


CARNEGIE MELLON UNIVERSITY LIBRARIES

BOUNDless





SPRING 2019

two A Conversation with: Huajin Wang

three Enhance + Expertise: Carnegie Mellon's Inaugural Open Science Symposium

four Enhance + Services: Libraries Speaker Series

five Build + Services: Building a Skillset with Software Carpentry

six Steward + Collections: Celebrating Ada Lovelace Day

seven Insight: Keith Webster

eight Enhance + Services: The Center for Midnight

nine Build + Expertise : Meet the New Associate Deans

ten Qatar: Solving Mysteries with Maps

Front cover:

Carnegie Mellon University Archives, "Campus Week Auto-Parade," circa 1921.

Back cover:

In his "Orphan Morphin" presentation for the Spring 2019 Libraries Speaker Series, filmmaker and curator Craig Baldwin "surfs the wave of obsolescence" using spoken word, graphics and motion picture clips to advocate for the radical re-working of marginalized film material in the forms of collage, compilation-doc and speculative fiction. This event took place in IDeATe Studio A on February 28, 2019.

Keith G. Webster, *Dean of University Libraries*

Contact the Dean's Office

412.268.2447

library.cmu.edu



Huajin Wang

Librarian Huajin Wang joined the University Libraries in 2017. A cell biologist by training, with more than 10 years of research experience, she is also a member of the AIDR 2019 Program Committee.

What is AIDR 2019?

AIDR stands for Artificial Intelligence for Data Discovery and Reuse. It is a conference that aims to bring together everyone whose work is related to using AI or machine learning to facilitate data discovery and reuse. It takes place May 13-15 at Simmons Auditorium at Carnegie Mellon University.

What does “data discovery and reuse” mean and why is it important?

“Data discovery and reuse” means finding existing data that are out there and reusing these data to solve a new problem or give an old problem a deeper look. Scientists generate lots of data every day, at a cost of millions of dollars. These large and complex datasets often contain lots of information, making it impossible for a single investigator to extract all of the useful material. So, it makes sense for multiple investigators with different expertise to look at the data from different angles.

Unfortunately, the reality is that most datasets only get used once, in the original publication. After that, these datasets often either live on the PI’s server or in data repositories, and few of them get used again. All the rich information contained in the datasets that was so expensive to produce in the first place gets buried. As you can see, facilitating the reuse of data would allow science to move more quickly and be more economical.

How does AI help with data discovery and reuse?

With the recent advances in machine learning and AI, it is possible to train computers to learn certain tasks, and find optimal solutions to a problem, such as integrating different datasets. Machine learning and AI are already being used extensively to build search engines, databases, and to facilitate data analytics and automation in almost every discipline. It’s about time that people working in all these disciplines come together, benefit from mutual expertise, and address these challenges together, using the power of AI.

How did you end up bringing this conference to CMU?

My interest in data discovery and reuse started a long time ago when I was a biologist working in a big lab. After joining the Libraries as a research liaison, I felt that I was in an ideal position to help researchers with data problems. When I saw a funding opportunity from the National Science Foundation on data reuse, I knew this was a perfect opportunity. It happened that Nick Nystrom, my long-time collaborator at Pittsburgh Supercomputing Center (PSC) had the same idea. We collaborated with Dean Webster and Paola Buitrago, Artificial Intelligence and Big Data Group Leader at PSC, to submit a grant application together. And we got funded!

Why is CMU the right place to have this conference?

CMU has a strong community for AI research, almost everyone working at CMU, from every college, has some connection to AI. Our library is a leader in many national and international efforts to advance the state of research data management. And PSC is a national leader for scientific computing and hosts many important community datasets. We are a perfect team.

More information about AIDR 2019:

 <https://events.library.cmu.edu/aidr2019>



Carnegie Mellon's Inaugural Open Science Symposium

CMU's inaugural Open Science Symposium convened researchers from different backgrounds to generate awareness, support, and innovative ideas for open science.

The two-day Open Science Symposium, held October 18-19 at the Mellon Institute Library, brought together students and researchers from a wide variety of departments at CMU and the University of Pittsburgh to discuss open science, a growing movement that aims to make all research products, including data, code, and publications, freely available.

Despite receiving funding from federal tax dollars, science research output has historically been difficult to access and reuse. It is often published in journals with very expensive subscription costs, typically paid for by university libraries, and the data used to generate figures in publications are commonly not shared or only shared by request. These practices have made it difficult for scientists to access, reuse, and reproduce the work of others, and have in part led to a widely reported “reproducibility crisis” in science.

Open access publishing, funder and publisher mandates for data sharing, and a rising number of open science tools and platforms are rapidly changing both the practice and dissemination of scientific research. The inaugural Open Science Symposium tapped into this increasing enthusiasm with panels, workshops, and plenty of opportunities for networking.

On the first day of the symposium, panelists from a variety of scientific disciplines discussed the many benefits to practicing open science,

including better visibility of their work, increased citations, and enhanced opportunities for collaboration, all of which positively impacted their research and careers. Challenges discussed included gaps in infrastructure for open science, the time and effort it takes to make work shareable, and a lack of incentive. The additional challenge of a large variety of data types and standards has also led to a slower adoption of data sharing in some fields.

Coffee breaks and a Scientific Speed Dating event, which encouraged researchers from different disciplines to informally discuss their work and seek collaboration opportunities, contributed to an atmosphere of visibility and transparency, with the goal of fostering multidisciplinary collaborations.

The second day consisted of a series of hands-on workshops on popular and reliable tools and platforms for practicing open science, including Code Ocean, Bioconductor, PSC (Pittsburgh Supercomputing Center), BenchSci, and Open Science Framework. This was followed by a deposit-a-thon for researchers to learn more about KiltHub, CMU's institutional repository that is available for anyone at CMU to share finished research products, from datasets to publications.

The Open Science Symposium was a joint event by Mellon College of Science and University Libraries, supported by funding from the DSF Charitable Foundation. It was organized by Melanie Gainey, Ana Van Gulick, and Huajin Wang of University Libraries and Eric Ytri, assistant professor of Biological Sciences and Center for Neural Basis of Cognition (CNBC).



LIBRARIES SPEAKER SERIES

A new speaker series at the Libraries has welcomed special guests to talk about the history of the Carnegie International and experimental filmmaking.

The new University Libraries Speaker Series was created to highlight speakers from a wide range of backgrounds who embody at least one of the Libraries' core values of context, curiosity, or access. One speaker event will take place each semester. The Series, for which there is no charge, strives to deliver experiences that inform, delight, and enlighten the CMU community, contributing to a campus atmosphere that enriches the university's mission.

The inaugural event was held November 2 in the Posner Center with Ingrid Schaffner, curator of the 57th Carnegie International, and Carnegie Museum of Art Associate Registrar Elizabeth Tufts Brown. The Carnegie International had opened on October 13 at the CMoA, just a few weeks before this event. It is the oldest North American exhibition of contemporary art from around the globe. Attendees heard Schaffner and Tufts Brown speak about the International: its history, its artists, its files, and the research that make this exhibition so dynamic.

For the spring installment of the Speaker Series, Craig Baldwin, a filmmaker, artist and curator from San Francisco, whose creative interests involve the repurposing and remixing of found imagery, delivered his "Orphan Morphin" presentation. Highlighting "orphan" works, copyrighted material that no longer have an owner or sponsorship, thereby falling into the public domain, Baldwin used spoken word, graphics and motion picture clips to advocate for the radical re-working of marginalized film material in the forms of collage, compilation-doc and speculative fiction.



Building a Skillset with Software Carpentry

On a snowy January day, 40 attendees gathered in Hunt Library to learn carpentry. But they didn't bring tool belts, just their laptops.

If you're looking to nail down your computer programming skills, Software Carpentry can give you the tools you need.

Software Carpentry is a non-profit organization that teaches basic computing skills to researchers. Workshops are designed to teach foundational computing skills, including automation of workflows, version control of code and files, and basic programming in Python or R. The workshops are hands-on and interactive, with an emphasis on building practical skills and confidence to learn more programming.

University Libraries became institutional members of the Carpentries last September, which allows Libraries employees to organize workshops with instructors from across the country and provide an opportunity for faculty and staff to participate in instructor training.

In January, about 40 graduate students, postdocs and staff attended a workshop on Python. Colleen Schneider, a doctoral candidate in psychology, jumped at the opportunity to participate after hearing about it from Ana Van Gulick, librarian and program director, Open Science.

“Programming skills are becoming so important in many fields, but learning those languages

can be overwhelming if you are not a computer scientist,” Schneider said. “The Software Carpentry class was a great way to learn the ropes and also learn about more resources for further programming success!”

The accessibility is what appealed to Cheryl Rozinski, a master's degree candidate in public policy and management.

“The workshop provided an opportunity to dive into new programs at a pace that is comfortable in a low stakes environment,” Rozinski said. “For me, learning software in a masters class setting was very stressful as teachers often skip steps or do not fully explain the why. This setting is not only safe, affirmative, and enjoyable, but you leave with new skills that can apply to your everyday work.”

Rozinski plans to apply the skills she learned at the workshop to her work at CMU — she is using GitHub for file sharing — and after she earns her degree.

Offering software and data carpentry workshops is one way the Libraries are expanding their support for Open Science and Research Data at CMU. A team of more than 10 data experts in the Libraries is focused on expending research support through digital tools, web resources, trainings and consultations, and offering core computing skills to the CMU community through Software and Data Carpentry is key to that mission.



Celebrating Ada Lovelace Day

In celebration of Ada Lovelace Day, the Libraries invited the students of Women@SCS to a tea party in the Fine and Rare Book Room and private viewing of the historic computing machines – including two Enigma Machines – of the Traub-McCorduck Collection. Ada Lovelace Day is a celebration of women in STEM (science, technology, engineering, mathematics) held every year on the second Tuesday of October in recognition of the first female computer programmer. Senior Librarian Mary Catharine Johnsen and CLIR Fellow Andrew McGee hosted the student visitors and shared stories about the history of computing.

Portrait by Alfred Edward Chalon



Insight

As we pursue our mission to advance research, learning and knowledge sharing across the university, the digital technologies that have transformed the dissemination of academic books and journals have also brought profound change to the nature of the scholarly record. At the Libraries, our focus on capturing the formal outputs of research, typically published in scientific articles and monographs, has today expanded to the curation, sharing, and re-use of an array of artifacts generated during the research process.

We are able to capture details of experimental protocols, field notes, community conversations on research projects, and open reviews and commentaries on the outcomes of research. The technologies and approaches that make all of this possible are often viewed through the lens of open scholarship, an environment in which the research process and its outcomes are shared as widely as possible.

Last October, we welcomed over 100 delegates to the Open Science Symposium, where speakers and panelists shared their work, and the digital tools they use, to advance their research practice. Open scholarship offers

libraries great opportunities to add innovative services that impact the research communities they serve. Carnegie Mellon's librarians are at the heart of this movement, and are willing to help faculty and students from many disciplines explore solutions. One of our most popular developments has been the introduction of software carpentry workshops, events that allow any member of the CMU community to enhance their skills in working with data and other research outputs.

Perhaps the most visible part of the evolving scholarly record is research data. I have written previously about our KiltHub repository which provides a platform in which research data can be stored, preserved, and used by other scholars. As the volume of available research data grows, the potential of artificial intelligence to support fresh analysis becomes more apparent. We were delighted to receive a grant from the National Science Foundation, in partnership with the Pittsburgh Supercomputing Center, to organize the AIDR 2019 conference, which will take place on campus May 13 – 15.

Finally, we were all thrilled to welcome our two new Associate Deans, A.M. Salaz and Brian Mathews, who joined us at the end of last year. A.M. moves to Pittsburgh from the Carnegie Mellon Qatar campus, and Brian joined us from Virginia Tech. Please join me in welcoming them to the Libraries.

A handwritten signature in black ink that reads "Keith G. Webster". The signature is written in a cursive, slightly slanted style.

Keith G. Webster
Dean of University Libraries

The Center for Midnight

Can a dozen artists, technologists, and scholars collaborate with each other and with machines to produce a readable, interesting story in under 12 hours?

With the Frank-Ratchye STUDIO for Creative Inquiry acting as his temporary home base, bestselling author and technologist Robin Sloan led a pop-up writing collective of students, artists, and scholars through a three-day experiment of generative fiction.

Sloan, whose first book “Mr. Penumbra’s 24-Hour Bookstore” has been called a love letter to digital humanities, visited CMU for an extended artist’s residency at the invitation of the STUDIO for Creative Inquiry, the Department of English, and dSHARP.

His most recent work experiments with human/computer collaborative writing, with the development of a machine that reads a selection of written text and suggests possible words to come next in the sentence or phrase. The machine does this by reaching into its model of language, a recurrent neural network trained on whatever collection of text seems appropriate, and trying to find sensible endings to the sentence a writer has started.

During Sloan’s visit, an assembled team of artists and scholars gathered for four hours a day for three days, connected with Sloan’s generative text editor, and attempted to assemble a readable short story in the space of 12 hours.

Inspired by David Markson’s “The Last Novel,” Sloan decided to compile the short story in 1-3 sentence snippets, which would allow people to contribute as much or as little as they were able. The story would be about a yet-unnamed artistic movement, so he pre-trained his recurrent neural network on the biographies of artists, and the group got to work.

The ensuing narrative took the form of a short account of the Center for Midnight, a fictional artistic movement of the late twentieth century. The results exceeded expectations, resulting in intriguing sentences and phrases such as:

“The golden age of lithography”

“She embroidered the ideas of Laura de Gioste on a seaside tree.”

“When he died, she is reported to have said, ‘He became a response to himself.’”

On the final day, Sloan put order to the group’s words and replaced a few proper nouns to solidify the narrative thread. The contributors marked any confusing passages and spent time fixing problem sections: making the narrative flow, removing tangents, and tightening the text.

Somehow, amidst the chaos of machine prose and a rotating group of amateurs, the group had assembled a story with a narrative arc, enjoyable prose, and a coherent (if strange) plot.

robinsloan.com/center-for-midnight/



MEET THE NEW ASSOCIATE DEANS

Carnegie Mellon University Libraries has strengthened its leadership team supporting CMU teaching, learning, and research efforts with the addition of two new associate deans.

A.M. Salaz, formerly a senior librarian at the Carnegie Mellon University in Qatar Library, has joined the University Libraries as the associate dean of research and academic services and Brian Mathews, formerly the associate dean for learning and outreach at Virginia Tech, has joined as the associate dean for innovation, preservation, and access. Salaz and Mathews assumed their new roles on December 1.

“The rapid evolution of the 21st century library demands innovative, creative leaders who bring a user-centric approach to our work. A.M. and Brian will be critical to our efforts to support student success inside and outside the classroom, and to strengthen the work of our academic community to ensure its transformative impact on our campus and beyond,” said Keith Webster, dean of University Libraries and director of emerging and integrative media initiatives.

In her role as associate dean of research and academic services, Salaz will oversee and provide strategic direction for the Libraries’ team of twenty liaison librarians and information specialists, all of whom are members of the Faculty. She will also continue to advance her own scholarship and publishing,

which focuses on adult learning, digital reading, transnational higher education, and information literacy.

Salaz founded the Digital Information Behavior Lab at the CMU-Q campus and is a past-chair of the board of directors of the Information Literacy Network of the Gulf Cooperation Council. She serves as a contributing analyst for the Academic Reading Format International Study (ARFIS) based out of UCLA.

As the associate dean of innovation, preservation, and access, Mathews will supervise and provide strategic direction for the sixteen faculty and staff in the departments of Scholarly Communications and Publishing, Data Management, Digital Scholarship, Digital Humanities, Digitization and Archives & Unique Collections. This vital role brings together diverse departments that preserve unique materials, provide access to research, and innovate new methods for scholarship and information discovery.

Mathews comes to CMU from Virginia Tech, where he was associate dean for learning and outreach. In this role, he integrated library resources and services into new general education requirements, guided campus-wide initiatives for digital literacy, and managed over \$6 million in library renovations. He participates in the Association for Research Libraries Leadership Fellows Program and is the author of “Encoding Space: Shaping Learning Environments that Unlock Human Potential.” He formerly wrote a blog for the “Chronicle of Higher Education” called “The Ubiquitous Librarian.”

Solving Mysteries with Maps



Historian Christopher Alario traced the history of the Qatar peninsula on maps at a lecture at Carnegie Mellon University in Qatar.

On February 11, the CMU-Q Library hosted a presentation called “Qatar National Library Presents: Solving Mysteries with Maps in Heritage Library.” At the event, attended by about 40 faculty, staff and students, Qatar National Library (QNL) Historical Maps Librarian Christopher Alario demonstrated the kinds of research projects made possible by the rich collections at the Heritage Library of the QNL, with particular attention to the maps of Arabia and the Persian Gulf. He highlighted mysteries presented in the documents – such as why Qatar vanished from maps for 230 years between 1596 and the 1820s – and offered clues to their solutions.

“Historical maps are like time machines,” Alario said, “they tell you how our ancestors viewed the world.”

During his talk, Alario noted that Qatar has been a known place since the time of Ptolemy, nearly 2,000 years ago, and the presence of the peninsula was represented in maps from 1548 to 1596. Oddly, the Qatar peninsula—although not necessarily the towns and settlements—disappeared almost completely

from maps until 1813.

Crucial to solving the mystery is to understand that the waters surrounding Qatar are dangerously shallow for sailing vessels.

Alario pointed out that in the 1500s, the Arabian Gulf was explored primarily by the Portuguese, and later by Dutch and British ships.

“The British and Dutch ships had a deep draft,” he said, “so they would have stayed far from the coastline so they wouldn’t run aground.” Alario believes at that distance, the peninsula would not have been visible. “The Portuguese caravels, however, had a shallower draft, and they were more inclined to use local craft.”

The maps from the 1500s appear to come from Portuguese sources, whereas later maps are derived from British and Dutch explorations, which may account for the disappearing peninsula.

Throughout his presentation, Alario referred to dozens of historical maps of the region. Most of these maps are in the Qatar National Library collection, and many are on permanent display at the library.

With its location directly across the street from QNL’s new building, the CMU-Q Library is uniquely poised to host joint events, such as this.

Carnegie Mellon University
Libraries

Office of the Dean
5000 Forbes Avenue
Pittsburgh, PA 15213

CARNEGIE MELLON UNIVERSITY LIBRARIES



BOUNDless