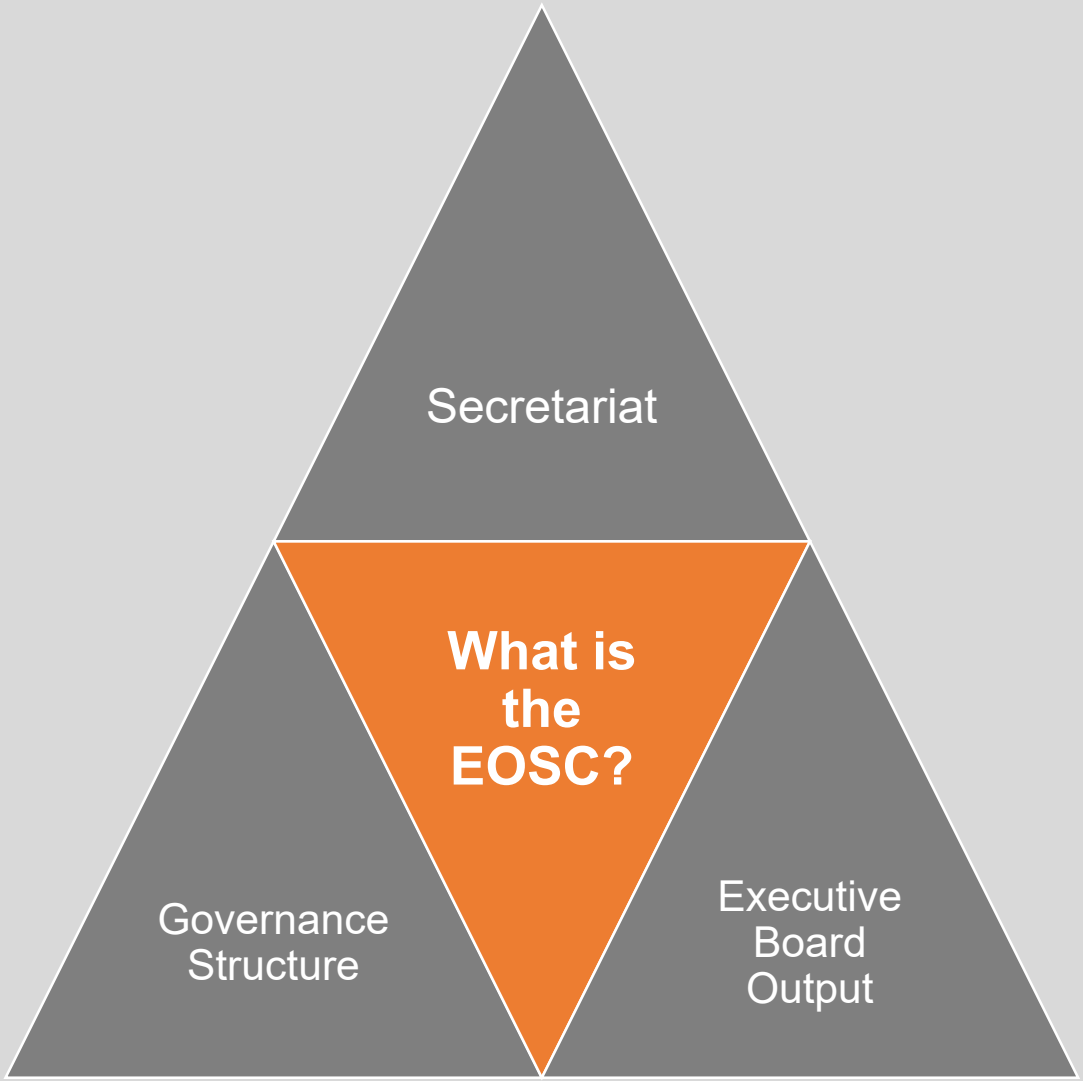
EOSC
European Open Science Cloud

National RDM – June 15, 2021

Outline - EOSC Unmasked



Richard Dennis (ricd@kb.dk)
Special Advisor – Data Steward
Research Support
Copenhagen University Library

EOSC - What is it?

- Environment for **hosting and processing research data to support EU science.**
- Aims to **develop a trusted, virtual, federated environment that cuts across borders** and scientific disciplines to store, share, process, and re-use research digital objects (like publications, data, and software) following FAIR Data principles.
- **Brings together institutional, national and European stakeholders, initiatives, and data infrastructures** to develop an inclusive open science ecosystem in Europe.

This is expected to lead towards:

- **new insights and innovations**
- **higher research productivity and**
- **improved reproducibility in science.**

Richard Dennis (ricd@kb.dk)
Special Advisor – Data Steward
Research Support
Copenhagen University Library

EOSC Secretariat

Addresses the need for the set-up of an operational framework supporting the overall governance of the European Open Science Cloud (EOSC)

Effectively delivers 360° support to the EOSC Governance (Executive Board and its Working Groups)

Works openly and inclusively together with communities to co-create an all-encompassing European Open Science Cloud

Richard Dennis (ricd@kb.dk)
Special Advisor – Data Steward
Research Support
Copenhagen University Library

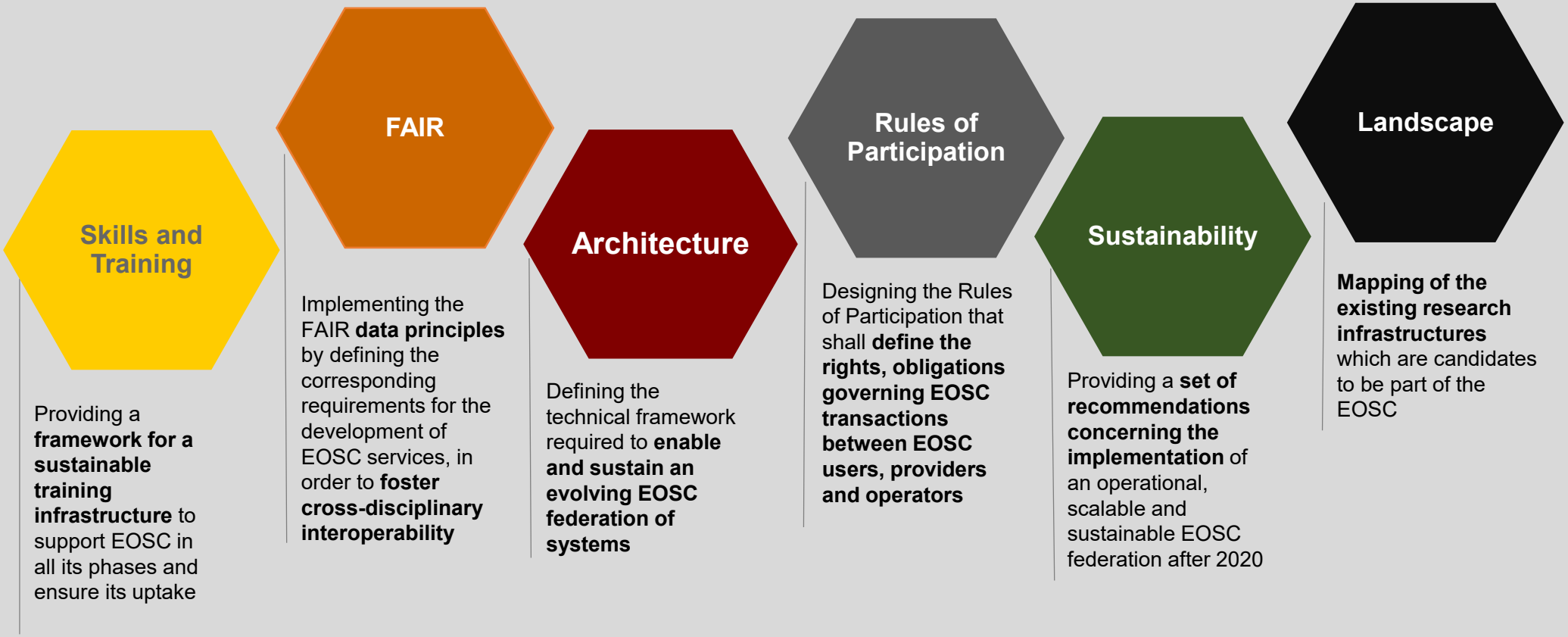
EOSC Governance and its structure

- The European Open Science Cloud (EOSC) is **governed by three constituent bodies**, as defined by the European Commission Staff Working Document Implementation Roadmap for the European Open Science Cloud.
- This governance structure is comprised of **representatives from** across the **user and provider community** as well as representatives **from EU Member States and the European Commission.**

*constituent bodies

Richard Dennis (ricd@kb.dk)
Special Advisor – Data Steward
Research Support
Copenhagen University Library

EOSC Working Groups



Richard Dennis (ricd@kb.dk)
Special Advisor – Data Steward
Research Support
Copenhagen University Library

EOSC Executive Board

The EOSC Executive Board and its Working Groups (2019-2020) produced 20 reports*
with recommendations in 6 priority areas for EOSC.
These are key to the EOSC Strategic Research and Innovation Agenda (SRIA) which provides a clear **roadmap for the next seven years to achieve the EOSC vision and objectives.**

Research-community centered	Multi-stakeholders	Openness	FAIR principles	Federation of infrastructures	Machine-actionable
EOSC will place research at the center of the initiative and will thus **prioritize engagement** with research communities to **understand their requirements, helping them and ensuring academic sovereignty of research data.**	**EOSC** will succeed if and only if it **follows a multi-stakeholder approach**.	EOSC will ensure that research **artefacts** are as **open as possible, as closed as necessary**.	EOSC will assemble research artefacts that are **findable, accessible, interoperable and reusable.**	EOSC will **federate existing and upcoming research Infrastructures.**	EOSC will strike the **right balance between machines and people in delivering the services** that will serve the needs of European scientists.

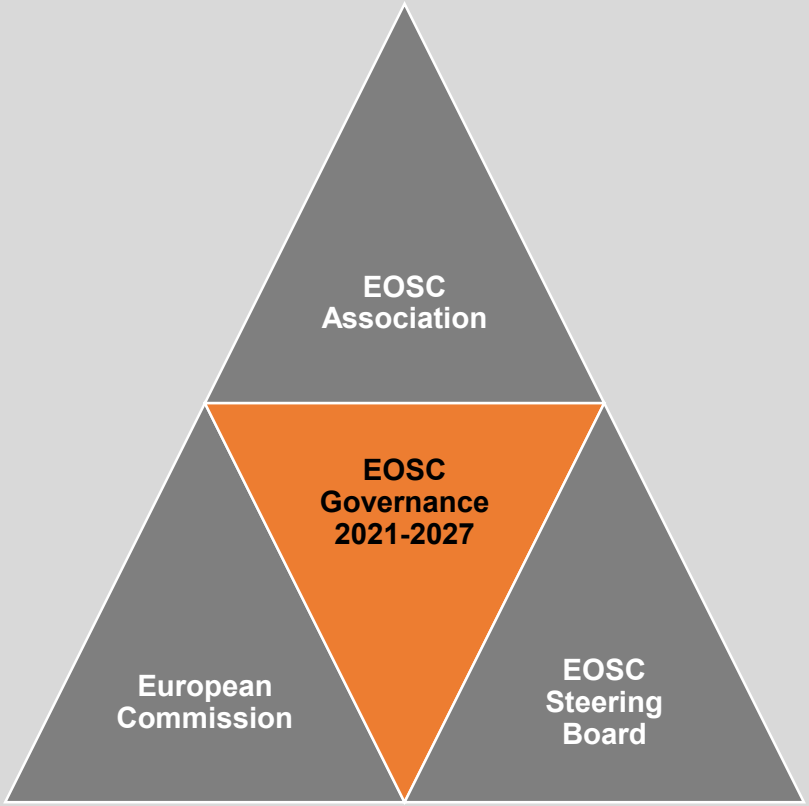
* 20 reports

Richard Dennis (ricd@kb.dk)
Special Advisor – Data Steward
Research Support
Copenhagen University Library

Outlook - EOSC Governance 2021-2027

EU countries and countries associated with Horizon 2020, represented in the EOSC Governance Board, agreed unanimously to **run the EOSC as a co-programmed European Partnership under Horizon Europe from 2021**.

The new governance model agreed with EU countries for the **current EOSC implementation phase after 2020 is tripartite** including:



Richard Dennis (ricd@kb.dk)
Special Advisor – Data Steward
Research Support
Copenhagen University Library

Thank you for your attention.

Question and Answers

Dataverse Presentation follows

Richard Dennis (ricd@kb.dk)
Special Advisor – Data Steward
Research Support
Copenhagen University Library

A Brief Overview of KUB and KU Data Repository Project:

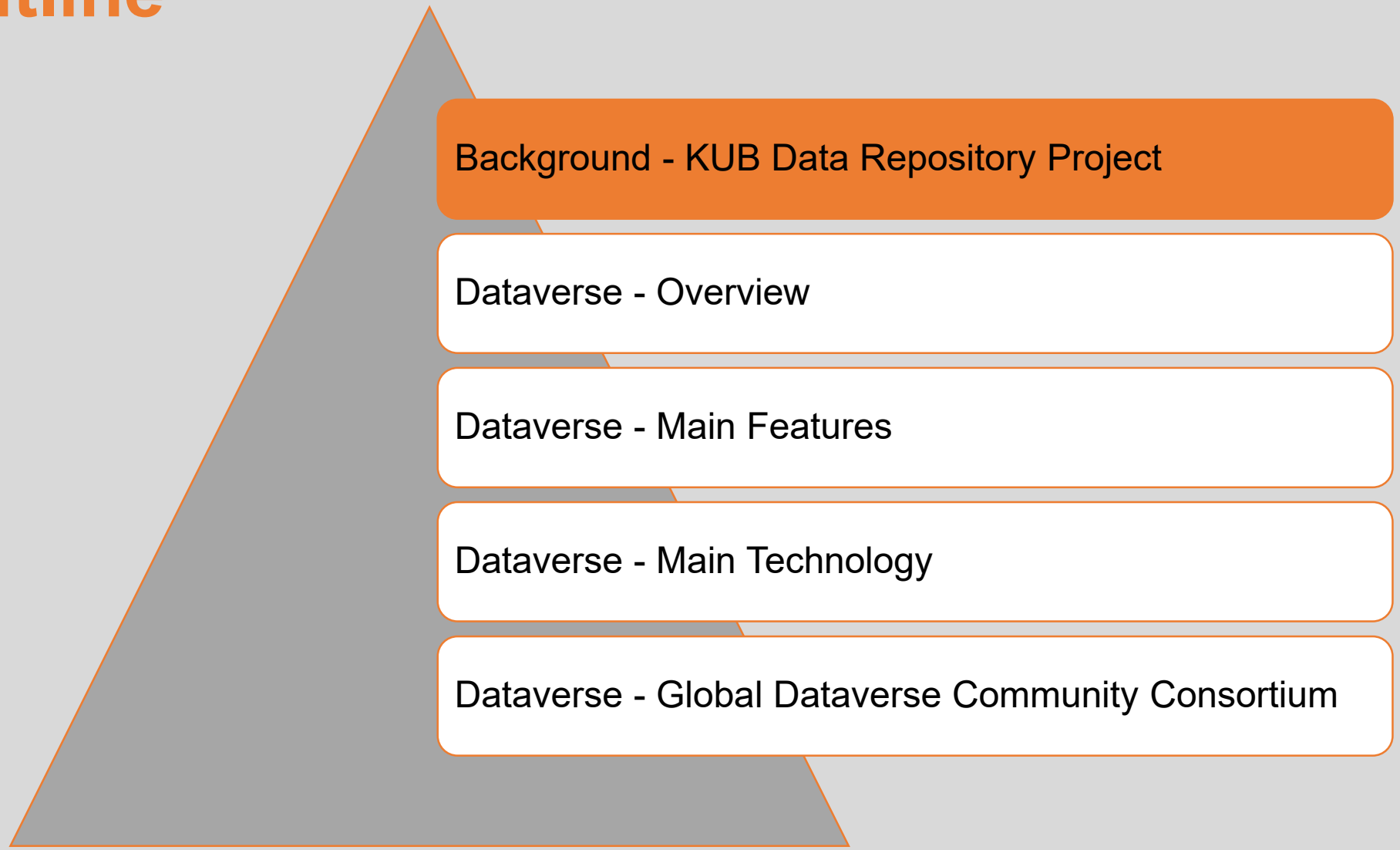
National RDM – June 15, 2021

Outline

- Background – KUB / KU Data Repository Project
- Dataverse - Overview
- Dataverse - Main Features
- Dataverse - Main Technology
- Dataverse - Global Dataverse Community Consortium

Richard Dennis (ricd@kb.dk)
Special Advisor – Data Steward
Research Support
Copenhagen University Library

Outline



Richard Dennis (ricd@kb.dk)
Special Advisor – Data Steward
Research Support
Copenhagen University Library

Dataverse - KUB / KU Background

- The KU Data Repository aims at providing a **platform for publishing and archiving research data** from the University of Copenhagen.

- The platform will enable all researchers (including Ph.D. students) to make their **research data and documentation thereof searchable, citable, and accessible** for reuse via a public web portal.

- UCPH Data Repository will be an **essential element of the university's central research infrastructure supporting Open Science** and the FAIR research data management principles.

Richard Dennis (ricd@kb.dk)
Special Advisor – Data Steward
Research Support
Copenhagen University Library

Dataverse

Dataverse - KUB / KU Background

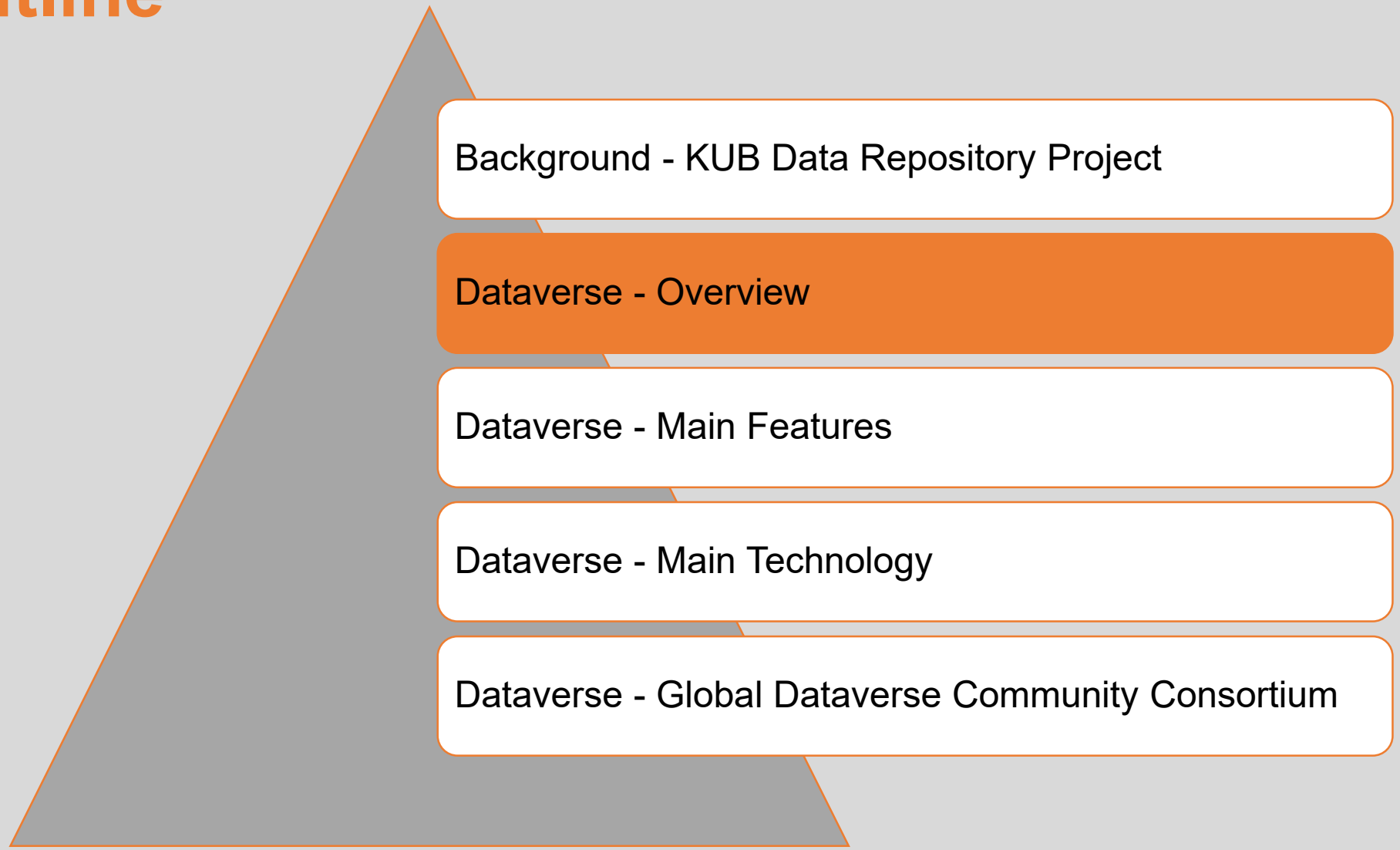
- It will help the university's researchers fulfill external requirements for **open and FAIR research data**, implement **best practices for responsible research**, and **adhere to UCPH's policy** for research data management.

- The **solution comprises the installation and operation of the open-source software** Dataverse on servers at the IT department for **secure and permanent data storage and metadata, controlled access management, and public availability** of the repository's web portal.

- The solution is a **collaboration between the IT department, Research and Innovation and Copenhagen University Library.**

Richard Dennis (ricd@kb.dk)
Special Advisor – Data Steward
Research Support
Copenhagen University Library

Outline



Richard Dennis (ricd@kb.dk)
Special Advisor – Data Steward
Research Support
Copenhagen University Library

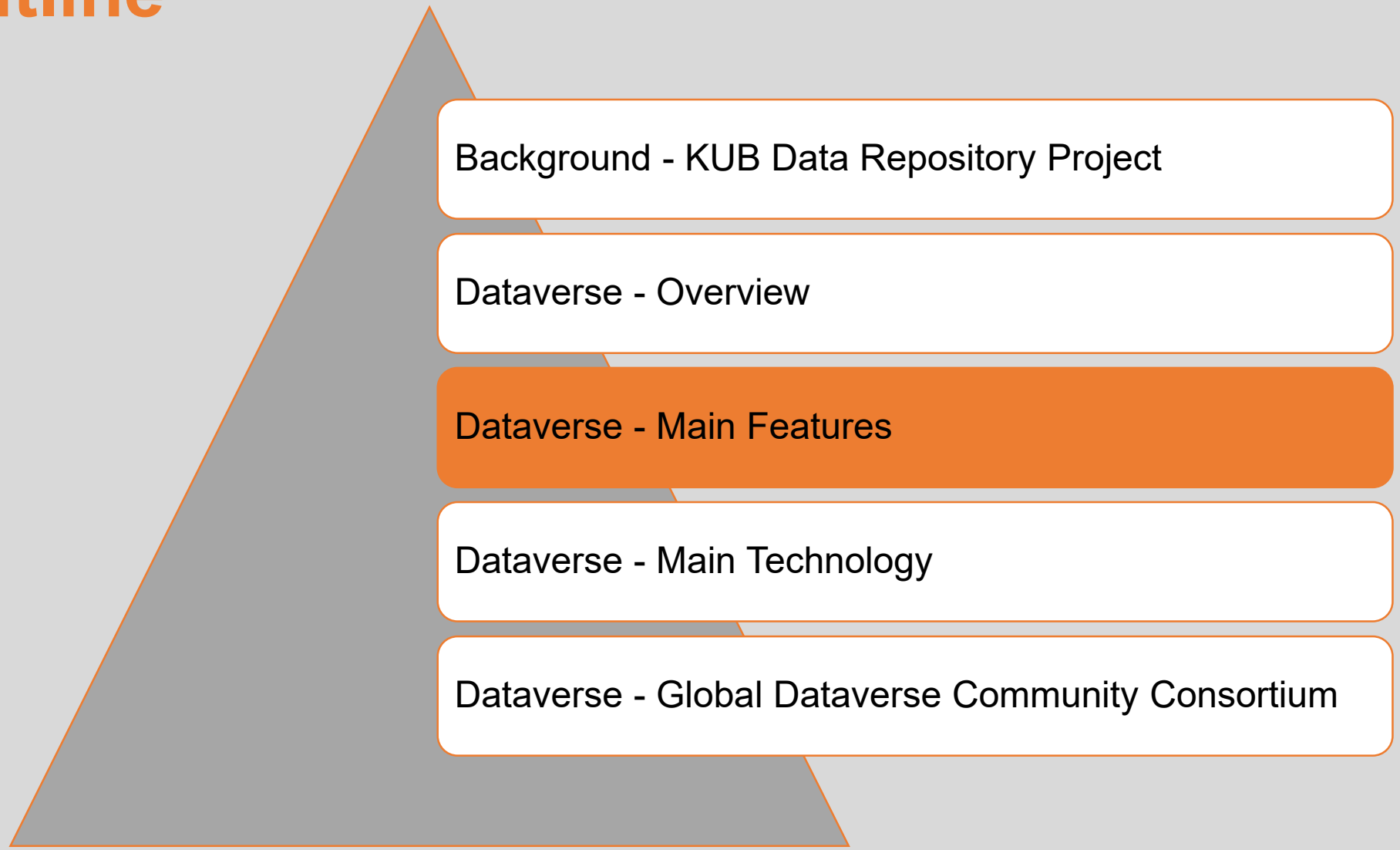
Dataverse

Dataverse - Overview

- Dataverse is an **open-source repository platform used for storing, describing, sharing, citing and exploring research data.**
- Designed to **support multiple types of data, users, and workflows**.
- Metadata and files **support different file formats**.
- Roles and Permissions for users to **control access**
- Developed at **Harvard's Institute for Quantitative Social Science** (IQSS) since 2006
- Development funded by IQSS and with **grants, in collaboration with institutions around the world**
- 12 on the core team - **developers, designers, UI/UX, metadata specialists, curation team, leadership team**

Richard Dennis (ricd@kb.dk)
Special Advisor – Data Steward
Research Support
Copenhagen University Library

Outline



Richard Dennis (ricd@kb.dk)
Special Advisor – Data Steward
Research Support
Copenhagen University Library

Dataverse

Dataverse - Main Features

Data Management

- **Persistent Identifiers:** Support for DOI and Handle at dataset and file level
- **Citation:** Generated references at dataset and file level
- **Versioning:** Versioning of datasets
- **Metadata:** Schemas for general metadata (e.g. Dublin Core) and domain- specific metadata (e.g. DDI), customizable metadata schemas
- **Embargo:** Restrict file access for a period of time (in development)
- **File storage:** Different systems: Local, Swift (OpenStack), S3 (Amazon)
- **Security:** Support for sensitive data (DataTags) (in development)

Adapted from Durand (2020)

Richard Dennis (ricd@kb.dk)
Special Advisor – Data Steward
Research Support
Copenhagen University Library

Dataverse - Main Features

User Management

- **Sign-in:** Multiple sign-in options: Native, Shibboleth, OAuth (ORCID, Github, Google, Microsoft), Open ID Connect
- **Collections:** Possible to create collections and sub-collections
- Branding and widgets

Adapted from Durand (2020)

Richard Dennis (ricd@kb.dk)
Special Advisor – Data Steward
Research Support
Copenhagen University Library

Dataverse - Main Features

Workflows

- Different curation and publishing workflows may be configured
- Private URLs for access to unplublished datasets, e.g., for peer review
- Data upload / download options:
  - Browser / FileUploader
  - Dropbox
  - Rsync (for big data “packages”)
  - Remote Storage (TRSAs)

Adapted from Durand (2020)

Richard Dennis (ricd@kb.dk)
Special Advisor – Data Steward
Research Support
Copenhagen University Library

Dataverse - Main Features

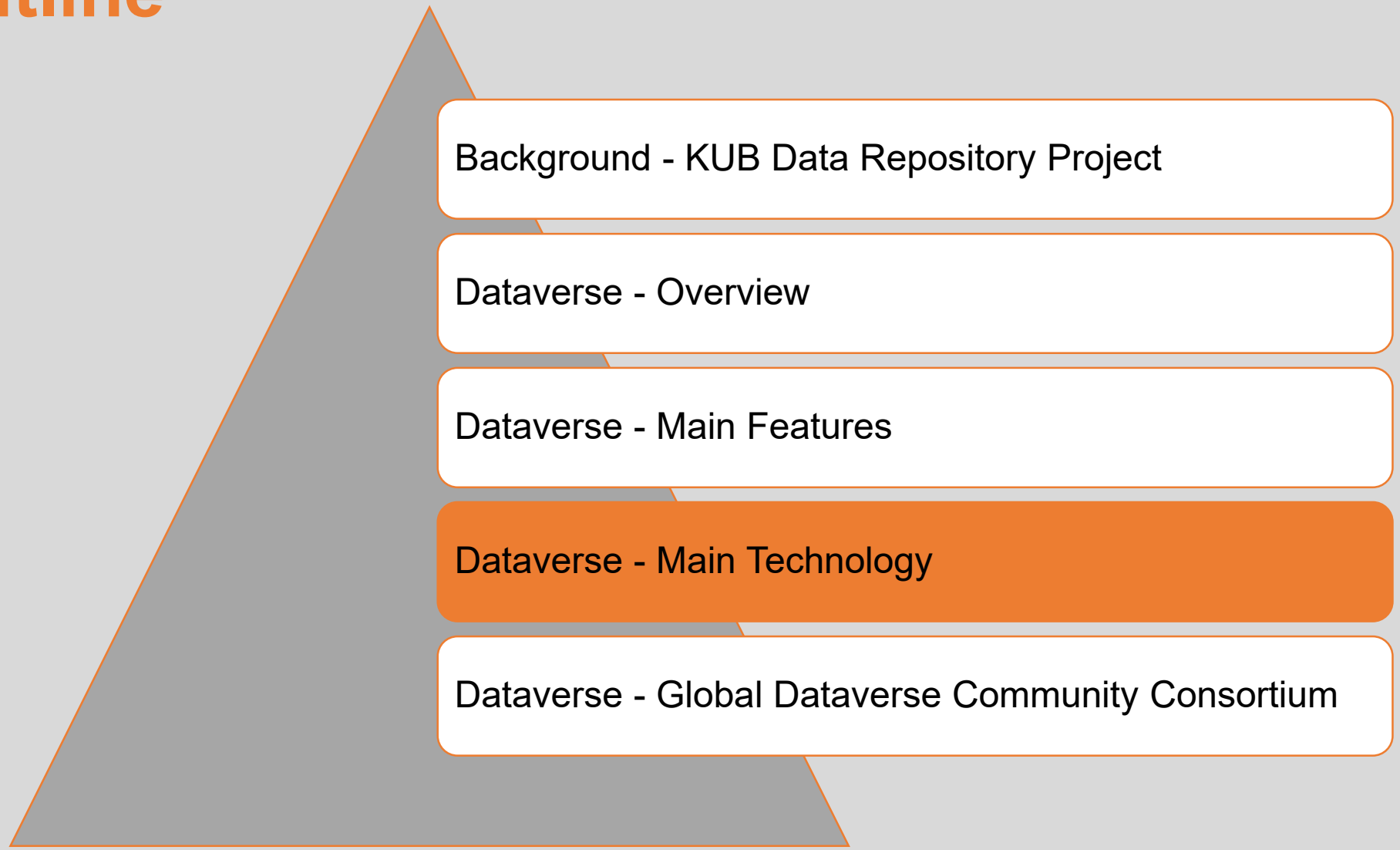
Interoperability

- **APIs**
  - SWORD
  - Native
  - Metrics
- **Harvesting (OAI-PMH)**
  - Server
  - Client
- **Modular external tools**
  - Explore and configure
  - Scope: Dataset / datafile

Adapted from Durand (2020)

Richard Dennis (ricd@kb.dk)
Special Advisor – Data Steward
Research Support
Copenhagen University Library

Outline



Richard Dennis (ricd@kb.dk)
Special Advisor – Data Steward
Research Support
Copenhagen University Library

Dataverse

Dataverse - Main Technology

Linux: RHEL/CentOS Operating System
Glassfish/Payara 5.20
Java SE8
Java EE11

- Presentation: JSF (PrimeFaces), RESTful API
- Business: EJB, Transactions, Asynchronous, Timers
- Storage: JPA (Entities), Bean Validation

Rserve: run as daemon to execute R code

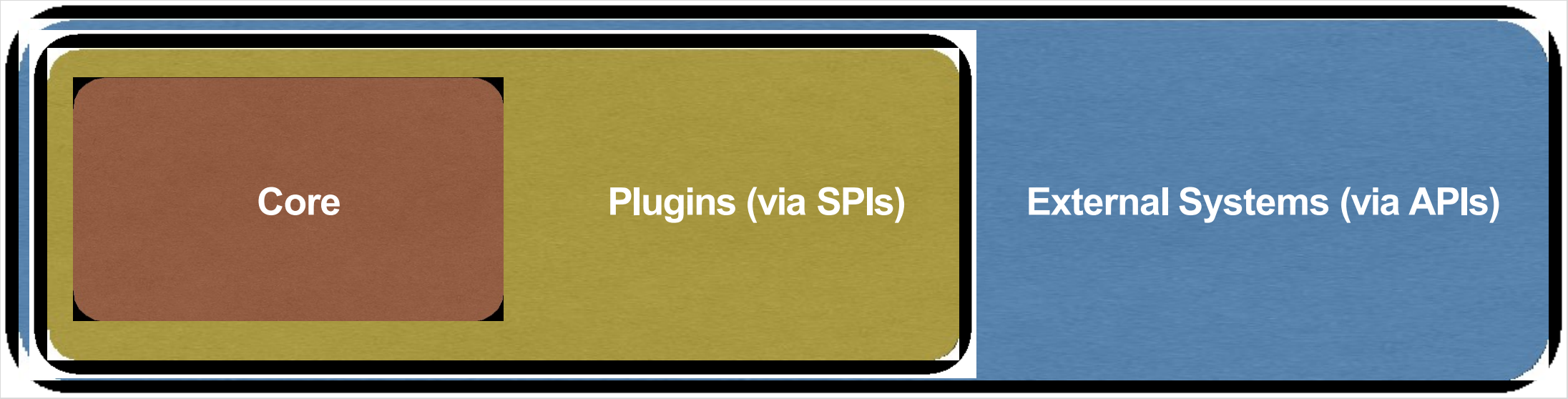
R: Dataverse uses R to handle tabular data files

Storage: PostgreSQL, Solr, File System / Swift / S3

Adapted from Durand (2020)

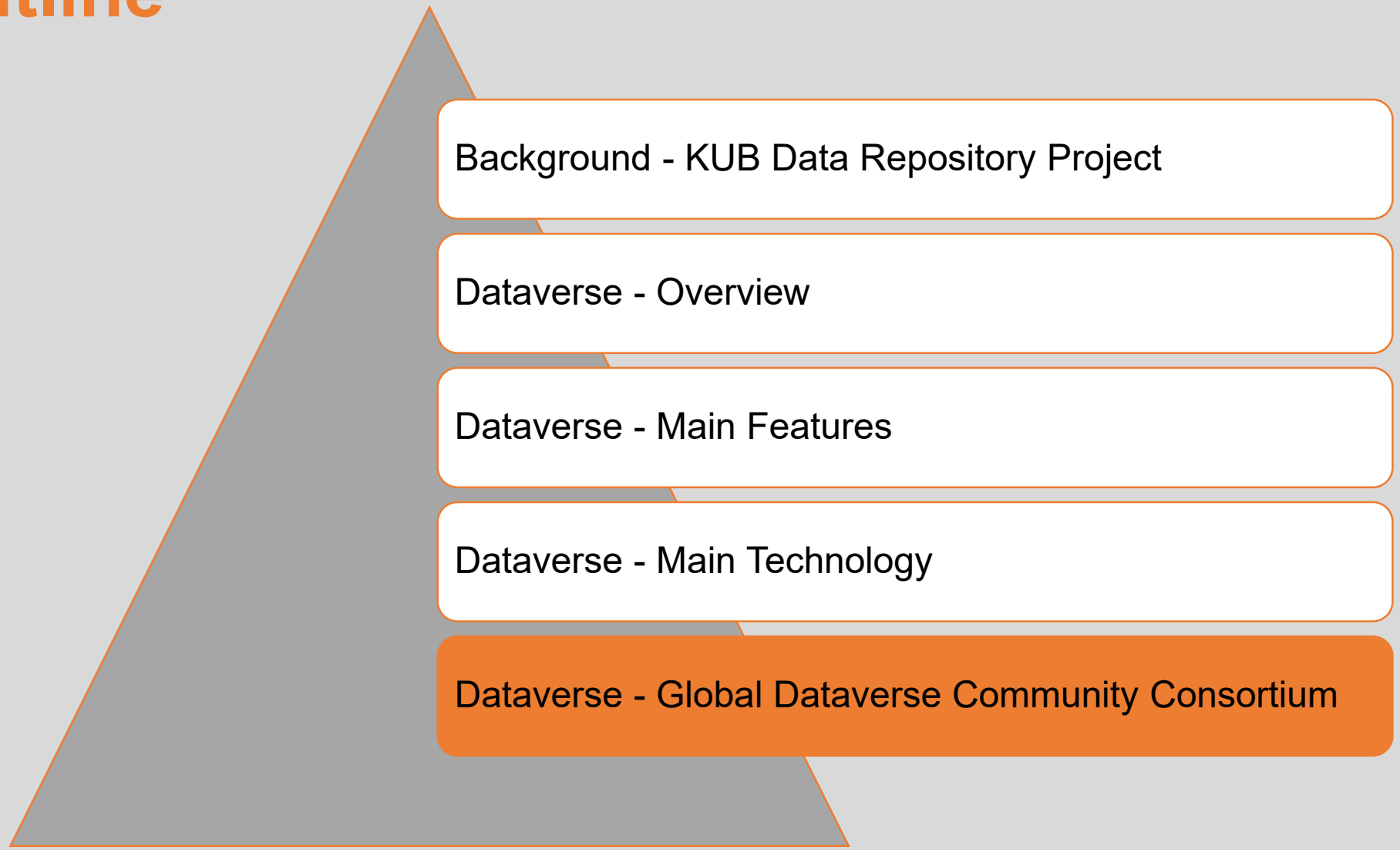
Richard Dennis (ricd@kb.dk)
Special Advisor – Data Steward
Research Support
Copenhagen University Library

Dataverse - Main Technology Ecosystem



Richard Dennis (ricd@kb.dk)
Special Advisor – Data Steward
Research Support
Copenhagen University Library

Outline



Richard Dennis (ricd@kb.dk)
Special Advisor – Data Steward
Research Support
Copenhagen University Library

Global Dataverse Community Consortium

The Global Dataverse Community Consortium (GDCC) is dedicated to providing international organization to existing Dataverse community efforts and will **provide a collaborative venue for institutions to leverage economies of scale in support of Dataverse repositories around the world**.

http://DataverseCommunity.Global

Richard Dennis (ricd@kb.dk)
Special Advisor – Data Steward
Research Support
Copenhagen University Library

Dataverse Community

Who is contributing?

- Developers, researchers, librarians, data scientists (several hundreds in total)

How are they contributing?

- Code (100+ contributors)
- UI/UX testing & interviews
- Almost daily discussing issues in the Dataverse Google Group
- Participating in Dataverse Community Calls every second week
- Dataverse Community Meeting once a year at Harvard
- Workshops & trainings, e.g., European Dataverse Workshop 2021

Adapted from Durand (2020)

Richard Dennis (ricd@kb.dk)
Special Advisor – Data Steward
Research Support
Copenhagen University Library

Expanding Membership

Australian Data Archive	Australia
Consorcio Madrono	Spain
DANS	Netherlands
DataverseNO	Norway
Fudan University	China
Göttingen eResearchAlliance	Germany
Harvard University	United States
International Centre for Research in Agroforestry	Kenya
John Hopkins University	United States
Nanyang Technological University	Singapore
Syracuse University	United States
Texas Digital Library	United States
University of California Los Angeles	United States
University of Campinas	Brazil
University of North Carolina Chapel Hill	United States
University of Virginia	United States

Richard Dennis (ricd@kb.dk)
Special Advisor – Data Steward
Research Support
Copenhagen University Library

Thank you for your attention.

Questions and Answers

Richard Dennis (ricd@kb.dk)
Special Advisor – Data Steward
Research Support
Copenhagen University Library

References

Crosas, Mercè. 2020. “Fair Principles and Beyond: Implementation in Dataverse”. *Septentrio Conference Series*, no. 2 (March). https://doi.org/10.7557/5.5334.

Conzett, Philipp. 2019. “Disciplinary Case Study: The Tromsø Repository of Language and Linguistics (TROLLing)”. https://doi.org/10.5281/zenodo.2668775.

Durand, Gustavo. 2020. “Dataverse’s Approach to Technical Community Engagement”. *Septentrio Conference Series*, no. 2. https://septentrio.uit.no/index.php/SCS/article/view/5424

GÉANT, and UNINETT. 2019. ‘Why TROLLing Is the Thing to Do for Linguists’. In The Field. May 2019. https://www.inthefieldstories.net/why-trolling-is-the-thing-to-do-for-linguists/.

Hüser, F., Rasmussen, H. R., Dennis, R. (2021). Copenhagen University Library. Architectural Assessment for UCPH Data Repository. Internal document: unpublished

Wittenberg, Marion, and Vyacheslav Tykhonov. 2020. “Dataverse in the European Open Science Cloud”. *Septentrio Conference Series*, no. 2 (March). https://doi.org/10.7557/5.5421.

Richard Dennis (ricd@kb.dk)
Special Advisor – Data Steward
Research Support
Copenhagen University Library