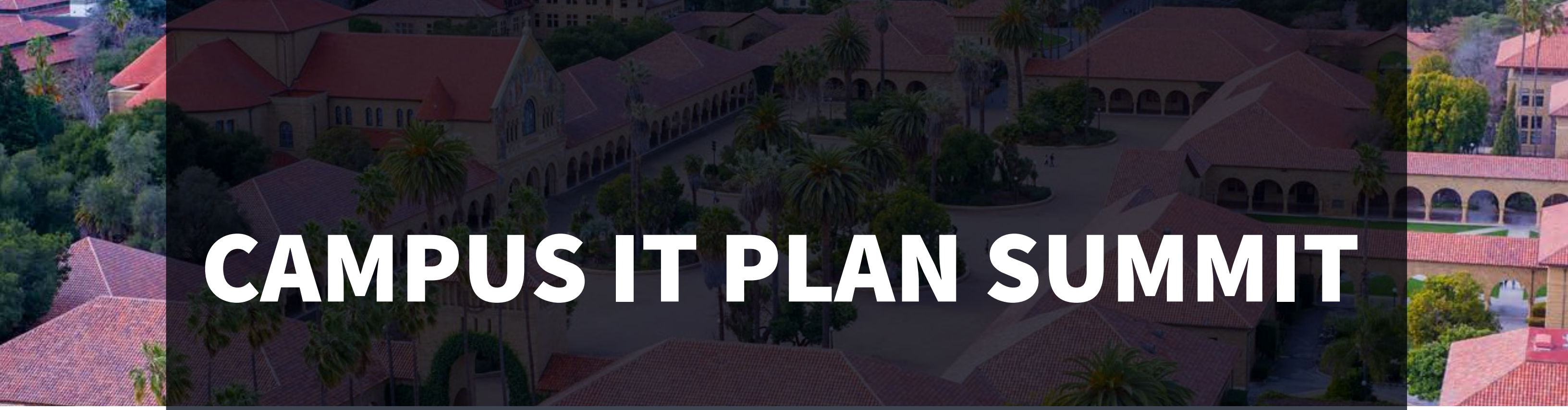

CAMPUS IT PLAN SUMMIT

A CAMPUS-WIDE VIEW OF RESEARCH DATA

NOVEMBER 2-3, 2022

Keynotes

DAY 1

Using Big Data to Solve New Global Health Challenges in Real-Time

Catherine Blish, MD, George E. and Lucy Becker Professor in Medicine, Stanford Medicine

Pandemic responses require rapid acquisition and dissemination of data. This session will cover how Stanford Medicine leverages datasets around the world, and how this enables the research enterprise.

Planning a Data Infrastructure for a New School

John Freshwaters, Chief Information Officer (CIO), Stanford Doerr School of Sustainability

With the opening of the new Stanford Doerr School of Sustainability, the technology strategy that got us here won't necessarily get us there. In this session, learn about the approach the Stanford Doerr School of Sustainability is taking to create a new IT strategy that aligns with the future needs of the new school, in collaboration with our community to support this new era of Sustainability both at Stanford and beyond.

NIH FINAL Data Sharing and Management Policy – What You Need to Know

John Borghi, Manager, Research and Instruction, Stanford School of Medicine
Scott Edmiston, Director of Research Data Governance and Privacy
Rochelle Lundy, Director, Office of Scholarly Communications, Stanford Libraries

A panel discussion on the NIH FINAL Data Sharing and Management requirements (effective 1/25/23) and other new federal data sharing initiatives such as the White House OSTP memo directing federal funders to make data publicly available by 2026. This session will summarize changes and invite participants into a lively discussion on the implications for research data at Stanford throughout the research lifecycle.

KEYNOTES

GSB Social Science Research Data and Faculty Interests

Matt Marostica, Director, Research Data Services & Collections Library, Stanford Graduate School of Business (GSB)
Brian Chivers, Senior Data Engineer, Stanford GSB
Todd Hines, Research & Discovery Manager, Research Subject Librarian, GSB Research Hub
Matt Hutchinson, Data Curation Specialist, GSB Library
Sean Kaneshiro, Manager, Collection Development and Research Discovery, GSB Library

Supporting social science research requires a multi-disciplinary team. We have found that supporting faculty research interests includes -- identifying data sources; negotiating data contracts; working with Stanford contracting offices; and bringing the data to an appropriate campus repository for data-cleaning, analysis and, ultimately, deposit in support of publication and future research. This process requires expert librarians with broad subject matter expertise, data contracts experts who understand the research process in order to represent faculty interests, data scientists to clean and store the data and to make it available to as many researchers as possible. At the GSB, most research data is proprietary which imposes additional complications. We propose a panel discussion led by GSB Research Hub staff who will demonstrate the deep interconnectedness of social science research data support.

DAY 2

Through the Lens of Stanford Professor and Researcher: Research Data Perspectives

Michal Kosinski, Associate Professor in Organizational Behavior, Stanford University Graduate School of Business

A Stanford Professor's experiences with the IRB, commentary on the issues of privacy, anonymisation vs. de-identification, and machine learning in research data.

Building Support for Research Data on Campus

Peter Leonard, Assistant University Librarian for Research Data Services, Stanford Libraries
Ruth Marinshaw, Chief Technology Officer, Research Computing, Stanford University
Julie Williamsen, Assistant Dean and Executive Director of the Research Hub, and Director of the Stanford GSB Library

The support ecosystem for research data on campus includes multiple institutions operating at the intersection of data, methodological expertise, and computation. In this presentation, representatives from Stanford's Libraries, Research Computing Center, and Graduate School of Business Research Hub will discuss services and offerings in many stages in the research data lifecycle.

KEYNOTES

Stanford Data Farm Panel Discussion

Vijoy Abraham, Assistant Director and Head, Center for Interdisciplinary Digital Research, Stanford Libraries
Kate Barron, Research Data Curator, Stanford Libraries
Brian Chivers, Research Analytics Scientist, Stanford GSB
Bella Chu, Associate Director, Data Core, Center For Population Health Sciences
Ian Matthews, Co-Founder and Chief Executive Officer, Redivis Inc.

Panel Discussion on the Stanford Data Farm, a SaaS product from Redivis Inc. that provides access, exploration and analysis of tabular and non-tabular datasets. Panel members include staff from GSB, Population Health Sciences (SoM), Stanford Libraries, and Redivis Inc.

Demystifying the Data Risk Assessment (DRA)

Amy Steagall, Chief Information Security Officer, Stanford University IT
Shawn Kim, Acting Director of ISO Consulting, Stanford University IT
Tad Perillo, Information Security Officer, Stanford University IT

During this keynote panel session, the Information Security Office will discuss the current process of Data Risk Assessment and the future plans to improve it.

Breakout Sessions

DAY 1

National Security and Research Data – A Window Into Stanford Research Policy and Integrity (RPI) Services

Data Ethics and Compliance Track

Scott Edmiston, Director of Research Data Governance and Privacy
Ronda Anderson, Director of Conflict of Interest and Conflict of Commitment
Jessa Albertson, Global Engagement Review Director

A panel discussion on the nexus between national security and research data reflected guidance such as NSPM 33. The topic will be used to introduce services offered by the Research Policy and Integrity program of the Vice Provost and Dean of Research (RPI) and how it works across Stanford offices to address conflict of interest, export controls, privacy and more. Participants will learn about the services and be invited to discuss collaborative models and strategies.

What Policy We Need to Support Web Archiving and Using Archived Websites as Data

Data Ethics and Compliance Track

Peter Chan, Web Archivist, Stanford Libraries

The library developed a policy to archive web sites, preserve the files at Stanford and provide access from Stanford controlled servers. The policy is based on the assumption that the archived websites are viewed individually by human researchers. When researchers archive websites with no intention to provide access to individual sites, should the same policy apply?

Persistent Identifiers for Data Citation and Beyond – Practical Tips and the Road Ahead

Data Sharing and Collaboration Track

Zach Chandler, Director Of Open Scholarship Strategy, Stanford Data Science Initiative

The modern research data ecosystem increasingly relies on persistent identifiers (PIDs) as a foundational technology for scholarly citation. Zach Chandler, Director of Open Scholarship Strategy will present practical steps researchers can take to apply PIDs when preparing research materials for publication, that ensures that datasets and open source code are cited properly, and their reuse is reflected in the scholarly record. He will show how PIDs can save time and effort, ensure compliance, and enhance findability and integrity.

BREAKOUT SESSIONS

Lattice: Extending the Lifecycle for Single Cell Data

Data Sharing and Collaboration Track

Jason Hilton, Lattice Project Lead & Data Wrangler, Cherry Lab, Genetics Department

The Lattice data coordination project supports the sharing of single cell genomics and imaging data as part of the Human Cell Atlas. We maintain a private database with a rich data model and metadata standards. The Lattice DB validates metadata and data and exports data to community resources once the researchers are ready to publish, ensuring that data are shared in a meaningful way towards reuse (lattice-data.org).

Machine Intelligence, Research Reporting and Stakeholder Involvement, A Pragmatic Approach

Data Sharing and Collaboration Track

Amy Price, Senior Research Scientist And Bmj Editor, Working In Cross Disciplinary Engagement And Research Methods

Machine intelligence relies on algorithms that reason about observed data to make predictions or decisions about our healthcare. To expand empathy and implementation for research and healthcare we need to involve and co-produce analysis and solutions with end-users or stakeholders and include their preferences and lived expertise in deep learning models. Currently, deep learning algorithms are applied in interactive settings for gaming and in decision making tasks, where model predictions have consequences on future inputs. This can be compared to observing a hamster on a treadmill, we learn about the task limitations but not about the hamster. These models learn by the quality and content of what they are exposed to. We will share ways machine learning models change healthcare and we will report on advances for empathy and implementation in the US and the UK. Finally we will share co-production can fit within the existing workflow. Great machine learning works at an ever-increasing scale in computation while retaining flexibility to develop new models.

The Development of a Searchable Metadata Platform for Research Data Assets

Data Sharing and Collaboration Track

Matt Hutchinson, Data Curation Specialist, GSB Library

Metadata about research datasets at the GSB are stored in multiple use-specific systems. The data stored in each platform is only the data useful for that user or that team that maintains the system. The GSB library is engaged in developing a web platform to unify records from multiple sources into a single index to track all dimensions of research metadata. This presentation would address the challenges faced by library staff, the progress made so far and ideas for future development.

BREAKOUT SESSIONS

STAnford Medicine Research data Repository (STARR): An Overview

Stanford Services Track

Priya Desai, Manager, Biomedical Informatics R&D

Clinical Informatics is at the confluence of rapid advancements in cloud computing, new techniques in data mining and ML, and more than a decade worth of electronic medical records data! Stanford has responded to this opportunity by creating STAnford medicine Research data Repository (or STARR, starr.stanford.edu), a single integrated data lake containing clinical data of different modalities from the two Hospitals.

Providing Digital Object Identifier (DOI) Services at Stanford

Stanford Services Track

Amy Hodge, Science Data Librarian, Stanford University Science And Engineering Resource Group
Hannah Frost, Associate Director, Digital Library Systems and Services, Stanford Libraries

Stanford Libraries' membership in DataCite enables us to provide DOI (Digital Object Identifier) services for the Stanford community. The need for DOIs is increasingly common for today's scholars who are actively publishing. In this presentation, we'll discuss the service options available for obtaining a DOI. The easiest way to get a DOI for a single work is to deposit that work into the Stanford Digital Repository using our online, self-deposit application., which automatically mints DOI for deposited works. Other options for obtaining a DOI include requests for one-off singleton and bulk DOI minting, as well as subscribing to DOI services that allow for access to the DataCite API for integration with systems at your project, group, or research center. We'll also talk about the latest addition to DataCite DOI services -- the ability to create IGSN IDs for physical research samples.

Where Is Your Data? Spatial Data Collections and Augmentation Services Provided by the Research Data Services Division of Stanford Libraries

Stanford Services Track

Stace Maples, Assistant Director of Geospatial Collections & Services, Stanford Geospatial Center & Branner Library Map Collections

Satellite imagery, Satellite tasking, Geocoding, Network Analysis, Spatial Data Discovery and more....

BREAKOUT SESSIONS

Data Studio

Stanford Services Track

John S. Tamaresis, Biostatistician, Biomedical Data Science

The Data Studio is a collaboration between the Stanford Center for Clinical and Translational Research and Education (SPECTRUM) and the Department of Biomedical Data Science (BDS). Data Studio features BDS faculty and staff who offer the following statistical/data science consulting services: workshops, office hours, and one-to-one consultations. These services are open to the Stanford community engaged in biomedical research. We also offer Data Studio as a class (BIODS 232) because it has educational value for students and postdocs interested in biomedical data science.

DAY 2

Cloud Object Storage for Researchers: Options and Considerations to Make to Avoid Expensive Surprises

Data Ethics and Compliance Track

Noah Abrahamson, Director of Cloud Security, Stanford Information Security Office Operation

In this breakout session, you’ll learn more about cloud object storage for researchers. You’ll learn about terms like “availability” and “reliability”, and how cloud service providers calculate monthly charges. You’ll learn about which options to select under different conditions and use cases. This session will also cover UIT’s negotiated discounts and enterprise agreements. By attending this session, researchers will be armed with enough knowledge to help avoid a potentially costly surprise.

REDCap: Useful Methods and Techniques

Data Ethics and Compliance Track

Dinara Bogetic, Data Manager, Sean N. Parker Center For Allergy And Asthma Research - Clinic

REDCap (Research Electronic Data Capture) is a browser-based HIPAA-compliant application for designing and managing clinical and translational surveys and databases. During this session, we will be covering strategic and effective approaches to support research coordinators and data managers with study design and data collection using REDCap.

BREAKOUT SESSIONS

ClinicalTrials.gov Results - Planning for Success

Data Sharing and Collaboration Track

Scott Patton, Clinical Trials Manager, Clinical Research Quality (CRQ)

Certain data elements needed for ClinicalTrials.gov results reporting are consistently missed in the sources that are provided for the project, often analyses performed with a manuscript in mind. This means that additional analyses are needed from the statistician, and the clinical team is impacted as they manage the results project to completion. Understanding what is needed and developing a plan before the study begins can bring attention to strengths and weaknesses in study preparations and help make the ClinicalTrials.gov results submission more useful and meaningful.

Research Methodology and Analysis of Administrative Databases in Medicine and Surgery

Data Sharing and Collaboration Track

Lakshika Tennakoon, MD, MSc, Data Scientist Department of Surgery, Trauma and Acute Care, Stanford University

This presentation is based on Healthcare Cost and Utilization Project (HCUP) administrative databases in research. Understanding the nature, infrastructure, and cataloging variables and outcomes, national representation, interpretations, and implications. Justification of usage of HCUP databases as a platform for common and rare diseases in the domains of Medicine and Surgery. This will Include dealing with big data-related strategies, challenges, strengths, and weaknesses.

AI for Genetic Discovery

Data Sharing and Collaboration Track

Gary Peltz, Professor of Anesthesiology, Perioperative and Pain Medicine

No model organism has contributed more than the laboratory mouse to improving human health. Many genetic factors were initially discovered or characterized in mice, and many therapies for human diseases were tested in mice before they were used in patients. Although AI has been integrated into human healthcare, very few AI advances have been used to analyze the data produced using the model organism that has been the foundation for many healthcare innovations. I describe an innovative AI-based computational pipeline that could identify causative genetic factors for murine genetic models of biomedical traits and human diseases.

BREAKOUT SESSIONS

The Stanford DMP Service

Stanford Services Track

John Borghi, Manager, Research And Instruction, School Of Medicine - Lane Medical Library

Requirements from federal funding agencies to create and implement data management plans will soon affect a broad set of Stanford affiliated researchers. These plans, often called DMPs, are short documents in which researchers prospectively describe what data they plan to generate over the course of a given research effort and how that data will be managed, preserved, and ultimately made available to others. In this presentation, we will discuss how the new campus DMP service and the DMPTool platform will assist researchers in creating data effective plans and connecting plans with related research outputs, tools, institutions, and individuals for improved information management, reporting, and tracking.

How the Stanford Digital Repository (SDR) Supports Research Data Services at Stanford

Stanford Services Track

Amy Hodge, Science Data Librarian, Stanford University Science And Engineering Resource Group
Hannah Frost, Associate Director, Digital Library Systems and Services, Stanford Libraries

In operation at Stanford Libraries for over a decade, SDR is used by scholars across Stanford to archive and share research data. It is an increasingly important component of Stanford's larger research information ecosystem by fulfilling requirements of Stanford's open access policy as well as those set forth by federal research funding agencies. The SDR is in a period of especially active development and future planning. Session attendees will come away with a concrete understanding of SDR services and the role and benefits of SDR at Stanford in support of the research enterprise.

Teaching Basic Research Data And Computational Skills Through Live Coding: The Carpentries at Stanford Libraries

Stanford Services Track

Zac Painter, Interim Head, Terman Engineering Library

Novice learners of programming or coding often find sitting in front of a computer to be overwhelming if they have to follow instructions without guidance. Live coding is a teaching method of sharing information about how to perform tasks around research computing, in an active and participatory way. In this session, attendees will learn about this instructional method and how it can be applied in different settings.