Creating a trusted and secure research environment

Session 6 | Great Room 3 | 2.00-3.00PM

Dr Ewa Deelman

Research Director at the University of Southern California's Information Sciences Institute (ISI)

Dr Mohammad Nasser Eddine

Vice President of Programs and Planning for the Canada Foundation for Innovation

Professor Phil Quinlan

Professor of Digital Engineering and Health Informatics at the University of Nottingham

Michael Arentoft

Head of Unit, Open Science, Directorate-General Research and Innovation at the European Commission

Dr Ashley Vanderley

Senior Advisor for Facilities in the Division for Astronomical Sciences at the National Science Foundation (NSF)

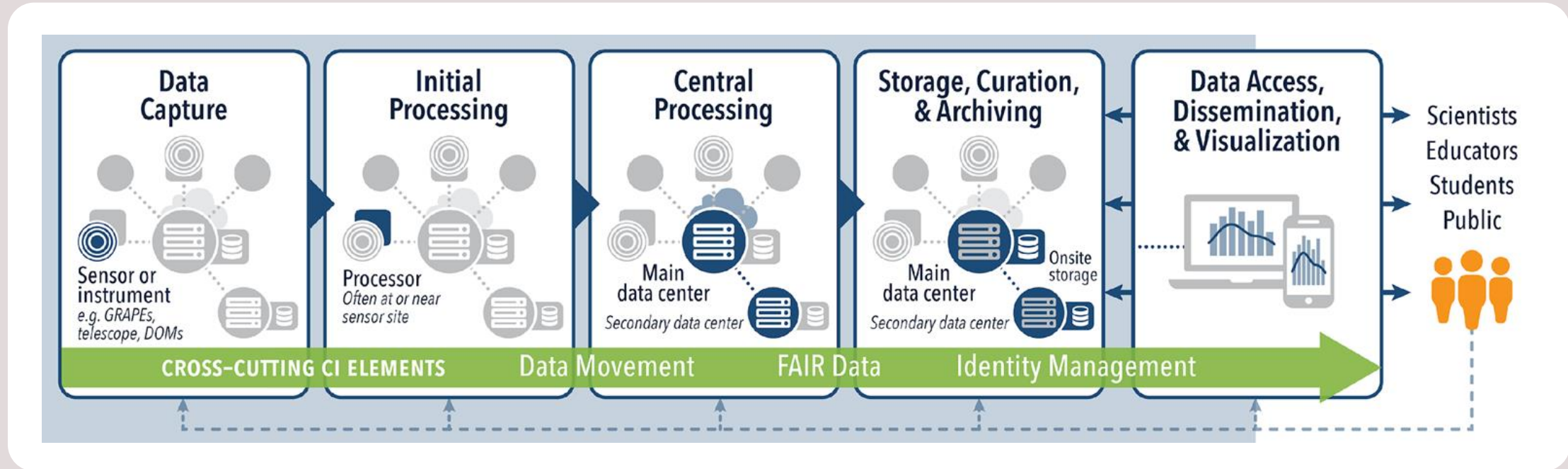# Creating a Trusted and Secure Research Environment

Ewa Deelman

University of Southern California
Information Sciences Institute

International Conference on Research Infrastructures

Brisbane Australia
December 3, 2024

CI Compass, The U.S. NSF Cyberinfrastructure Center of Excellence

CI Compass services focus on providing expertise and support to accelerate the research infrastructures' data lifecycle

Data in NSF Major Research Infrastructures

LIGO

Collective data

- Data is collected, calibrated, and analyzed by the collaboration
- The collaboration has all the details of the data collection and can influence how and what data is being collected
- Data is first disseminated within the collaboration for the purpose of research and after a period of embargo more broadly.
- Collaboration scientists are very familiar with the data
- Examples: high-energy physics, neutrino science, gravitational-wave physics

IceCube

Although very complex, full, detailed data and metadata have a single steward

Community data:

- Data collected by an RI for the purpose of generating data collections

- The RI does not directly conduct research on the data

- Data is used by scientists and the public who
  - Have limited influence on data collection or initial data analysis
  - Generally, have no insights into the data collection and analysis mechanisms beyond what is made publicly available


EarthScope

NEON
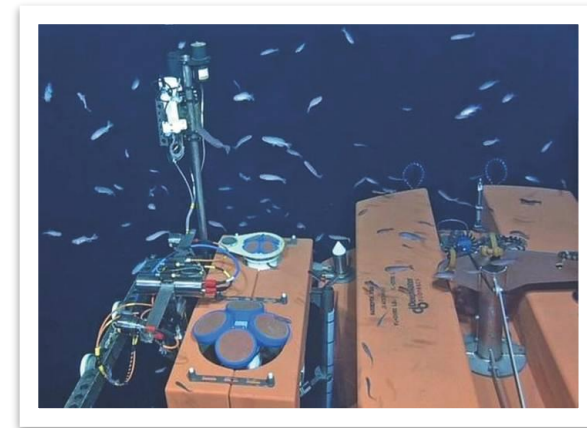
OOI

Although very complex, full, detailed data and metadata have a single steward

Principal Investigator (PI) data

- Data collected primarily by the PI and her/his team using the RI, and potentially additional instruments

- Data collected usually with a specific goal in mind (some information may not be captured or preserved) and for a specific domain

- Data products depend on information about the instrument

- The RI has no ownership of the PI data, or much influence over the data
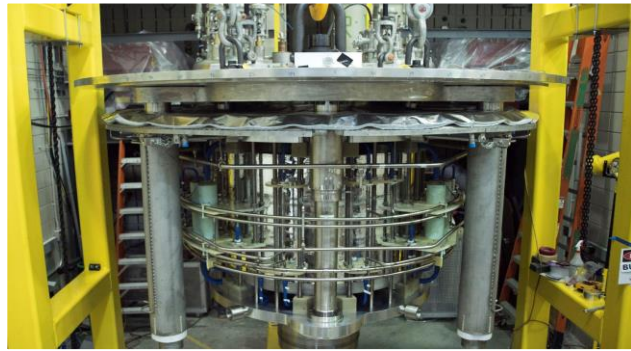
NHMFL, MagLab

NHERI Shake Table

CHESS

Metadata and data have multiple stewards and are distributed: sample preparer, experimenter, RI expert

PIs often do not have sufficient CI knowledge, tools, resources for dissemination

Some data is never meant to be shared in its raw form

SANE: secure environment for analysing sensitive data

SURF is the SURF is the ICT cooperative of Dutch education and research institutions. cooperative of Dutch education and research institutions.

https://www.youtube.com/watch?v=qxuA6GOc5Do

Attributes of Trustworthy Data

- **Availability**: accessible to authorized users as per defined policies.
- **Integrity**: data remains unaltered and free from corruption.
- **Authenticity**: clear provenance and verifiable origins.
- **Reproducibility**: enables replication of scientific results.
- **Confidentiality**: sensitive information is protected.
- **Credibility**: managed by trusted, authoritative sources.

Stakeholders and Their Roles

- **Data Users**: researchers, educators, the public
- **Data Providers**: entities responsible for ensuring data integrity, authenticity, and access.
- **Infrastructure Providers**: ensure secure storage, transmission, and processing.
- **Facilitation and Compliance Professionals**: ensure ethical use, regulatory compliance, and policy guidance.
- **Funders and policy makers**: influence data management practices by setting requirements for data sharing, security, and open science.

How do we create a trusted and secure research environment?

Michael Arentoft
Head of Unit of Open Science, in the Directorate-General Research and Innovation (R&I) at the European Commission

Dr Mohamad Nasser Eddine,
Vice President of Programs and Planning for the Canada Foundation for Innovation

Professor Phil Quinlan
Professor of Digital Engineering and Health Informatics at the University of Nottingham and an honorary director of Digital Health Research at Nottingham University Hospital's NHS Trust.

Dr Ashley VanderLey
Senior Advisor for Facilities in the Division for Astronomical Sciences and supports the Office of International Science and Engineering at the National Science Foundation (NSF).