

Workshops

Web of Science as a Research Dataset

Date:

November 14-15, 2016

Meeting Place:

[Social Science Research Commons \(SSRC\)](#), Indiana University

Woodburn Hall, Room 200

1100 East Seventh Street

Bloomington, IN 47405

Web [Indiana University Campus Map](#) »

Organizers:**Katy Börner**

Victor H. Yngve Distinguished Professor of Information Science
Director, CI for Network Science Center Curator, Mapping Science Exhibit
School of Informatics and Computing
Indiana University, Bloomington
katy@indiana.edu

**Eamon Duede**

Executive Director
Knowledge Lab
University of Chicago
eduede@uchicago.edu

**James Pringle**

Head of Industry Development
Clarivate Analytics

Mention the Workshop on Twitter

[Tweet #WoSRD](#)

Workshop Goals

This workshop brings together data scientists and data stewards from research centers that are using the Web of Science™ at scale. We will explore WoS from the perspective of a research dataset and work together on practical ways to better support our research in the future. While the main focus will be on the Web of Science, the results should be extensible to all similar metadata aggregations. This unique focus—bringing data stewards and data scientists from these centers together to work on shared needs in tandem with the Web of Science team—will enable us to redefine and fully repurpose WoS to fit our research goals. We intend to launch an ongoing community in which we will learn techniques and develop tools to improve the data that underlies our research.

Advance Preparations

- Data stewards will provide a short profile of how WoS as a dataset is being implemented in the context of their research center/university and the technical, content, and other challenges they are facing.
- Researcher data scientists will prepare a short profile of current research projects leveraging the WoS dataset, focusing on key

challenges such as linking, disambiguating, mining, etc. that, if solved, would offer greater research opportunities.

Schedule:

Monday, November 14, 2016

[View recording »](#)

12:00pm Arrival, Light Lunch

12:30pm **Welcome and Introductions**

1:00pm **Session 1: Web of Science “Outside the Box”**

Facilitator: Katy Börner

The Web of Science and similar metadata datasets are housed, maintained, and enhanced in local institutional enclaves. This session looks at how data scientists and managers set up the WoS dataset in an enclave, disambiguate the data (names, institutions, and geolocations), link it to other datasets, and build predictive models.

- Matt Hutchinson, Indiana University: **“Implementing Web of Science data for university-wide research”**.
- Yadu Babuji and Eamon Duede, University of Chicago: **“Web of Science data for secure, cloud-based cross-institutional access”**.
- Orion Penner, École Polytechnique Fédérale de Lausanne: **“Managing and exploring Web of Science data at the research group level”**.
- Nicholas Monath, University of Massachusetts Amherst: **“Entity Resolution and Clustering Author Disambiguation and Related Tasks”**.
- Vette Torvik, University of Illinois: **“Three applications of integrating PubMed with Web of Science”**.
- Lee Giles, Pennsylvania State University: **“Web of Science data in the context of open tools and infrastructures”**.

2:00pm Wrap Up: Discussion will focus on identifying key data challenges in these areas.

2:30pm *Break*

3:00pm **Session 2: Breakthrough Research**

Facilitator: Eamon Duede

A “lightning round” look at examples of new questions asked and answered in recent research across network science, science of science, scientometrics, and science policy. We will explore how WoS and related large-scale datasets enabled these new questions to be addressed and whether enhancements to the dataset could have better supported the research.

- **Approximately twelve participants to be confirmed.**

2:00pm Wrap Up: Discussion will focus on identifying key data challenges in these areas.

4:30pm Day 2 Preparation:

Establish “big questions” for the breakout teams. Set up for hackathon.

5:00pm *Adjourn*

6:30pm Dinner

Tuesday, November 15, 2016

8:00am Breakfast, Initial Breakout Team Meetings

9:00am **Session 3: Understanding Web of Science as Research Data**

Facilitator: Jason Rollins

For over 50 years, the Web of Science evolved as a dataset in response to changing research contexts and priorities. Today, more researchers are using the Web of Science “at scale” to ask and answer powerful new questions about the shape, dynamics, and veracity of science and scholarship. The Web of Science now appears both an object of inquiry its own right and a vast sensor network for discerning large-scale trends. What is changing in this dataset to support these new uses, and what could change further? Presentations and discussion led by Clarivate Analytics team.

- Jim Pringle: **“WoS Metadata as Research Data”**
- Ted Lawless: **“Web of Science Data Integration”**
- Ling Bai: **“Data Unification and Disambiguation: Institutions and Authors”**
- Sebastien Brien: **“Clarivate Analytics in the Cloud: Architecture and Analytics”**

10:30am *Break*

11:00am **Session 4: Hackathon Breakout Sessions**

Facilitators: Eamon Duede, Jason Rollins, and Ted Lawless

A mix of sessions determined by 3-4 “big questions” prioritized on Day 1, grouped as:

- A. **Technical Hackathon(s)**: Practical Focus on applying code across research centers in such areas as data disambiguation (names, institutions, geolocations), linking WoS data to other datasets, building models to predict gender, ethnicity, etc.
- B. **Topical Hackathon(s)**: Working across research centers on Authorship & Collaboration; Gender in Science; Topic

modeling and/or other topics defined by attendee interest.

C. *Community Hackathon*: Focus on establishing an ongoing community (e.g. setting up an enclave, tools & mechanisms for sharing code, citing and acknowledging contributions, and/or what is appropriate for cross-enclave sharing).

12:00pm Working Lunch
3:00pm Group Discussion
Facilitators: Katy Börner, Eamon Duede, and Jim Pringle
Group Discussion: Lessons, and results. What kind of community should we be?
4:00pm Wrap-up

Photos:



[View the photo gallery »](#)

Invited Participants



Titipat Achakulvisut

PhD candidate, Biomedical Engineering
Northwestern University



YY Ahn

Assistant Professor
Indiana University



Yadu Babuji

Chief Infrastructure Wizard
Computation Institute
University of Chicago



Linge Bai

Senior Data Scientist
Clarivate Analytics



Brenda Betancourt

Postdoctoral researcher in Statistical Science
Duke University



Sebastien Brien

Chief Analytics Officer
Clarivate Analytics



Grégoire Côté

President
Science-Metrix



James Evans

Professor of Sociology
University of Chicago



Diego Fregolente

Postdoctoral Fellow
Northwestern Institute on Complex Systems [NICO]



Lee Giles

Professor
Pennsylvania State University



Junming Huang

Postdoctoral research associate
Northeastern University



Matt Hutchinson

Data Manager
Network Science Institute
Indiana University



Ted Lawless

Solutions Specialist
Clarivate Analytics



Robert Light

Senior Systems Analyst, Database Administrator
Cyberinfrastructure for Network Science (CNS) Center
Indiana University, Bloomington



Lu Liu

Graduate Student
College of Information, Science & Technology
Pennsylvania State University



Jared Lorince

Postdoctoral Fellow
Northwestern Institute on Complex Systems [NICO]



Patricia L. Mabry

Executive Director and Senior Research Scientist
Network Science Institute
Indiana University



Helge Marahrens

Graduate Student, Sociology Department
Indiana University, Bloomington



Fil Menczer

Professor of Informatics & Computer Science
Director of The Center for Complex Networks & Systems Research (CNetS)
Indiana University, Bloomington



Staša Milojević

Associate Professor of Informatics
Indiana University



Nicholas Monath

MS/PhD Student
University of Massachusetts, Amherst



Hao Peng

Graduate Student
Indiana University



Orion Penner

Postdoctoral Researcher
Chair of Innovation and IP Policy
EPFL



Valentin Pentchev

Director of Information Technology
Network Science Institute
Indiana University, Bloomington



Adam Ploszaj

Assistant professor, Centre for European Regional & Local Studies (EUROREG), University of Warsaw & Visiting



Assistant professor, Centre for European Regional & Local Studies ECRULES, University of Warsaw & visiting professor at Cyberinfrastructure for Network Science (CNS) Center Indiana University, Bloomington



Jason Rollins

Senior Director of Innovation
Clarivate Analytics



Matthew B. Ross

Post-Doc
Ohio State University



Johanna Salazar

Assistant Director of The Johnson Center for Innovation and Translational Research
College of Arts & Sciences
Indiana University, Bloomington



Christian Schulz

ETH Zurich



Ben Serrette

Web Developer
Network Science Institute
Indiana University, Bloomington



Aditya Tandon

Graduate Student
The Center for Complex Networks & Systems Research (CNetS)
Indiana University, Bloomington



Vetle I. Torvik

Assistant Professor
School of Information Sciences
University of Illinois at Urbana-Champaign



Brian Uzzi

Richard L. Thomas Professor of Leadership & Organizational Change
Kellogg School of Management
Northwestern University



Emmanuel Thiveaud

VP, Research Analytics
Clarivate Analytics



Dashun Wang

Associate Professor of Management & Organizations
Kellogg School of Management
Northwestern University



Bruce Weinberg

Professor of Economics
Ohio State University, IZA, & NBER



Ian Wesley-Smith

PhD Student
University of Washington



Jevin West

University of Washington



Xiaoran Yan

Research Scientist
Network Science Institute
Indiana University, Bloomington



Ewa Zegler-Poleska

Information Science PhD student
School of Informatics and Computing
Indiana University, Bloomington



Sifan Zhou

Postdoctoral Fellow
Harvard University

Interested but could not attend.



Patricia Brennan

VP, Technology & Platform Product Management
Clarivate Analytics



Fil Menczer

Professor of Informatics & Computer Science
Director of The Center for Complex Networks & Systems Research (CNetS)
Indiana University, Bloomington



Tim Slaper

Director of Economic Analysis



Director of Economic Analysis
Indiana Business Research Center
Kelley School of Business
Indiana University, Bloomington



Aditya Tandon

Graduate Student
The Center for Complex Networks & Systems Research (CNetS)
Indiana University, Bloomington

Acknowledgments:

This effort is supported by **Clarivate Analytics**, the **Knowledge Lab at the University of Chicago**, the **Indiana University Network Science Institute (IUNI)**, and the **Cyberinfrastructure for Network Science Center at Indiana University**.

About Us

Our Mission
Current Team
Current Students
Collaborators
Visiting Artists
Funding

Outreach

Places & Spaces Exhibit
Open House

Research

Publications
Visualizations
Workshops
Presentations
CNS Talks

Videos

Project Demos
Research Talks
Miscellaneous

Development

Products
Data & Tools
Interactive Displays

News and Events

Latest News
All News
Events Calendar

Teaching

IU Courses
IVMOOC
Tutorials

Connect With Us

Contact Us
Visitor Info
Support CNS
Job Opportunities
Newsletter
Facebook
Twitter Google+

Thank you to our generous sponsors



LUDDY
SCHOOL OF INFORMATICS
COMPUTING, AND ENGINEERING

