Magic Meeting
May 6, 2020

Hosting and Working with Sensitive Data on High Performance Computing (HPC) environments

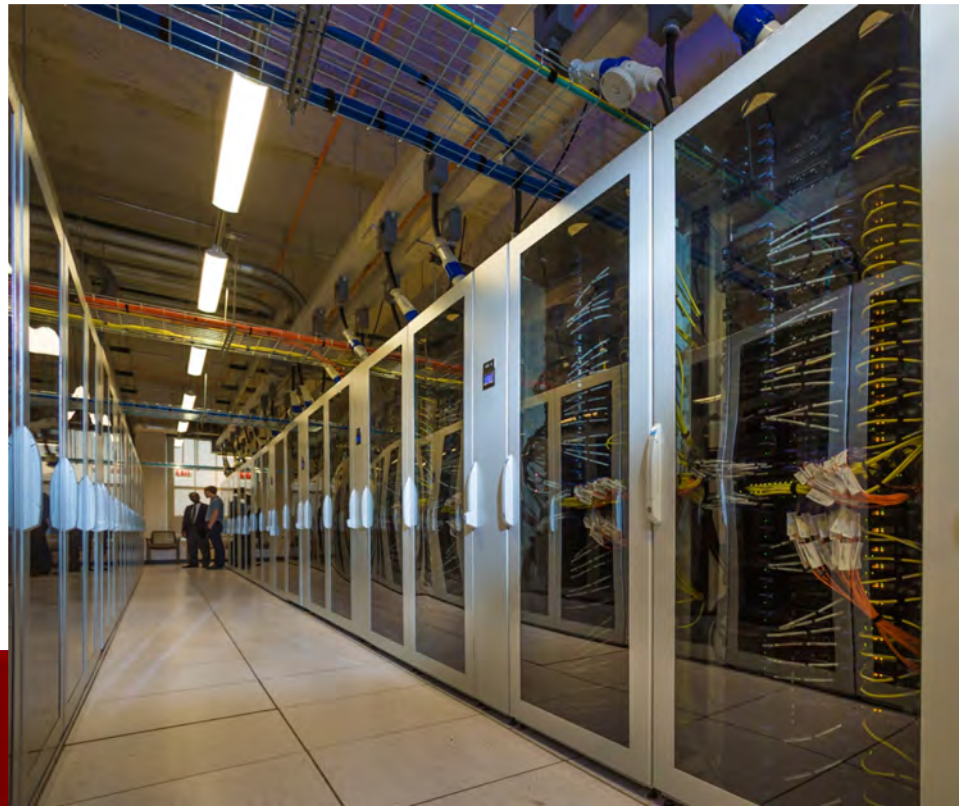
Hakizumwami Birali Runesha
runesha@uchicago.edu

The UChicago Research Computing Center (RCC)

A Unit in the Office of Research and National Laboratories

Enabling research and scholarship by providing access to **hardware** (high-end computing, storage and visualization resources), **software**, **training programs** and **Computational Scientists**

- High Performance Computing Center
- > 1,500 nodes - tightly coupled InfiniBand interconnected
- > 7 PB of Storage
- More than 600 Faculty
- More than 4,000 users



More info at rcc.uchicago.edu

Researchers today are increasingly working with data requiring compliance with a variety of data security and privacy standards

What did we observe and try to address?

- lack of standardized security, privacy, and operating procedures when receiving data sets provided by third parties, which may contain sensitive information.
- Increasing cyber security and data privacy requirements for human subjects research data collected by University researchers
- The deluge of data
- Availability of AI tools that question data privacy.
- increased number of data security breaches at peer institutions and associated risks in the event of a data breach involving research data sets.
- etc.

=> Need to support sensitive data at the RCC

A Variety of Regulations and Standards

ITAR

FISMA

DFAR

HIPAA

CSA

NIST 800-171

Illinois Statutes

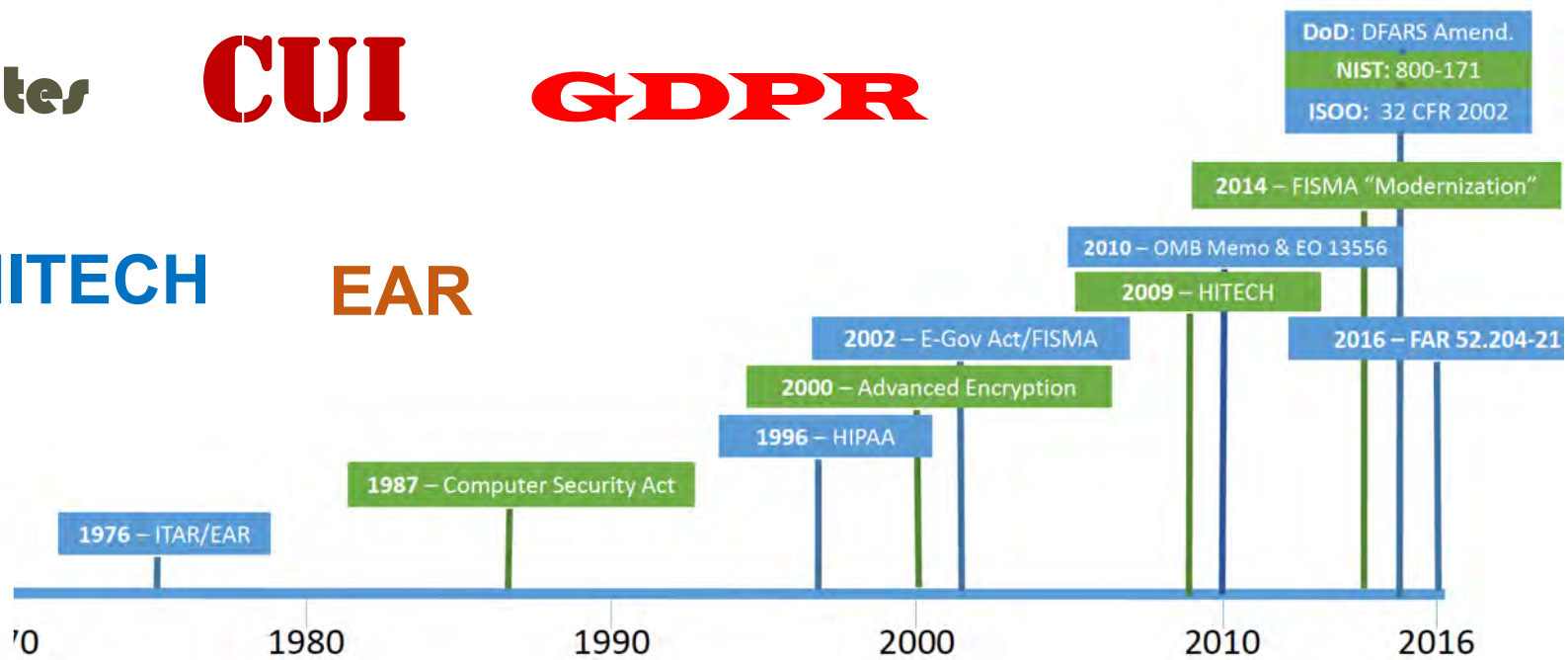
CUI

GDPR

HITECH

EAR

FAR



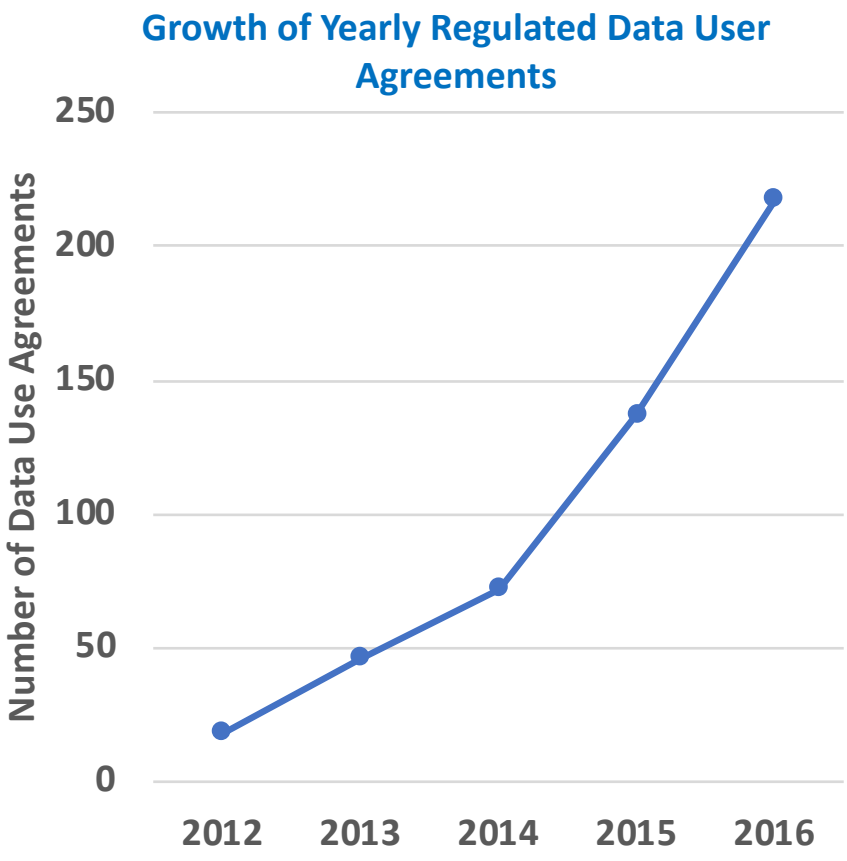
Source: Deumens

Data Privacy Considerations

- Ensure policies are in place to collect, share and use information appropriately
- Use and governance of personal data
- Align with evolving laws, regulations, and industry standards in the use
- Data privacy is not simply ensuring compliance with laws and regulations, it is also fostering ethical behavior and ensuring that data is managed and shared according to the values of the organization.
- 3 Aspects of privacy
  - **Legal and regulatory**
  - **Ethical**
  - **Value**

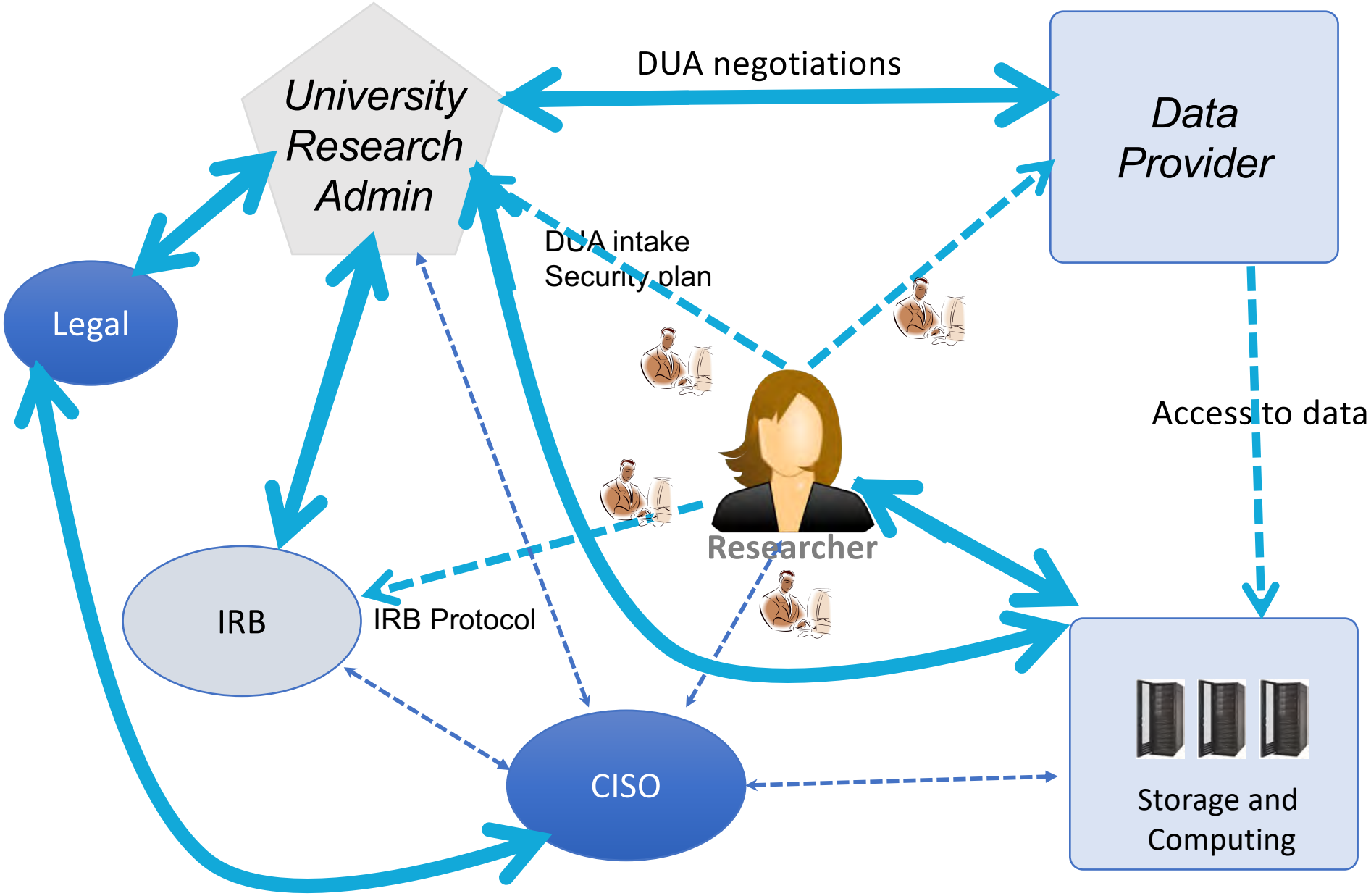
Source of Sensitive Data

- **Data providers** outside the University
  - Require a Data Use Agreement (DUA)
  - May require an IRB protocol
- **Generated** on campus
  - may require an IRB protocol
- **Procured**



IRB: Institutional Review Board

What happens in many cases …



Secure Research Data Strategy
srds.uchicago.edu

Secure Research Data Strategy (SRDS)

A partnership between the **Research Computing Center** (RCC), **IT Services** (ITS), **Office of Legal Counsel** and **University Research Administration** (URA) was established to provide secure computing environments for The University of Chicago researchers in order to access, store and analyze **sensitive** research data.

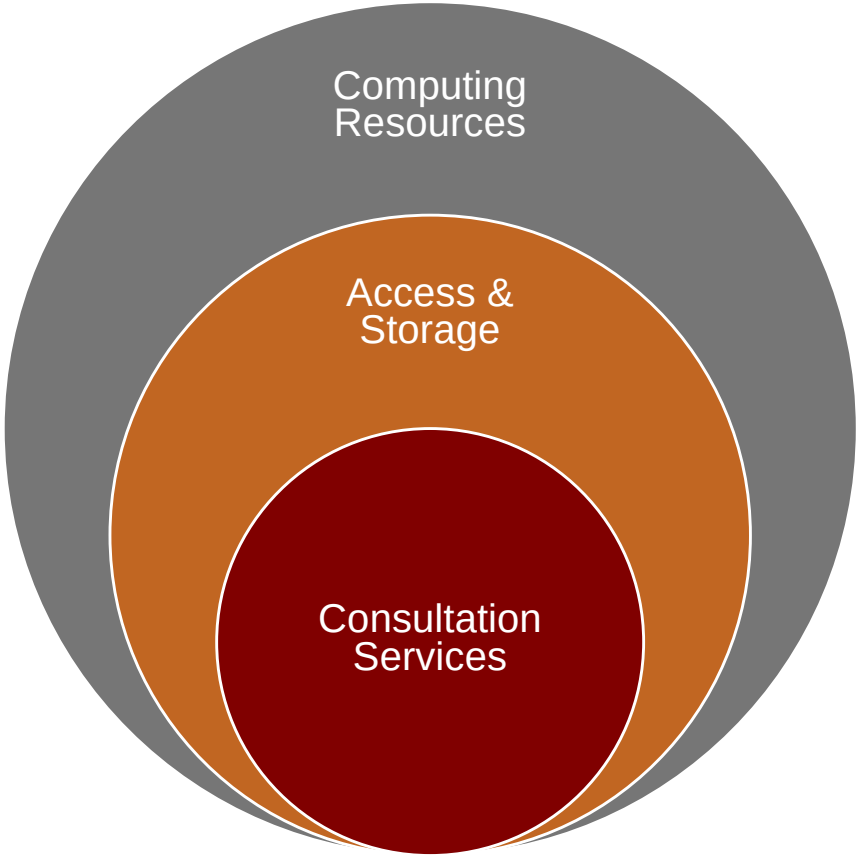
Strategy and Approach

- Developed the University Edition Cyber Security and Data Privacy policies
- Created the **Secure Data Enclave** with High Performance Computing capability to provide secure storage and processing
- Created a Workflow to bring risk assessment and consultation to faculty when they need it (including certification of processing environments PIs may wish to use)
- Created a Governance to manage the strategy over time and ensure consistent application across Schools, Institutes, and Divisions
- Create a process workflow and operating procedures
- Create the concept of an SRDS approved environment

The Secure Data Enclave (SDE)

The SDE is a secure, **centralized service** within the SRDS for faculty and researchers that work with sensitive research data that will include:

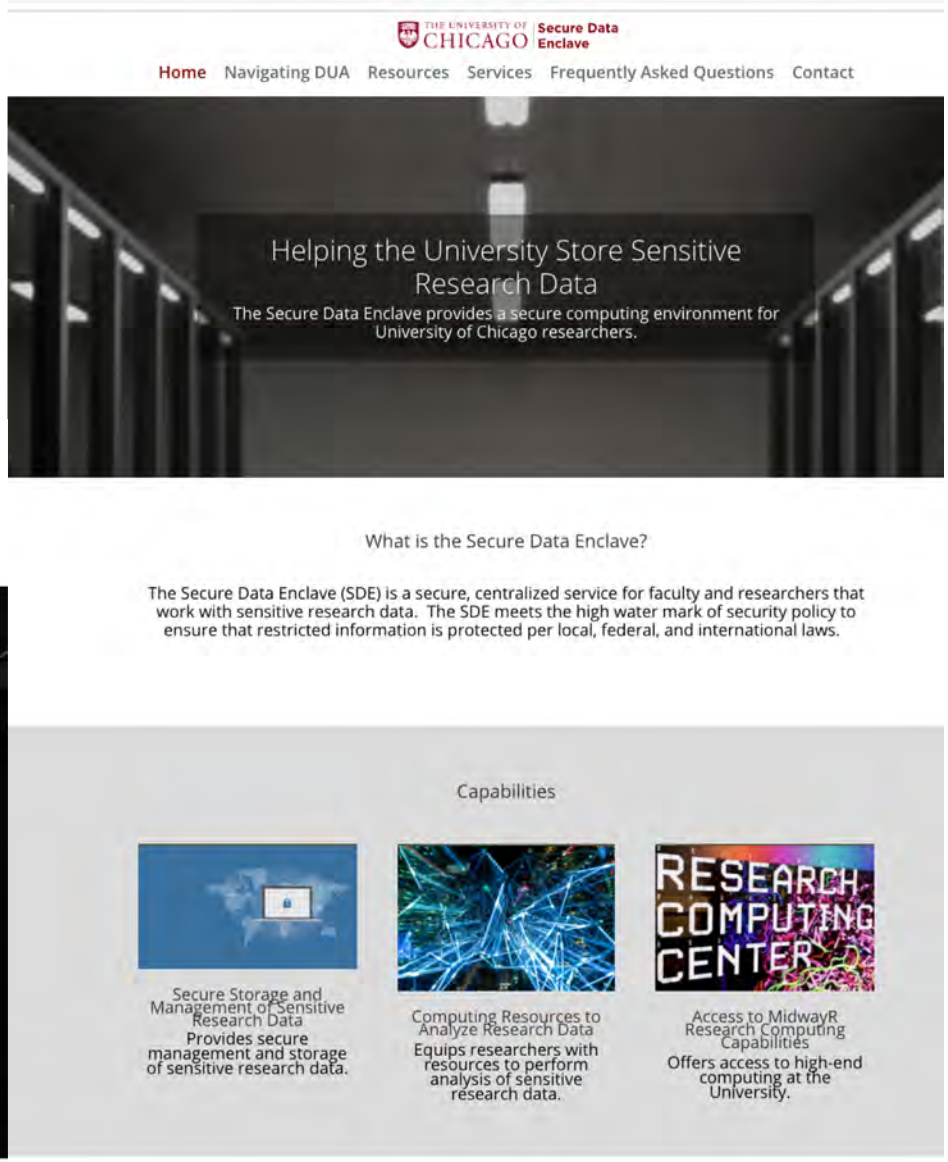
- Access and storage of sensitive data
- Computing resources to analyze, manage and report on sensitive research data
- Consultation services for understanding the obligations of handling sensitive data
- Consultation services for meeting sensitive data compliance



The SDE

MidwayR is an HPC system within the SDE

securedata.uchicago.edu



SDE Resources

MidwayR

- An HPC cluster that provides access to a secure platform environment for workloads requiring high-end computing or high performance storage

Commodity-Virtual Machines

- Provides access to a secure environment for workloads requiring lower computing and storage needs (including virtual machines, windows env.)

Secure Rooms

- Off the network

RCC HPC Ecosystem

Protection Level

Not restricted

Open data

Sensitive

MidwayR1

Low

MidwayR2

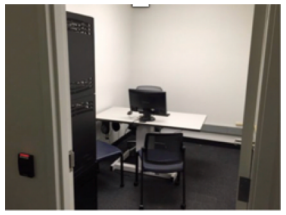
Moderate

MidwayR3

VM

High

Sensitive



Secure room off network

Controls …

- Access control: firewall blocking access , 2FA, etc
- Media control and encryption
- Change management
- System monitoring
- Log auditing
- Incident response
- Data ingestion

Looking ahead…

- Need for management systems and collaborative tools
- Need to develop a vocabulary for data obligations
- Need for common interpretation of security standards
- Need to develop standard operating procedures and workflows
- Need for a reference architecture and replicable processes
- Need for training

Acknowledgement

- Office of Legal Counsel
- ITServices
- University Research Administration
- Office of Research and National Laboratory
- The University of Chicago Research Computing Center

Thank you

"Any opinions, findings, conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the Networking and Information Technology Research and Development Program."

The Networking and Information Technology Research and Development (NITRD) Program

Mailing Address: NCO/NITRD, 2415 Eisenhower Avenue, Alexandria, VA 22314

Physical Address: 490 L'Enfant Plaza SW, Suite 8001, Washington, DC 20024, USA Tel: 202-459-9674, Fax: 202-459-9673, Email: nco@nitrd.gov, Website: https://www.nitrd.gov