

**NITRD Big Data Interagency Virtual Workshop on  
Pioneering the Future of Federally Supported Data Repositories**  
**Wednesday Jan 13th - Friday Jan 15th, 2021**  
<https://www.ornl.gov/bdworkshop21>

**Agenda**

**Jan 13<sup>th</sup> Day 1: Sharing Ideas**

1:00 - 1:35	Intro: Welcome and Purpose of the Meeting  Chaitan Baru (NSF) (co-chair) Kamie Roberts (NITRD NCO) Jerry Sheehan (NSTC Subcommittee on Open Science) Laura Biven (NIH) (co-chair)
1:35 - 2:35	Session 1: Invited Presentations Phil Bourne (University of Virginia): A SWOT Analysis of Federally Funded Resources – What Does it Tell Us? Rick Stevens (Argonne National Laboratory): Thoughts on a theory of data Michael Beauvais (McGill University): The Future of Federally Funded Data Institutions: Governance as Our North Star
2:35 - 2:45	Break
2:45 - 3:25	Session 2: Lightning Talks Georgia Tourassi (ORNL): Enabling data-driven science at the intersection of domains Margaret Levenstein (U Mich): COA3D: a new infrastructure to support diverse types of data over the research lifecycle Emiley Eloie-Fadrosch (LBNL): Towards a data science ecosystem for microbiome research Kerstin Lehnert (Columbia U): Sharing Infrastructure among Domain Repositories III: A Vision and Way Forward of the Council of Data Facilities Merce Crosas (Harvard): Dataverse: An open-source platform and community for building data repositories
3:25 - 3:35	Break
3:35 - 4:15	Session 3: Lightning Talks (continued) Stephen Burley (Rutgers): Cloud Data Storage and Computing Networks Enabling Federally Supported Data Resources Alexei Pevtsov (NSO): Digitization and preservation of historical astronomy data Noah Tobin (Georgia Tech): Current and Future Challenges of Cybersecurity Data for Machine Learning Rick Gilmore (Penn State): The human behaviorome project Monica Youngman (NOAA): From Sun to Mud: Evolving to Meet NOAA's Diverse Repository Needs

AGENDA (Times shown are ET)

## **NITRD Big Data Interagency Virtual Workshop on Pioneering the Future of Federally Supported Data Repositories**

<https://www.ora.gov/bdworkshop21>

	Jeff De La Beaujardiere (NCAR): Geoscience Data Exchange (GDEX)
4:15 - 4:55	Session 4: Federal Panel Jerry Sheehan (NIH/NLM): Desirable Characteristics of Repositories for Managing and Sharing Data Resulting From Federally Funded Research Kim Pruitt (NIH/NCBI): NIH Metrics and Lifecycle Work Group Michael Cooke (DOE): Public Reusable (PuRe) Data Resources Nathaniel Bastian (DOD): JAIC Data Management Platform for AI Development Bob Hanisch (NIST): NIST Research Data Management Framework (RDaF)
4:55 - 5:00	Recap of Day 1 and Prompt for Day 2

**\*\* Breakout sessions for invited participants only**

**AGENDA (Times shown are ET)**

## NITRD Big Data Interagency Virtual Workshop on Pioneering the Future of Federally Supported Data Repositories

<https://www.ora.gov/bdworkshop21>

### Jan 14<sup>th</sup> Day 2: Repositories of the Future

1:00 - 1:30	Intro: Overview and Goals for Day 2 Julia Lane (New York University)
1:30 - 2:30	Session 5: Invited Presentations Rob Knight (University of California San Diego): AI Needs to Handle Expanding Microbiome and Metabolome Data Resources--Especially for Global Pandemics Greg Beroza (Stanford University): STEAD: A Database for AI in Seismology RV Guha (Google): Data Commons
2:30 - 2:45	Break
2:45 - 4:00	Session 6: Breakout sessions** Breakout 1: Data and AI (Rick Stevens, Argonne National Laboratory) Breakout 2: Privacy and Ethics (Jules Polonetsky, Future of Privacy Forum) Breakout 3: Heterogeneous data, Federated data, and Interoperability (Kjiersten Fagnan, Lawrence Berkeley National Laboratory) Breakout 4: Characteristics of Repositories (Jerry Sheehan, NIH/NLM) Breakout 5: Dynamic data and real-time workflow (Ian Foster, Argonne National Laboratory and University of Chicago)
4:00 - 4:15	Break
4:15 - 4:45	Session 7: Recap Breakout Sessions
4:45 - 5:00	Recap of Day 2 and Prompt for Day 3

\*\* Breakout sessions for invited participants only

AGENDA (Times shown are ET)

## NITRD Big Data Interagency Virtual Workshop on Pioneering the Future of Federally Supported Data Repositories

<https://www.ornl.gov/bdworkshop21>

### Jan 15<sup>th</sup> Day 3: Learning and Working Together

1:00 - 1:30	Intro: Overview and Goals for Day 3 Deb Agarwal (Lawrence Berkeley National Laboratory)
1:30 - 2:30	Session 8: Invited Presentations Adam Rondinone (Los Alamos National Laboratory): The Society for Science at User Research Facilities <i>and</i> the Center for Integrated Nanotechnologies David Spirk (Department of Defense): DOD's Data Driven Transformation Barend Mons (University of Leiden): "Sorry but no one wanted to pay for the rocket launcher"
2:30 - 2:45	Break
2:45 – 4:00	Session 9: Breakout Sessions** Breakout 1: Community Building (Michael Cooke, Department of Energy) Breakout 2: Workforce development and metrics (Gerhard Klimeck, Purdue University) Breakout 3: Interoperability and training (Christine Kirkpatrick, UC San Diego) Breakout 4: Data Management Ecosystem (Bob Hanisch, NIST) Breakout 5: Interfacing with non-federal stakeholders (Shelley Stall, American Geophysical Union)
4:00 - 4:15	Break
4:15 - 4:45	Session 10: Recap Breakout Sessions
4:45 - 5:00	Closing Remarks
5:00	Workshop Adjourns

\*\* Breakout sessions for invited participants only

AGENDA (Times shown are ET)