TRENDS & ISSUES IN LIBRARY TECHNOLOGY

IFLA IT Section Newsletter

IFLA

In this issue:

Greetings from your new Editor!

Dear Colleagues,

This issue of the TILT newsletter is genuinely a team effort done with the help of all members of the IT communication team and Section officers.

I also want to thank the authors, and hope you enjoy reading the articles presented.

I look forward to your comments as well as contributions of articles to subsequent issues of the newsletter.

Alireza Tayarani, Editor
tayerani@gmail.com

IT Section 2020-21 Action Plan Spring update

The Action Plan was updated after the IT Section Standing Committee (SC)'s mid-term meeting on April 12, 2021. The agreed to continue with the series of IT related webinars, support the publication on "Robots in Libraries", proposed development of training manuals on digital skills for library staff, and research data guidelines for small academic institutions.

Several proposed activities were put on hold due to the ongoing pandemic situation or waiting for responses from joint activity partners including IT topics checklists, cybersecurity guidelines, and open interoperability standards.

Our proposal for a WLIC 2021 session on "Open Educational Resources: Where ICT meets education and scholarship" has been accepted by the IFLA committee.

The IT Section Action Plan and all our publications are available at: https://www.ifla.org/publications/62

Compassionate Computing: Reframing Technology Work for Cultural Change and Optimal Work Performance

Laurie N. Taylor
Senior Director for Library Technology and Digital Strategies University of Florida, US
laurien@ufl.edu

Todd Digby
Chair of Library Technology Services University of Florida, US
digby@ufl.edu

Introduction

In January 2020, the Libraries at the University of Florida created a new division, Library Technology and Digital Strategies, which brought together the Library Technology Services and Digital Partnerships & Strategies Departments, in recognition of the importance and pervasiveness of technology. The two departments have interconnected work, and the Libraries made this organizational change to enable increased effectiveness through changed ways of working. The new group was created to deeply connect technologies and socio-technical practices that embrace maintenance, minimal computing, ethics of care, mutual aid, collaborative practices, generous thinking, shine theory, and related practices.

Over the past year, we have utilized socio-technical philosophies and methodologies to enact cultural change across the Libraries as a whole through the frame of what we are calling "compassionate computing." In this article, we share about the prior context and what we mean by compassionate computing for enabling excellence in technological work and collaboration.

Coming Together as Library Workers

Before 2020, we would often hear "we're not librarians" or "you're not librarians" stated by technology support staff during discussions on collaborative work on technologies. These statements reflected the gap between our technical and subject experts where are technical staff saw themselves as primarily in support roles rather than taking ownership of projects or being part of project teams. We knew we needed to bridge this gap in order to have a better culture at work and to see optimal success for work deliverables.

When the Libraries at UF created Library Technology and Digital Strategies, the new Senior Director for the Division and the Chair of Library Technology Services immediately collaborated on strategies to foster cultural change. Our initial work included meeting with all department chairs and unit heads, to ask them to hold a conversation with us on goals and dreams, as they relate to IT, and any concerns. We let them know that we were working to ensure existing concerns can be addressed as quickly as possible and we then wanted focus on how IT can enable, enhance, and partner on goals and dreams across the Libraries. We asked them to lend us an hour of their time so that we could hear their thoughts and to develop this together.

In parallel to the dozens of hours of conversation with collaborators outside of the division, we held conversations with those within, asking on their concerns, goals, and dreams. Everyone had concerns, many of which we were able to quickly address and correct. The concerns people shared spoke to the need for changes, which we implemented, including stronger project management and full project portfolio management to ensure projects were done right and the right projects were done. Our questions on goals and dreams had fewer responses, and often people were not sure what even might be possible to imagine. We worked through these conversations to tease out different threads. In doing so, we learned that our core needs included changing our framing for how we talk and think about technologies.

In the past, work by Library Technology Services had been framed as "break/fix" or "putting out fires" in both the areas of supporting technology hardware and in the area of software or digital system development. This frame focused on urgent and tangible over the necessary longer-term and important. Implementing project portfolio management addressed much of the need, but we knew from our conversations that more was needed. From the conversations, we knew that the existing framing impeded technology and subject matter experts from coming together as partners and collaborators for needed work, and from being able to dream together for our best culture and best ways of working.

Compassionate Computing

We started to use the term compassionate computing to bring together known best practices from collaborative partnerships on technologies. These best practices include recognizing:

- All technologies are socio-technical, and thus include policies, practices, people, and communities who are using the technologies.
- For optimal work, we must consider how technologies will be used and maintained.
- Confronting complexity requires compassion and grace: we need established rules, practices, and patterns to ensure consistent and equitable operations.
- By focusing on technical optimization and compassion, we can best support everyone, including technical workers.

- By structuring our work as compassionate computing, we can best support current needs and future planning.

We used compassionate computing to reframe and reorient all aspects of our work. In doing so, we were able to see tremendously positive work products and culture change.

For examples, we spoke about compassionate computing as a way to bring folks closer together for a huge host of needs:
- First ever web migration (we had many web systems in use, but had never done a full web migration and information architecture, so this was a huge lift for technical and cultural change
- Needed migration of our digital collections system
- Need implementation of archival finding aids system
- Dealing with chicken and egg problems: how do we implement needed methods, practices, and patterns for stabilizing our work, as with project portfolio management, while we are awash in urgent and important work

Many more needs of various types: rolling out new services (e.g, UFLib Domains), process and system changes (OneDrive over SAN personal file spaces, changing how we handle accounts and updates for LibApps, AD cleanup), creating time and support for upskilling of technology workers, and more.

For this host of needs, we are achieving at a higher than expected level on all of them. This is because we are using compassionate computing to bring our groups together across the Libraries in a productive and generous manner. Critically, we have had higher than expected success on all areas even though all of these have taken place during the pandemic, with a host of new needs.

Thankfully, compassionate computing provided a frame that includes responsiveness overall, including enabling resilient and resistant operations. When the pandemic required major and urgent changes, we used the concept of compassionate computing to focus our community and to rapidly pivot and develop new methods.

One simple example comes from the pandemic and remote work. Prior to the pandemic, UF did not allow people to take equipment home. Then, we had a workforce who needed to work from home. We had no methods, and suddenly had to do it at scale. Following compassionate computing, we knew we had to fulfill our mission of supporting workers. That meant we had to support people in taking equipment home. We had to make it easy for them, and easy for us to comply with requirements for tracking equipment. Within a day, we changed how we process equipment requests, and sent people home with equipment. We expanded our internal helpdesk hours and began home computing support. In the past, UF would not buy equipment for home use, if covered by the office, with this otherwise being redundant. When workers took their workstations home, this included a five-foot network cable.

While a short cable is ideal in the office, we knew that moving workstations home would have problems. A short network cable means that our colleagues (and their kids and dogs) would be tripping over the cord, with dining rooms being new home offices. We knew we needed options, like longer cables and wifi USB cards. This previously "redundant" equipment was now essential for their work. We bought cards and cables as quickly as possible to meet needs.

With so many people buying these, along with headsets and webcams, and we could not source the items fast enough to meet needs, we collaborated with the Fiscal unit to support reimbursements when workers ordered their own, which often meant dramatically faster delivery.

Under a frame of break/fix and putting out fires, the shift to remote work would have been painful. Under the frame of compassionate computing, we had a collective mission and method, and that enabled us to quickly develop smooth and successful processes for a multitude of dramatic changes.

Empathy for Our Users, Our Systems, and Ourselves

Part of building out the frame for how we do compassionate computing included work to support those within the division. The Libraries at UF have conducted the ClimateQual: Organizational Climate and Diversity Assessment, which is an assessment of library staff perceptions concerning (a) their library's commitment to the principles of diversity, (b) organizational policies and procedures, and (c) staff attitudes. The Libraries conducted this survey in 2014 and 2019. The 2014 survey placed the Library Technology Services Department well below the other departments and units across the Libraries in many areas. Although the department saw significant increases in the 2019 results, there are still improvements to be made.

One area that we focused upon was to improve departmental communications and information sharing, which resulted in a series of strategic planning retreat meetings held in a remote fashion. We held the strategic planning sessions in summer 2020 where we covered several readings and improvements for work for project portfolio management. We also developed a statement of our shared values, which include:
- We work from a place of empathy for our users, our systems, and ourselves.
- We strive to do our best as proactive, aspirational, and collaborative partners.
- We are honest and accurate in estimating our time and impact, accepting the limits of what we will know in any given situation.
- We understand our work includes handling problems for things beyond our control.

(https://ufdc.ufl.edu/AA00078428/00004)

Stating values is necessary for collaborative work so that we can begin from places of understanding. In working on these values, one of our team members asked: will others see this; will they know that this is what we believe and want?

The question came up because the team wanted others to understand and agree with our values, which took place as we shared our values in the Joint Chairs library leadership meetings and in a library-wide facilitated discussion series on all of our work as compassionate computing. In working on cultural change across the Libraries more broadly, creating our values internally and sharing them broadly was critical work to ensure that our internal culture changed.

To further support cultural change, we also discussed our work in terms of the Triangle of Satisfaction, which depicts the three types of needs to achieve satisfaction: emotion, outcome, process. We asked our team, in our work as technologists, how often do we get to have folks be fully satisfied?

We have philosophies and methods like agile and scrum, project portfolio management, acceptance criteria, and more, to support the process, result, and emotion. Even with those supports, how often are our stakeholders satisfied? How about our technology workers? We hope that everyone has high levels of justice and high levels of satisfaction in their workplaces. At UF, we have made great strides; however, we had a major problem for our technologists because of communication. We needed better communication across groups who had not been productively connected in positive ways. To make the change, we knew we needed to build empathy and compassion, to be able to support problems as they arise, and to deal with them collaboratively, together, as equals.

The strategic planning retreat sessions helped foster an understanding of the various aspects of the work across the department and, also, led to actionable changes in how we document and communicate our work, both within the department and across the Libraries. This led to an increased sense of ownership and pride in the work that is done.

Recently, we also engaged in taking the Everything DiSC assessment to build a better understanding about what our personal workplace tendencies are and how others may approach activities differently. This was a key support in our abilities to understand those that we support and partner with, so that we can keep building on our compassionate computing approach.

Conclusion

We continue to see important cultural changes and tangible benefits from compassionate computing. In Fall 2020, the Library Technology Services Department won one of the annual Employee Excellence Awards, which is a peer-driven award program, and the award announcement focused on the array of work by the department and the frame of compassionate computing. Importantly, we have not heard the statements of "not librarians" and have instead heard questions and conversations that are orient using our stated values, DiSC types, and compassionate computing.

Good News

The IFLA IT Section has gone social!

Please follow our Twitter account (@ifla_it) for Section news and event information, calls for proposals, updates from the Big Data SIG, and other technology news.

Contact the Information Coordinator if you would like to promote an item via this account.

We will see you online!

Follow @ifla_it or browse tweets at: https://twitter.com/ifla_it.

Libtech in Iran: Open Innovation Platform in the National Library and Archives of Iran (NLAI)

Babak Zendehdel Nobari
Director of Planning and development
National Library and Archives, Islamic Republic of Iran (NLAI)
b-zendehdel@nlai.ir

Mozaffar Pasdar Shirazi
Deputy Director of Management and Resources Development and Head of Digital Affairs
National Library and Archives, Islamic Republic of Iran (NLAI)

Abstract

NLAI, as the Iranian national memory, is the largest Persian language database in today's world for the researchers in the area of businesses and data-driven technologies, the NLAI decided to play the role as the facilitator for the professionals of the field by a platform strategy. To this aim, using open innovation and digital transformation approach, not only released its valuable data in standard formats, but it also launched the Innovation and Data Driven Business Development Center to meet the requirements of Artificial Intelligence (AI) and Big Data Analysis in Persian in the world. This article tries to tell the story of innovation in the NLAI.

What is Open Innovation?

Open innovation, is the participation of different stakeholders of an organization (such as customers, competitors, suppliers, etc.) in the process of organizational innovation planning(H. W. Chesbrough, 2006). In fact, open innovation platform means creating an interactive and multifaceted network of different stakeholders of an ecosystem to enable the exchange of ideas in this network and cause synergy for the organization and all stakeholders. Open innovation was initially used more in industry but over time was also applied in the public and government sectors(H. Chesbrough & Crowther, 2006). Stakeholders' participation and supportive ecosystem are two vital dimensions in open innovation platforms implementation. (Henkel, Ilhan, Mainka, & Stock, 2018).

Open Innovation in Libraries and Archives

In recent years, the use of open innovation in libraries and archives has expanded. Maria Henkel et al. stated several examples of open innovation in libraries in their conference paper(Henkel et al., 2018). National Archives and Records Administration (NARA) developed a history hub platform as one of its strategies to achieve open government plan (2016-2018)'s targets to leverage the open innovation approach in order to enrich the collections.

Libtech Innovation Story:

In today's world, due to the unique role of national archives and libraries in the world, which are responsible for acquiring, processing, preserving, and providing access to the world's documentary heritage in the countries, traditional approaches alone cannot meet the growing needs of audiences.

Image1-The first Libtech event (VITA) organizers in the NLAI

The NLAI, has more than 15 million total items, and more than 250,000 registered members. In recent years, along with the exponential growth of information resources in Iran and the development of digital technologies, the need for a paradigm shift in the way of acquiring, processing, preserving, and providing access to resources was felt. For example, the number of books published in Iran has increased more than six times in the last twenty years, and one of the most important missions of the National Library is to update the National Bibliographic Bank, which is used by researchers. The use of automated processing techniques can help in this regard. The number of articles published in Persian has multiplied. Persian language data on social networks has become much more than before. In addition, in the last two decades, with the development of ERP systems in Iran, the growth rate of electronic records production has multiplied. All these cases showed the large volume of Persian language data and the need to use new technologies in the NLAI for overcoming the information explosion.

So, instead of considering the challenge in a discrete way, it was decided to find a systematic and consistent solution to it. Implementation and design of Open innovation cycle was the NLAI's systematic solution. Thus, in 2018, the NLAI planned to implement the open innovation platform for data-driven technologies (in the field of libraries and archives) in Iran called Libtech.

Libtech refers to the integration of emerging technologies such as AI, Blockchain, Deep Learning, Big Data Analysis, etc. into offerings by Library and Archive services to improve use and delivery to the consumer. Indeed, Libtech seeks to create a value-added service which works most in library and archive processes and services.

Its ultimate goal is not only to increase productivity but also ameliorate a user's experience with emerging technologies. Therefore, NLAI determined to launch a Libtech Innovation Center in order to develop new businesses in this field. The main objectives of the Libtech are:

1. Response to a change in need to adopt to new technologies by users in an open innovation collaborative platform;
2. Developing a value and network chain of Libtech ecosystem;
3. Involving users, researchers and entrepreneurs to improve service levels.

As stated in the definition of an open innovation platform, its two main dimensions include stakeholder participation and the support ecosystem. Regarding the support ecosystem, the following action were taken:

1. Establishing the Libtech policy making committee
2. Participation of a professional accelerator with experience in holding entrepreneurial events
3. Participation of experts and professors of AI (University of Tehran), Information Science (NLAI) and Business Development (specialized accelerator)
4. Preparation of software infrastructures (preparation of required sample structured and unstructured data and meta data of NLAI with regards to privacy policy, preparation of dedicated Libtech portal for registration and review of received proposals), hardware infrastructures (preparation of virtual machines and required processing resources on dedicated servers) and networking infrastructures (providing dedicated high-speed Internet for teams) for the first bootcamp (Vita event)
5. Preparing a list of technological challenges of the NLAI in order to submit to teams

The event took place in two stages from July 5th to August 5th, 2019. First, the resumes of the registrants were reviewed and nearly 350 participants were selected for the opening session to learn about the details and challenges of the event in different panels. Next, proposals were submitted by participants for the event over a week. Totally, 92 proposals were received and 42 of which were selected thereafter. In the next two weeks, the workplace and required infrastructures such as co-work space and sample data were provided for the startup groups. Moreover, there were workshops in the areas of AI, business development and NLAI value chain.

Finally, some of teams collaborated with the projects of NLAI. In other cases, the final products were purchased by NLAI and the others were supported by the accelerators for fundraising. The following four teams were selected to meet the expected targets in the stated challenges:

1. A Persian/Arabic Word Spotting system for historical handwritten manuscripts;
2. A Persian captcha for Persian and Arabic manuscripts and lithographs;
3. Abstracting and summarizing Persian articles, and representing extracted concepts in graph network;
4. An intelligent online chat bot for reference desk services.
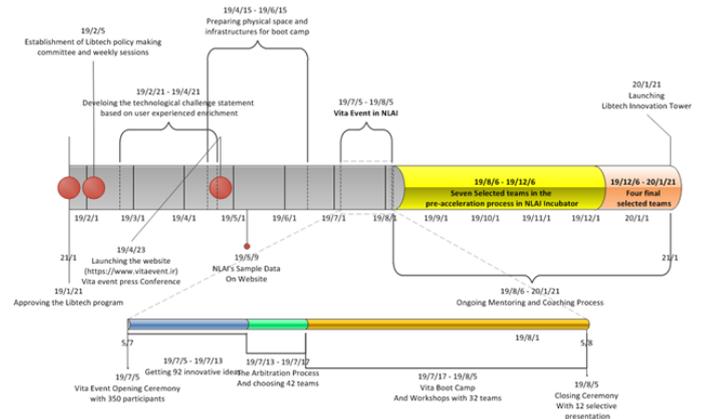
Image2-Timeline of the Libtech platform (VITA reverse pitch)

Image 3- workshops (Business Model Development, AI, Information Science, etc.) were running in parallel during VITA event.

Image 4- Startup teams in VITA event (boot camp)

Challenge statement in VITA reverse pitch

Given that VITA as the first Libtech event cycle (VITA) was to be held in reverse pitch, the challenge statement was prepared in cooperation with the NLAI's experts, and approved by the policy making committee. The enrichment of user experience and experts' opinions were the basis for the statement. The technological and business challenges of NLAI were promulgated at the website and communication channels in social media. The most important challenges were as follows (Table1):

Challenge Category	Challenge sub-category
Automatic Persian Text Processing	Automated Cataloguing System
	Catalogue Assistant
	Automatic Indexing (Keyword Extraction)
	Named Entity Recognition
	Related Texts Detection
	Sentiment Analysis/ Opinion Mining/ Opinion Extraction
	Automatic Text Summarization
	Topic Detection
	Automatic Scientometrics Analysis
Persian Web Data Extraction	Web Archiving
	Bulk Download
Image processing in Persian Resources	Results Modification in Persian OCR
	Handwritten Text Recognition
	Similar Images Detection
	Automated Image Tagging and Indexing
Information Representation	Search Optimization Using Fuzzy Logic
	Recommender Systems in Search Optimization
Digitizing Persian Resources	Automatic Book and Non-book Material Scanning
	Processing, Optimizing and Quality Control of Scanned Resources
Crowdsourcing	Developing a Platform for Information Gathering from People

Table 1- The most important challenges in the first Libtech cycle (VITA reverse pitch)

Table 2 also shows some workshops on Libtech conducted by university professors and NLAI experts. Of course, in addition to these workshops, team and individual counseling sessions were also held for attendees in the fields of business management, AI and information science during the VITA event.

Business Management Workshops	Specialized Technological Workshops
1. Customer Study	1. Big Data Management
2. Competition and Blue Ocean Strategy	2. Deep Learning
3. Solution Design	3. Data Visualization
4. Business Model Development	4. Pattern Recognition
5. Validation	5. Recommendation System
6. Legal issues	6. Social Media Analysis
7. Investor Pitch	7. Named Entity Recognition
	8. Blockchain
	9. Cataloguing Standards

Table 2- List of the workshops in VITA event

Image 5- Presentation by one of the VITA event selected teams at the closing ceremony

Image 7- Libtech Innovation Center cafe and recreation center

Launching Libtech Innovation and Data Driven Business Development Center

Eventually, the Libtech Center was established as NLAI innovation center. One of the old buildings of the NLAI was designated as the Libtech Innovation Center which was ready for operation in 2020. The building with an area 1800 m2 is located in the downtown of Tehran (close to the best large universities of Iran). Libtech innovation tower has six floors and each floor consists of the following:

- Two floors for co-working space for startup teams (Image 5),
- One floor for the Libtech studio,
- One floor for the Libtech Academy,
- One floor for the conference and event halls and
- One floor for the cafe and recreation center (Image 6).
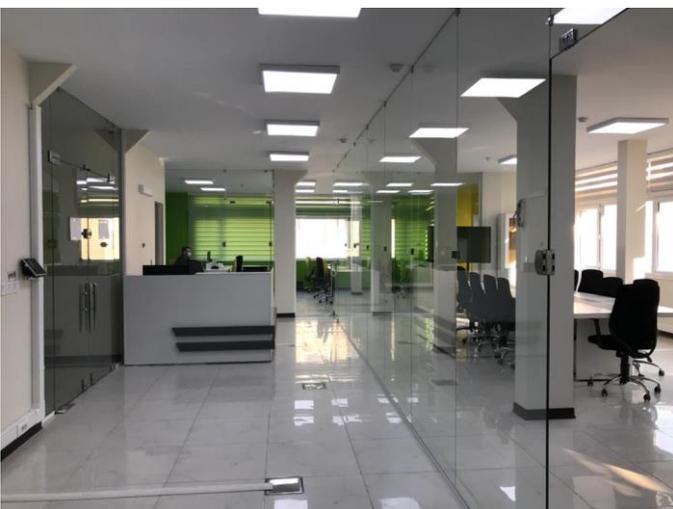
Image 6- Libtech Innovation Center co-working space

The following is one of the achievements of the first cycle of Libtech (in the Vita event), which is currently partnering with the NLAI.

A Persian/Arabic Word Spotting system for historical handwritten manuscripts (SANADJOO*)

A Persian/Arabic word spotting system was one of the most important developed projects in the Libtech platform in NLAI. The product (SANADJOO) was developed a new indexing/searching technology for historical manuscripts, a new paradigm to study Persian/Arabic historical heritage, in manuscripts, by using a full text search technology.

Making large-scale collections of digitized historical documents searchable is being earnestly demanded by many archives and libraries. Probabilistically indexing the text images of these collections by means of keyword spotting techniques is currently seen as perhaps the only feasible approach to meet this demand.
This project is divided in two phases:
1. **Indexing document (offline phase)**: indexing content of scanned document hierarchically: Whole document, page, line and word.
2. **Search keyword in online phase**: matching entrance query with indexed documents and show matched documents.

The recognition-free retrieval which is also known in the literature as **word spotting** or **keyword spotting** is the main subject of the work. The goal here is to retrieve all instances of user queries in a set of document images. Actually, the user formulates a query and the system evaluates its similarity with the stored documents and returns as output a ranked list with results which are most similar to the query.

The process is totally based on matching between common representations of features, such as color, texture, geometric shape or textual features, while conversion of whole documents into a machine readable format and recognition does not take place at all. Therefore, the selection and use of proper features and robust matching techniques are the most important aspects of a word spotting system.

In this case, document image word spotting techniques can be used to search the textual information from the digitized document images and make this information accessible to users. Word spotting is the task of locating specific words in a collection of document images. There are two principal approaches to spot document images. The first category is the traditional text search methods which necessitates efficient optical character recognition (OCR) techniques, that are right now unavailable for most of handwritten documents. These methods are referred to as OCR-based techniques. OCR-based techniques are not suitable choices for old printed and handwritten documents. This is due to the presence of several challenges in handwritten documents including: (1) poor quality documents, (2) writing style variability, (3) multiple writing styles and (4) word writing variations, etc. The second category is to use word spotting techniques to search in the image domain. In the last decades, word spotting has witnessed great interest as an emerging technology for document image retrieval applications and is becoming an eminent technique for this task. It has been introduced as an alternative to OCR-based techniques. The role of word spotting is especially emphasized in the case of historical document images because they are of poor quality and have a large writing style variability. Word spotting finds a specific keyword in document images by comparing features that are extracted from word images. It includes a matching process between word query image and a collection of word images stored in the database using the extracted features. Most of the proposed word spotting based techniques find keywords as follows: firstly, they build indexes based on low level features that are extracted from word images. Same features are also extracted from query images. Secondly, they search the entire database and measure how much the query image is similar to each word image in the database. Finally, the results are sorted based on the similarity matching level. Document image word spotting presents problems worth solving.

Manuscripts are among the most important evidences to our cultural heritage. Despite a large digitization, the wealth of their content remains largely inaccessible: current handwritten text recognition technology is not accurate enough to allow full text search. This situation raises the need of searching approaches specifically designed for text images. The sample of the system is shown in below.
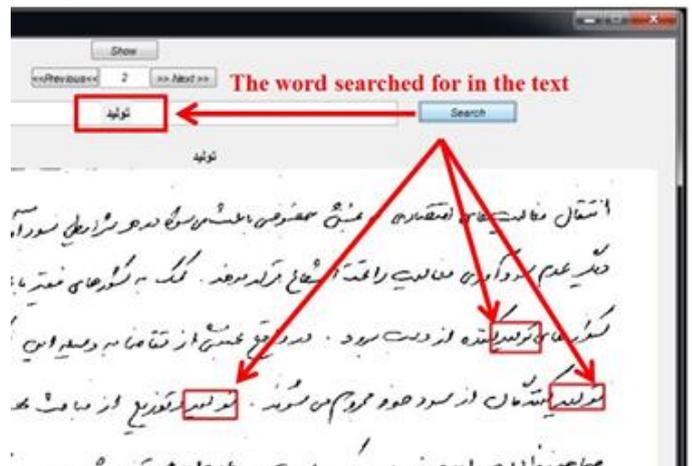
Image 8 – Sample of Word Spotting Project in Persian / Arabic Handwritten Manuscript (SANADJOO)

Conclusion

Nowadays platform thinking approach is critical to successful innovation in libraries and archives. Open innovation implementation is not a project with a beginning and end. It requires a continuous dynamic cycle. National libraries and archives, as policy makers and leaders of library and archival activities in each country, can establish and support an open innovation platform (Libtech) to accelerate and develop data-driven businesses. The NLAI is ready to share its experiences in this field with other countries.

References

Chesbrough, H., & Crowther, A. K. (2006). Beyond high tech: early adopters of open innovation in other industries. *R&d Management, 36*(3), 229-236.

Chesbrough, H. W. (2006). The era of open innovation. *Managing innovation and change, 127*(3), 34-41.

Henkel, M., Ilhan, A., Mainka, A., & Stock, W. (2018). *Open innovation in libraries.* Paper presented at the Proceedings of the 51st Hawaii International Conference on System Sciences.

* The Co-founders of the SANADJOO team who attended the Vita event: 1- Majid Iranpour Mobarakeh, 2- Ali Ghanbari Sorkhi

From MARC to Machine Learning: The Library of Congress and Digital Transformation

Leah Weinryb Grohsgal
Program Advisor to the Director of Digital Strategy
Library of Congress, Washington, D.C.
LC-Labs@loc.gov

Decades of Digital Work

The Librarian of Congress created the Library's Digital Strategy Directorate in 2018, with a charge of advancing the agency's digital transformation and housing the digital innovation lab. Located in the Office of the Chief Information Officer and known as LC Labs, we collaborate across the Library and with external partners on experiments, research, and outreach to open possibilities for the Library's digital future. We build on past and ongoing work using technology to make libraries more accessible and inviting.

In our own corner of the world, the Library of Congress has long dug into digital methods to preserve and provide access to its massive, varied collections. It was Library of Congress computer programmer and systems analyst Henriette Avram who, in the 1960s, developed the MARC format and ushered in an age of automation and sharing in libraries. In the 1990s, the Library embraced the internet, working with other organizations to compile an online resource with millions of freely-available items known as the American Memory Project. And in the aughts, the National Digital Information Infrastructure and Preservation Program (NDIIPP) worked with a network of partners on ways to preserve digital materials—especially those most at risk.

Looking out to work going on across the nation and the world, we are continuing this tradition. Our Digital Strategy, adopted in 2018, aims to achieve the Library's vision that "all Americans are connected to the Library of Congress." Our goals include opening the treasure chest, connecting, and investing in our future.

Throwing Open the Treasure Chest

The Library has been collecting digital materials for over three decades; with millions of treasures, we are exploring how to allow the greatest possible access to this content. Our Innovator in Residence initiative, kicked off in 2017, supports innovative uses of our collections. Last year, two innovators combined technology to spark interest in Library visual and audio content.

Benjamin Lee's Newspaper Navigator used machine learning technologies to harvest images, including photographs, illustrations, cartoons, and maps, from the Chronicling America historic newspaper site. Where the 16 million+ newspaper pages had been made text-searchable by optical character recognition, there was no similar tool to aid researchers looking for visual content.

Lee built on the Library's existing Beyond Words image identification experiment, using volunteer-generated data to train machine learning algorithms and extract an astonishing 100 million images.

Now, those images are available to the public for download. Additionally, a search interface also allows the public to search over 1.5 million photographs and execute real-time machine learning refinements. Newspaper Navigator is the Library's first machine learning project that allows the public to reimagine how they can search visual content using machine learning techniques, empowering users to train a machine learning algorithm by selecting photos that they are interested in.

In addition to newspapers, the Library holds major collections of sound and moving image recordings. Hip hop and other musicians work with audio samples to create music. Seeing this connection, Brian Foo created Citizen DJ. The application allows users to explore free-to-use sounds from Library collections, make beats with samples, and listen to full recordings on loc.gov.

Samples include clips from The Early Motion Pictures and Sound Recordings of the Edison Companies, the Variety Stage Sound Recordings and Motion Pictures, the Joe Smith Collection, the Free Music Archive, American English Dialect Recordings, and the National Screening Room. Foo developed an interface for users to explore free-to-use sounds, remix and combine with beats in a browser, and download sounds for use with other software. Because this work often involves novel use of library collections, Foo also worked with Library staff to create a guide to the legal and ethical issues involved in digitally sampling recorded sounds.

Connect

Connecting with and engaging users is the second component of the Library's Digital Strategy. Since its launch in 2018, the Librarian's signature crowdsourcing initiative, By the People, has offered thousands of pages for transcription and addition to Library collections, connecting with tens of thousands of volunteers across the nation in the process.

One of the aims of By the People is to inspire users to have an ongoing relationship with the Library. By inviting people to transcribe, review, and tag digitized pages from the Library's collections, the program welcomes everyone to take part in surfacing our nation's cultural heritage. Transcripts also improve searchability and access to documents, including people who are not fully sighted.

Volunteers create and review transcriptions, returning them to the Library's website for inclusion with the original digitized documents.

Over the past two-and-a-half years, By the People has hosted 21 campaigns, releasing nearly half a million pages for transcription from presidential papers; personal archives of the women's suffrage, abolition, and civil rights movements; works of writers like Walt Whitman; the musical and tradition documentation produced by folklorist Alan Lomax, and more. Last year, the flourishing program transitioned from experiment to permanent service, showcasing a path for sustaining digital experiments over the long term.

Invest in the Future

The final goal of the Digital Strategy asks that we look to the future, considering the tools and technologies most likely to play a role in 21st century libraries. Machine learning and artificial intelligence present interesting possibilities for helping libraries expand their digital reach. Before applying these technologies to our content, however, the Library needs more information and demonstration of their possibilities and pitfalls. To that end, we initiated a "season of machine learning" to explore the opportunities and challenges of applying this sort of artificial intelligence to library collections.

We collaborated with a major research group, the University of Nebraska-Lincoln's Project AIDA, which provided insights and recommendations about the gaps around machine learning and AI solutions and their applications in cultural heritage settings, the need for ongoing experimentation and demonstration via interdisciplinary research teams and creation of ground truth training data, and the role of establishing values statements and building on successes with crowdsourcing.

An expert state-of-the-field report by Northeastern University Professor Ryan Cordell provided context and guidance for future machine learning applications. The report incorporated an extensive environmental scan, interviews with practitioners, and analysis of risks and possibilities. Finally, LC Labs worked with Library partners to test the possibilities for enhanced description and processing of audiovisual content using commercially available speech-to-text transcription. The experiment tested a commercial speech-to-text service on archival audio content, using digital spoken word collections from the American Folklife Center.

The team found that the use of speech-to-text software continues to require human intervention; while reducing the work involved in making recordings accessible, there remain additional requirements due to variations in audio recordings, "noise" introduced due to age, and the necessity of dealing with differing dialects and languages.

Together, the results of this suite of explorations is helping chart a path for the Library in the adoption of new machine learning technologies. We have sought to invest in the future by sharing our examinations not only within the Library, but with other cultural heritage organizations seeking to provide wide access to content.

Building on the Past, and Looking Forward

These efforts and more build on decades of work at the Library and beyond to boost momentum and capacity for the agency's Digital Strategy. We are also looking forward to new explorations.

For example, we are experimenting with sharing digital collections at scale using a cloud computing environment, generously supported by the Andrew W. Mellon Foundation. Our fourth Innovator in Residence—writer, artist, and educator Courtney McClellan, will design a collaborative annotation tool to engage students of all ages with historic Library materials. A program aimed at uncovering the voices and stories of underrepresented groups has also been funded by the Mellon Foundation. "Staff innovators" are exploring tools and modes of access for born digital materials. And we are working to identify the next phases of crowdsourcing, prototype "humans in the loop" methods combining crowdsourcing and machine learning, and explore alternative discovery and access points for Library content.

The Library of Congress Digital Strategy Directorate continues to guide Library-wide conversations about growing and providing access to our content. Looking to our colleagues near and far for their innovative work and approaches, we look forward to continuing the iterative work of throwing open the treasure chest, connecting, and investing in a future that welcomes all Americans to the Library of Congress.

Big Data SIG Plans for WLIC 2021

Cory Lampert
Professor and Head, Digital Collections
University Libraries
University of Nevada, Las Vegas
cory.lampert@unlv.edu

The Big Data Special Interest Group is happy to announce that we will be supporting two sessions at the virtual 2021 WLIC.

First, we will be holding a SIG Case Studies Session with a slate for speakers from Asia, Australia, North America, and Europe who will discuss projects on the theme, "Better together: Creating solutions to the challenges of data use, reuse, and sharing". We hope that you will join us to learn more about the implications of data challenges for our individual users, our organizations, and our own roles as library professionals. This session will provide an overview of multiple case studies showcasing synergies in data use, reuse, sharing, documentation, description, access, and archiving among and between various communities.

The second session is a Presentation Session in partnership with the Science & Technology Section and Continuing Professional Development and Workplace Learning Section on the topic of, "Library Carpentries: A 360 Degree Review of Supporting Data Science Skills". This session will introduce what Carpentries are and how libraries have established programs to teach coding and offer data science skills to researchers and users globally. The goal of Carpentries is to empower users to use software and data in their work and to build capacity for librarians to feel empowered in data-centric roles and partnerships. Four speakers will address themes related to offering and supporting Library Carpentries and will share case studies and relevant experiences.

Please stay tuned to the IT Section Twitter (https://twitter.com/ifla_it or Twitter handle: @ifla_it) for more details on registration dates, the final program schedule, speaker biographies, and how you can engage with colleagues during these sessions. We look forward to seeing you this year online 17-19 August 2021.

IFLA Section

Information Technology

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 applications of 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 and workshops. 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 and resources regarding activities of the Section, session minutes, publications, and membership details.

The IFLA-IT email list provides a forum for members to exchange ideas and 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 December.

Primary Contacts

Chair
May Chang, University of Cincinnati, United States
may.chang@uc.edu

Secretary
Maria Kadesjö,
National Library of Sweden, Sweden
maria.kadesjo@kb.se

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

Standing Committee Members
•Edmund Balnaves, Prosentient Systems, Australia; ejb@prosentient.com.au
•Anna Tereza Barbosa da Silva, Oslo Police District, Norway; anna.tereza.barbosa.da.silva@politiet.no
•Leda Bultrini, ARPA Lazio, Italy; leda.bultrini@gmail.com
•Almudena Caballos Villar, Biblioteca de la Universidad Computense de Madrid, Spain; acaballo@ucm.es
•Lynn Kleinveldt, Cape Peninsula University of Technology, South Africa; lynn.kleinveldt@gmail.com
•Nthabiseng Kotsokoane, University of the Free State, South Africa; nthabik@gmail.com
•Peter Leinen, Deutsche Nationalbibliothek, Germany; p.leinen@dnb.de
•María Loretto Puga, Biblioteca del Congreso Nacional de Chile, Chile; mpuga@bcn.cl
•Sogoba Souleymane, University of Ségou, Mali; Sogoba.souleymane@gmail.com
•Elena Sánchez Nogales, Biblioteca Nacional de España, Spain; elena.sanchez@bne.es
•Ngozi Blessing Ukachi, University of Lagos, Nigeria; nukachi@unilag.edu.ng
•Astrid Verheusen, LIBER, Netherlands; a.verheusen@kb.nl
•Evviva Weinraub Lajoie, University at Buffalo, United States; evvivawe@buffalo.edu
•Jianyong Zhang, National Science and Technology Library of China (NSTL), China; zhangjy@nstl.gov.cn
•Katherine Zwaard, Library of Congress, United States; kzwa@loc.gov
•Maria de la Peña, Fundación I.E. Library, Spain; maria.delapena@ie.edu
•Emina Čano Tomić, National Library of Serbia, Serbia; emina.cano.tomic@nb.rs

Convenor Big Data
•Cory Lampert, University of Nevada Las Vegas, United States; cory.lampert@unlv.edu

Corresponding Members
•Zainab Ajab Mohideen, Universiti Sains Malaysia, Malaysia; zainab@usm.my
•Michelle Jiyon Ryu, Sinai Health System, Canada; michelle.ryu@sinaihealth.ca
•Wouter Klapwijk, Stellenbosch University, South Africa; wklap@sun.ac.za
•Alireza Tayarani, Library of Iran Parliament, Iran; tayerani@gmail.com