



# FALL 2023 SCHEDULE

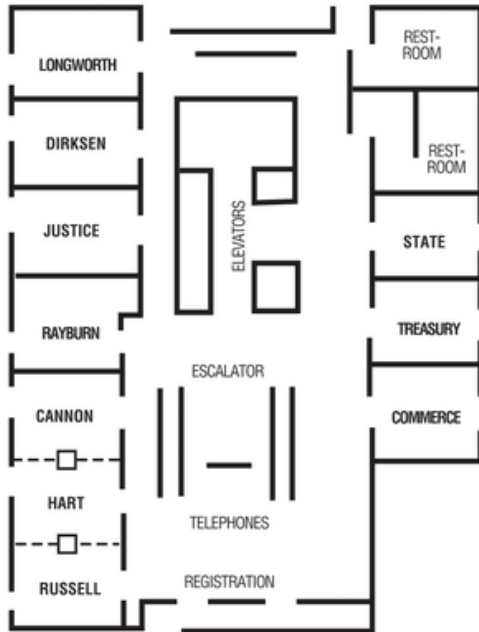
COALITION FOR NETWORKED INFORMATION

---

MEMBERSHIP MEETING  
JW MARRIOTT DOWNTOWN, WASHINGTON, DC  
DECEMBER 11-12, 2023

# JW MARRIOTT WASHINGTON, DC

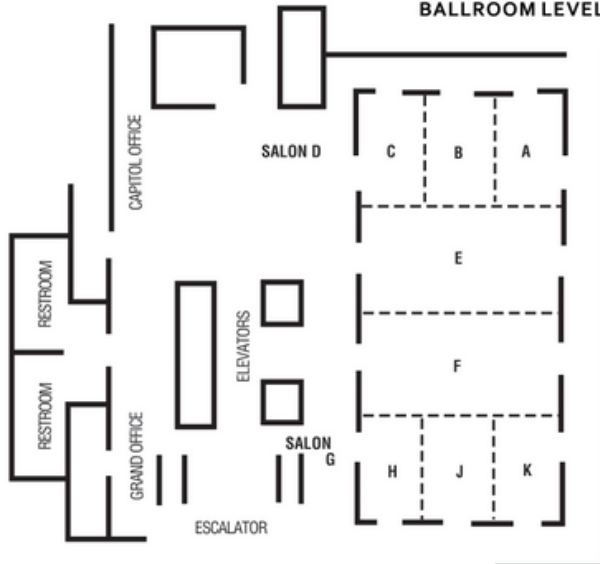
MEETING ROOM LEVEL



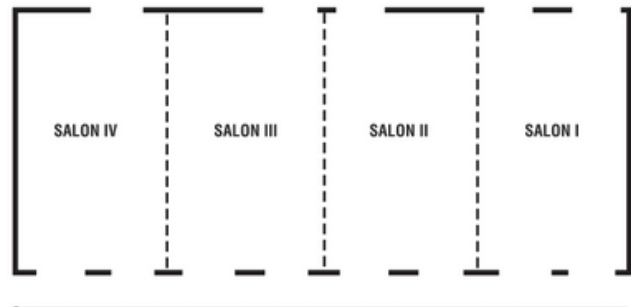
PENNSYLVANIA AVENUE/LOBBY LEVEL



BALLROOM LEVEL



BALLROOM LEVEL





Coalition for Networked Information

# Coalition for Networked Information Fall 2023 Membership Meeting

December 11–12, 2023

Washington, D.C.

#cni23f

**Network:** JWMarriott\_CONFERENCE

**Wi-Fi Passcode:** CNIfall23meet

For the most up-to-date information visit the digital, mobile-friendly  
schedule:

<https://cnifall23mtg.sched.com>

#### **CNI Code of Conduct**

CNI is committed to maintaining a welcoming and inclusive environment for inquiry, constructive disagreement, and intellectual freedom and honesty. We do not tolerate personal attacks, harassment of any kind, violence, or disruptive behavior. Please be respectful of our community's diversity and generous of others' views. Please bring concerns to our attention by contacting a member of the CNI staff.

[cni.org](https://cni.org)

# CNI Fall 2023 Schedule-at-a-Glance

## MONDAY, DECEMBER 11

- 11:00am Registration Opens (*Grand Foyer*)
- 11:15am First-time Attendees (*Salons D & E*)
- 12:00pm Refreshment Break (*Grand & Capitol Foyers*)
- 1:00pm Opening Plenary: Clifford Lynch (*Salons I & II*)
- 2:15pm Refreshment Break (*Grand & Capitol Foyers*)

2:45pm

### PROJECT BRIEFINGS

- |  |                          |
|--|--------------------------|
| 1.1 Legal Literacies for Text Data Mining - Cross-Border   | <i>Salons I &amp; II</i> |
| 1.2 To Increase or Decrease Capacity: The What, How, & Why of 21st Century Library Skill Development | <i>Salons D &amp; E</i>  |
| 1.3 UCSD Library & Göttingen State & University Library: Update on a Long-Term Collaboration         | <i>Salons F &amp; G</i>  |

3:15pm

### Passing Break

3:25pm

### PROJECT BRIEFINGS

- |  |                          |
|--|--------------------------|
| 2.1 Models for Sustainable & Inclusive Data Science Consulting & Collaboration in Higher Education | <i>Salons I &amp; II</i> |
| 2.2 The Ecosystem for Research Networking: Exploring Democratized Access to Research Instruments   | <i>Salons D &amp; E</i>  |
| 2.3 Multi-Custodial Approaches to Digital Preservation of Scholarship                              | <i>Salons F &amp; G</i>  |

3:55pm

### Refreshment Break (*Grand & Capitol Foyers*)

4:25pm

### PROJECT BRIEFINGS

- |   |                          |
|---|--------------------------|
| 3.1 Generative AI & Research Integrity  | <i>Salons I &amp; II</i> |
| 3.2 Partnerships in Research & Data Services: High Performance Computing, Collocation, & Facilitation | <i>Salons D &amp; E</i>  |
| 3.3 The Federal Year of Open Science  | <i>Salons F &amp; G</i>  |

## MONDAY, DECEMBER 11 Continued

5:10pm Passing Break

5:20pm Lightning Rounds (*Salons I & II*)

6:10–7:30pm Reception (*Salons III & IV*)

## TUESDAY, DECEMBER 12

7:30am Breakfast (including discussion roundtables) (*Salons III & IV*)

### Roundtable topics:

- Cyberinfrastructure to Support Exchange of Sensitive/Proprietary Metrics
- EDUCAUSE Research Computing & Data Community Group
- HELIOS
- Information Infrastructure & Grand Challenges
- Improvisational Yet Coordinated Approach to Campus Research Data Support & Services
- LEADING Next Steps
- LIBER (Association of European Research Libraries)
- Making AI Generative for Higher Ed
- More may be added; check online schedule for final list

8:45am PROJECT BRIEFINGS

4.1 Navigating the AI-Driven Academic Frontier: Tools & Initiatives	<i>Salons I &amp; II</i>
4.2 Blue Core: A Community-Operated, Shared BIBFRAME Data Store	<i>Salons D &amp; E</i>
4.3 The Data Core at Weill Cornell Medicine: A Secure Computational Enclave for Sensitive Data Analysis	<i>Salons F &amp; G</i>

9:15am Passing Break

9:25am PROJECT BRIEFINGS

5.1 Researcher & Institutional Impact of Data Management & Sharing Policies	<i>Salons I &amp; II</i>
5.2 RDA-US Pilots the Targeted International working GRoups: US (TIGRUS) Program	<i>Salons D &amp; E</i>
5.3 Open Book Collective: Collective Paths Toward an Open & Sustainable Ecosystem for Monographs	<i>Salons F &amp; G</i>

## TUESDAY, DECEMBER 12 Continued

9:55am Refreshment Break (*Grand & Capitol Foyers*)

### 10:25am PROJECT BRIEFINGS

6.1 GPT Tools That Provide Source Information	<i>Salons I &amp; II</i>
6.2 A Radical New Approach to Data Citation: Cook the Carrots, Burn the Sticks	<i>Salons D &amp; E</i>

10:55am Passing Break

### 11:05am PROJECT BRIEFINGS

7.1 Updates from Funders: Priorities & Trends	<i>Salons I &amp; II</i>
7.2 Bringing Digitized Special Collections into the Research Workflow through JSTOR: Outcomes of a Three-year Pilot	<i>Salons D &amp; E</i>
7.3 Working Session to Vet the Products of the ACLS Commission on Fostering & Sustaining Diverse Digital Scholarship	<i>Salons F &amp; G</i>

11:50am Lunch (*Salons III & IV*)

### 1:00pm PROJECT BRIEFINGS

8.1 The NSF Public Access Initiative, Projects Funded, & Catalytic Aims of the Program	<i>Salons D &amp; E</i>
8.2 Duke's Research Support Initiative: Assessment, Recommendations, & Implementation	<i>Salons F &amp; G</i>

2:00pm Passing Break

2:15pm Closing Plenary: Open Access, Open Scholarship, & Machine Learning: A Panel & Community Conversation (*Salons I & II*)

3:30pm MEETING ADJOURNS

# CNI FALL 2023 Membership Meeting

MONDAY, DECEMBER 11

---

1:00–2:15 pm  
Salons I & II

## Opening Plenary: Clifford Lynch

*Clifford Lynch, Executive Director, Coalition for Networked Information*

CNI Executive Director Clifford Lynch will begin his customary December plenary with a survey of some recent key developments and trends in the networked information and research landscapes. With this as context, he will discuss the evolution of CNI's programs, strategies, and plans for the future and talk briefly about initiatives for program year 2023-2024 and beyond. There will be time for questions and discussion.

5:20–6:10 pm  
Salons I & II

## LIGHTNING ROUND

Back-to-back presentations

**Cyberinfrastructure to Support the Scalable Exchange of Sensitive and Proprietary Usage and Impact Metrics Across Public and Private Stakeholders**

*Christina Drummond, University of North Texas*

**LEADING Next Steps: Evaluating the Sustainability and Impact of Post-graduate Professional Development and Mentorship Programs**

*Erik Mitchell, University of California, San Diego*

**Ecosystem for Next Generation Infrastructure (ENGIN)**

*Sayed Choudhury, Carnegie Mellon University*

**The Research Data Support Landscape: Findings from a National Inventory of University Services**

*Dylan Ruediger, Ithaka S+R*

**The Stacks Platform: A System for Onsite Access to Rights Restricted Digital Content at the Library of Congress**

*Trevor Owens, Library of Congress*

**Unexpected Opportunities Illuminated by Yale's LUX Project**

*Robert Sanderson, Yale University*

**LIBER - Association of European Research Libraries**

*Julien Roche, LIBER*

Full presentation descriptions are available on the CNI website:

<https://www.cni.org/mm/fall-2023/project-briefings-breakout-sessions-f23>

7:30–8:45 am  
Salons III & IV

## **Breakfast Discussion Roundtables**

During breakfast, attendees may join lightly facilitated roundtable discussions on well-focused topics. These will be located at tables throughout the dining area, marked by table signs. The discussion tables are optional. If you do choose to join one (or more), you do not need to RSVP and may come and go as you please.

**Cyberinfrastructure to Support the Scalable Exchange of Sensitive and Proprietary Usage and Impact Metrics Across Public and Private Stakeholders,**  
*Christina Drummond, University of North Texas & Charles Watkinson, University of Michigan*

**EDUCAUSE Research Computing and Data Community Group**  
*Susan Ivey, North Carolina State University*

**HELIOS**  
*Alicia Salaz, University of Oregon*

**Information Infrastructure & Grand Challenges**  
*Donald Waters, Coalition for Networked Information*

**Kind of Blue: Improvisational Yet Coordinated Approach to Campus Research Data Support and Services**  
*Todd Grappone, University of California, Los Angeles & Salwa Ismail, University of California, Berkeley*

**LEADING Next Steps**  
*Erik Mitchell, University of California, San Diego*

**LIBER (Association of European Research Libraries)**  
*Julien Roche, LIBER*

**Making AI Generative for Higher Education**  
*Dylan Ruediger, Ithaka S+R & Karim Boughida, Stony Brook University*

More topics and facilitators may be added.



2:15–3:30 pm  
Salons I & II

## Closing Plenary: Open Access, Open Scholarship, and Machine Learning: A Panel and Community Conversation

*Rachael Samberg, Program Director, Office of Scholarly Communications, University of California, Berkeley*

*Heather Sardis, Associate Director for Technology and Strategic Planning, Massachusetts Institute of Technology*

*Richard Sever, Assistant Director, Cold Spring Harbor Laboratory Press; Co-Founder, bioRxiv and medRxiv*

*Moderator: Clifford Lynch, Executive Director, Coalition for Networked Information*

Recent issues surrounding the use of the scholarly literature as training data for machine learning (ML)-based systems, particularly large language models (LLMs) and various generative artificial intelligence (AI) applications driven by these systems, are raising new questions about what the community seeks to enable and accomplish through commitments to open access (OA) and open scholarship. These questions are shaped in part by assumptions about the extent to which these new ML-driven systems will be centralized and controlled by a few large technology firms, as opposed to being diffused and distributed much more widely through open-source developments.

The emergence and adoption of ideas surrounding OA and open scholarship have largely been driven by visions of desirable futures and values-based choices for institutional investments; in recent years, this has enjoyed a relatively broad consensus despite some ongoing disagreements about specific tactics to advance these goals. Suddenly, as we see examples of where generative AI may take us and the highly extractive nature of these generative AI systems, the consensus on ideal futures seems less clear. What will most effectively advance scholarship and the creation of new knowledge?

Historically, there has been clear consensus extending from the freedom to read scholarly literature to the freedom to compute on this corpus through text mining and natural language processing technologies. This had been recognized in funder OA policies and mechanisms like Creative Commons licenses. Newer ML-based technologies create persistent and reusable computational artifacts, and there seems to be less comfort with a powerful, broad-based set of mandates to enable these developments. Yet the implications of blocking or impeding these developments also remain unclear. And whatever is chosen must scale.

The panel will consider the current state of developments, particularly in light of the seemingly insatiable demand for training data by LLM-driven systems, and will focus on both what the desirable outcomes should be in light of the values and goals shaped by OA and open scholarship and the mechanisms that might be adopted to advance these outcomes.

# CNI Fall 2023 Membership Meeting

## PROJECT BRIEFINGS

MONDAY, DECEMBER 11

---

2:45–3:15 pm  
Salons I & II

### **1.1 Legal Literacies for Text Data Mining - Cross-Border ("LLTDM-X")**

*Thomas Padilla (Internet Archive)*

*Rachael Samberg (University of California, Berkeley)*

This session will identify and explain the key legal and ethical issues that challenge US digital humanities (DH) practitioners engaging in "cross-border" text data mining (TDM) research. The National Endowment for the Humanities (NEH)-funded project Legal Literacies for Text Data Mining - Cross-Border (LLTDM-X) builds upon the previous NEH-sponsored institute, Building Legal Literacies for Text Data Mining (Building LLTDM), which provided guidance and strategies to DH TDM researchers on navigating legal literacies for text data mining (including copyright, contracts, privacy, and ethics) within a US context. A common challenge highlighted during Building LLTDM was the fact that US TDM practitioners encounter expanding and increasingly complex cross-border legal problems when: (i) the materials they want to mine are housed in a foreign jurisdiction, or are otherwise subject to foreign database licensing or laws; (ii) the human subjects they are studying or who created the underlying content reside in another country; or, (iii) the colleagues with whom they are collaborating reside abroad, yielding uncertainty about which country's laws, agreements, and policies apply. LLTDM-X was designed to better understand the cross-border issues that DH TDM practitioners face, with the aim of using them to inform prospective research and education. The session will include a briefing about round tables held in 2023 with cross-border TDM practitioners and law and ethics experts from six countries, including session takeaways. It will also include the case study we developed to serve as initial education for TDM practitioners on these issues. The session will conclude with a discussion of the broader efforts needed to integrate LLTDM literacies into library services and DH curricula to reduce barriers to domestic and cross-border DH TDM research.

<https://buildinglltdm.org/lltdmx>

<https://buildinglltdm.org/>

2:45–3:15 pm  
Salons D & E

### **1.2 To Increase or Decrease Capacity: The What, How, and Why of 21st Century Library Skill Development**

*Tony Zanders (Skilltype)*

*Keith Webster (Carnegie Mellon University)*

*Karim Boughida (Stony Brook University)*

Decisions abound for the modern library leader on the shape, size, and structure of their organizations. Central to the debate remains a perennial question of which components of librarianship should the library continue to develop capacity amidst the ever-shifting technology and business landscapes. In this session, community members will share their perspectives on managing organizational development in today's challenging environment, and they will discuss their experiences pursuing alignment with institutional goals.

<https://www.skilltype.com>

2:45–3:15 pm  
Salons F & G

### **1.3 University of California, San Diego Library and Göttingen State and University Library: Update on a Long-Term Collaboration**

*David Minor (University of California, San Diego)*

*Jan Brase (Göttingen State and University Library)*

In 2016 the University of California, San Diego Library and the State and University Library Göttingen in Germany signed a memorandum of understanding (MOU) for strategic cooperation. Based on this MOU, the two libraries have pursued a wide range of collaborations and discussions across their organizations. Work has included staff exchanges, comparisons of service models, and technical explorations. This presentation will give an overview of similarities and differences between the libraries as examples of research-oriented university libraries of equivalent size and philosophy in different countries. It will also address how the collaboration has benefited each institution, the challenges experienced, and plans for future cooperation in 2024 and beyond.

3:25–3:55 pm  
Salons I & II

### **2.1 Models for Sustainable and Inclusive Data Science Consulting and Collaboration in Higher Education**

*Mara Blake (North Carolina State University)*

*Emily Griffith (North Carolina State University)*

This briefing will provide an update on the "Models for Sustainable and Inclusive Data Science Consulting and Collaboration in Higher Education" project generously sponsored by the Alfred P. Sloan Foundation. The project leveraged the successful collaboration between the Libraries, the Data Science Academy, and the Department of Statistics at North Carolina State University (NCSU) to provide interdisciplinary data science consulting to the campus. The NCSU model utilizes a mix of staff, faculty, and graduate student staffing to serve users across all the colleges of the university. Driven by an interest in approaches to structuring data science consulting in higher education and a relative dearth of recent literature on the topic, the presenters convened a cohort of experts from a variety of institution types to explore the topic. The cohort enthusiastically engaged with different facets of administrative models for data science consulting to identify considerations for structuring programs. The project briefing will share the

successful structure of the project, the initial considerations identified by the cohort, and details on forthcoming output from the project.

3:25–3:55 pm  
Salons D & E

## **2.2 The Ecosystem for Research Networking (ERN): Exploring Democratized Access to Research Instruments**

*Maureen Dougherty (Ecosystem for Research Networking)*

*Barr von Oehsen (Pittsburgh Supercomputing Center)*

*Forough Ghahramani (Edge)*

Multi-institutional collaborations are on the rise and are creating new pressures to support shared access to institutional resources including research instruments, computational resources, and data. Supporting and managing access to special purpose instruments, the massive data sets produced by these instruments, computational and storage infrastructure, and the related expertise, pose a significant challenge for these collaborations, especially for under-represented and non-R1 institutions. This session will discuss how the Ecosystem for Research Networking (ERN) is designing and implementing a prototype open source, edge computing platform for democratizing access to scientific research instruments; providing networking research computing resources including quantum computing; handling massive data sets produced by research instruments such as cryogenic electron microscopes (cryo-EM); supporting computational and storage infrastructure; and building the expertise to support compute-intensive research, which presents significant challenges. The presentation will conclude with an open community discussion about what impacts such a compute model could have on collaborations, approaches for achieving adoption by resource owners/national centers, attracting community support, and stimulating scientific discovery.

<https://www.ernrp.org>

<https://www.ernrp.org/cryoem-remote-instrument/>

3:25–3:55 pm  
Salons F & G

## **2.3 Multi-Custodial Approaches to Digital Preservation of Scholarship**

*Jefferson Bailey (Internet Archive)*

*Gaëlle Béquet (ISSN International Centre)*

This session will describe several multi-institutional efforts to ensure persistent accessibility of open access (OA) scholarship, especially at-risk scholarly publications that have few resources for long-term preservation. In 2021, the paper "Open is not forever: A study of vanished open access journals" highlighted the lack of preservation infrastructure for many smaller academic journals. Representatives from ISSN and the Internet Archive (IA) will describe local projects, their organizational partnership to pursue this work, and collective efforts to mitigate the possible ephemerality of smaller online academic publications. Starting in 2018, the IA pursued a large-scale project to build a complete collection of OA scholarly outputs that are published on the web, the IA Scholar

project. While the program originated to solve the challenges of a lack of preservation infrastructure for many "long-tail" publishers, it also has illuminated the benefits of multi-custodialism, i.e. that digital archives have affordances that enable their preservation and accessibility at a variety of complementary stewardship organizations. This talk will describe the current status of the IA Scholar project, how related data-sharing efforts have advanced a multi-custodial approach to ensuring both preservation and access to essential scholarly knowledge, and how these archives can be shared and stewarded by a diverse set of organizations. The IA is a partner in the Keepers Registry service managed by the ISSN International Center. This service aggregates information on the long-term preservation of digitized and digital journals provided by leading archiving agencies around the globe. It has expanded rapidly in recent years, with new partnerships involving archiving agencies such as Biblioteca nacional de España and TIB - Leibniz Information Centre for Science and Technology in Germany. Moreover, the service has evolved to offer statistics to measure archival coverage by country, as well as a service for diagnosing and retrieving the archival status of periodical collections. The elaboration of an ambitious international development program for the years to come will build on these achievements to attract archiving agencies from the Global South. Both presenting organizations are also participants in Project JASPER, a combined effort of ISSN IC, the Directory of OA Journals, IA, CLOCKSS, and the Public Knowledge Project, to build technical integrations that ensure at-risk journals can access digital preservation services, free of charge, that are embedded into their existing workflows. The session will highlight collaborative and strategic work for the digital preservation of OA scholarship and also discuss the work of Project Jasper to provide archiving and access services to scholarship from underrepresented nations.

<https://scholar.archive.org/>

<https://keepers.issn.org>

<https://doaj.org/preservation>

<https://doi.org/10.1002/asi.24460>

4:25–5:10 pm  
Salons I & II

### **3.1 Generative Artificial Intelligence (AI) and Research Integrity**

*Chris Bourg (Massachusetts Institute of Technology)*

*Heather Sardis (Massachusetts Institute of Technology)*

*Erin Stalberg (Massachusetts Institute of Technology)*

Generative AI is thoroughly disrupting the traditional ways in which we maintain and signal the trustworthiness and integrity of research. With public trust in academia and research waning, and in the face of recent high-profile instances of research misconduct, the scholarly community must act swiftly to develop policies, frameworks, and tools for leveraging the power of generative AI in ways that enhance, rather than erode, public trust in science, academia, and research. Despite the much-discussed risks, generative AI also offers the opportunity to design tools and develop policies to address some of the known failures of the current system. In September 2023, the MIT Libraries was among 27 teams

awarded seed grants to probe the social implications of generative AI and deliver impact papers for publication in 2024. This presentation will focus on articulating next-step strategies for increasing the likelihood that generative AI will impact research integrity in transformationally positive ways. Specifically, it will address: the potential of generative AI to mine research data and scholarly output at a massive scale, guidelines for the ethical and responsible use and citation of both paywalled and openly available research articles in generative AI, risk assessment and risk mitigation when using generative AI in the creation of a research article, uses of generative AI to improve the quality and availability of research data, and mechanisms for accelerating the open availability of scholarly literature and data to increase the quality of the data being used in generative AI tools.

<https://news.mit.edu/2023/mit-scholars-awarded-seed-grants-generative-ai-0918>

4:25–5:10 pm  
Salons D & E

### **3.2 Partnerships in Research and Data Services: High Performance Computing, Collocation, and Facilitation**

*Lauren Geiger (Mississippi State University)*

*Michael Navicky (Mississippi State University)*

*Jason Clark (Montana State University)*

*Doralyn Rossman (Montana State University)*

#### **Collaborative Approach to Research Computing and Data Management**

Mississippi State University (MSU) conducted the Research Computing and Data (RCD) Capabilities Model assessment developed by the Campus Research Computing Consortium (CaRCC) several months ago. The assessment uncovered several shortfalls from the current, siloed approach to research computing and data management. It also identified opportunities through overlapping and often not well-publicized capabilities that would create a broader collaborative approach to research computing and research data management. To address these issues, representatives from the MSU Libraries, Information Technology Services, the Office of Research Security and Compliance, and High Performance Computing formed a team to identify cross-campus needs and create a plan to address them. As a result, this team submitted an Area 7 (Planning) Campus Cyberinfrastructure grant. By continuing to work together with campus partners to build collaborative relationships, the institution can leverage computational capacity, data, software, and domains to maximize research outcomes.

#### **Montana State University's Research Alliance: Creating Intentional Library and University Partnerships and Spaces for Research and Data Services**

Montana State University (MSU) has seen an increase in the number and complexity of researchers' data service and research facilitation requests—from writing data management plans to analyzing research and data sharing networks, or from storing data to high performance computing and machine learning applications for research data. In response to these evolving needs, MSU formed a unique partnership, the Research Alliance, made up of campus entities that support faculty, student, and staff research and research data needs. The partners

reside in a newly created space in the MSU Library, and they include the Center for Faculty Excellence; the Office of Research Development; the MSU Library Research Optimization, Analytics, and Data Services (ROADS); the Undergraduate Scholars Program; and Research Cyberinfrastructure (University IT). This project is unusual in that much preparation work went into forming a truly collaborative partnership before the launch of the Research Alliance space, and also because the Library is a fully present partner in the initiative. Throughout the project, the Library convened working and planning groups, facilitated service design, and provided leadership in the formation of the Alliance. This presentation will discuss the partnership formation, it will reflect on the successes and challenges in the early days of the Alliance's formation, and it will include discussion of where shared data science projects and instructional collaborations are gaining traction between partners.

<https://www.montana.edu/research-alliance/>

<https://www.montana.edu/research/ord/>

<https://www.montana.edu/uit/rci/>

<https://www.montana.edu/facultyexcellence/>

<https://www.montana.edu/usp/>

Background information from Rebecca Bryant of OCLC

Research: <https://hangingtogether.org/social-interoperability-at-montana-state-university/>

4:25–5:10 pm  
Salons F & G

### **3.3 The Federal Year of Open Science**

*Martin Halbert (National Science Foundation)*

*Maryam Zaringhalam (White House Office of Science & Technology Policy)*

*Ashley Sands (Institute of Museum and Library Services)*

*Brett Boley (National Endowment for the Humanities)*

2023 was designated as a Year of Open Science by the White House Office of Science and Technology Policy (OSTP) and 16 US federal agencies. This multi-agency initiative across the federal government has coincided with planning and implementation efforts associated with the 2022 OSTP Nelson Memorandum. The Year of Open Science initiative has aimed to spark change and inspire open science engagement through events and activities that will advance the adoption of open, equitable, and secure science. This panel will discuss synergistic activities undertaken as part of this initiative and desired outcomes.

<https://open.science.gov/>

# CNI Fall 2023 Membership Meeting

## PROJECT BRIEFINGS

TUESDAY, DECEMBER 12

---

8:45–9:15 pm  
Salons I & II

### **4.1 Navigating the Artificial Intelligence-Driven Academic Frontier: Tools and Initiatives**

*Elias Tzoc (Clemson University)*

*Joelen Pastva (Carnegie Mellon University)*

*Benjamin Shaw (University of Maryland)*

*Leo Lo (The University of New Mexico)*

Less than a year ago, ChatGPT revolutionized the art of writing. This powerful tool unleashed a whirlwind of possibilities, transforming the way many people summarize, draft, and create different types of content, including emails, outlines, and even scholarly writing. The change has impacted every industry, including higher education. Generative artificial intelligence (AI) like ChatGPT and other AI technologies are powerful tools that can be used to improve academic research and writing, but it is important to use them responsibly and ethically. This presentation will include discussion of AI resources and tools that the academic community has created or adopted to proactively assist learners and researchers in navigating an increasingly AI-powered environment. The presentation will highlight examples from four institutions that have implemented initiatives such as AI literacy modules, AI tools for academic research and writing, and AI awareness campaigns.

8:45–9:15 pm  
Salons D & E

### **4.2 Blue Core: A Community-Operated, Shared BIBFRAME Data Store**

*Tom Cramer (Stanford University)*

*Simeon Warner (Cornell University)*

As BIBFRAME works its way to production at libraries across the world, leveraging linked data for library cataloging remains as promising as ever. By converting from "strings to things," incorporating data from the World Wide Web, and using a flexible ontology that builds on but is not constrained by MARC, library description can jump to modern standards. Until recently, however, BIBFRAME has been a challenge to adopt and incorporate into institutional workflows. Gaps in tooling and historical cataloging practices have created an expectation for copying and modifying data locally. As part of the Mellon Foundation-funded Linked Data for Production (LD4P) projects, and working in concert with the Library of Congress and others, we imagine the development of a new library metadata environment with linked data at its core. This environment will be cooperatively shaped and owned, with an open data license, allowing linked data to work at scale and for libraries to build off each other's work rather than duplicating effort and expense. With working integrations among existing components in the ecosystem (FOLIO,



Alma, Sinopia, ShareVDE, etc.), we are poised to make incremental progress toward widespread BIBFRAME adoption. This presentation will present the findings from the LD4P grants as well as the next steps with a growing consortium of adopters around this shared linked data store.

8:45–9:15 pm  
Salons F & G

### **4.3 The Data Core at Weill Cornell Medicine: A Secure Computational Enclave for Sensitive Data Analysis**

*Sarah Ben Maamar (Weill Cornell Medicine)*

Researchers at Weill Cornell Medicine (WCM) have an immediate need for secure environments to analyze massive datasets containing sensitive information, such as patient health data from records or external data from entities like the Center for Medicaid and Medicare Services (CMS). To meet this demand, WCM's Samuel J. Wood Library, in conjunction with the WCM infrastructure and cloud engineering teams, has established a secure computational data enclave known as "Data Core." The Wood Library team administers the environment, ensuring all data governance agreements are current and approved by the institutional review board. The Wood Library team also serves as the data custodian, interacts with data providers, and ensures appropriate data governance for the data it stewards. Data Core provides computing power in a controlled environment, along with scientific software and scalable memory, catering to research projects with substantial computational requirements. Initially, it primarily served Population Health Science researchers dealing with extensive datasets for healthcare improvement. Today, Data Core supports nearly 100 projects from various departments, accommodating individual users, research teams, and entire classrooms. Researchers are billed for their usage, covering expenses related to staffing, software, and servers, ensuring sustainable support for the service. The Data Core has been certified by CMS to host its massive patient health datasets for both on-premise and cloud environments, providing cost-effective choices depending on researcher needs. This project briefing will discuss the administration and value of a secure platform for the analysis of large and sensitive datasets.

<https://datacore.weill.cornell.edu/>

9:25–9:55 am  
Salons I & II

### **5.1 Researcher and Institutional Impact of Data Management and Sharing Policies**

*Jake Carlson (University at Buffalo)*

*Jonathan Petters (Virginia Polytechnic Institute and State University)*

*Joel Herndon (Duke University)*

The last 15 years have seen a marked growth of data management and sharing policies among federal agencies in the US and Canada. While these policies have an undeniable impact in terms of increased publicly available datasets, they have also impacted the research practices of funded researchers and the services and

infrastructure provided by institutions. Researchers and institutions alike share the responsibility to align practices with funding agency requirements concerning data management and sharing, but each stakeholder group has responded in ways that may not align with one another. This presentation delves into research resulting from the National Science Foundation-funded Realities of Academic Data Sharing (RADS) Initiative and provides a comprehensive comparative analysis of services and infrastructure of six academic institutions, as well as an overview of the overall impact of these policies for researchers and institutions. Insights into services, infrastructure, and impact can lead to the creation of streamlined pathways for enhancing institutional efficiencies in data management and sharing.

9:25–9:55 am  
Salons D & E

## **5.2 Research Data Alliance US (RDA-US) Pilots the Targeted International working GRoups: US (TIGRUS) Program**

*Beth Plale (Pervasive Technology Institute, Indiana University)*

This presentation will reintroduce the CNI community to the newly restructured Research Data Alliance US (RDA-US). It will include a review of the objectives of RDA-US, and through discussion, establish points of intersection where RDA-US can best facilitate future needs of the US research library community.

The Research Data Alliance (RDA) has long been a partner with research libraries and has served as a venue for globally informed consensus on best practices for publishing and sharing the products of research. It is a consortium committed to accelerating the sharing of research data across barriers with members from academia, industry, and government who work through focused working and interest groups to create recommendations (standards) that address data infrastructure roadblocks. Its national nodes, which include RDA-US, amplify these groups' activities to data management communities within their specific regions. The newly re-envisioned RDA-US has already made forward strides, identifying program officers, task force leaders, and steering committee members. RDA-US executive director Beth Plale is one of the dozen international researchers that founded the now 10,000+ member RDA a decade ago. Plale is also a founder of the HathiTrust Research Center.

RDA-US is a recognized advocate of the new forms of research and scholarship that are enabled by RDA recommendations and is facilitating solutions to US-based issues in data sharing to RDA through working groups. The landscape of science and scholarship in the US is evolving rapidly, with a growing emphasis on data sharing, collaboration, and standardized practices. However, challenges persist in achieving widespread adoption of principles and shared practices. RDA-US and the Pervasive Technology Institute (PTI) at Indiana University have joined forces to accelerate standardization efforts, administer US-funded programs, and advance the development of solutions to open science barriers. RDA-US, with the support of PTI, launched the Targeted International working GRoups - US Program, TIGRUS. TIGRUS addresses US data challenges, including those within the national

library community. Along with the TIGRUS program, RDA-US is targeting initiatives in workforce development, resource curation, and community building.

<https://www.rd-alliance.org/>

9:25–9:55 am  
Salons F & G

### **5.3 Open Book Collective: Collective Paths Toward an Open and Sustainable Ecosystem for Monographs**

*Lidia Uziel (University of California, Santa Barbara)*

*Livy Snyder (Open Book Collective)*

Built on the successes of the COPIM (Community-led Open Publication Infrastructures for Monographs) project (2019-23) and established in May 2022, the Open Book Collective (OBC) is an international partnership and a collective of open access (OA) publishers, infrastructure providers, libraries, and other non-profit organizations. The OBC's mission is to create a new OA book publishing ecosystem that is equitable, community-governed, and built on sustainable business models and community-owned infrastructures. The OBC is part of a larger holistic vision for the OA book ecosystem supported by Arcadia and Research England, who have together committed up to \$7 million to continue the work of the COPIM project, including financial support for the OBC and Thoth, under the remit of the Open Book Futures (OBF) project (2023-26). By pioneering new community-owned OA business models—including new funding and publishing models, a new collaborative platform, and collaboration with open, community-led infrastructures—the OBC charts a new collective path toward an open and sustainable monograph future. The session will include discussion of the OBC's community-led governance structure and will highlight the project's principal challenges in reshaping the larger open knowledge ecosystem. It will also include discussion of a strategic partnership focused on amplifying bibliodiverse and equitable community-led approaches and expanding critical infrastructures for OA book publishing. The principles of 'scaling small' will be discussed, as well as the importance of seeking strategic partnerships with publishers, universities, and infrastructure providers in diverse national and linguistic contexts beyond the existing networks in the UK and North America and, more particularly, in Africa, Australasia, Europe, and Latin America.

<https://openbookcollective.org>

<https://thoth.pub>

10:25–10:55 am  
Salons I & II

### **6.1 GPT Tools that Provide Source Information**

*Gary Price (infoDOCKET & ARL Day in Review)*

*Peter Brantley (University of California, Davis)*

Some of the most popular generative pre-trained transformer (GPT) resources do not provide source and citation data. This can be problematic not only in the short term (citation issues) but also in the long term (fiction becoming fact). This presentation will focus on online tools that do provide source information for the

facts supplied in the text. It will also include discussion of several productivity tools, as well as a discussion about these tools and issues that arise from their use, focusing on how our community can enhance them with the skills we have and the many other tools we offer. Tools to be discussed include Perplexity, You.com, Phind, Google Bard, Bing Chat, Docalysis, and Humata AI.

10:25–10:55 am  
Salons D & E

## **6.2 A Radical New Approach to Data Citation: Cook the Carrots, Burn the Sticks**

*Jamie Wittenberg (University of Colorado Boulder)*

*John Chodacki (California Digital Library)*

*Kristi Holmes (Northwestern University)*

Librarians and other stakeholders, through the Make Data Count initiative, have worked to advance data citation adoption among researchers and publishers by leveraging a variety of incentives (carrots) and regulations (sticks). Though much progress has been made, DataCite and Crossref systems show that across millions of scholarly outputs, structured 'data citations' are only present in tens of thousands of records. Often, this is because researchers mention underlying data without creating a structured citation or because publishers do not support structured citations for datasets in a paper's references. The Make Data Count initiative devised a new strategy that does not rely on researchers or on publishers to assert the relationship between a paper and its underlying datasets. With funding from Wellcome Trust, DataCite has worked with the Chan Zuckerberg Initiative to develop a machine-learning algorithm that extracts references to underlying data from full journal articles and preprints without the inclusion of structured data citations. This model has been applied to the full text of hundreds of millions of articles, resulting in the Open Global Data Citation Corpus—a trusted central aggregate of all references to research data across articles, preprints, government documents, and other outputs. This corpus will fundamentally change the way libraries, bibliometricians, research administrators, software systems, and funders measure the impact of scholarly research.

<https://makedatacount.org/data-citation/>

11:05–11:50 am  
Salons I & II

## **7.1 Update from Funders: Priorities and Trends**

*Brett Boley (National Endowment for the Humanities)*

*Ashley Sands (Institute of Museum and Library Services)*

*Martin Halbert (National Science Foundation)*

*Moderator: Katherine Klosek (Association of Research Libraries)*

In this update, representatives of major funding organizations will discuss the current status of programs, goals, and processes relevant to the CNI community and will offer observations on trends and priorities in the fields they monitor. Ample opportunities will be allowed for audience questions and discussion.

11:05–11:50 am  
Salons D & E

## **7.2 Bringing Digitized Special Collections into the Research Workflow through JSTOR: Outcomes of a Three-year Pilot**

*Bruce Heterick (ITHAKA)*

*Kevin Guthrie (ITHAKA)*

Libraries use a variety of systems and tools to build, manage, provide access to, preserve, and distribute digital versions of their special and primary source collections. As more of these valuable materials have been digitized and become discoverable, opportunities have exploded for those collections to be used by students and researchers, not only locally, but around the world. However, institutions are also facing the reality of increased costs associated with maintaining and preserving these important resources, particularly as accessibility mandates need to be met and research practices evolve (e.g., text and data mining, generative artificial intelligence, etc.). Over the last three years, ITHAKA has supported a pilot to test whether the infrastructure created to support JSTOR and Portico could be used to provide more cost-effective and impactful access to and preservation of these collections. During that pilot, 300 institutions published nearly 1,700 digitized collections to JSTOR with the hypothesis that: (1) putting their digitized collections on the JSTOR platform would expand the reach of those collections for research and teaching; and (2) having those primary source collections available alongside the important secondary literature already on the JSTOR platform could improve the research process. We will review what we learned in that pilot, providing evidence of what worked, what did not, and how this pilot work has evolved into an emerging stack of infrastructure services from JSTOR and Portico.

<https://about.jstor.org/whats-in-jstor/infrastructure/>

<https://www.jstor.org/site/brigham-young-university/>

11:05–11:50 am  
Salons F & G

## **7.3 Working Session to Vet the Products of the American Council of Learned Societies (ACLS) Commission on Fostering and Sustaining Diverse Digital Scholarship**

*Gabriela Baeza Ventura (University of Houston)*

*Carol Mandel (Council on Library and Information Resources)*

*Kenton Ramsby (Howard University)*

*James Shulman (The American Council of Learned Societies)*

Supported by the National Endowment for the Humanities and the Mellon Foundation, the 21-member commission has held 23 focus groups that have formed the basis of a report that will be released in early 2024. This session will discuss the contours of the report along with major recommendations. Attendees are invited to respond to the key elements of the report before the work is finalized.

<https://www.acls.org/digital-commission-sustaining-diverse-scholarship/>

1:00–2:00 pm  
Salons D & E

## **8.1 The National Science Foundation Public Access Initiative, Projects Funded, and Catalytic Aims of the Program**

*Martin Halbert (National Science Foundation)*

*Kathleen Fitzpatrick (Michigan State University)*

*David Elbert (Johns Hopkins University)*

*John Chodacki (California Digital Library)*

*Bill Branan (Johns Hopkins University)*

The National Science Foundation (NSF) Public Access Initiative (PAI) has funded a wide range of projects to advance open science, FAIR data guiding principles, and other scientific practices related to the proposition that the outputs of publicly funded research should be publicly accessible, notably, the Findable Accessible Interoperable Reusable Open Science Research Coordination Networks (FAIROS RCN) projects. This session will feature presentations by various projects funded by the PAI in recent years, including awardees from the 2022 FAIROS RCN competition that received funding contributions from all the scientific directorates that comprise NSF. The projects represented in this session provide perspective on the wide range of scientific issues addressed by the NSF PAI and the catalytic nature of the program's goals.

<https://new.nsf.gov/public-access>

<https://pass.jhu.edu/>

<https://stemedplus.commons.msu.edu>

1:00–2:00 pm  
Salons F & G

## **8.2 Duke University's Research Support Initiative: Assessment, Recommendations, and Implementation**

*Timothy McGeary (Duke University)*

*Rebecca Brouwer (Duke University)*

*John Board (Duke University)*

Beginning in February 2022, Duke University Information Technology Advisory Council (ITAC) launched a comprehensive assessment of research IT needs across Duke faculties in order to identify the actual support needs and existing service gaps in meeting the needs of Duke's modern, computationally intensive research enterprise. The assessment was motivated by a recognition that research methods and domains had changed dramatically over the years, while IT services and support structures tended to change more incrementally. The final findings and recommendations were developed as a Phase 1 report, and the process culminated with six findings and 10 recommendations, categorized into the areas of People, Process, and Technology, across the university. Following the release of the Phase 1 report, Duke's vice president for information technology, vice president for research and innovation, and vice provost for library affairs teamed up with others to sponsor Phase 2 of the effort, developing service proposals to meet the Phase 1 needs expressed by faculty, which encompass much more than IT. From January–May 2023, six cross-functional staff teams, each with faculty

representation, drew up 39 proposed services to address Phase 1 expressed needs. After further service refinement and faculty and sponsor feedback, 12 services were recommended for implementation, comprising three overarching service clusters: (1) Better Support for Researchers by Adding Personnel, Improving Coordination, and Easing Service Discovery, (2) Enhance Computational Services and Build Capacity for Data-Intensive Research, and (3) Balance Security and Compliance Requirements with Flexibility Needed to Support Different Types of Researchers. This presentation will discuss all three phases of the initiative: Assessment, Recommendations, and Implementation.

<https://itac.duke.edu/resource/research-it-needs/>

## CALENDAR OF KEY MEETINGS

**International Digital Curation Conference**, Edinburgh, Scotland–February 19–21, 2024

**CNI 2024 Spring Membership Meeting**, San Diego, CA–March 25–26, 2024

**IS&T's Archiving 2024**, Washington, DC–April 9–12, 2024

**Open Repositories**, Gothenburg, Sweden–June 3–6, 2024

**CNI 2024 Fall Membership Meeting**, Washington, DC–December 9–10, 2024

**Joint Conference on Digital Libraries, 2024**, TBA

**Designing Libraires 2024**, North Carolina State University, Fall 2024, exact dates TBA

**CNI 2025 Spring Membership Meeting**, Milwaukee, WI–April 7–8, 2025

**CNI 2025 Fall Membership Meeting**, Washington, DC–December 11–12, 2025

For more information on CNI and its programs visit [www.cni.org](http://www.cni.org)



# IMAGE CREDITS

Front (left to right; top to bottom)

Ozone hole over the South Pole on Oct. 5, 2022

Credit: NASA Earth Observatory image by Joshua Stevens; edited by MIT News

Courtesy: National Science Foundation

Chrome Labyrinth

Credit: Michael Woodson, University of North Carolina

Courtesy: National Science Foundation

Etruscan Statuette, 4th century B.C. in Etruria

Artist unknown

Courtesy: Getty Museum

CC0 1.0 Universal

Mouse retina

Credit: Kenyoung Kim, Wonkyu Ju and Mark Ellisman, National Center for

Microscopy and Imaging Research, University of California, San Diego

NIH funding from: National Institute of General Medical Sciences (NIGMS)

Courtesy: National Institutes of Health

CC BY-NC 2.0

Sapphire Pool, Yellowstone National Park

Credit: Detroit Photographic Co.

Courtesy: Library of Congress, Prints & Photographs Division

Mosaic of a Rabbit, Syria in the 5th-6th century

Artist unknown

Courtesy: Getty Museum

CC0 1.0 Universal

Back:

University of Arizona Libraries Maker Studio within CATalyst Studios

Credit: Chris Richards, University of Arizona Communications

All images were slightly modified for design purposes.

Descriptions are from image sources unless otherwise noted.



**cni**

Coalition for Networked Information

21 DUPONT CIRCLE, SUITE 800  
WASHINGTON, DC, 20036  
202-296-5098  
[www.cni.org](http://www.cni.org)