

Data Archiving and Networked Services

The Reference Model for Social Science and Humanities Data Infrastructures: towards a common architectural model for Research Infrastructure architectures.

Mike Priddy, Maarten Hoogerwerf, John Shepherdson, Johan Fihn, Birger Jerlehag, Timo Gnadt

EDDI2016, Cologne, December 6 2016.





What is a Research Infrastructure

“The term ‘research infrastructures’ (RI) refers to **facilities, resources** and related **services** used by the **scientific community** to conduct **top-level** research in their respective fields, ranging from social sciences to astronomy, genomics to nanotechnologies.... RIs may be ‘**single-sited**’ (a single resource at a single location), ‘**distributed**’ (a **network** of distributed resources), or ‘**virtual**’ (the service is provided **electronically**).”

http://ec.europa.eu/research/infrastructures/index_en.cfm?pg=what



RIs in Social Sciences and Humanities

Research infrastructures communities have been setup to support Social Science and Humanities disciplines with **services** such as resource identification, data collection and data management. The **ESFRI** infrastructures have turned into formal **ERICs** (or equivalent)





An Architecture vs A Model

- **Architecture** - fundamental **concepts** or **properties** of a **system in its environment** embodied in its elements, relationships, and in the principles of its design and evolution¹.
- **Enterprise architecture** - an architecture in which the system in question is the **whole enterprise**, especially the business processes, technologies, and information systems of the enterprise².

1. <http://www.iso-architecture.org/ieee-1471/>

2. <https://msdn.microsoft.com/en-us/library/bb466232.aspx>

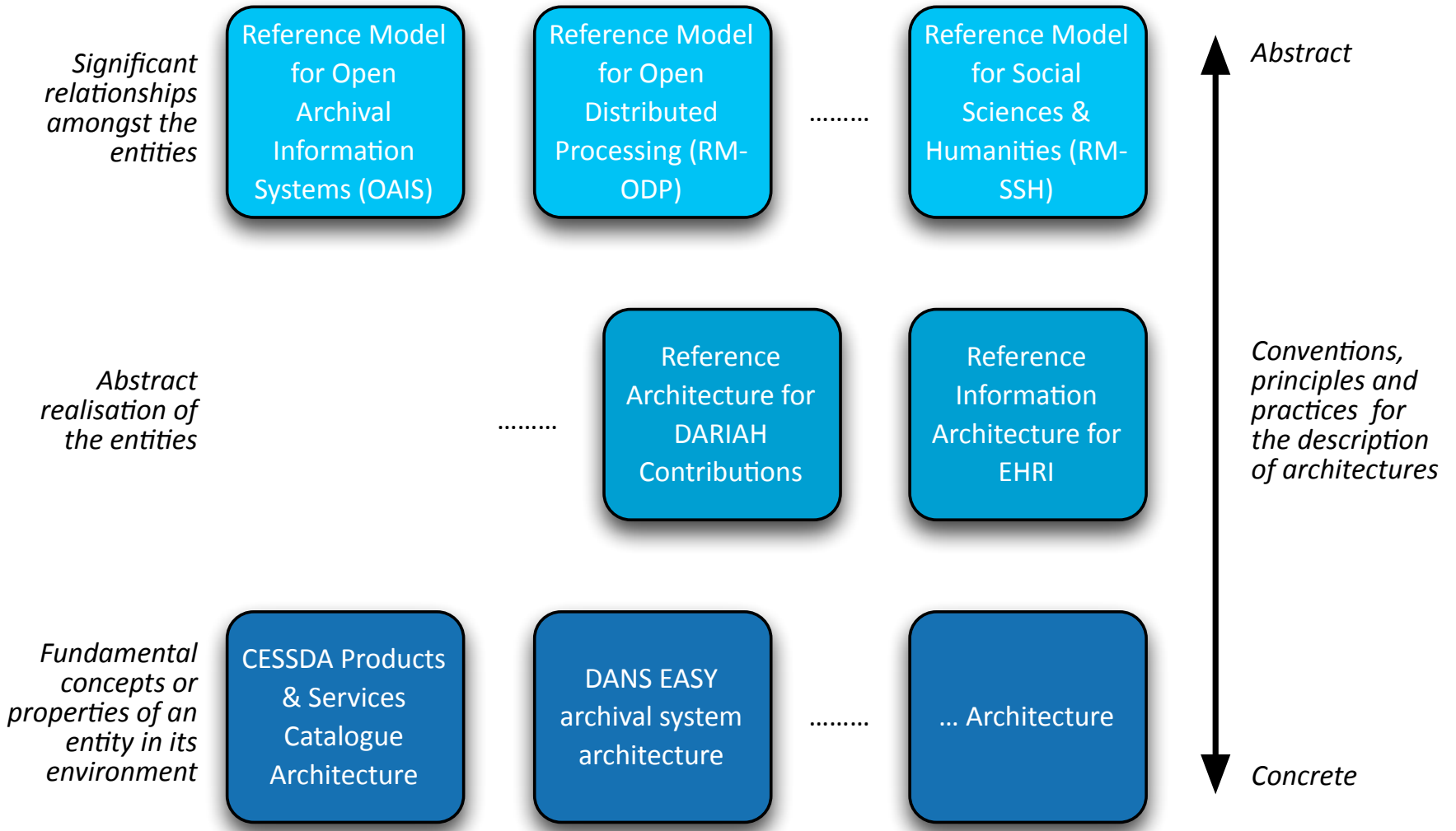


An Architecture vs A Model

- **Reference architecture** - models the **abstract elements** in the domain of interest independent of the technologies, protocols, and products that are used to implement a **specific solution** for the domain¹.
- **Reference model** - an **abstract framework** for understanding **significant relationships** among the entities of some environment².

1. <http://docs.oasis-open.org/soa-rm/soa-ra/v1.0/soa-ra.html>

2. <https://docs.oasis-open.org/soa-rm/v1.0/soa-rm.html>





Why a Reference Model for the Social Sciences and Humanities

A **reference architecture** requires **existing** architectures from which it **can be derived**, however the majority of the ERICs are in the **initial phase** of development and do not have a well-developed architecture for their research infrastructure.



Why a Reference Model for the Social Sciences and Humanities

A **reference model** identifies the the **communities**, relationships, **roles** and **interfaces** that could be present. For the SSH RIs this would **help to identify commonalities and similarities** from where it is possible to develop cross-cutting collaborations and instigate interfaces of knowledge, information & data.



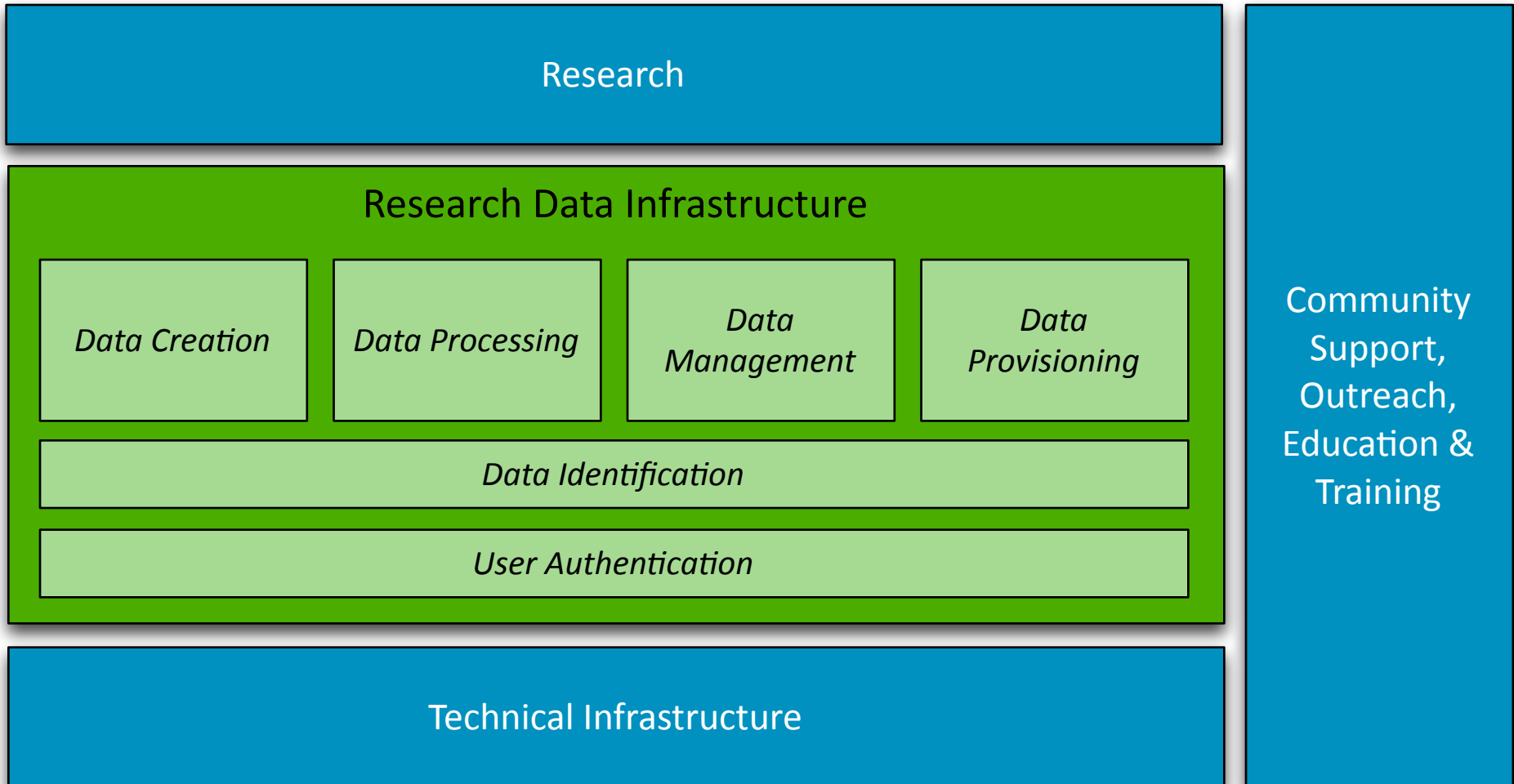
Modelling Approach

- All of the **SSH-RIs** considered by this reference model are **complex** systems that are **distributed** across Europe.
- Each SSH-RI has **differing aims** and **ambitions** along with **solutions** to tackle the challenges of distribution across countries, institutions and responsibilities.
- A reference model consists of a **minimal set of unifying concepts, axioms and relationships** within a particular **problem domain**¹. In addition, it should describe a **clear** problem that is solved².

1. <https://docs.oasis-open.org/soa-rm/v1.0/soa-rm.html>

2. https://en.wikipedia.org/wiki/Reference_model

Modelling Approach





Modelling Approach: RM-ODP

- Reference Model for Open Distributed Processing (RM-ODP) provides 4 elements: an **object modelling** approach, a way to specify a system in **interrelated but separate viewpoints** (enterprise, computational, information, engineering and technology), a system infrastructure that allows distribution transparencies and a framework to assess conformance.
- The viewpoints currently modelled in the RM-SSH are:
 - **Enterprise** viewpoint
 - **Information** viewpoint
 - **Computational** viewpoint



RM-SSH: Enterprise Viewpoint

The enterprise viewpoint describes **interrelated communities** as this allows common **purposes** and **policies** within the system to be grouped. These communities are described by **community contracts** which contain the **objectives, roles, policies, behaviour** and **enterprise object types** of a community.

RM-SSH: Enterprise Viewpoint

Data Creation

...to create/collect research data using an instruments or methodology.

Roles

Data Creator
Instrument Designer
Instrument
Data Collector

Behaviour

Deploy Instrument
Collect Data

Data Management

...to maintain the required quality of research data.

Roles

Data Owner
Data Manager
Preservation Manager
Data Admin System

Behaviour

Ingest data
Curate data
Administer data

Data Provision

...to make research data discoverable and accessible.

Roles

Data Provider
Data Consumer
Data Discovery System
Data Retrieval System

Behaviour

Data Publication
Data Discovery
Data Retrieval

Data Processing

...to transform research data for a specific use.

Roles

Process Designer
Processor
Process Consumer

Behaviour

Deploy Process
Collect Data

Data Identification

...to identify and locate research data.

User Authentication

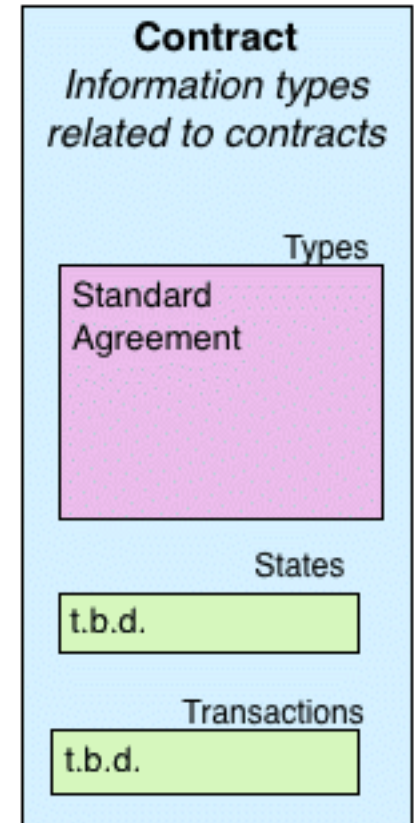
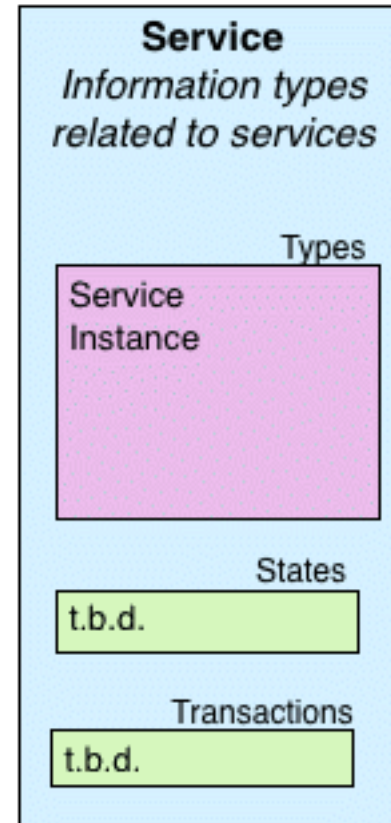
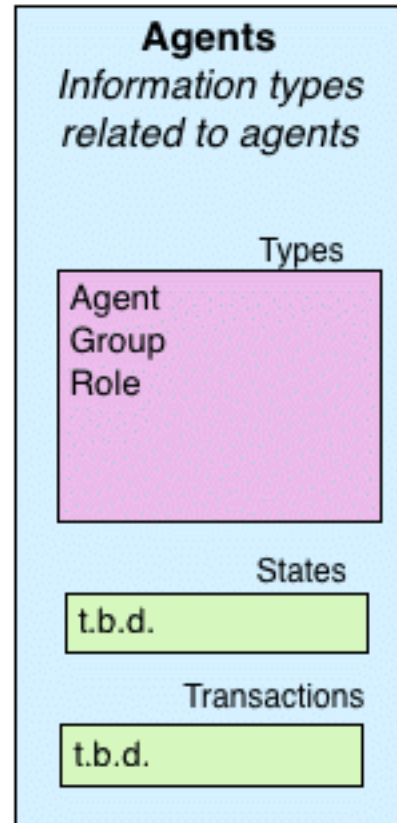
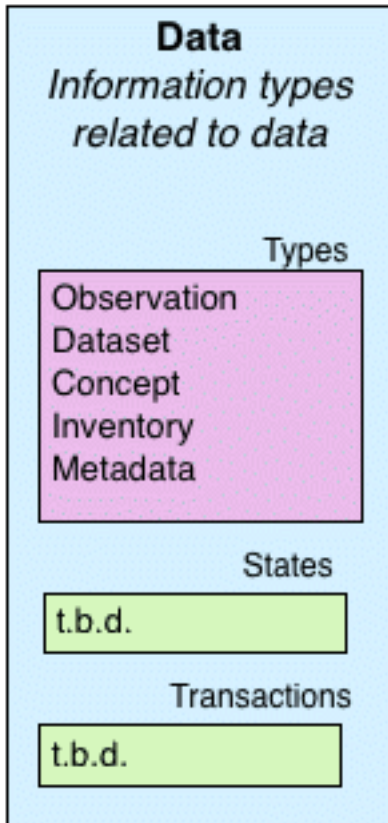
...to authenticate users.



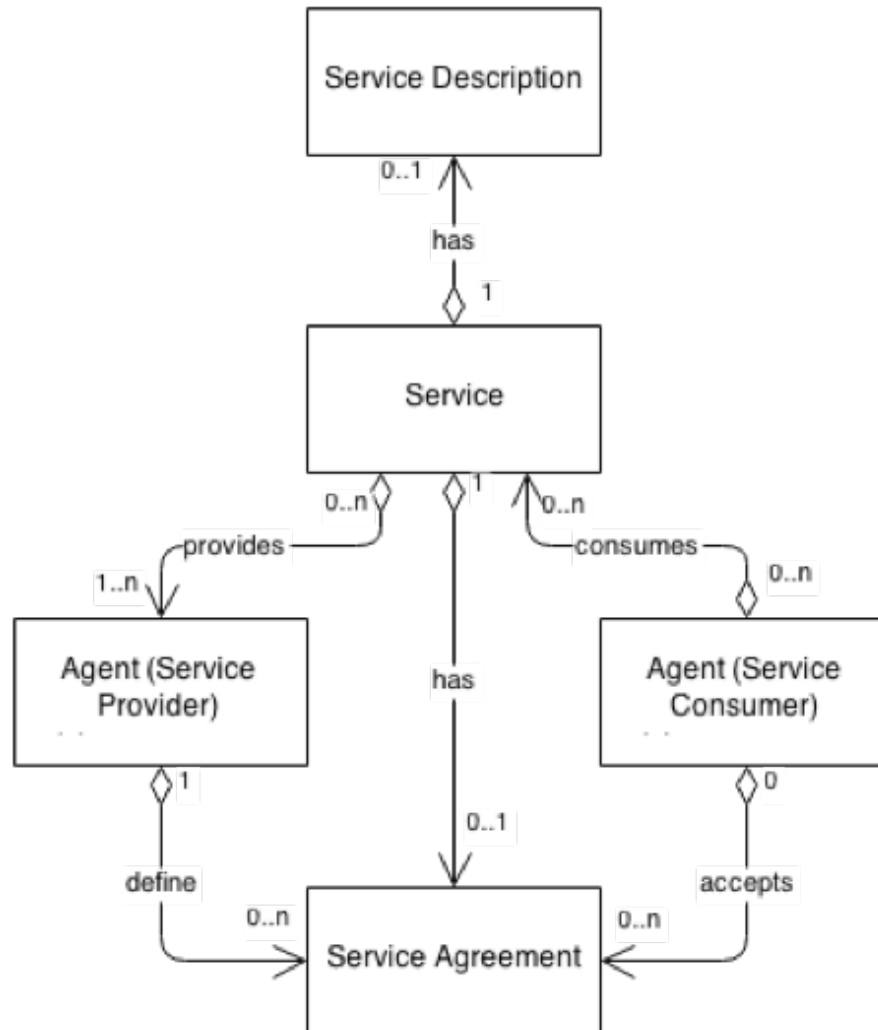
RM-SSH: Information Viewpoint

The **information viewpoint** focus is on the information **without** considering representation, implementation or distribution details or platform-specific execution details to create a **common abstract model** for the **shared** information in the system. This viewpoint is **independent** from **functions that transform** and manipulate the **data** and the **computational interfaces**. This viewpoint defines a configuration of **information objects**, their **behaviours**, **actions** that can be performed upon the information object and **constraints**.

RM-SSH: Information Viewpoint



RM-SSH: Information Viewpoint

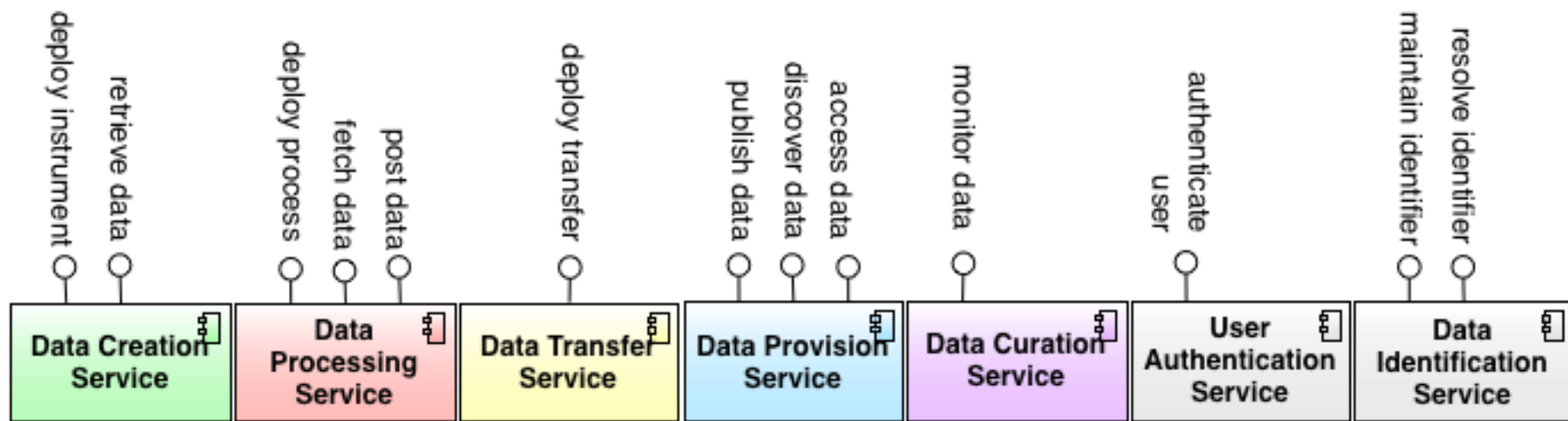




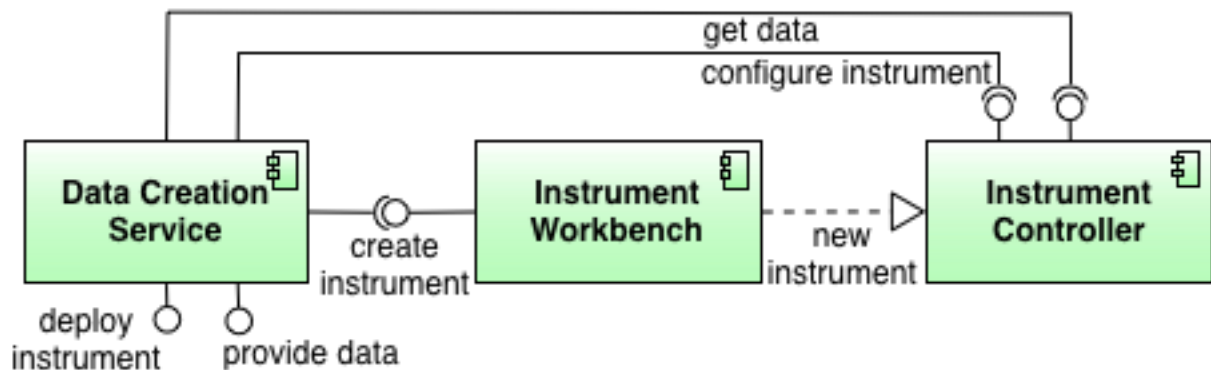
RM-SSH: Computational Viewpoint

The computational viewpoint identifies the prime **computational objects** and the **interfaces** that can be utilised for interaction with the object. Here we consider a computational object to be a **set of connected services** and **resources** to support **interaction** between members of the various communities as well as researchers' interaction with data. The infrastructure can here be seen as a chain of interactions through interfaces provided by the computational objects.

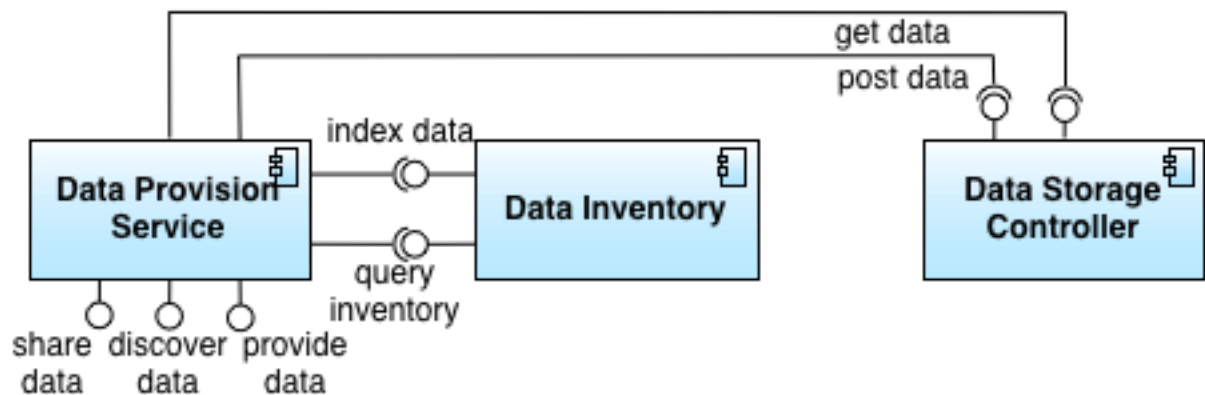
RM-SSH: Computational Viewpoint



RM-SSH: Data Creation Subsystem



RM-SSH: Data Provision Subsystem

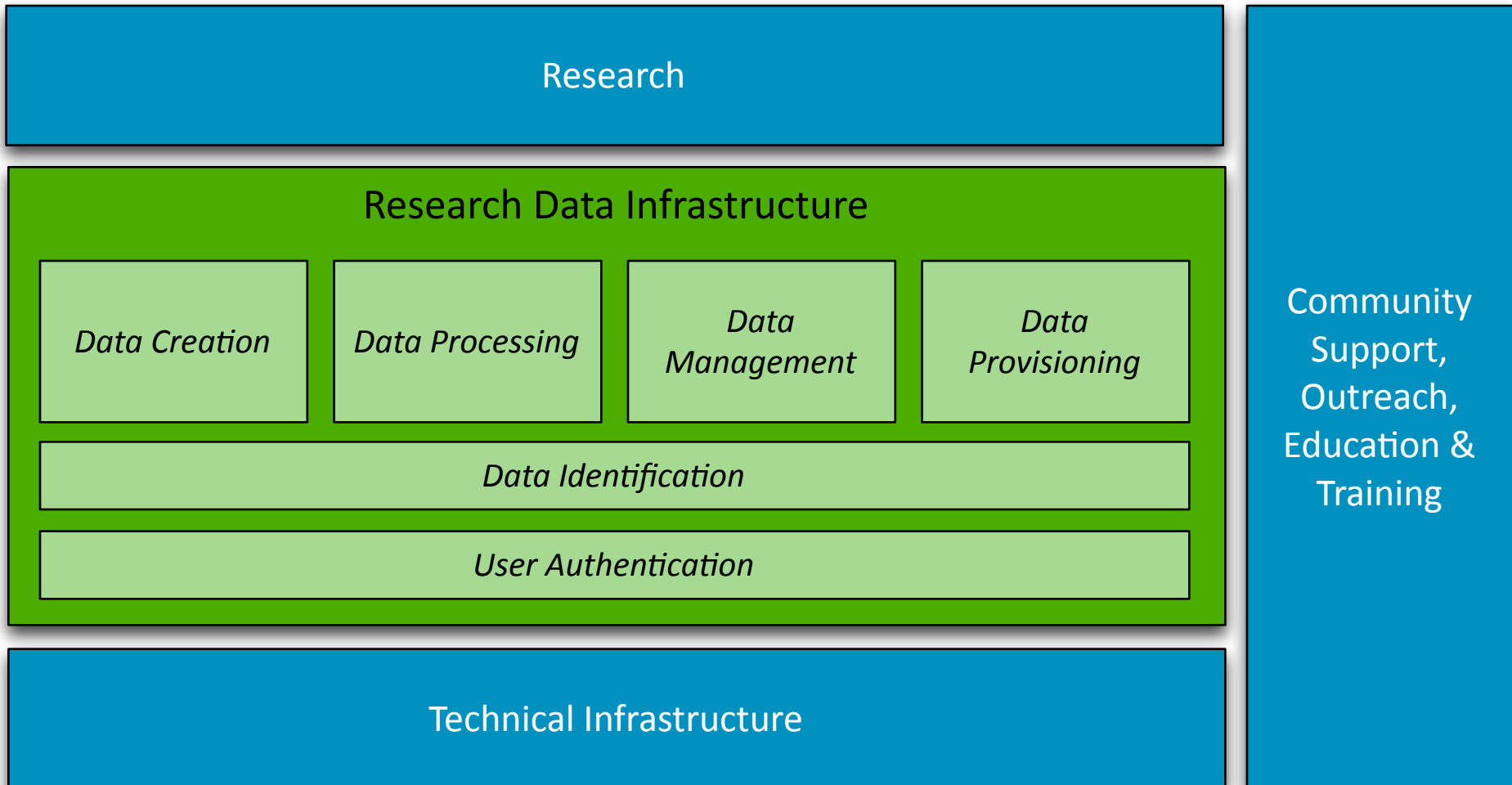




So what?

- RM-SSH should provide support for describing reference architectures and architectures in SSH domain.
- A common language across the domain for RIs.
- High abstraction but should cover all examples of research data infrastructures.
 - from 1 researcher with a laptop.
 - to an internationally distributed full data lifecycle supporting RI.
- However, currently incomplete for Research Infrastructures...

Research & Data Infrastructures





What next

- SSH RIs still in early stages and still (public) architectures do not exist.
- Currently being used on two projects
 - DARIAH Humanities at Scale project for reference architecture for in-kind contribution descriptions
 - European Holocaust Research Infrastructure for the “EHRI Information Architecture.”
- These efforts will contribute to and test the RM-SSH model.
- **<https://sites.google.com/a/dans.knaw.nl/reference-model-for-ssh-data-infrastructure/home>**



Mike Priddy - DANS -Data Archiving & Networked Services - mike.priddy@dans.knaw.nl

Maarten Hoogerwerf - DANS -Data Archiving & Networked Services

John W. Shepherdson - UKDA - UK Data Archive

Johan Fihn - SND - Swedish National Data Service

Birger Jerlehag - SND - Swedish National Data Service

Timo Gnadt - SUB - State & University Library, Georg-August-Universitaet Göttingen.

Thank you for your attention.
Any questions?

<https://sites.google.com/a/dans.knaw.nl/reference-model-for-ssh-data-infrastructure/home>