# DeiC Storage
# &
# DeiC Sensitive Storage

## The annual DeiC conference

Comwell Kolding, 7 November 2023

Session host:
**Philippe Bonnet**
Professor at the Department of Computer Science
University of Copenhagen

DeiC

AARHUS UNIVERSITY

UNIVERSITY OF COPENHAGEN

# The Session

1. Introduction

2. Storage Use Case: Asger Svenning (Ecoscience, AU)

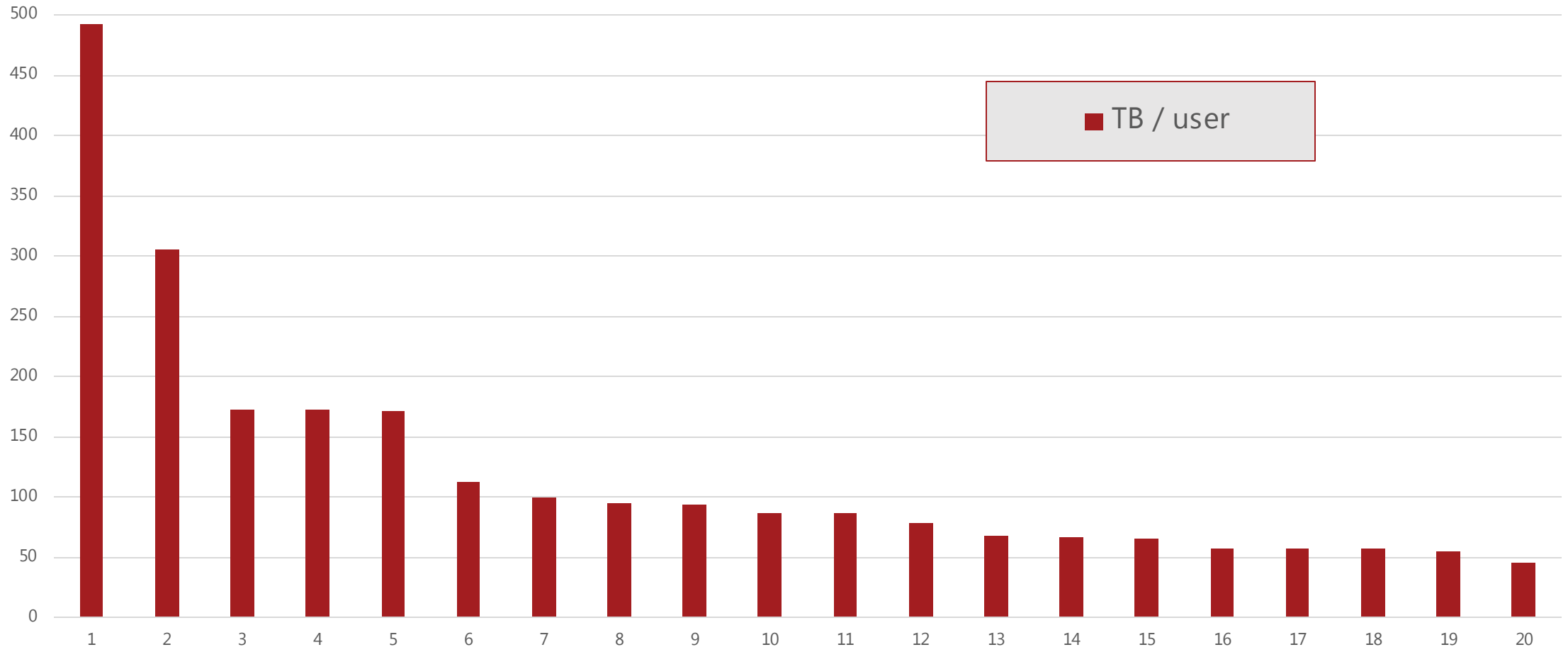
3. Sensitive Storage Use Case: Anders Tolver (Mathematical Science, KU)

4. Discussion

# The Need for Deic Storage

## Top 20 ERDA users at KU

(262 users with more than 1TB)



# The Need for Deic Sensitive Storage

The following personal data is considered 'sensitive' and is subject to specific processing conditions:

- personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs;
- trade-union membership;
- genetic data, biometric data processed solely to identify a human being;
- health-related data;
- data concerning a person's sex life or sexual orientation.

## References

- Article 4(13), (14) and (15) andArticle 9 and Recitals (51) to (56) of the GDPR

Traditional Data science:

- Data cleaning
- Data derivations
- Data analytics

**Regulated processes for users, infrastructure admins and legal/policy admins.**

# What to expect

A national data storage system open to all researchers at Danish universities consisting of two services that are based on the same technology.

**DeiC Storage**
Secure solution for storing research data. approved for general scientific data and *not* for highly sensitive data.

**DeiC Sensitive Storage**
Secure solution for storing and sharing sensitive data. Particularly personal data requiring special care under the EU General Data Protection Regulation (GDPR).

DeiC

AARHUS UNIVERSITY

UNIVERSITY OF COPENHAGEN

# DeiC Storage and DeiC Sensitive Storage

## An open national solution for storing research data

The services provide:

- Storage capacity combined with a system where researchers can store, share, exchange, and reference data.
- A solution that functions in conjunction with institutional and domain-specific solutions, as well as with international infrastructures.
- A general data storage system for research data and facilitate data exchange with DeiC's HPC and other storage facilities and data management services.
- A possibility to invite collaborators from the global research environment to .
- A technical solution for storing sensitive data under GDPR (phase 2).

The services are developed in a consortium between the University of Copenhagen (SCIENCE and KU-IT) and Aarhus University. The services are open to all Danish universities through DeiC.

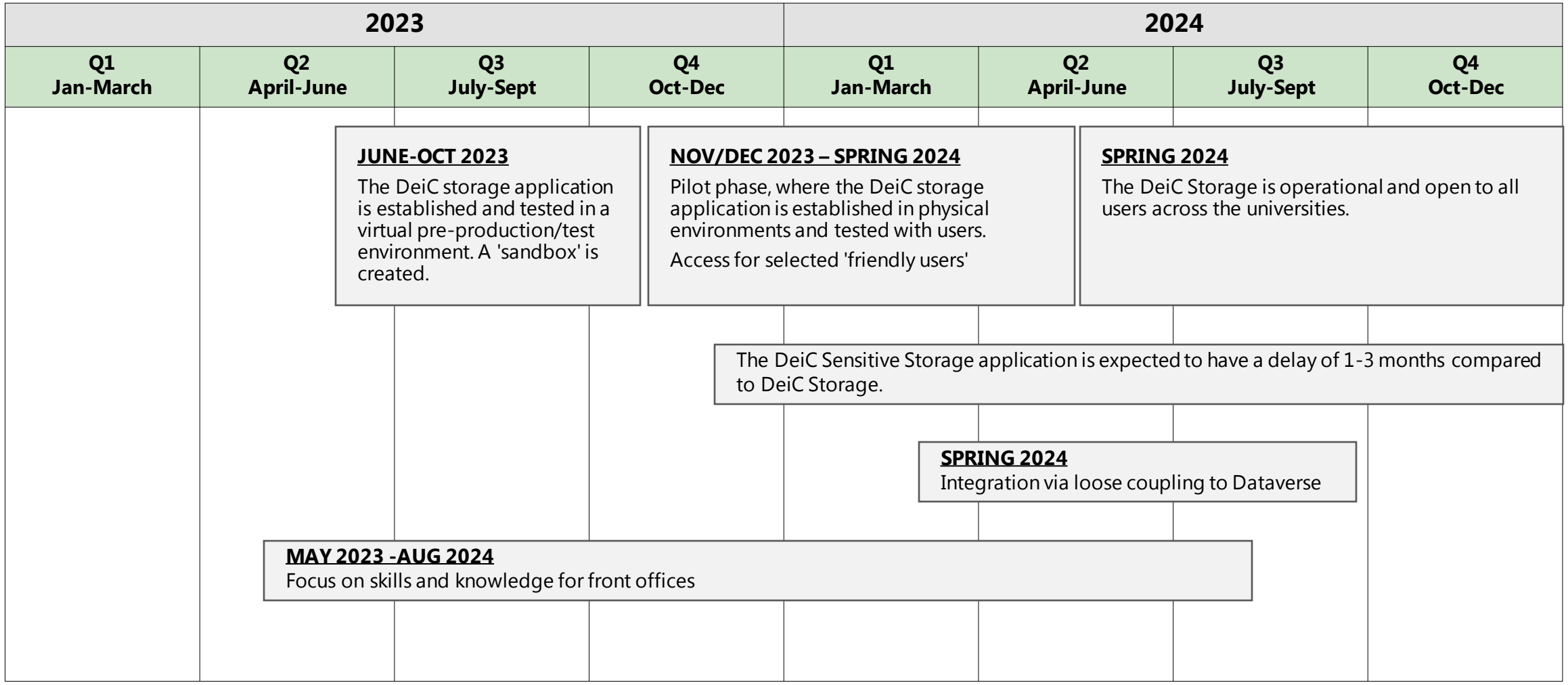
# My First Question

1. How do I access DeiC Storage?
   - Collection of files
   - Secure File Transfer Protocol (SFTP): file transfer (put: upload/get: download), Resume transfer (reput, reget), Directory listing (ls), remote file removal (rm)
     - Transfer file between local machine and remote server
     - Mount (as network drive) via SSHFS (linux, macos, windows)
     - Session on command line vs. scripts of SFTP commands as batch files
     - Public key authentication
   - Web interface
     - WAYF MFA authentication

# My Second Question

2. How do I get a DOI for my data set?
   - Mutable files are stored on DeiC storage for X years (policy decision)
   - Deic Storage makes it possible to *freeze* a file
   - A frozen file is immutable and permanent
   - DeiC Storage issues **a persistent link to a frozen file**
   - How about metadata?
     - The process for collecting metadata is defined/supported in Dataverse
     - DOI is obtained via Dataverse
     - The metadata generated in the process is stored in DeiC storage, frozen with a persistent link
   - No persistent link (or DOI) for data stored on sensitive storage

# Timeline


2023
2024
Q1 Jan-March
Q2 April-June
Q3 July-Sept
Q4 Oct-Dec
Q1 Jan-March
Q2 April-June
Q3 July-Sept
Q4 Oct-Dec
JUNE-OCT 2023
The DeiC storage application is established and tested in a virtual pre-production/test environment. A 'sandbox' is created.
NOV/DEC 2023 – SPRING 2024
Pilot phase, where the DeiC storage application is established in physical environments and tested with users.
Access for selected 'friendly users'
SPRING 2024
The DeiC Storage is operational and open to all users across the universities.
The DeiC Sensitive Storage application is expected to have a delay of 1-3 months compared to DeiC Storage.
SPRING 2024
Integration via loose coupling to Dataverse
MAY 2023 -AUG 2024
Focus on skills and knowledge for front offices

# Use Cases

## Deic Storage

Asger Svenning, AU

ERDA

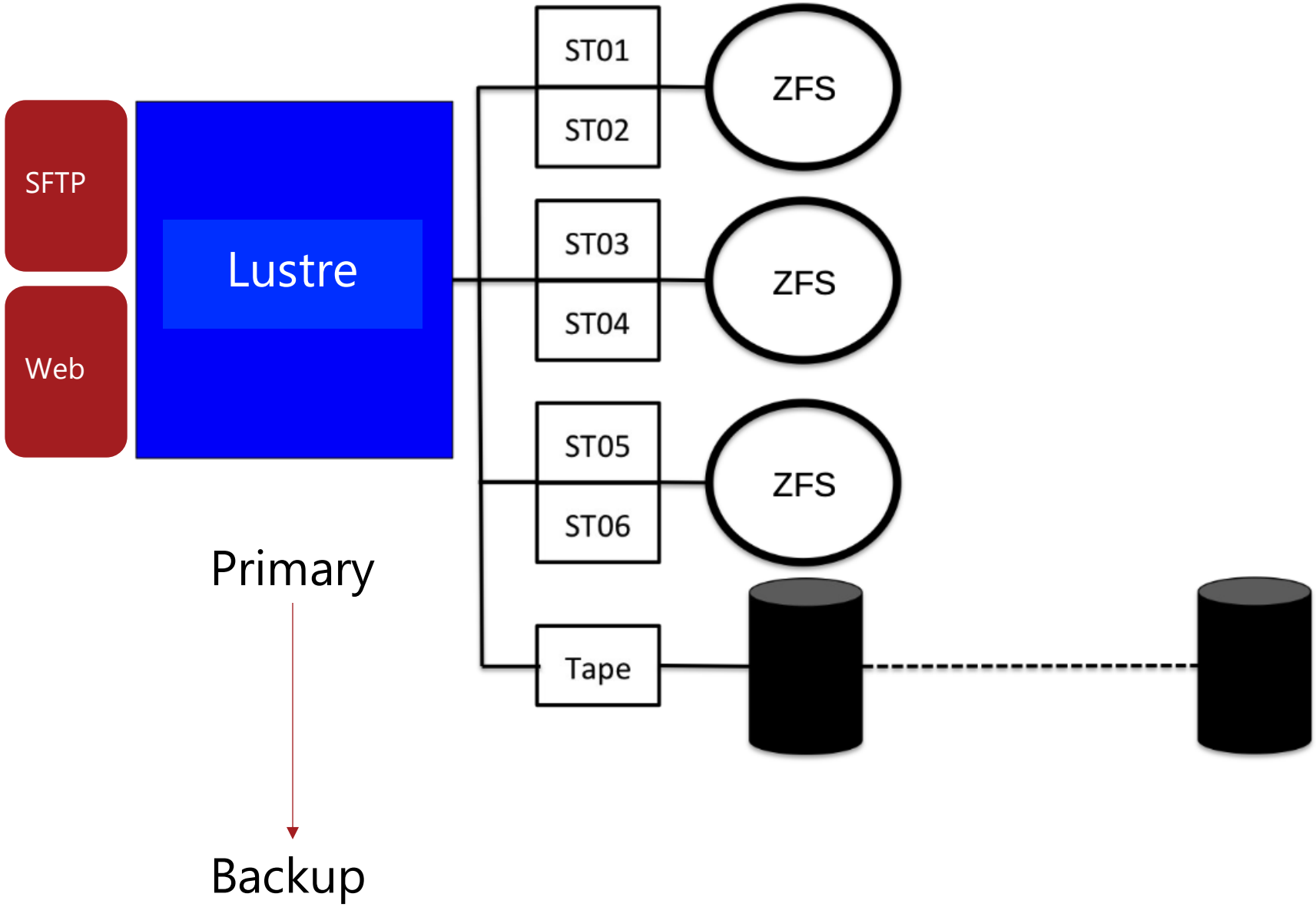
## Deic Sensitive Storage

Anders Tolver, KU

SIF

# Extra Slides

# Architecture



# ISO 27001

- Identify and assess information security risks
  - Risk management processes
- Security controls
  - Access control, cryptography, physical security, and incident management
- Practices based on continual improvement
  - Regular monitoring, performance evaluation, and periodic reviews

https://www.iso.org/standard/27001
https://digst.dk/sikkerhed/iso-27001/hvad-er-iso-27001/

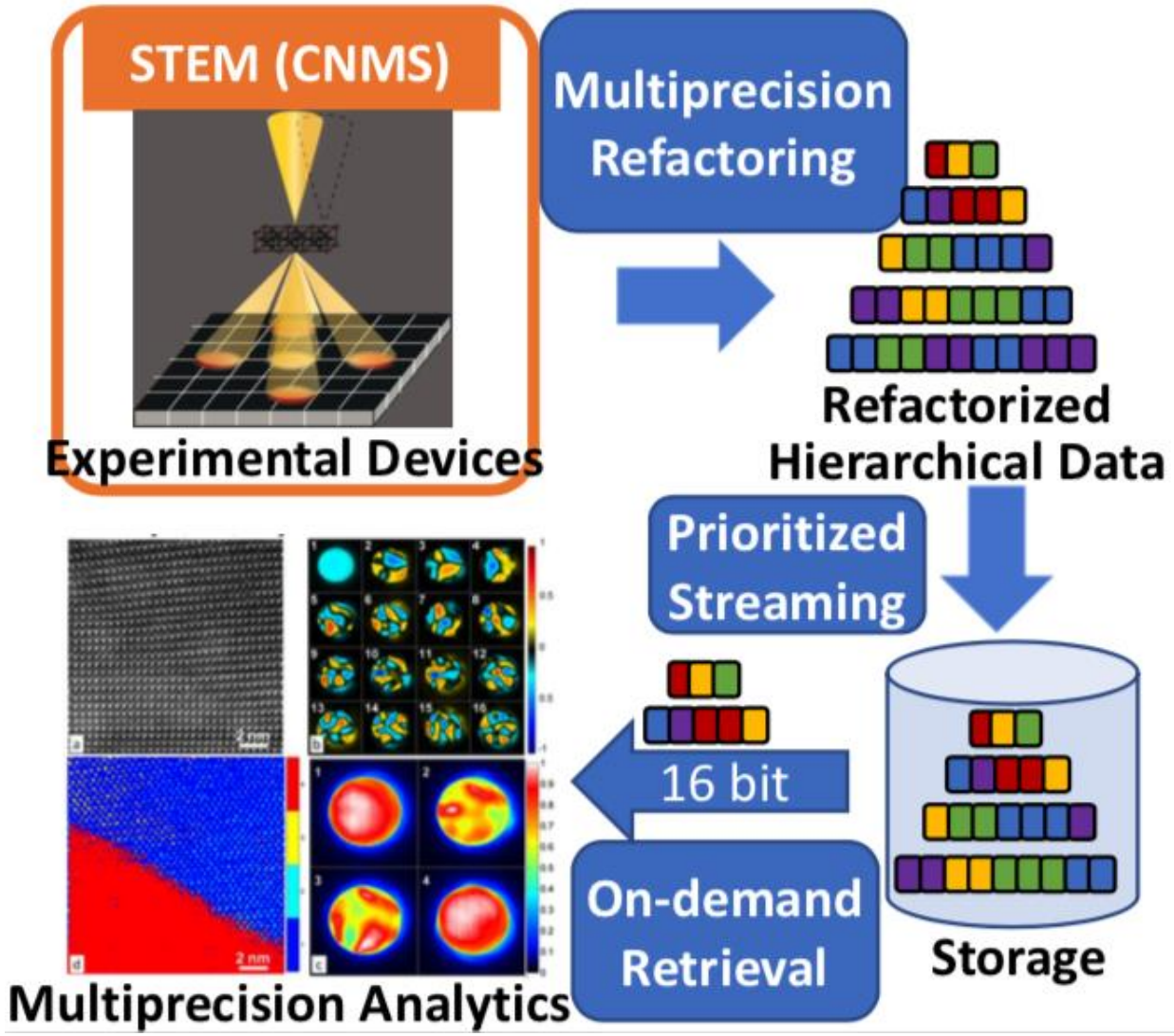
# EOSC Roadmap 2025-27

EOSC should be a federation of existing and planned research data infrastructures, adding a soft overlay to connect them and making them operate as one seamless European research data infrastructure

- PID services should be interoperable
- Semantic artefacts are **machine readable models of knowledge** such as controlled vocabularies, thesauri, and ontologies which facilitate the extraction and representation of knowledge within data sets using annotations or assertions
- Further support is needed to support the **development, archiving, sharing and reuse of research software**

https://eosc.eu/sites/default/files/2023-01/MAR_2025-27_draft.pdf

# Science-driven Data Management




Sirius 2 @ ORNL