounidou and Doropoulos', 'OAVA, Open Audio-Visual Archives', an ambitious project to develop a new Greek audio-visual archive (OAVA), aggregator.
Through the project, the largely Greek universities group explores the potential of AI to create a better search and retrieval mechanism for Greek language and related audio-visual material. The application of deep learning AI Models are applied to large Greek audio-visual datasets. Algorithms are trained and developed to perform speech recognition from Greek and English audio-visual datasets. This AI centered project serves to develop a better search mechanism. This occurs through both previous aggregate audio-visual metadata and the audio-visual materials' newly speech-recognized linguistic Greek and English content. Similar to the National Library of Spain's natural language processing models and training for Spanish on WARC files (Web Archive Files), the Greek group utilizes AI NLP techniques on audio-visual Greek and Greek-related material. This use of AI Natural Language Processing potential creates better searchable content through speech recognition processing of audiovisual material. The deep programmatic complexity and algorithmic detail of the project speaks for itself. Like other keystones presented here, the project makes strong headway into less trodden pathways. The future modelling potential is for other linguistic-focused audio-visual aggregators from various other national libraries and country global language centers to follow and continue to develop these models.

Rounding out the spectrum of AI papers, Andrew Cox, from the University of Sheffield, UK, approaches the topic from a different but equally important angle. Cox reflects in his article on largely overlooked ethical valences of artificial intelligence and advocates towards the necessity and importance of thinking about ethics for information professionals and information management.

Cox begins considerations with an overview of various important ethical parameters to which artificial intelligence's newer realized branch of computer science heralds.

Ethics Scenarios for AI, Andrew Cox

Cox then advances notions of the utility of AI-related ethical scenarios for thinking through complex ethical implications of AI systems. He puts forward a set of ethics-based AI case studies which may be utilized as forms of model curricula for teaching new information professionals. These may also be used as a useful point of departure for organizations beginning to think seriously about implications of AI design and implementation. AI systems have necessary ethical considerations for libraries and in this work Cox provides further links to eight well developed AI ethics scenarios as springboards and example for further reflection and discussion.

Conclusions

Together, this 2022 issue's grouping is a stellar constellation of leading edge AI library articles illuminating where information management professionals are taking AI. The articles also give an excellent indication of future possibilities, challenges and areas necessary for further development.

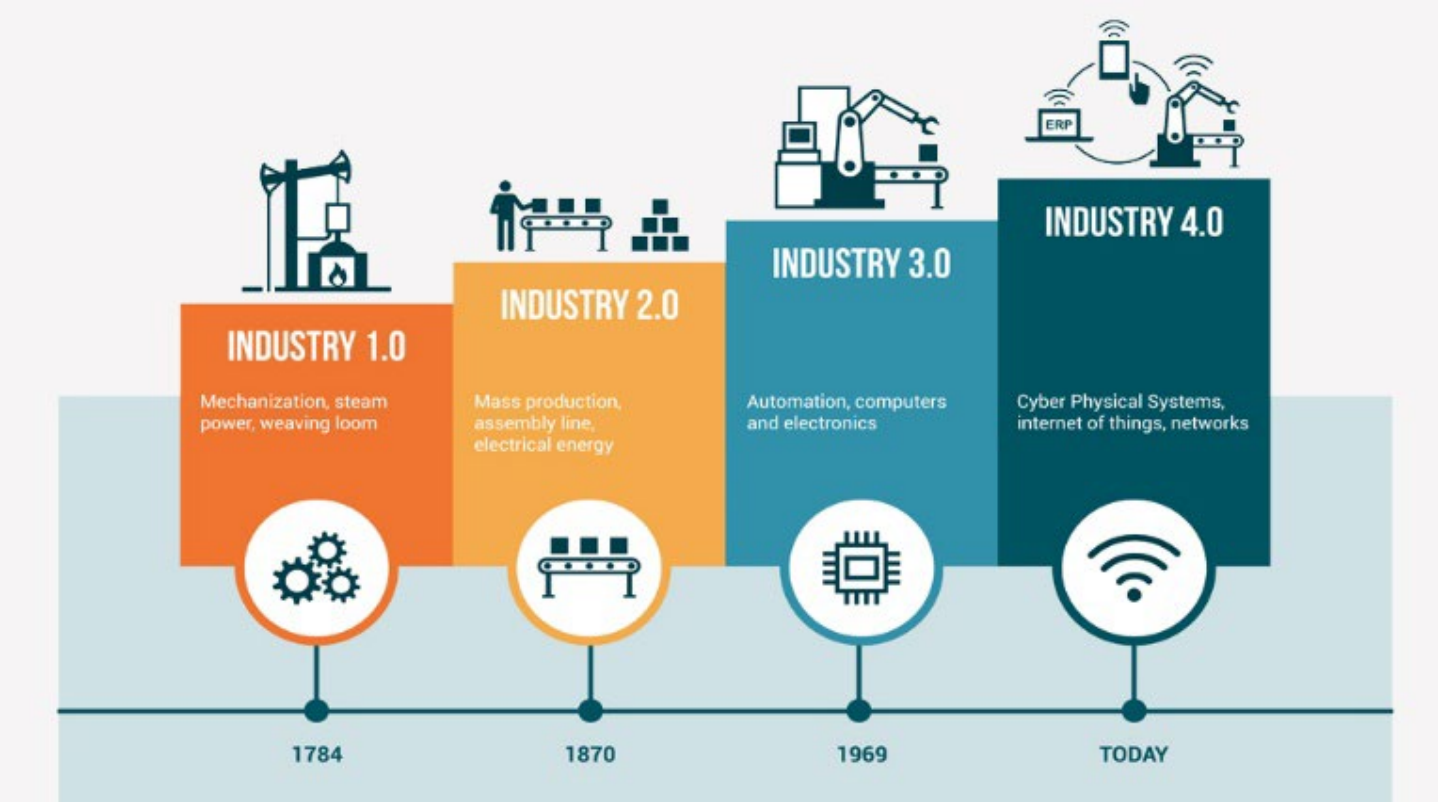
We have come a far way from the ancient Library of Alexandria's *kata-logos* (list of words), but our AI programs still find ancient antecedent with the logic of Aristotle. Our organization and training of new AI Deep Learning-models is largely in the global Cloud now, but we are also a global village that may reap the benefits from this shared environment.

We must now work and network together to solve these future AI challenges for the betterment of our global community. It is our duty to continue to help move the ancient '*kata-logos*' forward and continue to transform archives of information and data into insight, knowledge and hopefully wisdom towards future generations and our collective good.

The future does await but we build in the present. The shadows of our collective past infrastructures remain. This issue's content hopefully provides an inspirational glimpse of the future in this global grouping of excellent articles which follow.

AI: Already in Libraries?

Juja Chakarova, juja.chakarova@gmail.com
International Atomic Energy Agency (IAEA), Vienna Austria



Industrial revolutions from 1784 to AI and Today

Four industrial revolutions

Human development has striven to understand nature's forces and increase its chances for survival, thus being involved in scientific research and implementing results in technology. The term industrial revolution (IR) was first used by Arnold Toynbee (1852-1883) to describe Britain's economic development from 1760 to 1840, but later applied more broadly as a process of economic transformation.[1] The last two and a half centuries have witnessed four 'industrial revolutions'. The first IR (1784-) was powered by steam and water and transformed society from an agrarian society to an industrial one. The steam engine, developed by James Watt was introduced commercially in 1776. This was followed in the early 1800s by the steam-powered locomotive of Richard Trevithick. Steam-powered ships were transporting freight across the Atlantic and locomotives connecting Britain's industrial hubs. The second IR (1870-) was powered by oil and steel. This era witnessed the invention of the telephone, light bulb and internal combustion engines developed by Nikolaus Otto in 1864 and Rudolf Diesel in 1892. Electrical power was also introduced leading to industrialization, standardization and mass production. The third IR (1969-) was and continues to be marked by the development of electronics, information and communication technologies. We witness ubiquitous computing as predicted by the now famous Moore's law. Moore's law states that computer power doubles every 18 months. This means it grows exponentially. Other famous predictions failed, like the one by IBM's early President, Thomas J. Watson who said in 1943 **"I think there is a world market for maybe five computers."** The fourth IR will combine the physical, digital and biological worlds, enabled by quantum computing, genetic engineering, robots, cyber-physical systems and artificial intelligence (AI). Are we on the verge of it or already there?

Artificial intelligence (AI)

Although the term 'artificially intelligence' (AI) was introduced in mid- 20th century, the pillars of this concept were erected throughout many earlier centuries. Out of the pléiade of philosophers and scientists, let us mention a few, whose earlier contributions were crucial. The Greek philosopher **Aristotle** (384 - 322 BC) developed in his *Analytica Priora* the first principles of reaching a conclusion from a set of premises, guided by rules known as syllogisms.

Here is a familiar example:

All humans are mortal.
All Greeks are humans.
Therefore: All Greeks are mortal.

Aristotle was thus the founder of formal logic, which is still widely used in IT and AI. The If-Then principle, for example, is a core of computer programming and loops.

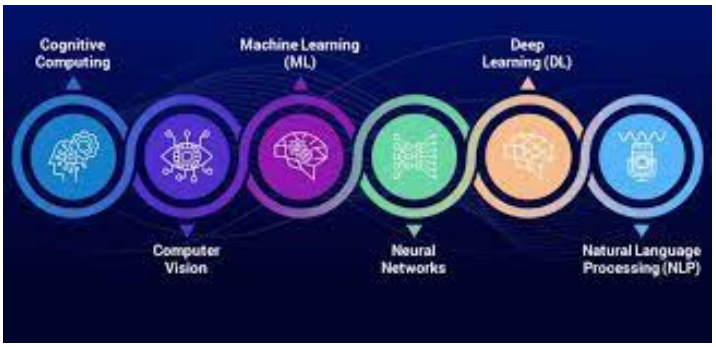
Alan Turing (1912 – 1954) was an English mathematician and computer scientist, widely considered to be the father of theoretical computer science and artificial intelligence. As a cryptologist he broke the code of the German naval Enigma in 1939. In 1950 Turing addressed the problem of artificial intelligence, and proposed an experiment that became known as the Turing test, an attempt to define a standard for a machine for it to be called "intelligent". The idea was that a computer could be said to "think" if a human interrogator asked questions to both a computer and a human and could not tell them apart.

John McCarthy (1927 – 2011) was an American computer and cognitive scientist. In 1955 McCarthy drew up the Dartmouth Workshop Proposal to the Rockefeller Foundation. For the workshop he envisioned a 10-person study of artificial intelligence and thinking machines. This six week workshop took place at Dartmouth College in Hanover, New Hampshire in 1956 and became a seminal event. The most remarkable result of the Dartmouth proposal was the success of the term AI, originally an engineering discipline whose relationship to biological organisms is mostly metaphorical and inspirational. Even now, there is an ongoing debate whether machines can think or just simulate thinking.

Douglas Hofstadter (1945-) is an American cognitive scientist. In 1977 at Indiana University he launched a research program called 'artificial intelligence research', but later renamed this to 'cognitive science research'. Even though Hofstadter stated having no interest in computers, his extraordinary works inspired many students to start careers in AI.[2]

AI Definitions

Numerous definitions exist to capture the very meaning of AI. According to McCarthy (1955) it is "… making a machine behave in ways that would be called intelligent if a human were so behaving". AI is also defined as "the study and design of intelligent agents where 'intelligent agent' is a system that perceives its environment and takes actions which maximizes its chance of success.[3] But how can we define intelligence? Debates are still ongoing to define human intelligence. Highly advanced machines like Deep Blue and others beat humans in chess, Jeopardy or Go. Are they intelligent, or just fast processors? They possess enormous datasets and simulating thinking but are they really thinking?



AI Subfields

AI Subfields

There are many subfields of AI. One researcher even listed 87 of them. This list is shorter, but contains the main subfields. **Machine learning** is the science of training a device or software to perform tasks. Its capabilities may be fine tuned by feeding it more data, so it can 'learn' over time. **Deep learning** is a subset of machine learning. Several layers of neural networks are connected in order to make a prediction. Deep learning is used in areas from autonomous cars to robotics. **Computer vision** (CV) is the process of analyzing visual images, so that computers can 'see'. When you use Google search for images now, the engine searches the metadata associated with the images, but this will change with the development of Computer vision CV. The computer can process for example flat, 2-dimensional images of a scene from different angles and reconstruct a 3-dimensional model. CV can be used to analyze sound, as the sound chart is also a visual image.

Speech Recognition or speech-to-text, is a capability which enables a program to process human speech into a written format[4]. Analyzing and transcribing speech faces different problems, like separating 'signal' from noise. As you might have discovered when learning a new language and listening to native speakers, there is no obvious break between words. The signal is continuous and the intonation of the speaker can change the meaning completely.

An earlier approach of **Natural Language Processing** (NLP) focused on codifying word categories and syntax. It wasn't fully successful though due to the many exceptions existing in any language. NLP now uses statistical machine learning methods. For example, Google translate, Trados and DeepL step on big data – the official translations of documents, thus confirming existing rules and staying flexible to 'learn'. Translating is just an example, but NLP is used for indexing, creating summaries and answering questions. Time is our most expansive commodity and saving it creates a niche for NLP.

An artificial **neural network** is a computer program inspired by certain principles of a real neural network, such as human brain. Multiple types of **robots** have been developed ranging from pre-programmed to humanoid to autonomous, teleoperated and augmented. Robots have various uses in industry, social work, elderly care, police, libraries and even as composers of music.

The future of the professions

Machines have made human life more secure and comfortable. We have seen this throughout the historical industrial revolutions. There is a parallel trend though, of creating fear and job insecurity. The beginning of the 19th century saw for example the movement of the Luddites in England destroying the machines in textile industry. With the vast implementation of robots and AI, those fears are revived on a new scale. Richard Susskind (1961-), a British author and professor specializing in legal technology, predicts the decline of the professions as we know them today in his book "The Future of the Professions"[5]. Susskind provides the example of the barista in Italy – a traditional profession of blending and serving coffee – which was suddenly made obsolete with the introduction of coffee machines and capsules containing special coffee blends. Disruptive technologies and services lead to the decline of some professions and create the need for new ones. Susskind predicts that 'increasingly capable machines will take on many of the tasks, that have been the historic preserve of the professions". As part or the total of our tasks will be delegated to 'increasingly capable machines', our professions will be dismantled incrementally'.

Susskind reflects on the professions of lawyers, teachers, tax advisors and consultants. He argues that the traditional knowledge gatekeepers will gradually delegate some of their tasks to intelligent systems – first the repetitive ones, but later some more sophisticated[6]. Robots and AI will in the future be able to even take decisions. Susskind is the IT adviser for the Lord Chief Justice of England and Wales. He also promotes online courts, asking the question: "Is the court a place or a service?'.

Robots in Libraries

AI and robots have seen implementations in different sectors, apart from the increased usage in industry.[7] Robots perform the profession of doctors, policemen, composers, elderly care, and even build a museum of robotics.[8]

It is certain that AI and robotics will also affect the profession of librarians as the traditional custodian of knowledge. In fact, there are numerous AI implementations in libraries, archives and museums, namely in computer vision, NLP, robots (RFID inventory, items delivery and storage, etc.), machine learning (search engines, indexing) and others.

Why do libraries delegate some tasks to machines or systems? Since the introduction of the OPAC our catalogues are available 24/7. With the implementation of RFID based self-check systems, libraries can be open to users 24/7 with no need of librarians to be present. Our readiness to delegate tasks is motivated by our dedication to serve our constituency better and to deliver continuous and seamless services. Is there a limit to this trend, or are we going to contribute to the annihilation of our own profession? Shall we contribute to the disruptive process or allow libraries to be disrupted? Shall we lead the process, be part of it, or be part of the collateral damage?

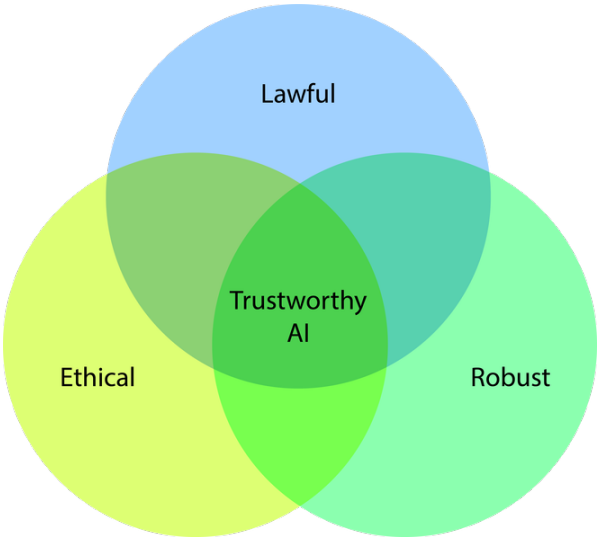
European Commission Ethics Guidelines for Trustworthy Artificial Intelligence

AI and Ethics

AI raises concerns not only regarding the loss of jobs, but also concerns about ethical issues. International organizations, as well as national bodies, defined the framework of AI development and ethical standards. The European Commission published its "Ethics Guidelines for Trustworthy Artificial Intelligence"[9] in April 2019. According to the Guidelines, "trustworthy AI has three components, which should be met throughout the system's entire life cycle.

AI should be:

(1) lawful, complying with all applicable laws and regulations
(2) ethical, ensuring adherence to ethical principles and values
(3) robust, both from a technical and social perspective".



Trustworthy AI: Lawful, Ethical, Robust. European Commission, 2019.

UNESCO adopted the Recommendation on the Ethics of AI[10] in November 2021 to contribute to the international regulatory framework aiming to achieve human-centred AI. Italy launched also in November 2021 its Strategic Programme on AI 2022-2024[11], aiming at becoming a global hub of AI research and innovation. We have witnessed some beginning AI implementations in libraries already. However, the process is not only costly, but also involves many iterations – hence the risk of errors and failure. It is necessary to combine our efforts and share the risk and the cost of library AI projects. We should invest in the new education required of AI and adopt the long view of our profession's future. We should also begin to plant seeds in a field we may never live to harvest.

Endnotes

[1] Britannica, The Editors of Encyclopaedia. "Industrial Revolution". Encyclopaedia Britannica, https://www.britannica.com/event/Industrial-Revolution. (Last accessed 19 March 2022)
[2] Gödel, Escher, Bach (1979); I am a strange loop (2007); Surfaces and Essences: Analogy as the Fuel and Fire of Thinking (2013)
[3] Wikipedia, Artificial Intelligence, last visited 20.3.2022
[4] Speech recognition.- IBM Cloud Education, 2020. Available at: https://www.ibm.com/cloud/learn/speech-recognition (Last accessed 3.4.2022)
[5] The Future of the Professions : How Technology Will Transform the Work of Human Experts. Oxford University Press, 2015. (with D. Susskind)
[6] Yuval Harari argues that 'AI can be better at jobs that demand intuition about other people", as our emotions and desires are no more than biochemical algorithms. See his book '21 Lessons for the 21st Century", Penguin (2018)
[7] According to the World Robotics 2021 report of the International Federation of Robotics, 3 million industrial robots operate in factories around the world.
[8] Robots Science Museum, Seoul. See: https://archello.com/project/robot-science-museum (Last accessed 3.4.2022)
[9] European Commission, Directorate-General for Communications Networks, Content and Technology, *Ethics guidelines for trustworthy AI*, Publications Office, 2019, https://data.europa.eu/doi/10.2759/177365
[10] *Recommendation on the Ethics of AI*, UNESCO, 2021. https://unesdoc.unesco.org/ark:/48223/pf0000380455
[11] *Strategic Programme on AI 2022-2024*. Italian Government, 24.11.2021. https://assets.innovazione.gov.it/1637777513-strategic-program-aiweb.pdf

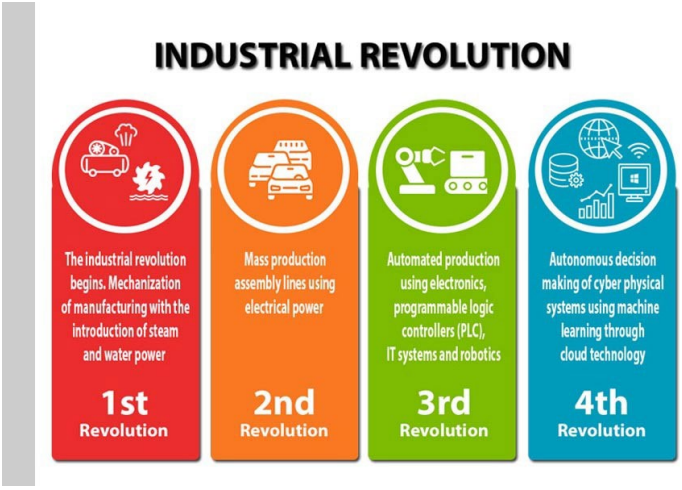
Smarter higher education learning environments through AI: What this means for academic libraries

Lynn Kleinveldt, Kleinveldtl@cput.ac.za
Department of Business and Information Administration, Cape Peninsula University of Technology

Introduction

The Covid-19 pandemic and lockdown calls for an embrace of Artificial Intelligence (AI) and robotics in the 'new normal' classroom. However, this poses the question: what does this all mean for academic libraries to support the 'new normal' learning, teaching and research at universities? To address this, AI needs to be placed in context as to what AI is and why it cannot be ignored in higher education. The shift towards smart learning environments implies embracing emerging technologies such as AI in higher education.

It is anticipated that by 2025, AI and robots will be utilised more for domestic jobs such as cleaners, nannies. In media it has also been reported that "robots are already writing new articles" (Pew Research Centre, 2021). A situation has unfolded where, as Singh et al. (2021) puts it, "the entire world is now striving to combat this pandemic with the aid of different, emerging technologies". This is evident in China's healthcare processes as robots are utilised for "disinfecting hospitals and delivering medical supplies" (Singh et al., 2021). This is but one example of the dilemma faced globally with unemployment rates on the rise. The tight competition between "university graduates, computers and bots for jobs" place further demands on higher education institutions to ensure graduate attributes. This means a creative, flexible and analytical thinking achieved through a technology-transformed curriculum (Piggott, 2021).

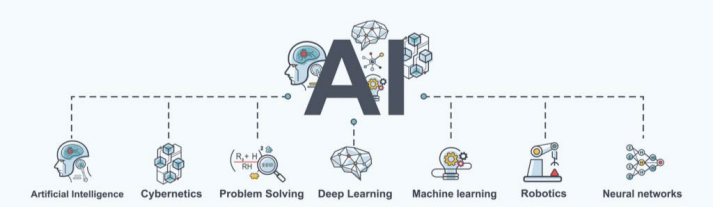


Historical Eras of Industrial Revolutions

The human element will form a key role in the equation, and the Fourth Industrial Revolution (4IR), will assist in creating innovative and dynamic ways to embrace technologies by unpacking the fear of the 'unknown'. AI, will complement research, teaching and learning activities, in a world where it is apparent that the only constant is 'change'. It is worth noting that the knowledge processes involved in the four industrial revolutions start with knowledge formulation (1st industrial revolution), knowledge evolution (2nd industrial revolution) and knowledge distribution (3rd industrial revolution). The 4th industrial revolution is based on cyber physical systems, networks and artificial intelligence and described as the knowledge mutation phase (Xing, Marwala and Marwala, 2018, 175).

Basic education around the world have been at the forefront of introducing AI-enabled robots to support learning. However, integration of robotics and Artificial Intelligence (AI) in the classroom in primary or pre-primary school levels are found to have both positive and negative impacts. A robot gives AI "a face" and is found to be beneficial in teaching simple programming to children as reported in Swiss schools. A robot also presents the danger of children becoming too emotionally attached to technology (Raaflaub, 2021). Therefore, the AI curriculum frameworks for grades K–12, such as the 17 competencies of AI Literacy are fundamental for integrating into basic education (UNESCO, 2022).

Technological advancement provides academics the opportunity to facilitate innovative problem-based learning. This advancement calls for universities to embrace the Fourth Industrial Revolution to enhance the learning experience. Considering the positive outcomes reported in basic education, AI has the potential to complement Teaching and Learning activities. AI also supports research at higher education institutions especially during times of uncertainty. Integrating emerging technologies into library services and museums has shown promise, ranging from stocktaking to orientation to enhancing reference services (Ocholla & Ocholla, 2020). Although information professionals boast about libraries constantly adapting to change, scholars argue that libraries take too long to incorporate emerging technologies such as AI into support services (Okunlaya, Abdullah & Alias, 2022; Wheatley & Hervieux, 2019). What makes matters worse for many organisations is that there are many different definitions of AI. Although Artificial Intelligence can simply be defined as a computer program developed to mirror human intelligence (Wheatley & Hervieux, 2019), it is far more than that.



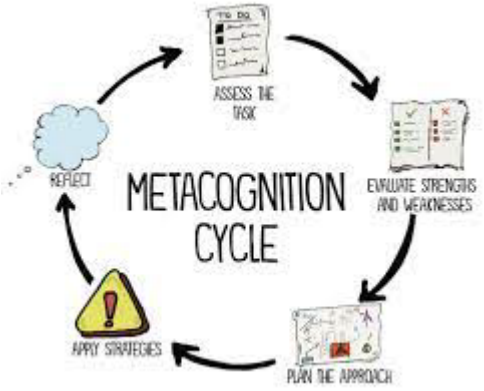
Interdisciplinary Forms and Subsets of AI

Okunlaya, Abdullah and Alias (2022) describe AI as being an "all-inclusive discipline" drawing from computer science, linguistics, information science, neuroscience, cognitive science and many more other disciplines. They further explain that AI makes sense of a vast amount of data. Through analysis, AI also personalise the user experience in various environments.

Organisations resisting to embrace AI will find it challenging to remain relevant in future (Okunlaya, Abdullah & Alias, 2022). It is therefore fundamental to understand why and how best to embrace AI in higher education to achieve the learning outcomes and meet the expectations of the workplace.

Why do we need to embrace AI in Higher education?

Over time, aspirations of citizens have become far more complex. This has contributed to a diverse economy and means that life-long learning and upskilling are an imperative. Obtaining a PhD is just not enough anymore in a rapidly changing world. Two curricula trends worth noting are the need for renewing general education and a focus on interdisciplinary research. It is therefore a time to "learn how to learn" through introducing a research component at undergraduate level that will result in a deeper understanding of "how we know what we know". Scholars further elaborate that "liberal education should give people the skills that will help them get ready for their sixth job, not their first job" (Lee & Yuan, 2018: 98). Cyber-physical systems (CPS) of the future, which include "collaborative robots" are envisaged to be highly intelligent, electrified, and connected (Gürdür, Boman, Törngren, 2021). Maintaining the quality of higher education in the era of the Fourth Industrial Revolution remains fundamental. This means that the focus needs to be on skills for learning beyond information transfer (Gürdür, Boman & Törngren, 2021). Linking metacognition to Information Literacy is considered "an essential element of intelligence" (Ridley and Pawlick-Potts, 2021: 2).



Metacognition Cycle

Another aspect for higher education institutions to ponder is the adoption of "AI thinking" or "AI literacy". Scholars speak about AI literacy developing due to algorithms and artificial intelligence being closely connected. Both form parts of the 'literacies of the digital' as illustrated in Figure 1. What is common in these literacies are references made to "computer, internet, information, computation, and algorithmic" literacy. There is "an especially strong and complementary connection between computational literacy and information literacy" (Ridley and Pawlick-Potts, 2021: 2). The significance of Algorithmic literacy is that the main purpose is about "knowing and meaning making through the processes of internalizing and externalizing information" (Ridley and Pawlick-Potts, 2021: 2). This presents academic libraries an opportunity to consider incorporating Algorithmic literacy into information literacy training.

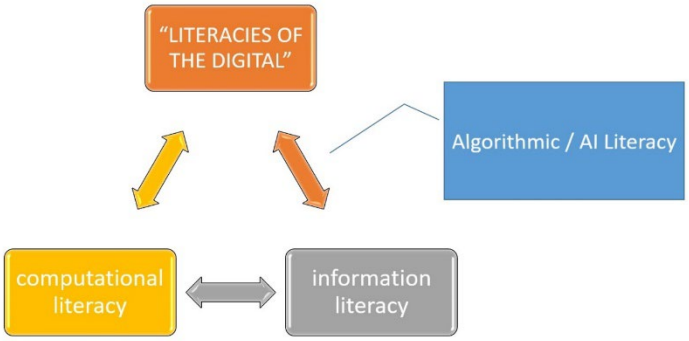


Figure 1 Design adapted to Algorithms and Artificial Intelligence (Ridley and Pawlick-Potts, 2021: 2)

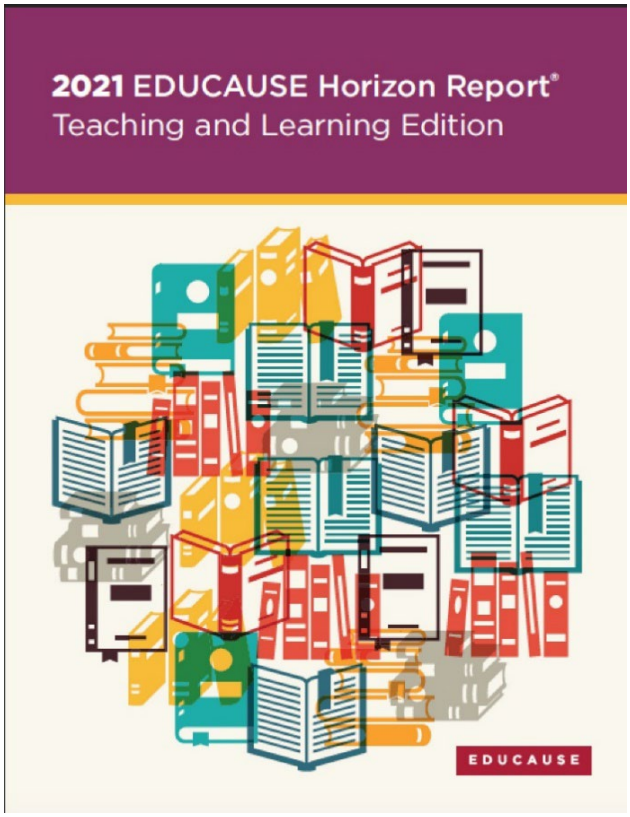
AI certainly plays a part in higher education institutions building smart learning environments. The role of a smart learning environment is to incorporate technology in such a way that the learning experience becomes simpler and enriched. In Helsinki for example, smart learning environments introduce learners to virtual reality applications. These applications allow learners to navigate places like the sea floor. Augmented reality technologies enable pupils at school to design their own school grounds (Mehtelä, 2021). These technologies have the potential to enhance student-centred learning approaches in higher education.

It is worth noting that in 2021 AI was placed at the top of the technology trending list (EDUCAUSE Horizon Report, 2021; Simplilearn, 2020). Simplilearn (2020) recommended the top ten trending technologies for organisations to consider in 2021 if they want to have a competitive advantage. AI recommendations are also in line with earlier emphasis placed on future relevance of organisations dependent on embracing these technologies (Okunlaya, Abdullah & Alias, 2022). What adds value to their list, are the training and skills to be acquired for Artificial Intelligence (i.e. programming, linear algebra, probability, big data, machine learning algorithms). This needs to be incorporated into the higher education curriculum for workplace preparedness.

The Covid-19 pandemic has to a large extent become the driving force for many organisations. This is quite evident in retail businesses which must embrace the Fourth Industrial Revolution technologies at a fast pace, such as the AI-enabled apps for mobile shopping, resulting in many e-commerce advantages that are convenient to everyone. The key AI technologies incorporated into mobile apps today are found to be "translation apps with neural networks, releasing new AI-powered offline language translator apps for iOS and Android, Speech Recognition Technology, Chatbots, Natural Language Technology, Machine Learning, Biometrics, Emotion Recognition, Image Recognition and Text Recognition" (Kanada, 2020).

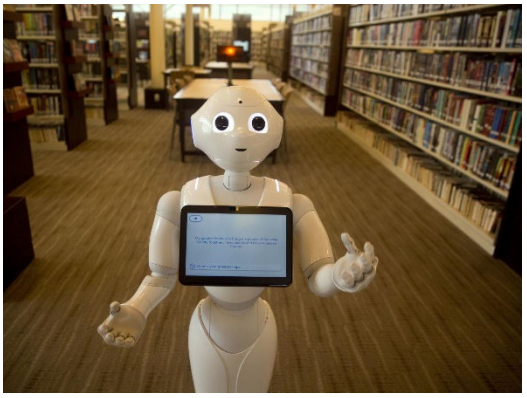
How is AI being used in higher education

According to the Educause Horizon Report (2021), AI is mainly used in higher education to address challenges in teaching and learning that will improve student success. The projects highlighted in the report described how AI incorporated in higher education support student success. One project at Durham University in the UK, is an AI-enabled "student engagement platform" called Holly, used to improve the student enrolment experience. At the University of Illinois Chicago, a chatbot called Socrates, acts as a digital assistant for helping students with navigating Blackboard (Educause Horizon Report, 2021).



2021 Educause and Horizon Report Teaching and Learning Edition

Another way AI is gradually being incorporated in higher education, according to the report, is the opportunity for re-curricularization across programmes that will accommodate "Generation AI." The Institute of Electrical and Electronics Engineers (IEEE) explains that "the next generation of children will be born into a world surrounded by technology, including solutions powered by artificial intelligence (AI). Generation AI will rely on artificial intelligence to assist them through all the milestones in their lives" (Educause Horizon Report, 2021). AI-enabled robots are also showing promise in higher education. For instance, the robot, Nao is being used as a teaching assistant to improve university students' vocabulary. This speaks to language policies and provides insight into how Nao could be used to incorporate multilingualism into teaching and learning at universities (Banaeian, & Gilanlioglu, 2021).



Pepper Robot at Roanoke Library, Roanoke Times.

The Pepper robot has shown promise in supporting students with problem-based learning activities. Most students that engage with the robot session indicated that the guidance received could be applied to solve study challenges faced (Robinson, Ward & Kavanagh, 2021). Therefore, Pepper has potential to complement problem-based learning in disciplines such as LIS.

One study did not specify the type of robot that was used to examine whether students interacting with a robot cheated during tasks, but this was found not to be the case (Ayub et al.,2021). This is perhaps an option to consider for online assessments. In many cases the Learning Management System, Blackboard has a Lockdown Browser feature which cannot be used for all assessments across faculties due to license limitations.

What does the use of AI in learning, teaching and research mean for academic libraries?

It is perhaps the time to consider how AI literacies could complement Information Literacy training. Some strategies to consider for enhancing the teaching and learning experience through AI could include:

- Training for academics and librarians
- Library schools and computer science departments collaborating more closely
- Exploring Funding opportunities for conducting AI projects
- Librarian-faculty collaboration remains fundamental in student success
- Coding for students should perhaps be considered in IL modules

Many universities worldwide have already integrated areas of data science (big data) and machine learning into the curriculum. Scholars also caution that these modules do not end up overpowering the learning objectives. Graduates need to be prepared for the world of work where thinking critically and adapting to change is vital for innovation. The Artificial Intelligence Library Services Innovative Conceptual Framework (AI-LSICF) developed by Okunlaya, Abdullah and Alias (2022) allows "the university library to foster innovative change, supporting the digital transformation of university research and learning with digital technologies like AI". They recommended that the framework be considered in strategic planning and that it be explored operationally in academic library support services.

What remains key is to strike a balance. Identify the areas where AI could best fit in to complement teaching and learning activities. Free up time for academics and librarians to shift from mundane activities to creating new science and knowledge to be incorporated in the curriculum. This will prepare graduates not only for that 6th job, but also begin to accommodate Generation AI.

References

Ayub, A., Hu, H., Zhou, G., Fendley, C., Ramsay, C., Jackson, K.L. and Wagner, A.R., 2021, August. If you Cheat, I Cheat: Cheating on a Collaborative Task with a Social Robot. In *2021 30th IEEE International Conference on Robot & Human Interactive Communication (RO-MAN)* (pp. 229-235). IEEE.

Banaeian, H. & Gilanlioglu, I. 2021. Influence of the NAO robot as a teaching assistant on university students' vocabulary learning and attitudes. *Australasian Journal of Educational Technology*, 37(3): 71-87

EDUCAUSE Horizon Report. 2020. Emerging Technologies & Practices and Influential Trends, 2020. https://library.educause.edu/-/media/files/library/2020/3/2020_horizon_report_infographic.pdf?la=en&hash=66854F48297FEB53724C25A93E7475355EE6D4E3 [27 February 2021].

Gürdür Broo, D., Boman, U. & Törngren, M. 2021. Cyber-physical systems research and education in 2030: Scenarios and strategies. *Journal of Industrial Information Integration,* 21.

Kanada, P. 2020. How to Use Artificial Intelligence in Mobile Apps. https://readwrite.com/2020/09/13/how-to-use-artificial-intelligence-in-mobile-apps/?__cf_chl_managed_tk__=pmd_fTQdvwlHq3WAiYkYXorVYX_ToUSf7zrcLnulgl.BRKY-1634027593-0-gqNtZGzNA3ujcnBszRPl [26 September 2021].

Lee, R.M. and Yuan, Y.S. 2018. Innovation Education in China: Preparing Attitudes, Approaches, and Intellectual Environments for Life in the Automation Economy. [Chapter 5]. In: N.W. Gleason. *Higher Education in the Era of the Fourth Industrial Revolution*. Singapore: Palgrave MacMillan.

Mehtelä, K. 2021. Smart learning environments. https://hundred.org/en/innovations/smart-learning-environments#b233f99a [3 December 2021].

Ocholla and Ocholla. 2020. Readiness of academic libraries in South Africa to research, teaching and learning support in the Fourth Industrial Revolution. *Library Management*, 41(6/7): 355-368.

Okunlaya, R.O., Abdullah, N.S. and Alias, R.A., 2022. Artificial intelligence (AI) library services innovative conceptual framework for the digital transformation of university education. *Library Hi Tech*.

Pew Research Centre. 2021. Experts Say the 'New Normal' in 2025 Will Be Far More Tech-Driven, Presenting More Big Challenges. 18 February 2021. https://www.pewresearch.org/internet/2021/02/18/experts-say-the-new-normal-in-2025-will-be-far-more-tech-driven-presenting-more-big-challenges/ [6 June 2021].

Pigott, J. 2021. Less is more: education for uncertain times. *Globalisation, Societies and Education*, 1-11.

Raaflaub, C. 2021. Switzerland gears up to place robots in classrooms. https://www.swissinfo.ch/eng/switzerland-gears-up-to-place-robots-in-classrooms--/46402716 [2 Oct 2021]

Ridley, M., & Pawlick-Potts, D. 2021. Algorithmic literacy and the role for libraries. *Information Technology and Libraries, 40*(2): 1-15.

Robinson, N.L. & Kavanagh, D.J. 2021. A social robot to deliver a psychotherapeutic treatment: Qualitative responses by participants in a randomized controlled trial and future design recommendations. *International Journal of Human Computer Studies*, vol. 155.

Simplilearn. 2020. Top 10 Technologies To Learn In 2021. [YouTube Video, 11 Dec 2020]. https://www.youtube.com/watch?v=vPk-H7Cu9Lo&t=5s [28 February 2021]

Singh, K.R., Dash, S., Deka, B. and Biswas, S., 2021. Mobile Technology Solutions for COVID-19. In *Emerging Technologies for Battling Covid-19* (pp. 271-294). Springer, Cham.

UNESCO. 2022. *K-12 AI curricula: a mapping of government-endorsed AI curricula*. https://unesdoc.unesco.org/ark:/48223/pf0000380602 [21 March 2022].

Wheatley, A. and Hervieux, S., 2019. Artificial intelligence in academic libraries: An environmental scan. *Information Services & Use*, 39(4): 347-356.

Xing, Lufuno Marwala, and Tshilidzi Marwala. 2018. Adopt Fast, Adapt Quick: Adaptive Approaches in the South African Context. In: Gleason, N. W. (ed.) *Higher Education in the Era of the Fourth Industrial Revolution.* Singapore: Palgrave

Digital Transformation, Data Reuse and Heritage Collections at the National Library of Spain

Elena Sánchez Nogales, elena.sanchez@bnes.es
Digital Innovation and Reuse, **National Library of Spain**

Alicia Pastrana García, alicia.pastrana@bne.es
Digital Collections, **National Library of Spain**

José Carlos Cerdán Medina, josec.cerdan@bne.es
Online Publications Deposit, **National Library of Spain**

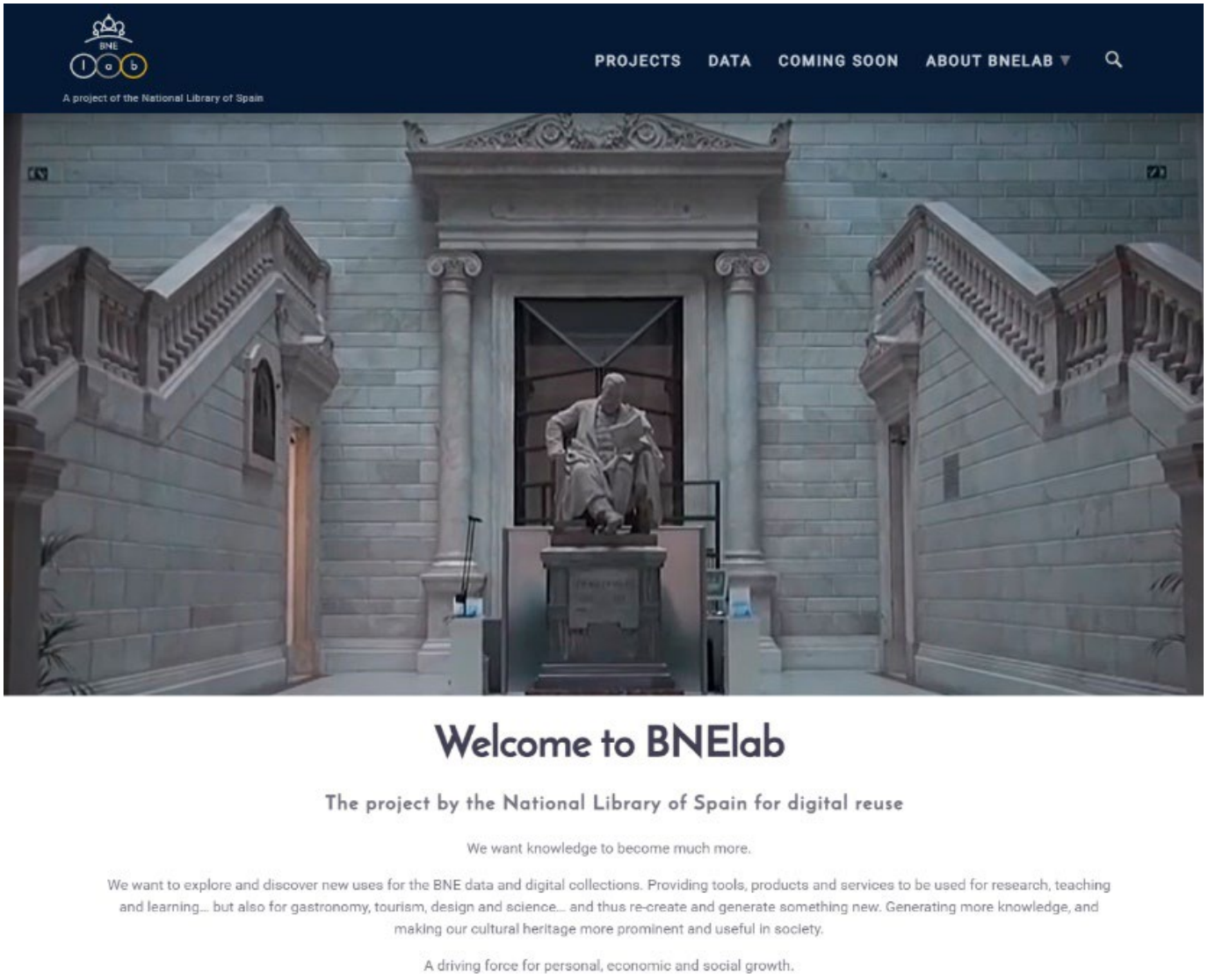


Figure 1. BNElab welcome page

With more than three hundred years of history and a collection of around 35 million documents, the National Library of Spain (Biblioteca Nacional de España, BNE) is the most important knowledge database in the country. The BNE has also undertaken a process of digital transformation over the last decade that has deeply and rapidly transformed many aspects and processes, from access and services to strategy and governance.

A rapidly increasing collection of digitized items include new digital resources such as web archives digitally born content, work on linked data and semantic enrichment of the catalogue (datos.bne.es). The process has also introduced significant internal considerations on how the Library complies with its mission and public service. How can the library relate to and engage with society? How should new digital ecosystems be used.

Within a very few years, the Library has created new forms of institutional cooperation through digital and web

environments and new forms of institutional cooperation. It has learned new narratives to share heritage and experimented with tools and ways to support learning, research and innovation. Importantly, the institution has integrated needs for new approaches for collaborative creation. Citizens should be engaged to bring new values to cultural heritage and thus make cultural heritage socially relevant and present in the digital era.

This path was early set as an institutional priority and strategic line in the 2015-2020 Strategy Plan which also followed the legislation and reuse of public information. A large opportunity came in 2016, where a comprehensive program and Digital Agenda was launched with the support of Red.es, a public corporate entity under the Ministry of Energy and Tourism.

BNElab

BNElab was created as a framework for all projects and initiatives under this strategy, covering various lines of action with a focus on cross-sector cooperation and public participation.

The common goal is to promote the use and reuse of the Library's data and digital resources, inspiring and encouraging others to use, create and recreate this digital heritage through:

Data: semantic enrichment, creating new open, reusable datasets and to harness and promote their use for various purposes (research, creative, etc.).

New 'products' for inspiration and search for new uses of the digital collections (i.e. gastronomy, design, fashion, tourism).

Service platforms and digital environments, with tools and specific resources, for researchers, teachers or scholars.

This gave birth to projects such as the crowdsourcing platform ComunidadBNE or BNEscolar. These promote the use and creative reuse of digital cultural heritage in primary and secondary education.

Figure 2. ComunidadBNE, the crowdsourcing platform

All about data

If there is one field of huge and still uncharted potential, that is data. As a national library, the BNE is a massive producer of standardized, interoperable high-value data. This has vast potential for research but also for software development, large scale and creative analysis. Bibliographic and authority catalogues can be created but also vocabularies, directories, linked data catalogues, national bibliographies, statistics and, of course, digitized texts themselves. This is also an immense and diverse knowledge resource, free and available for use and reuse, which has only begun to be used.

Catalogue data produced by the BNE was very early licensed as CC0 (in the early 2010's) and offered as complete sets for downloading in library formats. Under the BNElab framework this bibliographic and authority data, along with all the other possible data sources, was analyzed, harmonized, mapped and segmented. This make this manageable to convert into open reusable formats (csv, json, xml, txt, etc.) following standards for national and international open data recommendations. These full metadata records provide direct access to digital objects and full text. These records are also accompanied with full description of the mapping process and information on the context of the dataset's creation. This is important to help repeat users of data, process the data and evaluate whether the sets meet their needs.

Datasets are hosted in local servers, and referenced from BNElab and datos.gob.es. This national platform and aggregator for open data in the Spanish public sector hosts a total of 285 sets already available for download and reuse.

BNE's data a trustworthy source for new services and knowledge

Open data creates myriad opportunities in libraries, especially in national agencies. This includes new services for digital research, topic modelling, software development, machine learning, new discovery or visualization tools and artificial intelligence.

There is also an internal impact. The analysis of this data and the process of assessing, evaluating and labelling errors helps govern data use within an organization. Through data tools such as OpenRefine feedback from actual or potential repeat users of data is enabled. Open data standards and use therefore help across the entire ecosystem.

Along with data governance, new challenging questions arise as to how we select, organize and share our data. Any historical bias may unintentionally affect and pose a risk in massive analysis or training models. How should our organizations assess and evaluate these risks of inbuilt bias? How should they play an active role in bridging gaps between (open) data and trustworthy AI?

Here, the Spanish Web Archive brings a special case of massive application for artificial intelligence. The example shows how data can not only be used to create AI but also accelerate the development of new AI models.

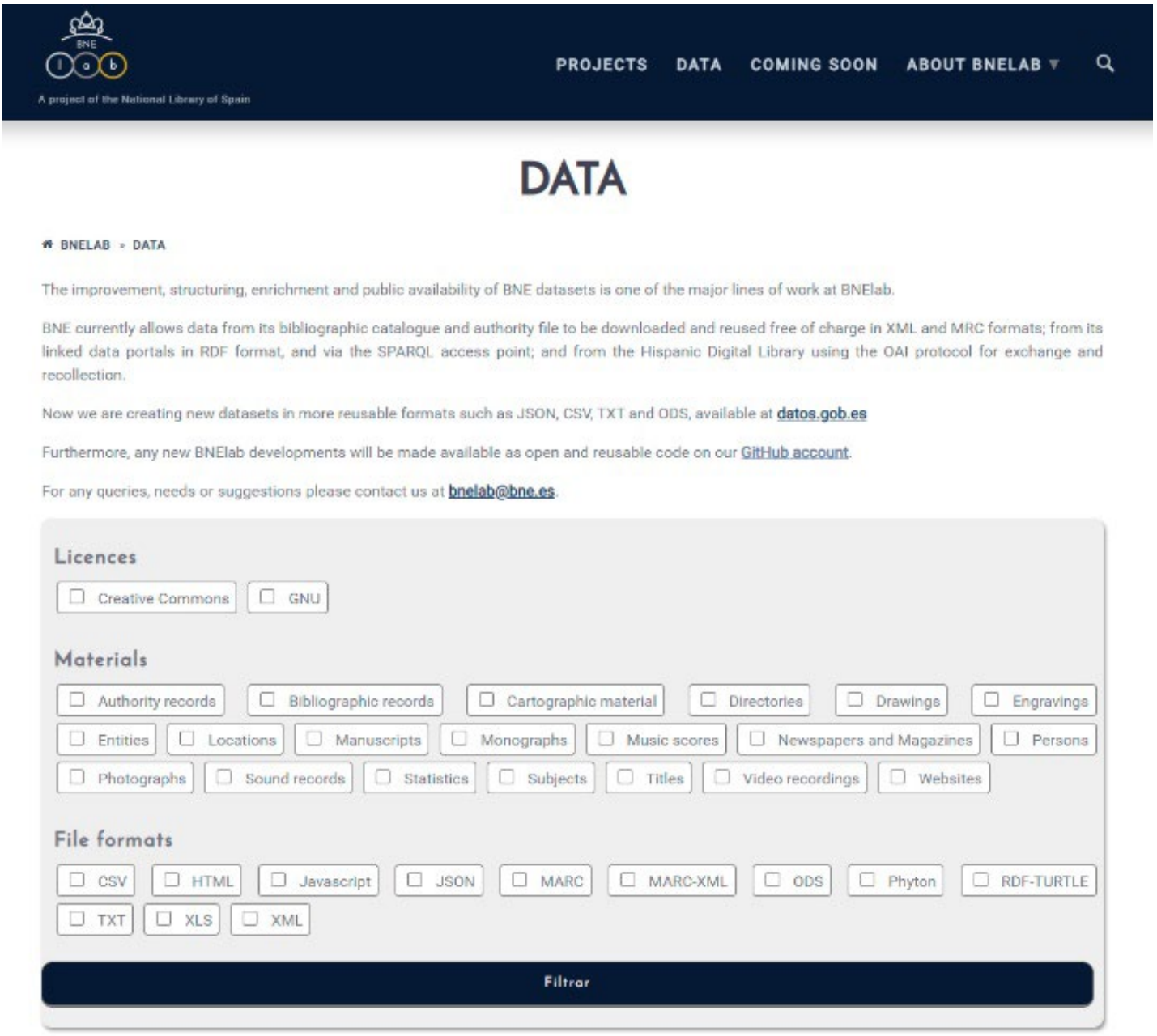


Figure 3. Searching for datasets in BNElab

The Spanish Web Archive as training field for Natural Language Processing models

The BNE has been harvesting information from the web for about a decade. It is quite a young archive, but already contains more than a Petabyte of information. This means it is one of the largest linguistic corpuses in Spain with huge potential for AI models' generation.

The opportunity to explore the creation of the first massive AI model of the Spanish language, using the Spanish Web Archive, came as a partnership with the Barcelona Supercomputing Center (BSC). This occurred under the framework of the Language Technologies Plan of the State Secretariat for Digitization and Artificial Intelligence of the Ministry of Economic Affairs and Digital Agenda of Spain. The Barcelona Supercomputing Center (BSC) is the leading supercomputing center in Spain. They offer infrastructures and supercomputing services to Spanish and European researchers.

Looking at the data

As most national libraries web archives, the Spanish one is based on a mixed model, combining broad and selective crawls. Broad crawls harvest as many Spanish domains as possible without going very deep in navigation. The .es domain is the main scope. Selective crawls complement broad crawls and harvest a smaller range or sample websites with greater depth and frequency. The sites are selected for their relevance to history, society or culture. Other domains such as .org, .com, etc. are included. Web Curators, from the BNE and Spanish regional libraries, select the seeds that will be part of these collections, using different relevance criteria in a heritage preservation context. For the project with the BSC, a collection of around 40,000 websites from selective crawls was considered.

Harvested websites are stored in WARC files (Web ARChive file format). For the purpose of this project the BSC uses

only the text extracted from the WARC files to train the language models, keeping the HTML text tags. Any other component such as audiovisual elements is discarded with the help of a parser script.

This first step of the process was carried out at the Library. Results are transferred to the BSC for a second cleaning task done with the supercomputer MareNostrum.

This is the most powerful computer in Spain and the only one capable of processing such a large volume of data in a short time frame. In the second phase, every not well-formed text was removed. This includes unfinished or duplicated sentences, erroneous encodings, and other languages. The result was only 'well-formed' texts in Spanish.

Spanish Language Models

A repository part of the MarIA project.

Corpora

Corpora	Number of documents	Number of tokens	Size (GB)
BNE	201,080,084	135,733,450,668	570GB

Models

- RoBERTa-base BNE: https://huggingface.co/PlanTL-GOB-ES/roberta-base-bne
- RoBERTa-large BNE: https://huggingface.co/PlanTL-GOB-ES/roberta-large-bne
- GPT2-base BNE: https://huggingface.co/PlanTL-GOB-ES/gpt2-base-bne
- GPT2-large BNE: https://huggingface.co/PlanTL-GOB-ES/gpt2-large-bne
- Other models: *(WIP)*

Fine-tunned models

- RoBERTa-base-BNE for Capitel-POS: https://huggingface.co/PlanTL-GOB-ES/roberta-base-bne-capitel-pos
- RoBERTa-large-BNE for Capitel-POS: https://huggingface.co/PlanTL-GOB-ES/roberta-large-bne-capitel-pos
- RoBERTa-base-BNE for Capitel-NER: https://huggingface.co/PlanTL-GOB-ES/roberta-base-bne-capitel-ner
- RoBERTa-base-BNE for Capitel-NER: https://huggingface.co/PlanTL-GOB-ES/roberta-base-bne-capitel-ner-plus (very robust)
- RoBERTa-large-BNE for Capitel-NER: https://huggingface.co/PlanTL-GOB-ES/roberta-large-bne-capitel-ner
- RoBERTa-base-BNE for SQAC: https://huggingface.co/PlanTL-GOB-ES/roberta-base-bne-sqac
- RoBERTa-large-BNE for SQAC: https://huggingface.co/PlanTL-GOB-ES/roberta-large-bne-sqac

Word embeddings

Word embeddings trained with FastText for 300d:

- CBOW Word embeddings: https://zenodo.org/record/5044988
- Skip-gram Word embeddings: https://zenodo.org/record/5046525

Datasets

- Spanish Question Answering Corpus (SQAC): https://huggingface.co/datasets/PlanTL-GOB-ES/SQAC

Figure 4. The MarIA Model in GitHub

The model comes into play once data files are prepared,. The BSC used a neural network technology based on Transformer. This was already tested with the English language and is now trained with Spanish. The result is an AI model that is able to understand the Spanish vocabulary, expression and writing rules at an expert level. This model is also able to understand abstract concepts and infer word meaning from the context of use.

The model is large and has proved more efficient than other models of the Spanish language. Its name is MarIA. It is open and available in GitHub.

This project represents a milestone in the application of artificial intelligence to Spanish language. It is also a great example of collaboration between national libraries and research centers. The potential uses of MarIA are multiple and diverse. These include language correctors or predictors, auto summarization apps, chatbots, smart searches, translation engines, auto captioning among others. These are all broad fields that promote the use of Spanish for technological applications. This also helps increase MarIA's presence in the world transforming the BNE's data resources and workforce into accessible technology for all.

The long-term results of this project are still to be explored. Currently, the National Library of Spain BNElab initiatives acts a central hub for innovative perspectives and collaborative experiences. Through these initiatives by the National Library of Spain (BNE), firm steps have been taken into deeply transformational paths of data governance and AI.

References

National Library of Spain, general web: http://www.bne.es/en/Inicio/index.html

BNElab: https://bnelab.bne.es/en/

BNElab video-presentation: https://www.youtube.com/watch?v=8tYTk2fJZwE (English subtitles)

BNE linked data catalogue: https://datos.bne.es/

Information about the project, model and technology in http://www.bne.es/en/Inicio/Perfiles/Bibliotecarios/DatosEnlazados/index.html

BNE's datasets in open and reusable formats: https://bnelab.bne.es/en/data/ and https://datos.gob.es/es/catalogo?publisher_display_name=Biblioteca+Nacional+de+Espa%C3%B1a

ComunidadBNE (crowdsourcing platform): https://comunidad.bne.es/

IFLA paper on the project: http://library.ifla.org/id/eprint/2560/ , presented at IFLA WLIC 2019 - Athens, Greece

BNEscolar: https://bnescolar.bne.es/

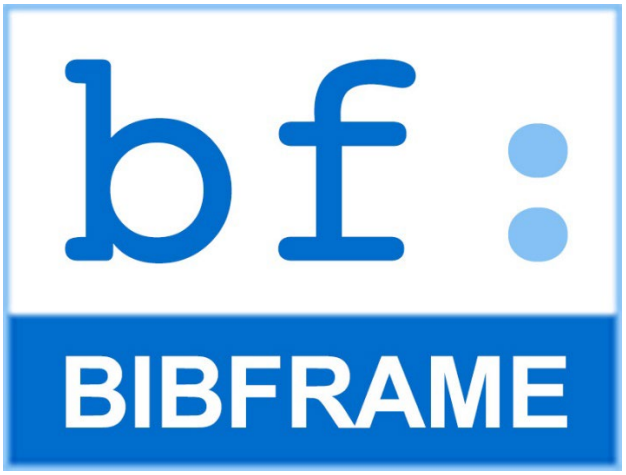
The Spanish Web Archive: http://www.bne.es/en/Colecciones/ArchivoWeb/index.html

Barcelona Supercomputing Center: https://www.bsc.es/

The MarIA model in GitHub: https://github.com/PlanTL-GOB-ES/lm-spanish

Conversion and visualization of MARC Registers to BIBFRAME in the Library of the National Congress of Chile

Marcelo Lorca, mlorca@bcn.cl
Information Analyst, Digital Resources
Library of the National Congress of Chile



Bibframe Logo

Introduction

Sharing data on the web has been on the rise since the beginning of the 21st century. With the evolution of the semantic web, new standards and tools have the potential to integrate different types of systems. Technologies have also been evolving in libraries to exchange bibliographic information. With the advent of the Functional Requirements of Bibliographic Records (FRBR) standards, the development of standards such as Resource Description and Access (RDA) has enhanced the fluent handling of bibliographic data. Working in a standard such as the Bibliographic Framework Initiative (BIBFRAME) allows for a smooth integration mechanism for bibliographic data with other collections and metadata. The visibility of bibliographic data in search engines such as Google allows for the making available of library collections on the web in a natural way for end users.

The Library of the National Congress of Chile (Biblioteca del Congreso Nacional de Chile, or BCN) supports that quality and relevance of information generated is fundamentally improved when bibliographic data is integrated with other data resources on the semantic web. Exploring and enhancing the processes of Machine Readable Catalogue (MARC) record conversion to BIBFRAME registers enables the integration of these different data resources with bibliographic registers in a linked data model.

Background

BIBFRAME is a framework for bibliographic registers which the United States Library of the Congress (LOCS) has been working on since 2011 and which prestigious institutions have now joined. BIBFRAME is proposed as the replacement of the MARC format. It allows the traditional library catalogue's records to be transformed into a compatible format with the semantic web. BIBFRAME takes a part of the RDA model and allows the bibliographic resources to be linked to each other and with other web resources. This brings those bibliographic registers to linked data. BIBFRAME works on a conceptual model that describes the intellectual properties and physical description of bibliographic registers. The model possesses helpful tools and resources available on a website describing how the model is to be implemented.

The main advantages of having the National Library of Chile (BCN) online catalogues on the semantic web, are to:

1) Retrieve and make visible a registry.
2) Standardize the data entries referring to bibliographic records.
3) Interconnect bibliographic data (online catalogue) with other types of information available in semantic web format.
4) Enrich bibliographic data with other open-data platforms.

Bibframe Implementation

Bibframe has been implemented by using the following resources and standards:

a) Language XML. A customizable data tagging language which gives semantic value to the description of data.
b) Resource Description Framework (RDF). A W3C standard for the semantic web which allows for the implementation of classes and properties defined in BIBRAME .
c) RDF(s). A W3C standard for the semantic web which allows for the description of BIBFRAME ontologies in a RDF structure.
d) Seamless integration of bibliographic resources from different institutions

The implementation of BIBFRAME in bibliographic metadata allows for the interaction of bibliographic data with other standards such as Dublin Core, the Metadata

Object Description Schema (MODS), or Simple Knowledge Organisation Systems (SKOS). This enables a scalability in the integration of our catalogs with available resources in linked data, as well as in digital libraries or repositories, wikis, and online encyclopedias.

Although BIBFRAME has only had a few years of development, it already has two versions. The second version corresponds to the work of different institutions led by the LOC from 2014 to 2016. This is the version format with which the BCN works. In this second BIBFRAME model version, one of the most remarkable milestones added is the identity ITEM. The ITEM identity enables the management of specimen data in a unified manner, achieving a better representation of specimen data. With this change, the model resembles the FRBR model. The number of institutions working in this implementation has expanded significantly since its inception. Among them, we can find:

a) German National Library (DNB)

b) National Library of Medicine (NLM)

c) Linked Data for Libraries (LD4L)

d) Music Library Association (MLA)

On the other hand, the FRBR Library Reference Model (FRBR-LRM) is an evolution of FRBR, the Functional Requirements for Authority Data (FRAD) and Functional Requirements for Subject Authority Data (FRSAD). These were the proposed models in the release of the first version of the bibliographic requirements in 1998. This new version unifies the three models in a single library reference model ratified by IFLA. The RDA implementation has been carried out according to these three models. There is continuity between MARC registers, FRBR, RDA and BIBFRAME. Every effort to implement these new models gives a strategic strength to all institutions which implement them.

Model of Conversion

The model of conversion has registers exported in MARC format from the Library Management System (Horizon), including ITEMS data. These registers are grouped in collections and then converted. One of the first collections converted and then exported had about 30,000 registers. An application takes these registers and carries them to a convert database. This database has entities associated to both FRBR-LMR and BIBFRAME. In another application module, the data are then exported to BIBFRAME format.

The application is a program developed in the C# Language and in the Visual Studio.NET2008 environment. It has three essential methods:

1) Loading MARC fields into the database.
2) Transforming and cleaning loaded data.
3) Exporting records in BIBFRAME format to a text file.

The process of conversion developed in the BCN consists of some important elements:

1) MARC Horizon Registers associated to some collections of the BCN.
2) Conversion table from MARC Registers to FRBR-LRM.
3) Application for the preparation of MARCEDIT, MARC Registers.
4) Conversion table from FRBR-LRM to BIBFRAME registers
5) Database conversion in MySQL, based on a proprietary data model trying to represent entities: BIBFRAME and FRBR-LRM.
6) Client server application developed in #C, connecting MARC registers with the relational database. The database behaves as a MARC record data container.

```
524 =LDR  00942fam  2200241 a 4500
525 =001  \\7746
526 =005  20170816000000.0
527 =008  860402s1983\\\\coc\\\\\\\\\\\00000dspa\d
528 =035  \\$aFFFO038
529 =040  \\$aclbc$cclbc$erda
530 =049  \\$ac$bn
531 =080  \\$a860-1(83)(082.2)
532 =245  00$aXXX aniversario :$b[ antologia] /$cCâirculo literario "Carlos Mondaca Cortâes".
533 =260  0\$aLa Serena - Chile :$bJuan Godoy Rivera ,$c1983$e(Coquimbo :$f8ilva)
534 =300  \\$a132 pâaginas :$bretratos ;$c21 cm.
535 =650  \7$aPoesâia chilena$xColecciones
536 =710  2\$aCâirculo Literario Carlos Mondaca Cortâes (La Serena, Chile)
537 =852  80$afg$h860-1(83);T787a;1983$xr=92.409
538 =949  \\$a8C$b8CMOMMN$c860-1(83) T787a 1983$fr=92.409$gMON
539 =336  \\$2rdacontent$atexto
540 =337  \\$anomediado$2rdamedia
541 =338  \\$avolumen$2rdacarrier
542 =952  \\$a860-1(83) T787a 1983$b8C$c8CMOMMN$f7746$i74$knd$nDonado a Liceo Penitenciario$oCDU$p3560700130548 2$yMON$zr=92.409
543
```

Fig 1 MARC Register (MARC Exports)

Data Model of intermediate database between MARC and BIBFRAME

The data model tries to reflect BIBFRAME 2.0 Model´s types and properties, as well as the entities associated with FRBR-LRM, relevant to specific converted registers.

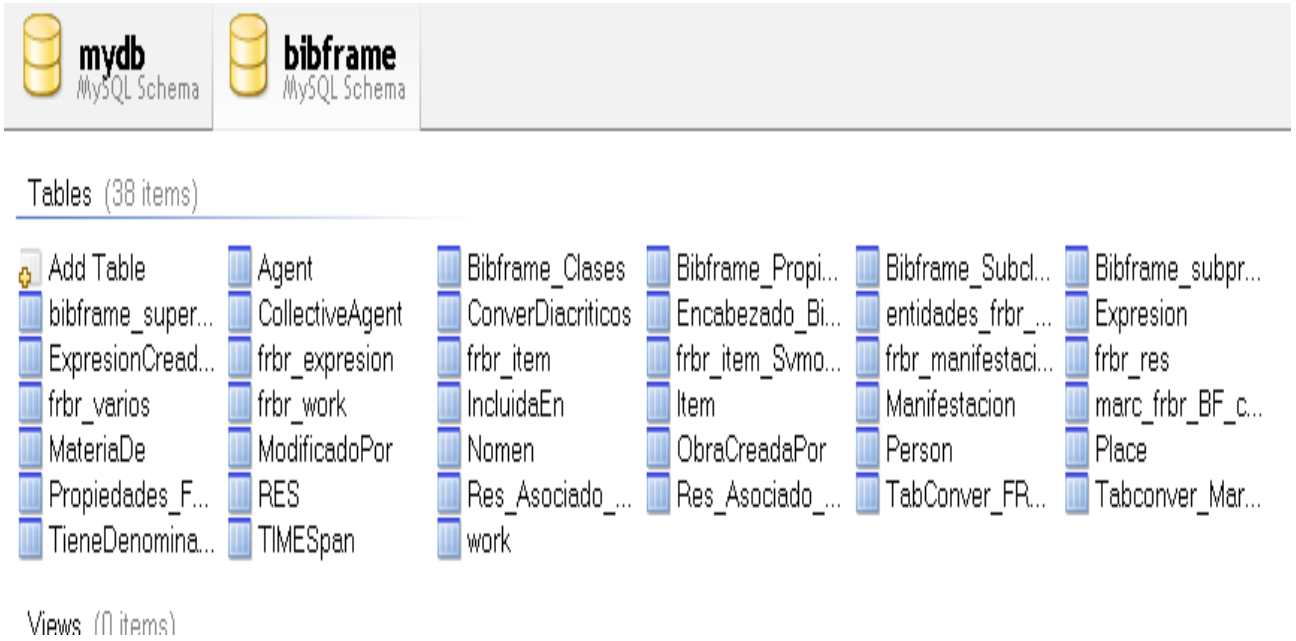


Fig 2. Database conversion tables

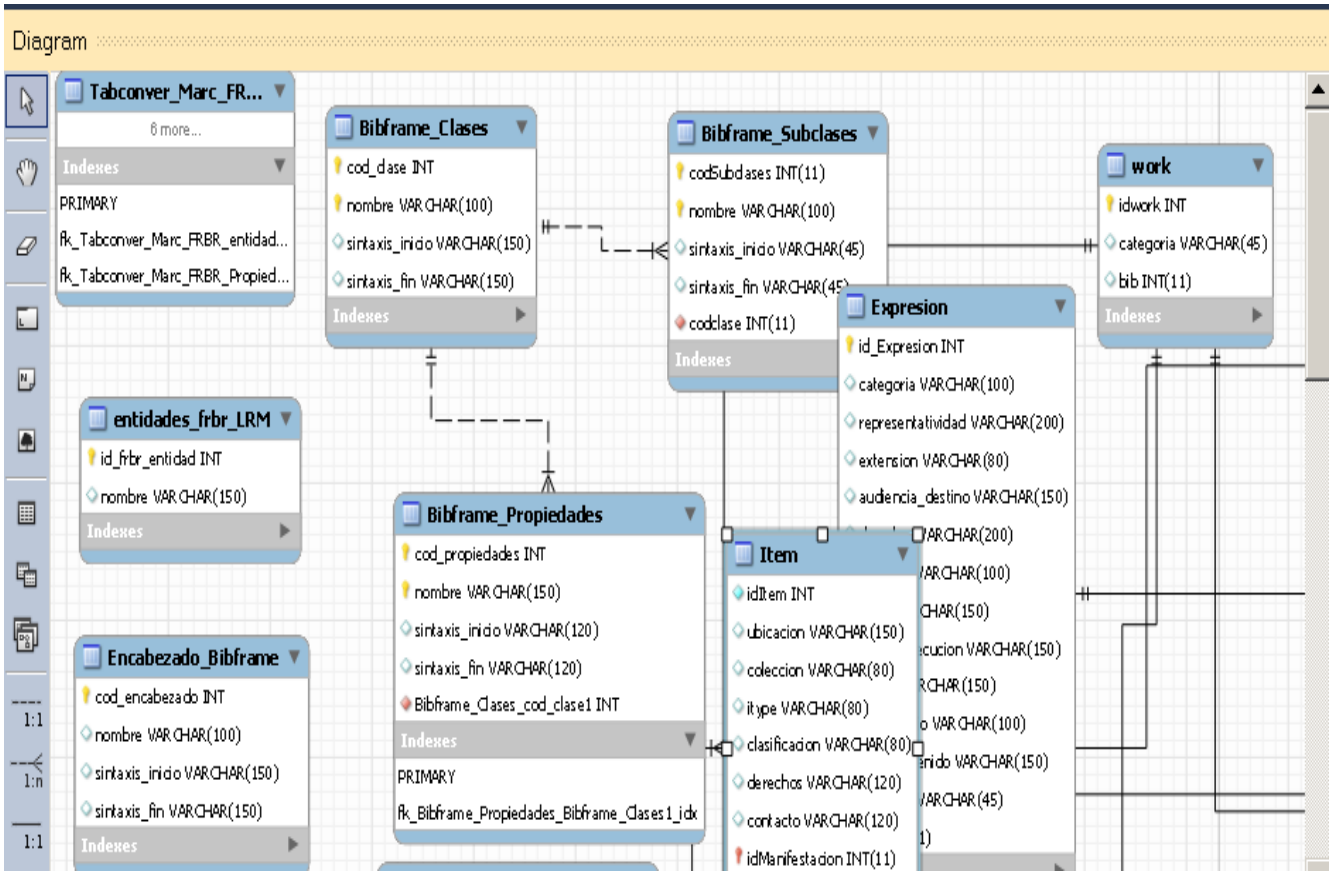


Fig. 3 Entities in conversion database diagram. (Below)

Methodology of the conversion process in the BCN

This methodology consisted of teamwork among professional cataloging experts and librarians specializing in data conversion. The following steps were taken:

1) Implementation of an intermediate conversion database (MySQL). This database stores FRBF-LRM format registers.
2) Implementation of an application created in C#, which interacts with converted MARC files and the aforementioned database.
3) Selection of BCN collections to convert.
4) RDA field settings to MARC registers of the selected collection.
5) MARC registers exported to the selected collection.
6) MARC registers conversion to exported MARC format.
7) Divide MARC registers files into files smaller than 10MB for fluent handling.
8) Apply the program indicated in point 2, and take the files to FRBR-LRM format.
9) Apply the program to convert registers from FRBR-LRM to BIBFRAME.
10) Conversion review and adjustments.

It is important to point out that data recovery from the relational database has been done according to SQL type queries. The most important query that brings the greatest number of registers has been optimized based on the generation of indexes, significantly improving conversion times.

Visualization Proposals

Bibliographic data visualization is one of the most complex aspects when registers are in MARC format. The advent of the FRBR-LRM standard allows for the integration of standardized bibliographic information structures in different library systems. However, the deployment for BIBFRAME requires specifications to enable the exchange of information with other types of data different to those used in the Library Management system. BIBFRAME information is structured in metadata associated to entities that group different types of bibliographic data.

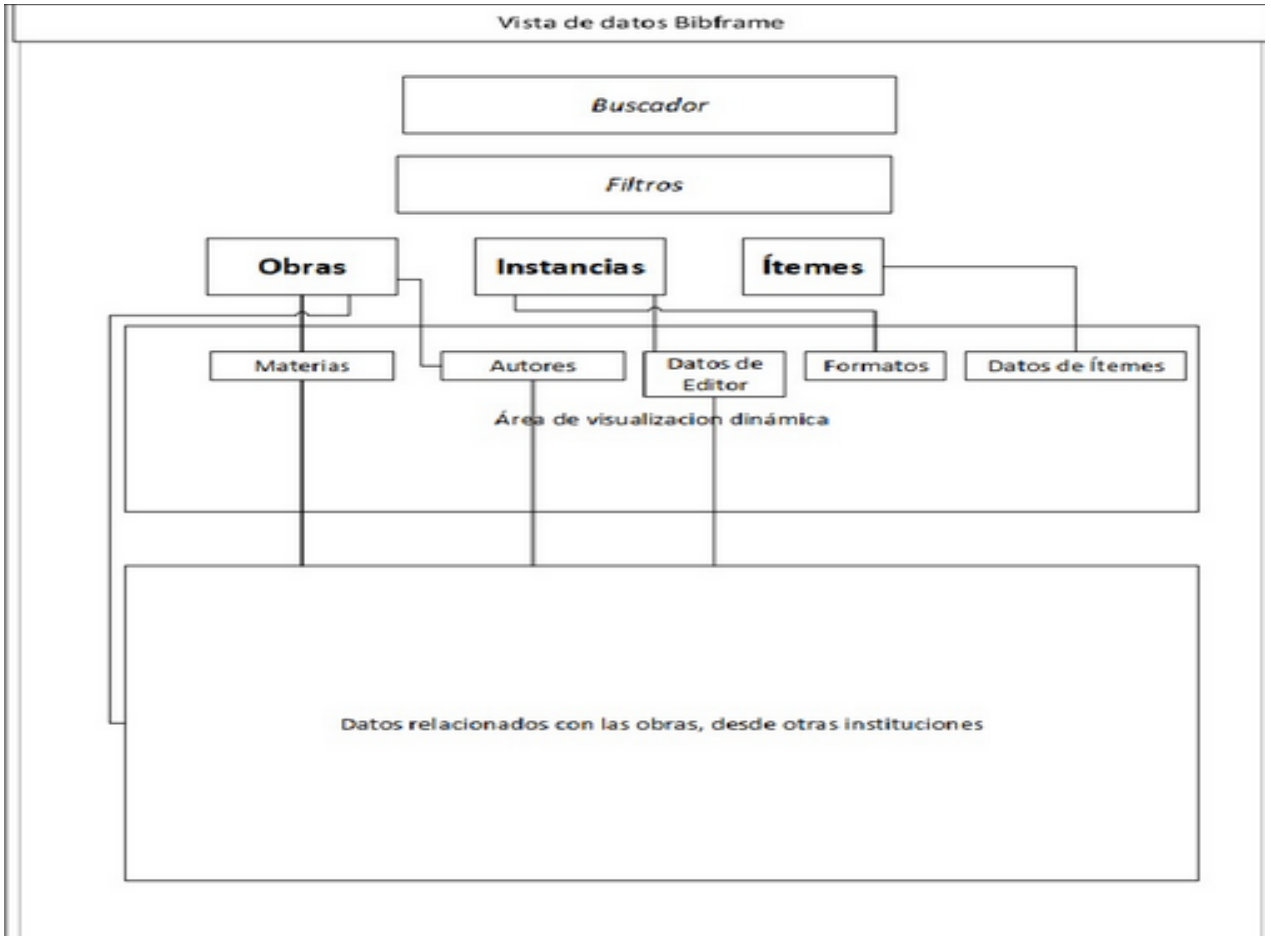


Fig. 4 Visual proposal of the dynamic management of bibliographic Data

Site data structures, instances, items, and authorities allow for a clear differentiation that enhances fluid data visualization and surfing.

We propose a scalable data distribution for its integration with other metadata sets. The proposal must be able to relate different types of works with authors and diverse collections. The works of an author, both personal and institutional, has to be integrated in unique and clear access points. Smooth surfing and a relevant recovery are essential to generate content maps. These allow related contents to be displayed on the foreground, generating filter systems that improve the relevance of the recovered material. This also raises future challenges with respect to the reconciliation of terms and their relationships in hierarchy and similitudes. It is important to generate tools for better management of authorities, clean terms, avoid synonyms and polysemes. Integrating authority registers with the SKOS enhances fluent management of subject language, especially at the level of Library Management Systems.

The next stage involves bringing bibliographic resources to a visualization aligned with graph-type structures and a dynamic relationship of bibliographic contents, from various institutions.

Conclusions

BIBFRAME development as a platform proposes new management in bibliographic cataloguing that involves deep changes in the new Library Management Systems. The integrated work of bibliographic records with other open data systems implies great challenges for teamwork. Strengthening discovery tools and the management of authorities will enable integration of catalogues with the Google-like search system.

The generation of ontologies that integrate diverse types of open data with bibliographic data will ask for creative work with respect to new services to develop in library systems, as well as documentation centers and web portals.

Enhancing processes of conversion from MARC registers to BIBFRAME will also involve strengthening transformations of these registers to the new RDA standard. Coordinated multidisciplinary work will enable prioritization appropriately to collections being converted.

Enhancing and implementing Editors in BIBFRAME involves deep transformations in the management of future cataloguing processes and classification. The form and workflows in technical processes will be transformed. This implies much more integrated work for new generation web resources. A more integrated way of working with different institutions will be required.

Challenges of converting our collections to Bibframe involves developing new paradigms and services. This will enhance the integration of bibliographic resources as a fundamental part of open data.

The challenges that BIBFRAME poses at the level of Latin American institutions are important. We should rethink our approach to issues of sharing our data and reusability. This is apart from incorporating the concepts of the semantic web and shared and related data in the education of the future generation of librarians. Our philosophy should focus on enhancing the exploitation of these registers, and integrating them with other semantic platforms.

Bibliography

(1)AGENJO, X. AND HERNÁNDEZ-CARRASCAL, F. Bibframe previews in 2014. TK, 9, 269. http://dx.doi.org/10.3145/thinkepi.2015.64

(2)Laura Po, Nikos Bikakis, Federico Desimoni, and George Papastefanatos. Linked Data Visualization: Techniques, Tools and Big Data, Morgan & Claypool, 2020.

(3)Martins Zviedris. Ontology Enabled Graphical Database Query Tool for End-Users. Institute of Mathematics and Computer Science, University of Latvia.

(4)Overview of the BIBFRAME 2.0 Model. https://www.loc.gov/bibframe/docs/bibframe2-model.html.

The Development of an audio-visual aggregator: the Open Audio-Visual Archives (OAVA) project in Greece

Afrodite Malliari, afmalliari@gmail.com
International Hellenic University, Department of Library, Archival & Information Studies

Ilias Nitsos, initsos@gmail.com
International Hellenic University, Department of Library, Archival & Information Studies

Sofia Zapounidou, zapounidous@gmail.com
Aristotle University of Thessaloniki, Library & Information Centre

Stavros Doropoulos, doro@datascouting.com
DataScouting

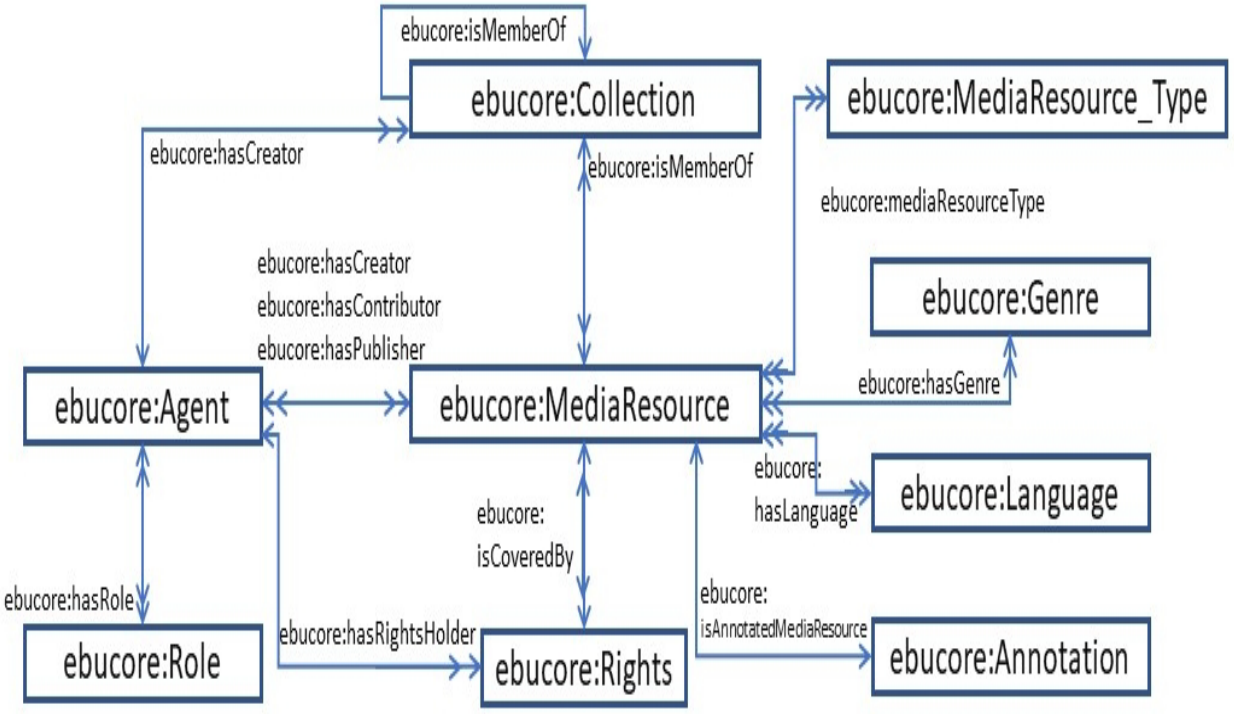


Figure 1. The OAVA Model depicting classes, and the relationships between them

Abstract

The available online audiovisual material maintained by public and private organizations and institutions is numerous and ranges from videos and narrations regarding historical and everyday life events to scientific, academic, and cultural events. Despite the apparent availability of audiovisual resources, finding useful resources is not as easy as one might think. Poor indexing by search engines and lack of metadata schemes are the main reasons for this inconvenience. The problem of dispersion of audiovisual resources and diversity of resource providers is largely solved by aggregation services.
In Greece there is no reference point for the search and access to audiovisual material. A Greek National Registry of audiovisual providers has not yet been implemented. The Open Audio-Visual Archives (OAVA) project aims to gather audiovisual material that is of Greek interest or contains Greek speech. The OAVA project provides a unified search mechanism not only to the aggregate metadata of audio-visual material but also to its searchable content. Through the application of deep learning models, algorithms are developed that perform Automatic Speech Recognition in Greek and in English.

During the project, 500 Greek-language audiovisual content providers were reviewed. 233 of them were found eligible according to specific criteria. The metadata used by each provider were mainly application schemes without metadata schemes. Those were recorded and mapped with Vufind and with EBUCore schema. The open-source software VuFind was configured to operate as the basis of the unified search platform.

This article briefly presents the objectives of the project, selection criteria, results for content and audiovisual providers in Greece. It overviews licensing issues, the OAVA metadata scheme and the basic functions of the search mechanism.

KEYWORDS

Audiovisual material, speech to text technologies, cultural heritage, open access, content aggregators

Introduction

The plethora of web content nowadays, including audiovisual content, is constantly creating challenges in terms of harvesting, indexing and searching. In the case of audiovisual content, much of the existing material remains unused, because it cannot be detected by potential visitors. In order to provide a single point of access, the creation of aggregation services is the only option for national or international organizations that wish to promote and make better use of online audiovisual resources. A typical example is Europeana (c2022), that focuses on providing European cultural heritage institutions with the necessary tools that will enable them to effectively share their collections (Europeana Foundation, 2019). Another effort is Trove (c2022), that aggregates content from "libraries, museums, galleries, the media, government and community organizations" in Australia (trove, c2022). Memobase (Memoriav, c2022) and DigitalNZ (National Library of Zealand, c2022) with similar examples in Switzerland.

The article examines the need for the development of an aggregator for content offered by Greek providers. It describes the steps to create this type of aggregator, as part of the OAVA project. Deep Learning models are used to create searchable text of the audiovisual content in Greek and English. An aggregator not only collects metadata, but also provides speech search mechanisms.

Research

The first step in creating the Open Audio-Visual Archives (OAVA) platform to provide access to various audiovisual resources was to create a multidisciplinary dataset. The dataset should contain resources, free of copyright limitations, with informative content excluding literary or artistic. The OAVA project aims to provide access not only to aggregated metadata but also to searchable content. It apples deep learning models and converts the speech contained in the audiovisual material into searchable text.

As there is no national searchable registry of audiovisual material in Greece, we gathered resources by:

a) conducting an online survey to Greek reference librarians who in some cases suggested and provided audiovisual resources to their users

b) searching with keywords on the web (e.g. "local history" AND "video")

c) browsing the websites of organizations that were expected to provide audiovisual content due to their activities and functions (Malliari, et.al., 2022).

A study of mapping collections of audiovisual resources in Europe (Klijn, & de Lusenet, 2008) was considered in order to select the websites.

The selection process took place in the first quarter of 2021. 500 providers of Greek-language audiovisual content were evaluated using the crAAp test (Currency, Relevance, Authority, Accuracy, and Purpose). If a website was eligible for further evaluation, its resources were evaluated with specific selection criteria based on:

(a) the Digital Library Federation and Council on Library and Information Resources study (Pitschmann, 2001) and (b) criteria found in other studies (Digital Public Library of America, c2022; Oesterlen, 2017; Scholz, 2019; Trove, c2020; Turnok, Kaye, & Carrasqueiro, 2010).

Criteria were grouped under context, content, form/use, process or technical, and metadata.

A total of 497 of the 500 candidate providers evaluated were found to be reliable by applying the crAAp test to their content. The review process was completed with the specific selection criteria mentioned. Eventually 233 providers were found eligible and included in the final trusted list containing: libraries, archives, museums, universities, governmental and non-governmental organizations and media organizations. More than half of them (58%) came from the public sector and universities. Universities contribute most of the content.

Approximately 80% of the total of 1710 resources offered came from universities. It is also interesting to mention that the majority of the resources were "open courses" (77%). Other resource categories were: "open educational material", "academic and/ or scientific events", "cultural events", "Interviews", "board meetings", "Campaigns", "archival material".

In terms of licensing, Greek audiovisual content providers tend to choose the same licensing method for all their resources. More than half of them chose to publish their resources in platforms. Their content can be reused under the terms of "Fair Use" (66%). 28% of the providers prefer to publish under "Creative Commons" licenses. 5% of providers have their own terms of use that do not match the existing licensing typology. Another small percentage (1%) use rights statements.

The Malliari, et.al. research (2022) revealed that most of the providers do not own any kind of publishing infrastructure. They utilize commercial streaming services, such as YouTube. In terms of metadata policy, almost every eligible provider (97%) uses an application profile developed according to their own needs and/or the capabilities of the information system they use. They do not utilize a well-known metadata schema, like DC. The dataset with the above-mentioned information regarding provider's name and type, channel publishing AV, resource genre and URL, licensing and metadata can be found at Zenodo[1].

OAVA schema

The OAVA project aims to collect online audiovisual content. Technical collaboration with stakeholders is not guaranteed. This is unlike other aggregation services such as Europeana and Digital Public Library of America . To aggregate and display the resources metadata into the OAVA platform, the metadata must conform to the OAVA schema.

[1] https://doi.org/10.5281/zenodo.5112283

The OAVA schema is based on the EBUCore Schema. The EBUCore Schema has been developed by the European Broadcasting Union for the description of broadcasting resources (EBU, 2020). It conforms to the EBU Class Conceptual Data Model (CCDM). It also takes other models into account (e.g., FRBR, Europeana Data Model) to enable mappings. The OAVA Schema uses only a subset of the EBUCore Schema. The descriptive needs are different in the OAVA project compared to the broadcasting resources on which the EBUCore Schema focuses.

The OAVA Schema uses 4 main classes: Media Resource, Collection, Rights, and Agent. The Media Resource class is used to describe the audiovisual resource providing information about the title, language, URL, abstract, thumbnail, etc. In case the Media Resource is part of a collection, the Collection class is used to provide the title and the description of the collection. The Rights class provides information about the rights by which a Media Resource instance is covered. The Agent class describes the agent. This is either the corporate body or person that has created, contributed to, or published the Media Resource. The OAVA Schema also uses other classes that use controlled vocabularies (i.e., Media Resource Type, Genre, Language). The remaining Annotation class is used to describe the enrichment of the metadata by the OAVA project. In the following figure, the OAVA model is displayed in Figure 1 above. Classes are depicted as rectangles.

The relationships between classes are depicted as arrows. The type of the relationship (1-1, 1-M, M-N) and cardinalities are also presented. To instantiate the model, metadata from providers are aggregated in a VUFind instance. They are then exported using the OAVA Schema conceptualizations. The metadata from providers are aggregated from publishers' webpages or systems. This often uses heuristic methods due to the lack of metadata standards. The aggregated metadata are enriched and mapped to the VUFind Schema that has been extended to conform to the OAVA Schema (see figure 2). Some examples of metadata enrichment involve the creation of a thumbnail for the resource, the assignment of type (e.g., video or audio), the provider's logo, the language of the resource, etc. The VUFind Schema consists of more than 100 elements, of which some are repeatable. For the needs of the OAVA project, these elements were analyzed to discover correspondences to the elements of the OAVA schema. When such correspondences were not found, the needed OAVA Schema elements (subset of the EBUCore Schema) were added to the VUFind schema. For the needs of the OAVA project, 22 elements from the VUFind were selected as corresponding to OAVA schema properties. 16 properties from the OAVA Schema were added (all under the ebucore namespace).

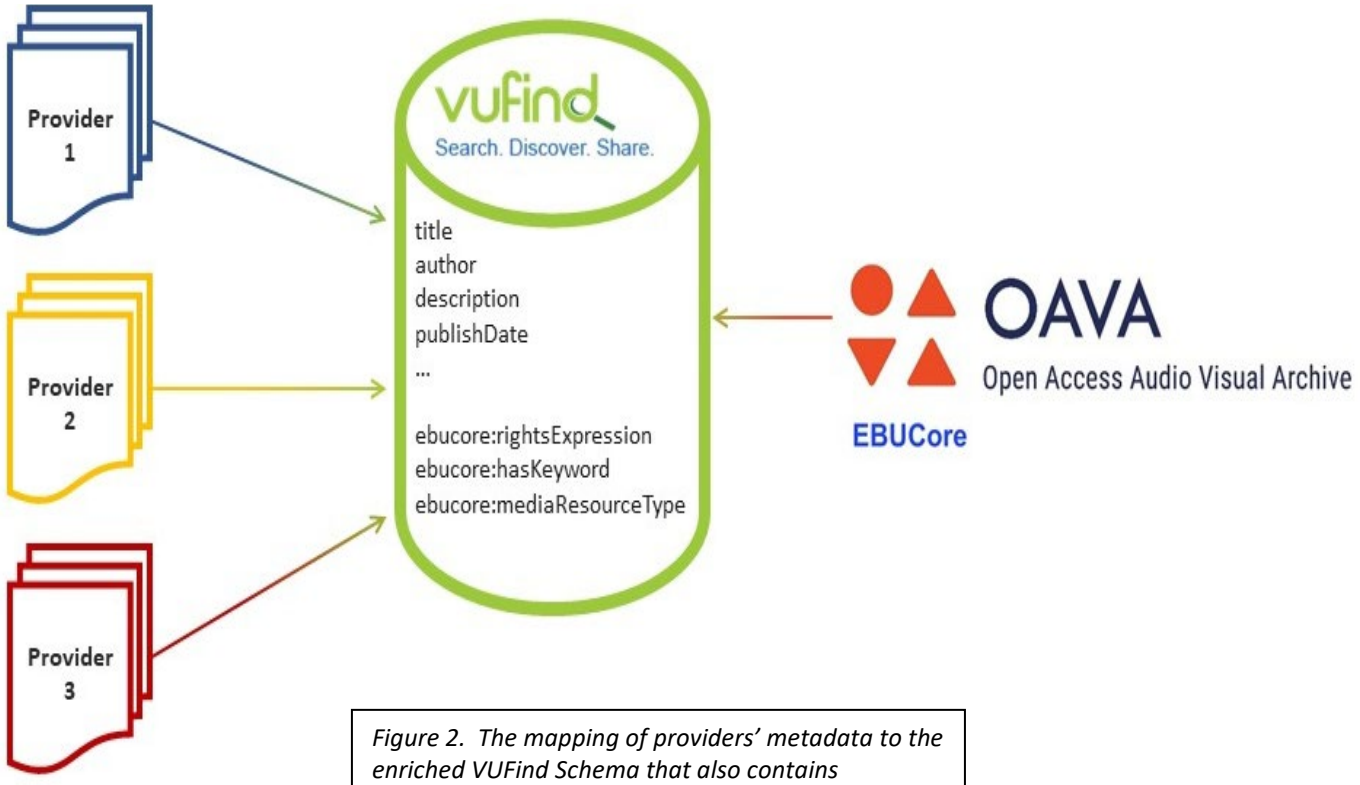


Figure 2. The mapping of providers' metadata to the enriched VUFind Schema that also contains properties from the OAVA Schema.

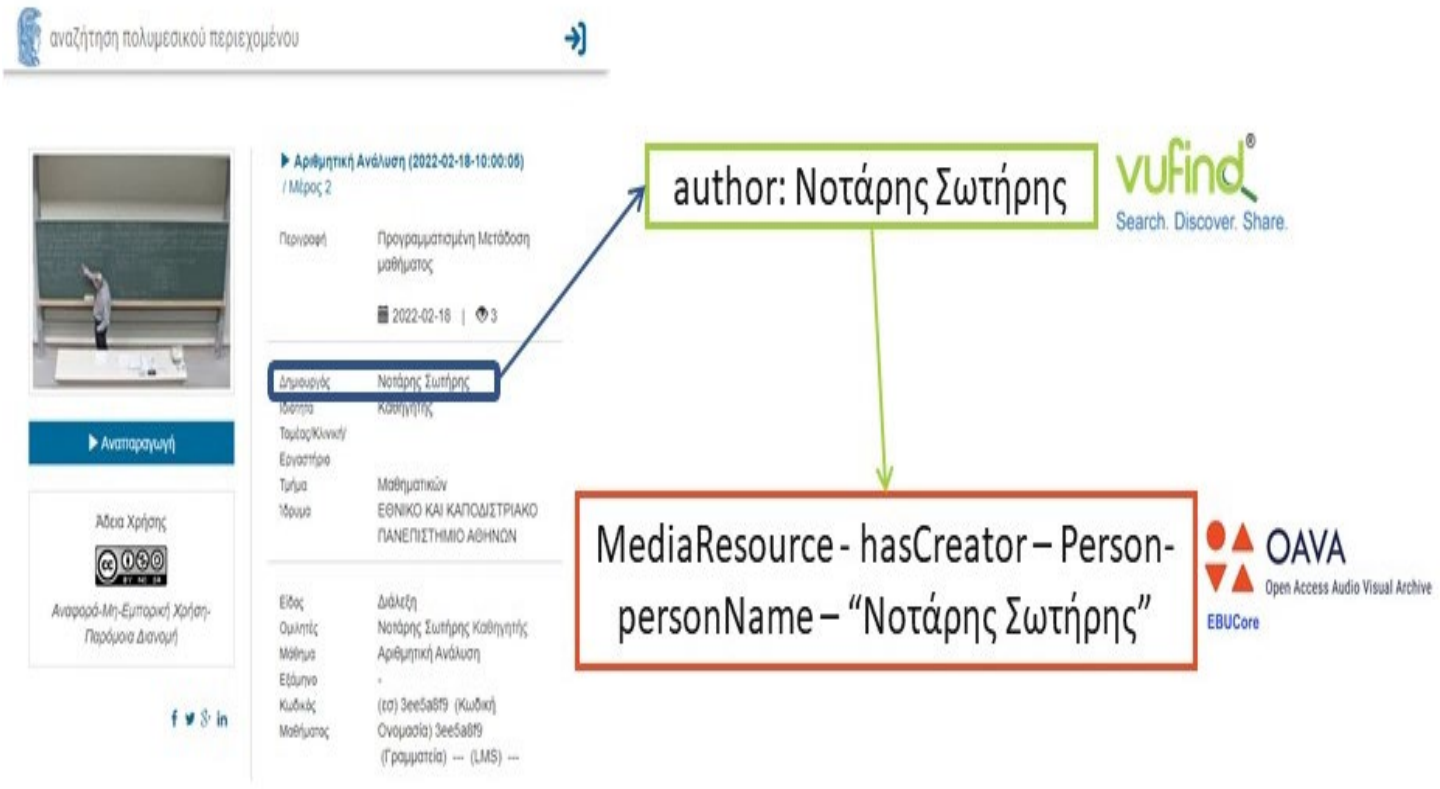


Figure 3. Author information from the provider's site is mapped to the VUFind element author. This piece of information triggers the creation of two triples MediaResource - hasCreator - Person and Person - personName - 'Νοτάρης Σωτήρης'

.

After the metadata is aggregated and stored in the VUFind Schema, it can be exported as triples using the OAVA Schema. Each record in VUFind describes a *MediaResource*. From the value of one property more than one triple can be created. In the following example, a lecture is described having as an author the *Person* 'Νοτάρης Σωτήρης'. After mapping to the OAVA Schema and exporting in triples, two triples are created: *MediaResource - hasCreator - Person* and *Person - personName* - 'Νοτάρης Σωτήρης' (Figure 3).

OAVA platform

A general overview of the OAVA platform architecture is presented below. Basic components and processing stages of the OAVA platform are now discussed. (Detailed information is included for the Automatic Speech Recognition approach, especially for Greek content).
The orchestrator is a basic component of the OAVA platform. It is used to monitor sources of audiovisual content. When the orchestrator finds new content that is missing from OAVA, it initiates a crawling procedure. The Worker is another basic component of the platform. It can crawl pages using either a traditional or headless browser that does not have a graphical user interface. While the crawling process is going on, the crawled pages are sent to a scraping procedure. A messaging agent, RabbitMQ, has been used to communicate between OAVA services. This allows OAVA to be horizontally scalable and distributed in different systems.
The scraping procedure plays a significant role in the functionality of the platform. The scraping determines whether the crawled page contains audiovisual material that needs to be indexed and analyzed. Once it finds such material it begins to extract useful information (title, description, authors, publishers, video links, etc.). The OAVA scraper maps each different field into a common model. Eventually, all the scraped information originating from different sources is mapped to the same format.
After the completion of the scraping procedure, the produced scraped information is sent back to the orchestrator. It is prepared for publishing and further enriched with metadata. The orchestrator then finds any multimedia links attached to this scraped information. It then forwards these links to a storage manager which the storage manager evaluates. The storage tries to find the best approach to get the multimedia content attached to that link. As soon as the downloading process of these multimedia files has been completed, the storage manager sends back the locally stored file paths to the orchestrator. The orchestrator then publishes the information containing the multimedia file paths to the main platform.

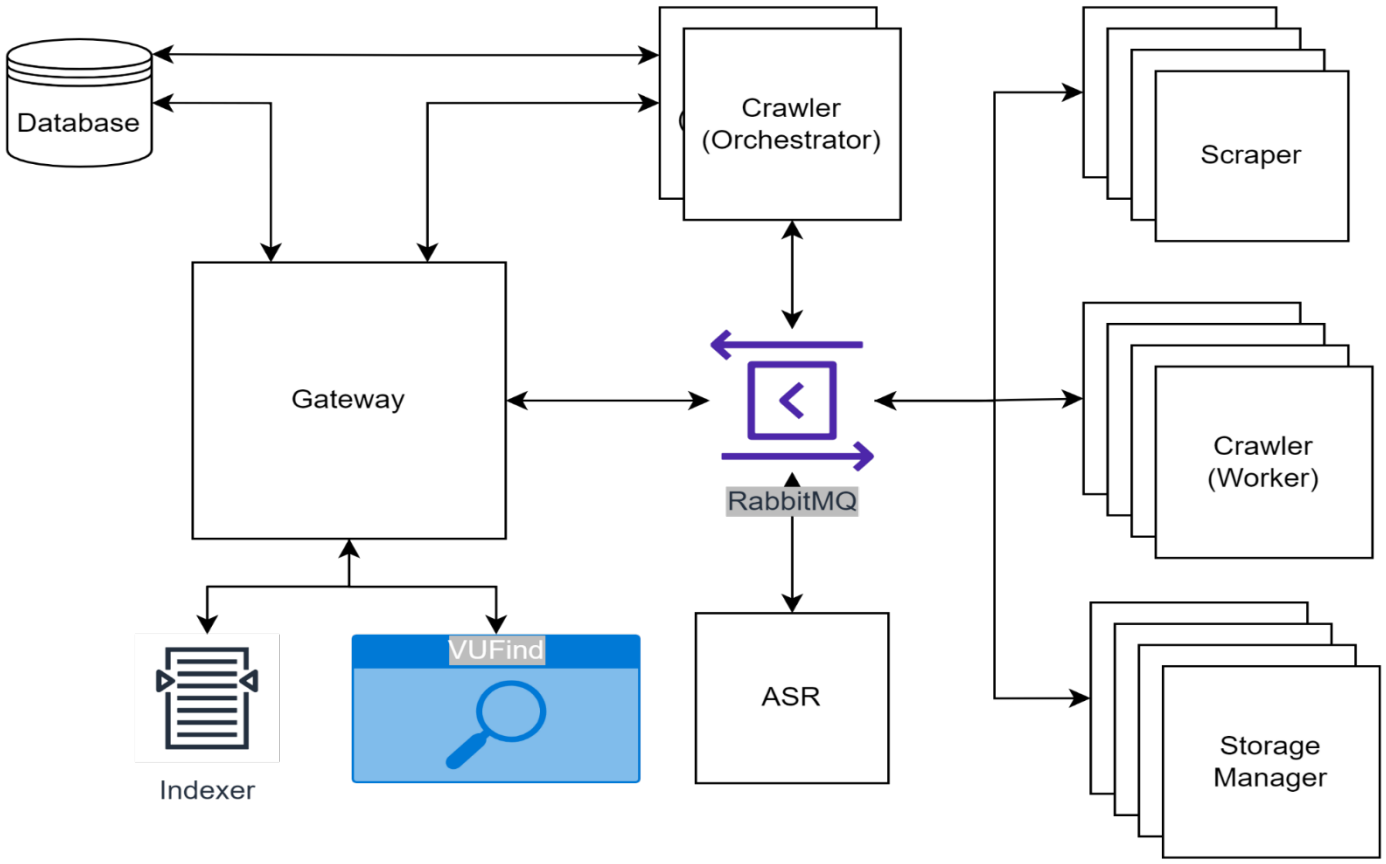


Figure 4. Overview of the OAVA platform architecture

Automatic Speech Recognition

One of the main jobs of the OAVA platform is to extract text. This is from the downloaded multimedia files using Machine Learning and more specifically Automatic Speech Recognition techniques. Automatic Speech Recognition models were developed for both English and Greek content. Models including datasets with over 7000 hours of oral speech in total were used Especially for the English language, (e.g., Librispeech, Fisher Corpus, Switchboard-1 Dataset, WSJ-0 and WSJ-1, National Speech Corpus - 1, Mozilla Common Voice). The top performing models also required a lot of resources for inference. For the English language we selected architectures that require less resources without significantly sacrificing performance. Examples of such models are CTC architectures (Graves et al., 2006) like Jasper (Li et al., 2019) and QuartzNet (Kriman et al., 2019). As a post-processing step we use a 4-gram language model to improve the output of ASR models.

For the Greek language there are no big datasets available. Training such models requires large scale speech recognition datasets. These consist of speech in audio format and the corresponding transcription. A few examples of Greek datasets are: Greek Mozilla Common Voice (22 hours), Transcribedbook (4 hours), Greek TEDx speeches (20 hours). It is apparent that the size of this corpus is very small compared to English datasets. To this end, transcribing additional speech data is required for training a Greek ASR model. In OAVA, more than 100 hours of additional data was transcribed and made available for training. This dataset includes: audio from the Greek parliament channel, Greek news channels, university open lectures, Onassis foundation events, NGO events and library events. After combining these datasets training of various state-of-the-art architectures took place for testing purposes. Note that for each experiment the weights of the English pretrained model were used as a starting point as preliminary experiments revealed better performance in the final Greek model when starting the training from an English pretrained model, instead of random weight initialization. For the evaluation dataset, the audio from the university open lectures was merged with Onassis foundation event, NGO events and library events. The focus was mainly on two architecture categories, namely the Connectionist Temporal Classification (CTC) models (Graves et al., 2006) and RNN-Transducers (RNN-T) (Graves et al., 2012). For CTC models, apart from Jasper and QuartzNet experiments were also conducted with Citrinet (Majumdar et al., 2021). For the RNN-T models experiments were conducted with ContextNet (Han et al., 2020) and Conformer (Gulati et al., 2020).

The overall WER of the transducer model was further improved by adding a language model as a post processing step. To do this a 6-gram language model was trained using millions of articles that were crawled from Greek blogs. A Greek language model was trained that was applied after the conformer transcription. The final WER was close to 0.20 in the evaluation set.

When audio arrives in the Automatic speech recognition module it is processed. The corresponding English or Greek model is then applied to produce transcription. As soon as the process is finished the new enhanced information enriches the existing one. The multimedia files used in the process are then permanently deleted. After all this process, the produced information is exported and mapped to VUFind.

Conclusions

This article has given a high-level outline of a research project aimed at developing a platform that will aggregate audiovisual resources by Greek providers. As search results give access not only to aggregated metadata (using the OAVA model), but also searchable audio content, resources should be free of copyright limitations. In this respect, only informative content has been selected (not literary or artistic). Deep learning models have been applied that convert the speech contained in the audiovisual material into searchable text to enhance the user search experience.

Funding

This work was supported by the European Regional Development Fund of the European Union and Greek national funds through the Operational Program Competitiveness, Entrepreneurship and Innovation, under the call RESEARCH - CREATE - INNOVATE [project code: T2EDK-00526].
Project partners: Department of Library, Archival and Information Studies, International Hellenic University; DataScouting, Software Development Company

References

Digital Public Library of America. (c2022). *Collection Development Guidelines*. Available from: https://pro.dp.la/hubs/collection-development-guidelines
EBU. (2020). *EBU Core Metadata Set (EBUCore): Specification v. 1.10. TECH 3293*. Available from: https://tech.ebu.ch/docs/tech/tech3293.pdf
Europeana. (c2022). *Europeana.* Available from: https://www.europeana.eu/
Europeana Foundation. (2019). *Europeana Publishing Framework 2.0.* Available from:https://pro.europeana.eu/files/Europeana_Professional/Publications/Publishing_Framework/Europeana%20Publishing%20Framework%20V2.0%20English.pdf
Graves, A. (2012). *Sequence transduction with recurrent neural networks*. arXiv preprint arXiv:1211.3711.
Graves, A., Fernández, S., Gomez, F., & Schmidhuber, J. (2006, June). Connectionist temporal classification: labelling unsegmented sequence data with recurrent neural networks. In *Proceedings of the 23rd international conference on Machine learning* (pp. 369-376). doi: 10.1145/1143844.1143891
Gulati, A., Qin, J., Chiu, C. C., Parmar, N., Zhang, Y., Yu, J., ... & Pang, R. (2020). *Conformer: Convolution-augmented transformer for speech recognition*. arXiv preprint arXiv:2005.08100.
Han, W., Zhang, Z., Zhang, Y., Yu, J., Chiu, C. C., Qin, J., ... & Wu, Y. (2020). *Contextnet: Improving convolutional neural networks for automatic speech recognition with global context*. arXiv preprint arXiv:2005.03191.
Klijn, E., & de Lusenet, Y. (2008). *Tracking the reel world: A survey of audiovisual collections in Europe*. Amsterdam: European Commission on Preservation and Access. Available from: https://www.ica.org/sites/default/files/WG_2008_PAAG-tracking_the_reel_world_EN.pdf
Kriman, S., Beliaev, S., Ginsburg, B., Huang, J., Kuchaiev, O., Lavrukhin, V., ... & Zhang, Y. (2020, May). Quartznet: Deep automatic speech recognition with 1d time-channel separable convolutions. In *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 6124-6128). IEEE. doi: 10.1109/ICASSP40776.2020.9053889
Li, J., Lavrukhin, V., Ginsburg, B., Leary, R., Kuchaiev, O., Cohen, J. M., ... & Gadde, R. T. (2019). *Jasper: An end-to-end convolutional neural acoustic model*. arXiv preprint arXiv:1904.03288.
Majumdar, S., Balam, J., Hrinchuk, O., Lavrukhin, V., Noroozi, V., & Ginsburg, B. (2021). *Citrinet: Closing the Gap between Non-Autoregressive and Autoregressive End-to-End Models for Automatic Speech Recognition*. arXiv preprint arXiv:2104.01721.
Malliari, A., Nitsos, I., Zapounidou, S., & Doropoulos, S. (2022). Mapping audiovisual content providers and resources in Greece. *International Journal on Digital Libraries*, 1-11. doi: 10.1007/s00799-022-00321-6
Memoriav. (c2022). *Memobase: Access to Switzerland's audiovisual cultural heritage*. Available from: https://memoriav.ch/en/memobase
National Library of New Zealand. (c2022). *About DigitalNZ.* Available from: https://digitalnz.org/about
Oesterlen, E.-M. (2017). *Aggregation Handbook, 3rd edition*. The Hague: Europeana, EUscreen. Available from: http://blog.euscreen.eu/wp-content/uploads/2017/03/Aggregation-Handbook_Revised-edition3_2017.pdf
Pitschmann, L.A. (2001). *Building Sustainable Collections of Free Third-Party Web Resources*. Washington, D.C.: Digital Library Federation, Council on Library and Information Resources. Available from: https://clir.wordpress.clir.org/wp-content/uploads/sites/6/pub98_57d70f70b208f.pdf
Scholz, Henning. (2019). *Europeana Publishing Guide v1.8: A guide to the metadata and content requirements for data partners publishing material in Europeana Collections*. The Hague: Europeana Foundation. Available from: https://pro.europeana.eu/files/Europeana_Professional/Publications/Europeana%20Publishing%20Guide%20v1.8.pdf
Trove. (c2020). *Technical specifications*. Available from: https://trove.nla.gov.au/technical-specifications
Trove. (c2022). *What is Trove.* Available from: https://trove.nla.gov.au/about/what-trove
Turnok, R., Kaye, L., & Carrasqueiro, L. (2010). EUscreen content selection policy. In *EUscreen Workshop, 23-24 June, Mykonos, Greece*. Available from: https://www.slideshare.net/EUscreen/turnok-kayecarrasqueiro-e-uscreen-content-selection-policy-euscreen-mykonos

Ethics case studies of Artificial Intelligence for library and information professionals

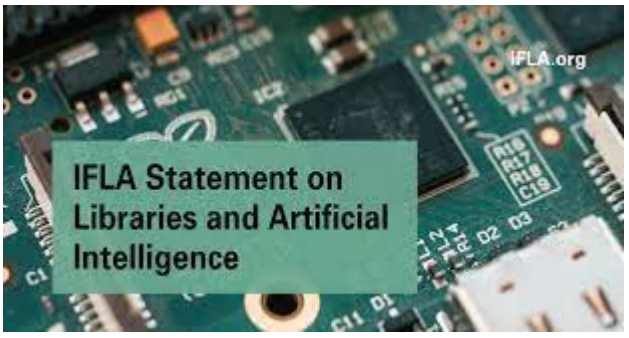
Andrew Cox, a.m.cox@sheffield.ac.uk
Information School, University of Sheffield, UK

Introduction

Artificial Intelligence (AI) offers exciting possibilities for library and information professionals and our users. It promises to increase access to knowledge by offering new ways to automatically describe and retrieve information from collections (Cordell, 2019; Cox, 2021). It could enable adaptivity and personalisation in information provision. AI driven chatbots and voice agents provide dialogic and supportive ways of accessing information. AI could also be applied to the analysis and prediction of user behavior. However, AI has raised a storm of public ethical concern, especially relating to bias, intelligibility and privacy (AINow 2018, 2019). In response, many organisations have published guidelines for ethical AI (Jobin, 2019). Among these are some useful sector specific guides about how to approach the development of ethical AI, e.g. for education (The institute for ethical AI in education, 2020; JISC, 2021). These are invaluable for information professionals working in that sector. This paper is written from the belief that librarians and information professionals have a unique value perspective on AI. This itself needs to be fully developed. Our calling to promote access to knowledge brings a unique perspective on the ethics of AI. Certainly, information professionals need more resources to help think through the ethics of AI in their practice.

Information Professionals, Ethics and AI Scenarios

One of the defining characteristics of a profession is its commitment to ethical conduct. But codes of ethics published by professional bodies are increasingly aspirational (Frankel, 1989): they articulate shared values at an abstract level, rather than explaining in detail how to behave in particular situations. Furthermore, they are not updated frequently enough to respond to new ethical challenges. Advancing technologies seem to be an important locus of such new ethical dilemmas, albeit it is debatable how often they raise fundamentally new questions (Ferguson, et al., 2016). Ferguson, et al.'s (2015) study of ethical awareness around RFID suggests that where technologies are concerned, information professionals may be over-reliant on vendors to ensure that their use is ethical. In this context other materials beyond broad ethics codes are needed to support professionals to think through the implications of AI in their work. Some very useful resources already exist. IFLA (2019) has published an insightful commentary on the ethical implications of AI. Padilla's (2019) report for OCLC offers a guide to responsible AI development. Yet there remain gaps in the tools we have as a profession to explore our ethical responses to these new technologies.



IFLA Statement on Artificial Intelligence, http://bit.ly/2FLDxpg

Ethics scenarios can play an important role here in posing dilemmas in relatable ways, stimulating us to reflect and debate how we think we ought to act. As essentially stories they are engaging ways to open up challenging topics. They can relate closely to everyday practical issues and are effective in promoting adult learning. We have good sets of ethics scenarios for information work in general such as in Buchanan and Henderson (2009), McMenemy, Poulter & Burton (2014) and Rösch (nd). However, AI is important enough to require its own set of scenarios. The purpose of this paper is to explain the thinking behind a set of scenarios of AI ethics for library and information professionals. The scenarios themselves can be accessed at https://doi.org/10.15131/shef.data.17081138.v1 and are available on a CC-BY-SA license so that they can be repurposed for the classroom or specific organisational contexts.

Ethics Scenarios:
https://doi.org/10.15131/shef.data.17081138.v1

Ethical aspects of AI

Our understanding of the ethics of AI is premised on our definition of AI itself. However, defining AI is challenging because it is a complex and evolving idea. In essence, AI is happening where computers do some of the thinking and decision making normally done by humans. But achieving this has been an aspiration for computer science since the 1950s. Our understanding of intelligence has evolved over the period since then. The technologies that seem to offer some form of intelligence have also changed. Even when we focus just on today's technologies, they are not necessarily simple to understand or define. For example, McKinsey's (2018) state that "AI capability" consists of: robotic process automation; computer vision; machine learning; natural language text understanding; virtual agents or conversational interfaces; physical robotics; natural language speech understanding; natural language generation; and autonomous vehicles. Each of these technologies creates its own potential ethical issues. A further complexity is that because the operation of many of these technologies is premised on data the ethics of AI encompasses many of the previous debates around "big data" ethics.

AI is much more than just a bundle of technologies, however powerful. Unlike many technologies, AI does have a strong place in the public consciousness through its popularisation in such forms as science fiction. AI is often presented as a dystopian rather than as a utopian possibility. Responses to AI cannot be disconnected to responses shaped by such imaginary visions. More immediately AI is also associated with a powerful discourse that has commercial value. Our view of AI is shaped by companies that seek to promote their products as having transformational capabilities for organisations. Some of this may be hype for functionality that is frankly not very new. It could also be seen as linked to culturally powerful notions of technological determinism (Mirza and Seale, 2017). These ideas present technological change as inevitable and able to unproblematically solve complex social problems. Rather than accept this idea of technology driven change as given, we should be thinking in ethically informed ways about our choices as individuals, communities and societies. We should be asking questions about whether the technology fixes the problems we need to fix in ways that consider human needs. In resistance to AI as a powerful marketing discourse are dystopian perspectives on AI. These are linked to issues around the power of high-Tech companies, the profound social implications of dataification and dataveillance, and sustainability issues around power demands of AI. These also point to the wider challenges that exist around AI.

At this level it is important to acknowledge ethical issues around how AI is developed and governed and its ultimate value, rather than purely about its technical aspects (Greene et al, 2019). Critics suggest that currently ethical implications are a secondary thought in the development process. Often, big tech companies' statements about ethics are no more than ethics washing. There is a need to embed ethical thinking more deeply into the whole process of developing and using AI. Many AI applications have a widespread impact on society. There is a need for all stakeholders to participate or be represented in the development process, rather than it be seen as an expert decision-making process purely during design. Perhaps legal regulation is the best way for society to protect itself, rather than relying on companies to behave ethically. Ultimately, a question to ask about AI is whether an application is not just ethical, with a neutral social impact, but whether it actively promotes social justice. Such reflections point to the difficulty of easily defining AI ethics in any context.

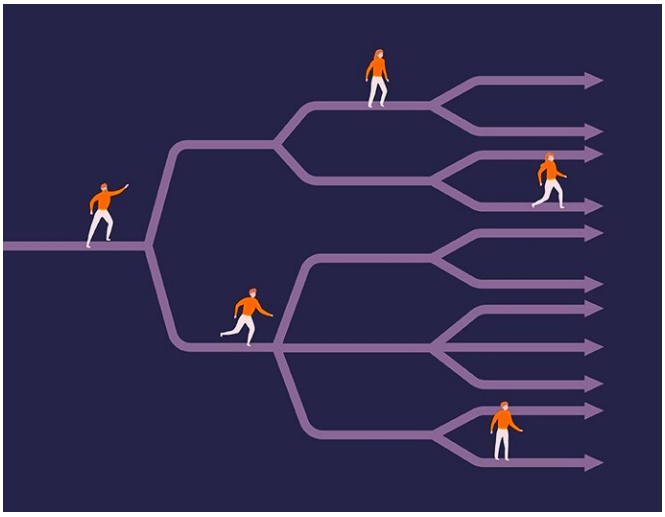
More directly much of the debate around AI ethics has focused on the following inter-related issues: bias; transparency, explainability and accountability; privacy; safety and security; and impacts on human choice and freedom (Jobin, 2019; Fjeld et al., 2020). The news has reflected a large number of cases where AI has been shown to be biased, such as when facial recognition fails to recognise non-white skin tones. To a large extent these problems appear to derive from biases in the historic data that has been used to train an algorithm. In other cases, bias seems to arise from the way that the AI industry workforce is predominately made up of young white males, so that their narrow assumptions are reflected in the AI they produce.

Another particular area of concern is around transparency, explainabilty and accountability. If AI tools learn to make decisions from data, rather than in ways determined by a human coder, then there is an immediate problem of how to explain the process and its outcomes. The workings of AI is not necessarily clear even to the designer of the algorithm, so explaining it is problematic. If the decision-making performed by AI is not transparent, there becomes an issue of accountability. Who is responsible when the AI makes a mistake? Indeed, how can truly informed consent be gained? Privacy is clearly another important issue with AI applications. Personalisation and adaptivity rely on connecting together personal data from many sources. Holding such personal data poses safety and security risks. AI also raises fundamental issues around human agency. What is the human role when computers are making decisions? The concept of 'nudging' is one example of the way that AI can be seen as a threat to human agency. An important dimension of automation is also its impact on work, including professional work. If it is more efficient AI may replace roles. AI could also be used to control human work. Equally it could leave scope for professionals to focus on the more rewarding and complex aspects of their roles. The impact on jobs is an important dimension of AI ethics.

In addition to these general ethical challenges, we should also consider issues related directly to information professional values. IFLA (2019) offers an insightful analysis of some of the main freedom of information and expression issues posed by AI. These are issues given somewhat less emphasis in the wider debate about the ethics of AI. They illustrate one way that the library and information profession contributes a distinctive viewpoint on the AI ethics debate. Personalisation and adaptivity to individual need are often central to accounts of what AI can offer. They can also create filter bubble effects. This effectively limits free access to information. AI's use in forum moderation could also limit freedom of speech, particularly when it is designed to cautiously block any material that might be any way controversial (Privacy International, 2018). If AI's use can block access to information in unnoticed ways, its use in creating deepfakes also has far reaching potential impacts on trust in information. IFLA (2019) also points to the way that AI's use of data can undermine privacy. This can create fear of surveillance with a chilling effect on what people search for and read. Thus, the growing use of AI can be seen as reinforcing risks of surveillance and dataveillance (Privacy International, 2018).

Reflecting on this discussion it can be summarised that the following aspects need to be reflected in AI ethics scenarios for the profession:

1) A sense of balance between the potential benefits of AI with its ethical challenges.
2) The range of AI applications relevant to information professionals, from direct uses in information services to its use in service management and the case where information professionals might be supporting organisations to deploy AI.
3) The range of different contexts of information work, e.g. health, libraries, commercial and legal sector etc.
4)The issues relating to data reflecting AI's reliance on "big data".
5) Reference to the gamut of ethical dilemmas that the wider literature emphasizes, especially bias; transparency, explainability and accountability; privacy; safety and security; and impacts on human choice and freedom.
6) Mention of the issues raised by the impact on jobs and professional roles.
7) Emphasis on the ethical concerns which are of particular importance to information professionals, such as around access to information and freedom of expression.
8) Reference to wider issues around participation by all stakeholders in the development and use of AI and its impact on social justice and sustainability.



AI Ethical Choices and Ethics Scenarios: https://doi.org/10.15131/shef.data.17081138.v1

Scenarios

Scenarios are widely used to encapsulate ethics dilemmas. They encapsulate the issues in a way that promotes discussion. They are accessible stories that seek to stimulate open-ended debate about ethical choices. In designing the first iteration of the scenarios developed in this paper the principles laid out by Institute of Business Ethics good practice guide (Bradshaw, 2012) were followed. These include the recommendation to set the scenarios in relatable professional contexts, yet avoiding excessive detail. They should be open-ended rather than implying an obvious moral.

Having produced an initial set of scenarios, a number of library and information experts were consulted who suggested various further elaborations particularly to the notes supporting discussion with each scenario. The result of this process was eight scenarios. They can be accessed at DOI https://doi.org/10.15131/shef.data.17081138.v1 Each consists of a short description of a situation. The reader is prompted to weigh up what the ethics issues are and try and decide what action they would take. Accompanying notes with each scenario unpack some of the issues raised without implying neat solutions. A short bibliography identifies useful references.

A short description of the final scenarios is offered to give the reader a sense of what the scenarios contain.

1. "**Supporting First Responders**" – set in a health context the scenario sees data managers voice objections to sharing personal data for a system to support first responders to improve interventions in an emergency situation. The scenario raises issues around areas such as consent, privacy, security and transparency. It also emphasises the dilemma where there are life saving benefits set against levels of risk around issues such as privacy and questions of legality: how does this interplay with ethics?

2. "**Nudges**" – set in a university context where the library is asked to contribute data to be processed by a tool that nudges students to change their behaviour to improve their well-being. This again raises issues of both about privacy and consent. The scenario poses issues around human agency, in the context of influencing human behaviour, even if it is for proven benefit to the user. Most fundamentally there is the question of where such an app fits in the wider strategy to support student mental health and well-being.

3. "**The Voice Assistant**" – A public library offers a voice assistant service to answer questions but meets a number of objections, including the risk directly to staff jobs or changing professional roles and potential loss of human contact for users. The scenario also raises a number of issues around bias and stereotyping (through the gendered naming of chatbots).

4. "**A Special Collection**" – A donation is predicated on enabling access to controversial content. This prompts an exploration of issues around bias and representation in collections.

5. "**Forum Moderation**" – imagines automation of moderation of a public forum creating issues around freedom of expression.

6. "**The Recommender System**" – is based around responses to an imagined recommendation tool, including some relating to the chilling effects. The recommender system generates a sense of surveillance, as well as lack of transparency, privacy and bias.

7. "**Stakeholders**" – explores issues of representation through involvement of stakeholder communities in an AI project. Like scenario 8 this is not about a specific AI, more about governance and stakeholder involvement.

8. "**Project Partners**" reflects concerns about power and ethics in a joint AI project. The scenario reflects the likelihood of library and information professionals being involved in wider organisational projects about AI and the challenges this raises.

The scenarios have been released on a CC-BY-SA licence so that library and information professionals can adapt them to their local needs, such as by adjusting the sector setting and to their own organisational context.

Conclusion

There is growing interest in AI both directly applied in library and information professional work and in the wider organisations in which they are embedded. There is also an urgent need for debate about the ethics of the technologies in the profession. Ethics scenarios provide a way to instantiate the dilemmas in relatable open-ended ways. They promote discussion and also increase understanding of AI. Seeking to encompass the range of issues eight scenarios were developed. By making these scenarios available it is hoped they can be adapted for local use in organisations and by educators.

References

AINow (2018). *2018 report*. AI Now institute https://ainowinstitute.org/AI_Now_2018_Report.pdf

AINow (2019). *2019 report*. AI Now institute https://ainowinstitute.org/AI_Now_2019_Report.pdf

Bradshaw, K. (2012). *Developing and using business ethics scenarios*. Institute of Business Ethics.

Buchanan, E. A., & Henderson, K. A. (2009). *Case studies in library and information science ethics*. McFarland.

Cordell, R. (2020). *Machine learning and libraries: a report on the state of the field*. Library of Congress. https://apo.org.au/node/307049

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

Ferguson, S., Thornley, C., & Gibb, F. (2015). How do libraries manage the ethical and privacy issues of RFID implementation? A qualitative investigation into the decision-making processes of ten libraries. *Journal of librarianship and information science*, 47(2), 117-130.

Ferguson, S., Thornley, C., & Gibb, F. (2016). Beyond codes of ethics: how library and information professionals navigate ethical dilemmas in a complex and dynamic information environment. *International Journal of Information Management*, 36(4), 543-556.

Fjeld, J., Achten, N., Hilligoss, H., Nagy, A., & Srikumar, M. (2020). *Principled artificial intelligence: Mapping consensus in ethical and rights-based approaches to principles for AI.* Berkman Klein Center Research Publication, (2020-1).

Greene, D., Hoffmann, A. L., & Stark, L. (2019, January). Better, nicer, clearer, fairer: A critical assessment of the movement for ethical artificial intelligence and machine learning. *In Proceedings of the 52nd Hawaii international conference on system sciences*.

IFLA (2019). *IFLA statement on libraries and artificial intelligence.* IFLA. https://www.ifla.org/wp-content/uploads/2019/05/assets/faife/ifla_statement_on_libraries_and_artificial_intelligence.pdf

The institute for ethical AI in education (2020).*The ethical framework for AI in education*. https://www.buckingham.ac.uk/research-the-institute-for-ethical-ai-in-education/

JISC (2021) A pathway towards responsible, ethical AI https://repository.jisc.ac.uk/8548/1/a-pathway-towards-responsible-ethical-ai.pdf

Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence*, 1(9), 389-399.

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

McMenemy, D., Poulter, A., & Burton, P. (2014). *A handbook of ethical practice: a practical guide to dealing with ethical issues in information and library work*. Chandos.

Milano, S., Taddeo, M., & Floridi, L. (2020). Recommender systems and their ethical challenges. *AI & Society*, 35(4), 957-967.

Mirza, R. and Seale, M. (2017). "Who Killed the World? White Masculinity and the Technocratic Library of the Future." In *Topographies of Whiteness: Mapping Whiteness in Library and Information Science*, edited by Gina Schlesselman-Tarango, 171–97. Sacramento, CA: Library Juice Press.

Padilla, T. (2019). *Responsible Operations: Data Science, Machine Learning, and AI in Libraries.* OCLC Research Position Paper. OCLC Online Computer Library Center, Inc. https://www.oclc.org/research/publications/2019/oclcresearch-responsible-operations-data-science-machine-learning-ai.html

Privacy International (2018). *Privacy and freedom of expression in the age of artificial intelligence.* https://privacyinternational.org/report/1752/privacy-and-freedom-expression-age-artificial-intelligence

Rösch, H. (nd) *Edili: Ethical dilemmas for librarians and other information workers: case studies* https://www.iws.th-koeln.de/edili/edili-about.htm

Thornley, C., Ferguson, S., Weckert, J., & Gibb, F. (2011). Do RFIDs (radio frequency identifier devices) provide new ethical dilemmas for librarians and information professionals?. *International Journal of Information Management*, 31(6), 546-555.

New Horizons for Artificial Intelligence in Libraries

IFLA WLIC 2022 Satellite Conference on Artificial Intelligence, 21-22 July 2022, Galway, Ireland

Hosted by National University of Ireland. See Satellite website at https://www.nuigalway.ie/ifla for more information, registration, accommodations and activities.

The IFLA Information Technology (IT) Section in collaboration with the National University of Ireland Galway is delighted to present a satellite conference in Galway, Ireland on Artificial Intelligence in libraries and information services. There is increasing use of AI technologies in many aspects of libraries, including research and reference services, classification, customer service through chatbots and discovery systems.

IFLA also recognises the importance of AI and an AI IT SIG has been newly formed in 2022 by the IFLA IT Section. This first historic in-person IFLA AI-centered Satellite conference will provide an opportunity to explore and discuss AI in libraries, including practical implementation, upcoming opportunities as well as ethics, transparency, unconscious bias and important groundbreaking connections between AI technologies and libraries in the 21st century from a wide variety of angles.

Program Day 1 – Thursday 21st July Presentations		
NUI Welcome and conference logistics – Emma Goode, NUI Galway		08:45
Opening remarks – Dr Edmund Balnaves (Chair, IFLA IT Section)		08:50
Introduction: John Cox, National University of Ireland, Galway		09:00
Keynote Speaker: Dr. Paul Buitelaar, Director, Data Science Institute, NUI		09:20
Stream 1: AI in	**Stream 2: AI Implementation**	
Emmanuelle Bermès,Neil Fitzgerald: AI4LAM	Dr Ray Uzwyshyn: Steps Towards Building Library AI Infrastructures: Research Data Repositories, Scholarly Research Ecosystems and AI Scaffolding	10:30
Tea break		11:15
Dr. Mojca Rupar Korošec Artificial intelligence is already in libraries, let's master it	Dr Edmund Balnaves: Libraries and AI - practical examples of AI in use and design considerations	11:30
	Jean-Philippe Moreau: AI and image analysis at the BnF: experience feedback	12:15

Lunch		13:00
Dr Lynn Kleinveldt: The use of AI to assist IL integration across subjects and levels to support students acquiring 21st century skills for the world of work	Itai Veltzman: The Ex Libris journey for Artificial Intelligence in Libraries	14:00
Panel discussion		14:45
Tea break		15:15

	Workshop	
Ethics and AI Dr. Andrew Cox, University of Sheffield		15:30
Walk/Taxi to Reception		17:30
18:00 Reception & Dinner (welcome from Ex Libris)		

Program Day 2 – Friday 22nd July

Presentations

Library tour		08:30
Welcome, May Chang, USA		09:10
Keynote Speaker: Dr. Andrew Cox		09:20
Stream 1: Ethics and AI in context	**Stream 2: Machine Learning**	
Josette Riep; Dr Annu Prabhakar: Toward Bias Conscious Artificial Intelligence for Student Success in Higher Education	Thomas Zaragoza; Aline Le Provost; Yann Nicolas: From text to data inside bibliographic records. Entity recognition and entity linking of contributors and their roles from statements of responsibility	10:30
Tea Break		11:15
Karolina Andersdotter: The AI and Libraries Study Circle: how 100 library professionals increased their AI literacy	Anna Kasprzik: Get everybody on board and get going – the automatization of subject indexing at ZBW	11:30
Galvia AI – Chatbot design	Martin Malmsten: Without heading? - fully automated linked subject systems creation	12:15
Lunch		13:00
	Florian Engel: Automatic indexing using AI methods – a project insight at the German National Library	14:00
Panel discussion		14:45
Tea break		15:15

	Workshops	
Building Open source Chatbots – Dr Edmund Balnaves and Imam Khamis		15:30
Closing remarks – Dr Edmund Balnaves & Emma Goode		16:30

IFLA IT 2-Day Library Carpentries Workshop @ Waterford Institute of Technology, Ireland

Waterford Institute of Technology
Telecomm ICT Research Centre

On July 22 and 23, 2022, a 2-day pre-conference "Library Carpentry Workshop" will take place at the Waterford Institute of Technology in Waterford (WIT), Ireland, 2 hours outside of Dublin. The WIT will be merging with the Institute of Technology Carlow (IT Carlow) to form a new technological university on May 1st, 2022. Organized in the open format of a hands-on workshop, the event offers a unique software skills training for library professionals.

Library Carpentry is a growing international community whose mission is to teach librarians the tools, techniques and best practices around working with data and using software to automate repetitive tasks. This goal is critical to building skills capacity in our profession and supports the IFLA Strategy. The goal of this workshop is to provide participants with basic knowledge on programming languages, data structures and data management. The focus is on basics of the Unix shell and OpenRefine, how to use Git and GitHub, and the Python programming language. The workshop is given by local The Carpentries Instructors and helpers with a background in librarianship and information science to help participants. There is a 60 euros fee for the workshop which includes coffee/tea break and lunch for both days.

Please mark the date on your calendar and review the official satellite website for registration and payment information: https://lirgroup.heanet.ie/index.php/2022/03/09/ifla-library-carpentry-workshop-july-22-and-23/

The event is organized by the Science and Technology Library (STL) Section in close cooperation with Higher Education Network Library Information Resources Group (HEANet LIR Group), Continuing Professional Development and Workplace Learning (CPDWL) Section, Information Technology (IT) Section and Big Data Special Interest Group.

IFLA IT Section

The Information Technology (IT) Section promotes and advances the application of information and computing technologies to library and information services in all societies, through activities related to best practices and standards, education and training, research, and the marketplace. The scope covers IT for creation, organization, storage, maintenance, access, retrieval, and transfer of information and documents for all types of libraries and information centers; IT for the operation of libraries and information centers; and related management and policy issues. Of primary importance are applicationsof IT for supporting access to and delivery of information. In recent years, the uses of use of technology in libraries have expanded to cover improved machine learning and AI techniques, digital humanities, and data analytics.

The section meets annually at the IFLA Congress; in between congresses, members collaborate with other Sections on programs andworkshops. There are election ballots every two years as members complete their 4-year term. The IT Section is one of the largest in IFLA with over 300 members from nearly 80 countries, all types of libraries, and a range of disciplines. We welcome all members (http://www.ifla.org/membership).

The IT Section's website at http://www.ifla.org/it has news andresources regarding activities of the Section, session minutes, publications, and membership details.

The IFLA-IT email list provides a forum for members to exchange ideasand experience in the use of information and communication technologies in libraries. The list address is ifla-it@iflalists.org, and subscription is at https://mail.iflalists.org/wws/info/ifla-it.

The Trends & Issues in Library Technology (TILT) newsletter is published twice a year in June/July and January.

Primary Contacts

Chair
Edmund Balnaves, Prosentient Systems, Australia; ejb@prosentient.com.au

Section coordinator
Elena Sánchez Nogales, Biblioteca Nacional de España, Spain ; elena.sanchez@bne.es

Secretary
Cory Lampert, University of Nevada Las Vegas, United States ; cory.lampert@unlv.edu

Information Coordinator
François-Xavier Boffy, Université de Lyon, France; francois-xavier.boffy@univ-lyon1.fr

Standing Committee Members

- Anna Tereza Barbosa da Silva, Forsvarets høgskole,Norway; adasilva@mil.no
- Sylvain Bélanger, Library and Archives Canada; sylvain.belanger@canada.ca
- Leda Bultrini, ARPA Lazio, Italy; leda.bultrini@gmail.com
- Almudena Caballos Villar, Biblioteca de la UniversidadComputense de Madrid, Spain ; acaballo@ucm.es
- Maria Kadesjö, National Library of Sweden; maria.kadesjo@kb.se
- Alenka Kavčič Čolić, National and University Library, Slovenia; alenka.kavcic@nuk.uni-lj.si
- Wouter Klapwijk, Stellenbosch University, South Africa; wklap@sun.ac.za
- Lynn Kleinveldt, Cape Peninsula University of Technology, South Africa; lynn.kleinveldt@gmail.com
- Yeon-Soo Lee, National Library of Korea; myth4ys@korea.kr
- Peter Leinen, Deutsche Nationalbibliothek, Germany;p.leinen@dnb.de
- María Loretto Puga, Biblioteca del Congreso Nacionalde Chile, Chile. mpuga@bcn.cl
- Sogoba Souleymane, University of Ségou, Mali;Sogoba.souleymane@gmail.com
- Ngozi Blessing Ukachi, University of Lagos, Nigeria;nukachi@unilag.edu.ng
- Ray Uzwyshyn, Texas State University Libraries, United States; ruzwyshyn@txstate.edu
- Qiang Xie, National Library of China; XIEQ@NLC.CN
- Katherine Zwaard, Library of Congress, United States;kzwa@loc.gov

Big Data SIG

Convenor
- Cory Lampert, University of Nevada Las Vegas, UnitedStates. cory.lampert@unlv.edu

Information Coordinator
- Patrick Cher, Singapore National Library. patrickcher@gmail.com