

# Information Architecture For Secure Trustworthy Data Services

Data Services of the Future (Session D3)

Hervé L'Hours: CESSDA Trust &  
UK Data Archive, UK Data Service,  
University of Essex

09 June 2022/ IASSIST  
2022

cessda.eu



**DATA BY DESIGN**

IASSIST 2022

Full (Original) Title

Information Architecture

for

Secure Trustworthy Digital Repositories

Quality Culture

*not* Standards Theater

# Work in Progress

Journey towards:

Disciplinary, Trustworthy & FAIR-enabling  
(Findable, Accessible, Interoperable, Reusable)

Repository and (meta)data Services

# Standards Theatre vs. Quality Culture

Standards theatre (cf: Security Theatre)

- Giving the impression of commitment to, and compliance, with standards practice.
- Superficial and presentational rather than practical

Quality Culture: **Service**-oriented

- Driven by mission, goals, outcomes delivered effectively and efficiently
- Depends on understanding of the wider context of why we do what we do

“Minimal documentation to some purpose”

# Information Architecture

- Organising, structuring and labelling content
- Filing and finding
- Managing change

✗ the repository data and metadata

✓ Business information

About people, processes, technology: policies, procedures, reference information, transactional information, sections, meetings, projects, groups.

**“Process-lead”**

More stable than organisational structure, sections, departments etc

# Wider Scoping for Repositories & Data Services

## Information Architecture

- Local organisational context

## Information **Infrastructure** Architecture

- Research Lifecycle and Ecosystem Context

# Idealised Research Lifecycle

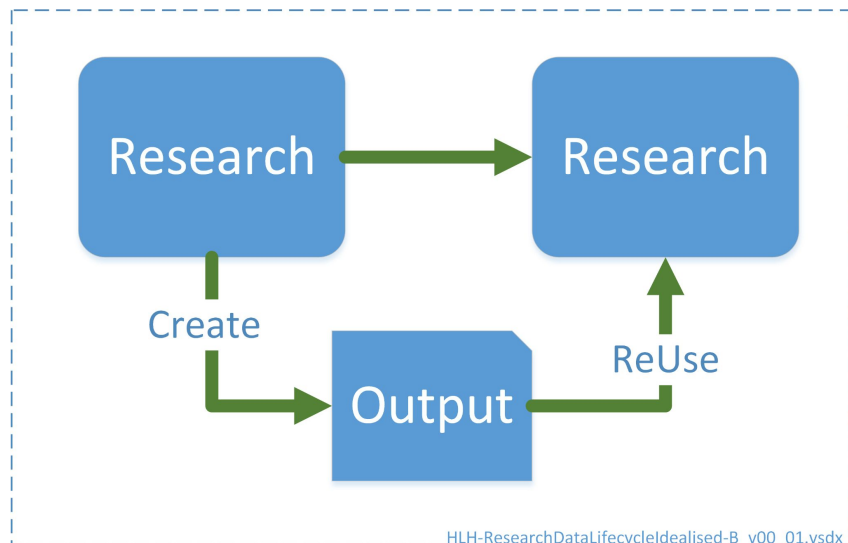
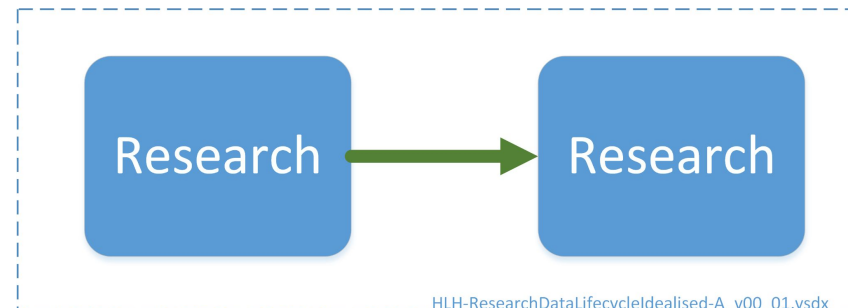
Researchers

Research > More Research

But...

Outputs for ReUse

Physical and Digital (focus on Digital)



# Outputs: Digital Objects & Curation

Digital Objects. Bits. meta(data) Data and/or Metadata.

Research “data” Management: Raw and processed data, papers & publications, instruments, software and code, controlled vocabularies, ontologies, methods

## Curation

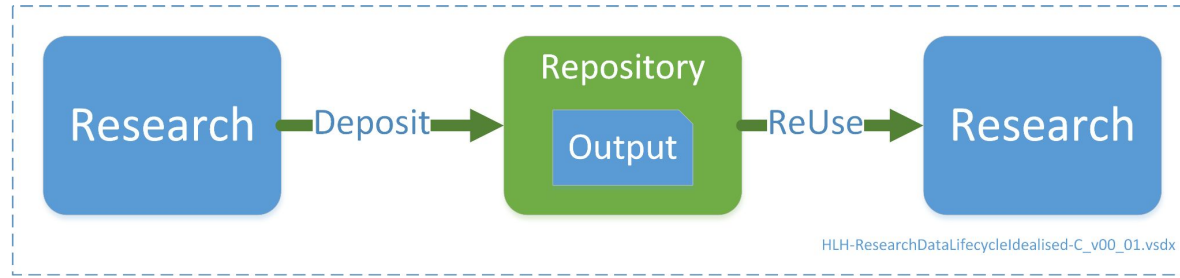
- **Context:** Storage, backups, access control of the bits
- **Structure:** File types, and formats and schemas: Logical, technical
- **Conceptual:** Content, meaning, understanding

Enter the Repository (also Galleries, Libraries, Archives, Museums etc)



# Curation

Deposit-Curate-Access-ReUse



Snapshot in time: ready for deposit, ready for access, ready for publication ready for reuse. Now!

**But:**

Time & Change: People, Processes, Technology

# Curation (Snapshot) vs. Preservation (+Time & Change)

EOSC Task Force  
Long-Term Digital Preservation  
**Minimalist Lifecycle Perspective**



# Preservation & Standards

OAIS Reference Model, Trustworthy Digital Repository (TDR) standards, ISO16363, DIN31644 and CoreTrustSeal.

## **CoreTrustSeal**

Research Data Alliance (RDA): single, “core” level, low barrier to entry TDR standard  
16 Requirements. Organizational Infrastructure, Digital Object Management, Technology/Security. V3.0 in progress\*

\*This presentation uses the changes currently proposed by the CoreTrustSeal Board

# CoreTrustSeal: A brief history

2008/2009

2013

2017

WDS Certification of  
Regular Members



Data Seal of Approval Certification  
of Trusted Data Repositories

RDA Repository Audit and Certification  
DSA–WDS Partnership WG



Expectation or prerequisite  
for, e.g.

- WDS Regular Members Accreditation
- CLARIN B Centres
- CESSDA Service Providers



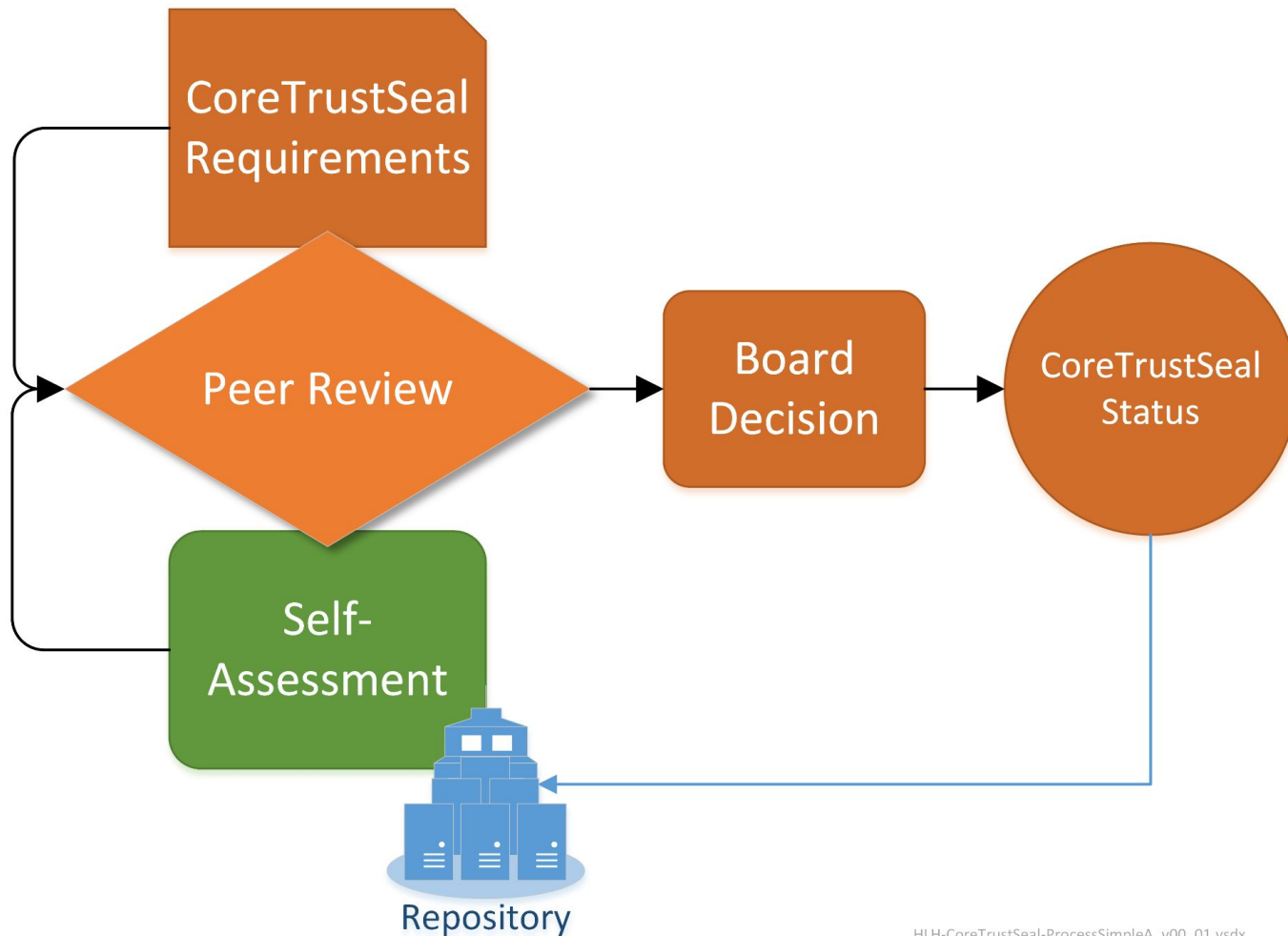
# Time & Change: People, Processes, Technology

**Sustainable Trustworthy Repository Data Services** offering *Active Preservation*. Necessary to maintain digital objects over time.

**Active preservation.** Usable and understandable by a defined 'designated community'

CoreTrustSeal adopted by: World Data System regular members, CLARIN B Centres, CESSDA Services Providers. 120+ Certified Repositories





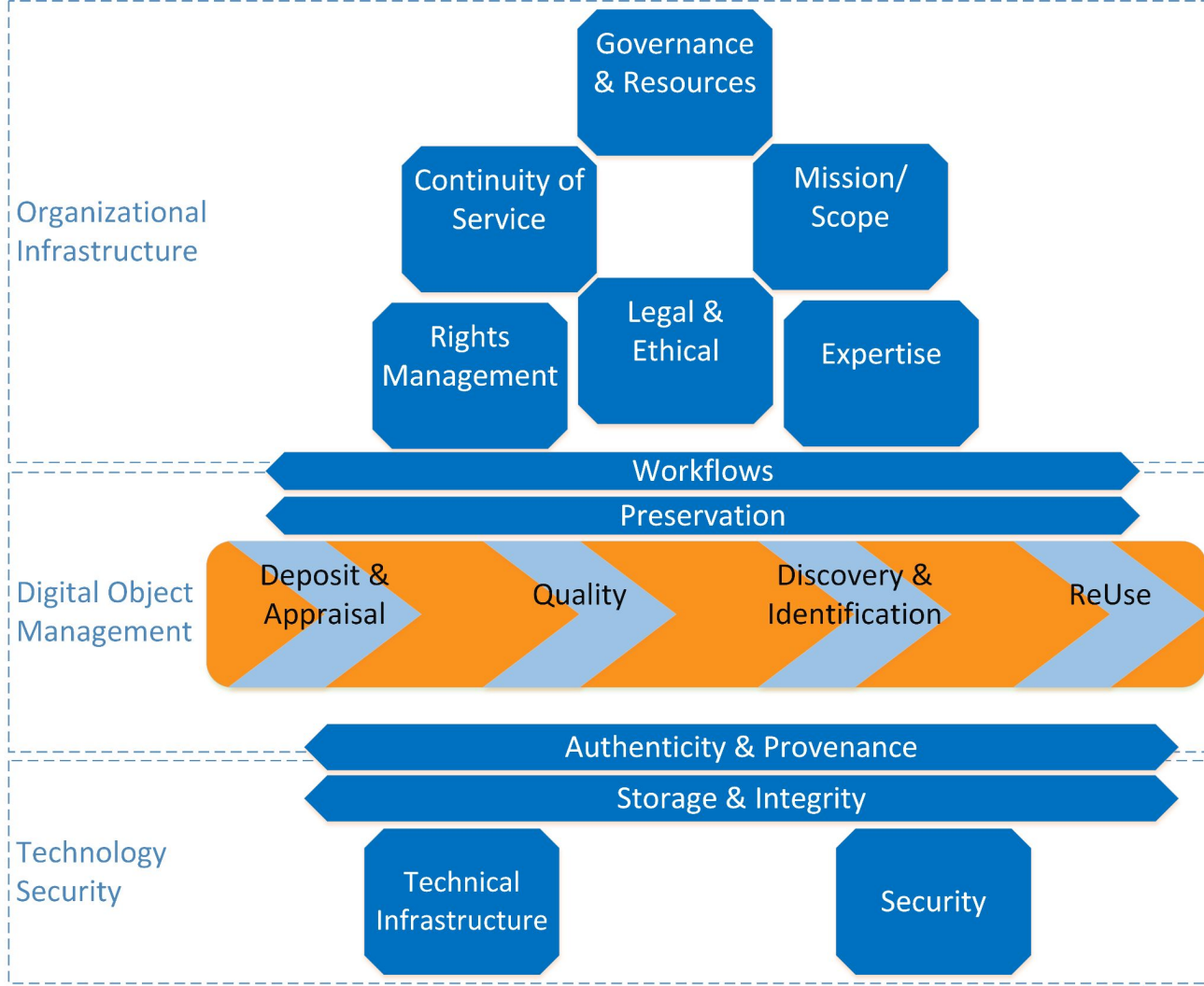
# CoreTrustSeal Requirements

## Version 2.0 vs Version 3.0 Proposals

### CoreTrustSeal v2.0 to v3.0 Changes Overview

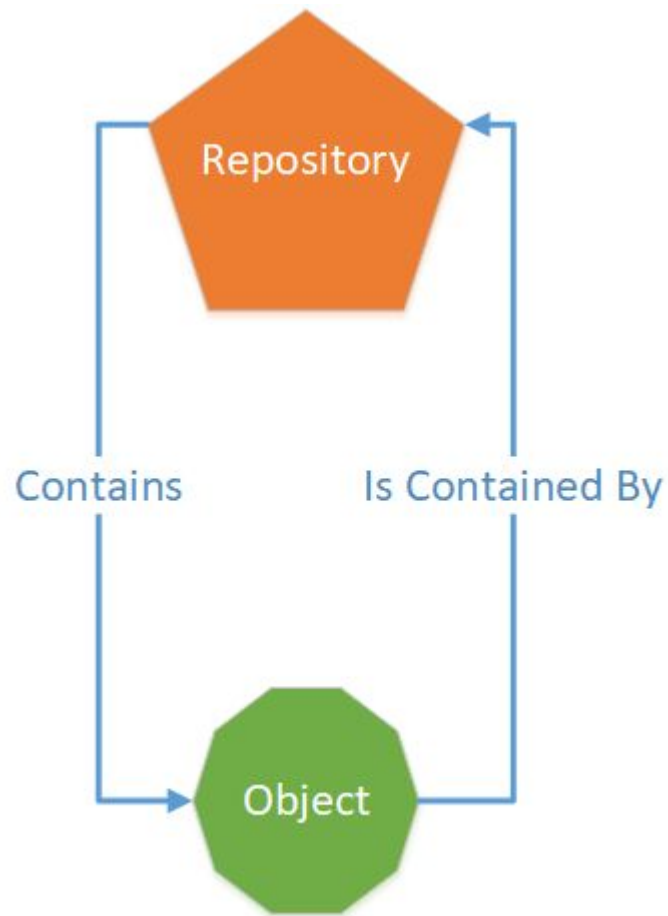


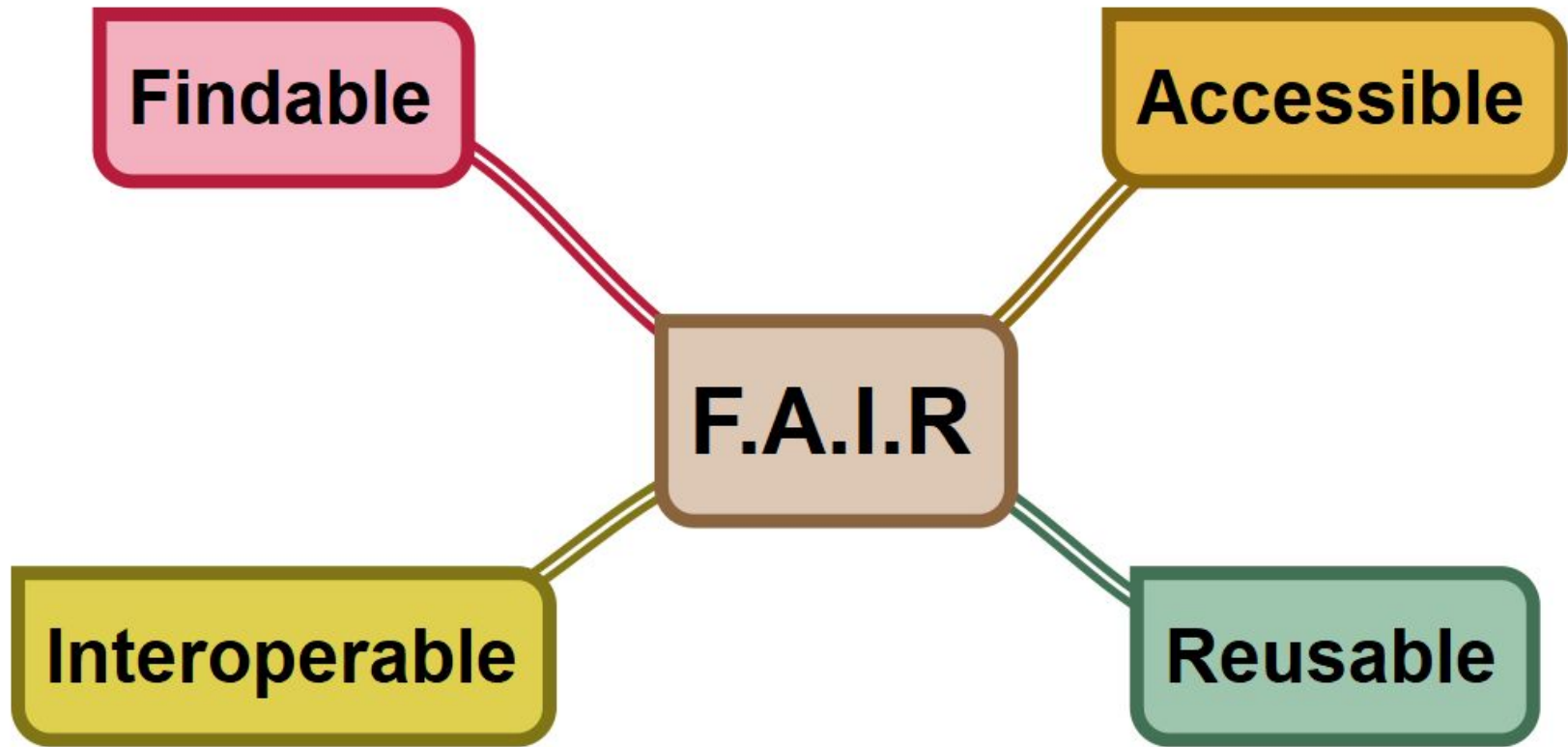
# CoreTrustSeal Requirements





# Repository: Objects in Context





**Findable**

**Accessible**

**F.A.I.R.**

**Interoperable**

**Reusable**

# FAIR Principles

## Findable

F1. Meta(data) Identifier:  
Globally Unique/Persistent

F2. Rich Metadata Descriptions

F3. Metadata: Specifies  
Data Identifier

F4. Searchable Resource:  
Indexed/Registered

## Accessible

A1. Meta(data):  
\*Retrievable via ID  
\*Standard Comms Protocols

A1.1. Protocol: Open,  
Free, Universal

A1.2. Protocol: Auth/Auth  
(if necessary)

A2. Metadata Remain Accessible,  
when Data not Available

## Interoperable

I1. Meta(data) Knowledge  
Representation:  
\*Formal  
\*Accessible  
\*Shared  
\*Broadly Applicable Language

I3. Meta(data) Qualified References to  
other (Meta)data

## Reusable

R1. Meta(data) Plurality of Attributes:  
\*Rich  
\*Accurate  
\*Relevant

R1.1 Meta(data) Data Usage License:  
\*Clear  
\*Accessible

R1.2 Meta(data) Provenance

R1.2 Meta(data) Domain-relevant  
Community Standards

I2. Meta(data) Vocabularies:  
Follow FAIR Principles

# Enter the FAIR Principles

Findable, Accessible, Interoperable ReUsable: so far, so familiar to repositories

- 15 Principles
- Indicators (RDA)
- Metrics (e.g. FAIRsFAIR)
- Tests (e.g. F-UJI tool)

Turning FAIR into Reality: Repository Context

Recognised **CoreTrustSeal** as an exemplar **standard** and assessment **process**, also noted future dependencies on **Registries**.

4 Letters  
F.A.I.R

15  
Principles

40+  
Indicators

??  
Metrics

??  
Tests

# FAIR & the European Open Science Cloud (EOSC)

Support for achieving CoreTrustSeal.

- EOSC-Nordic (regional)
- SSHOC (disciplinary)
- FAIRsFAIR (broad)

FAIR (snapshot) vs FAIR+Time (Requires Preservation)

CoreTrustSeal+FAIR Enabling Repositories (objects and their context)

- Network of FAIR Trustworthy Data Repositories
- EOSC Association Long Term Data Preservation Task Force
- FAIR IMPACT

# Mapping CoreTrustSeal (2.0) to the FAIR Principles



- F1. (meta)data are assigned a globally unique and persistent identifier.
- F2. data are described with rich metadata (defined by R1 below).
- F3. metadata clearly and explicitly include the identifier of the data it describes.
- F4. (meta)data are registered or indexed in a searchable resource.

## **R13. Data discovery and identification**

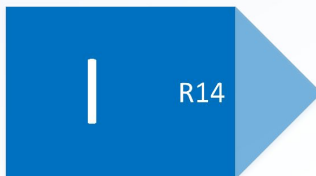


- A1 (meta)data are retrievable by their identifier using a standardized communications protocol.
  - A1.1 the protocol is open, free, and universally implementable (vs context)
  - A1.2 the protocol allows for an authentication and authorization procedure, where necessary.
- A2 metadata are accessible, even when the data are no longer available.

## **R15. Technical infrastructure**

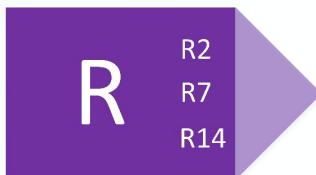
## **R16. Security**

## **R10. Preservation plan**



- I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
- I2. (meta)data use vocabularies that follow FAIR principles.
- I3. (meta)data include qualified references to other (meta)data.

## **R14. ReUse**



- R1. meta(data) are richly described with a plurality of accurate and relevant attributes.
  - R1.1. (meta)data are released with a clear and accessible data usage license.
  - R1.2. (meta)data are associated with detailed provenance.
  - R1.3. (meta)data meet domain-relevant community standards.

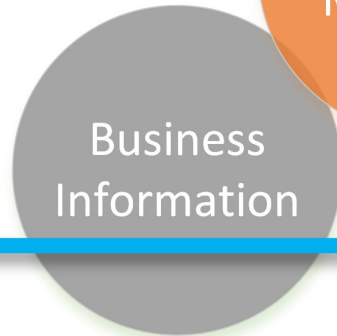
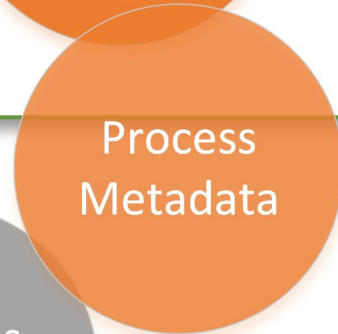
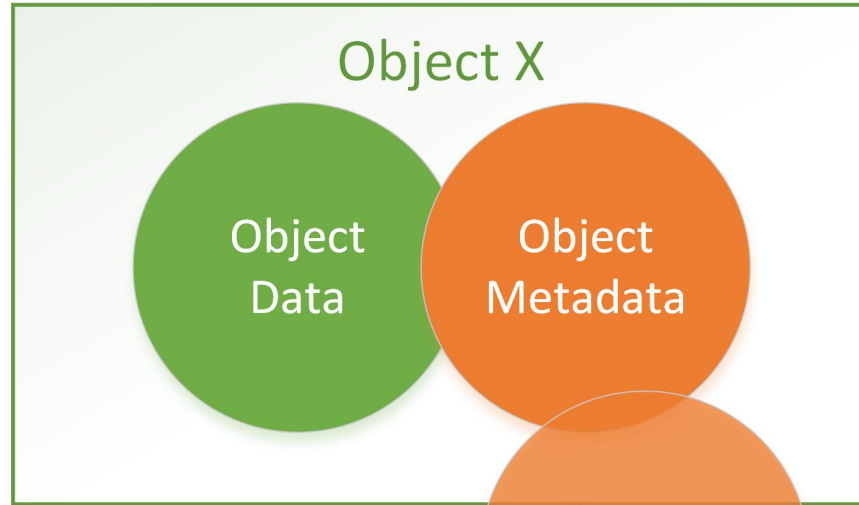
## **R14. ReUse**

## **R2. Licenses**

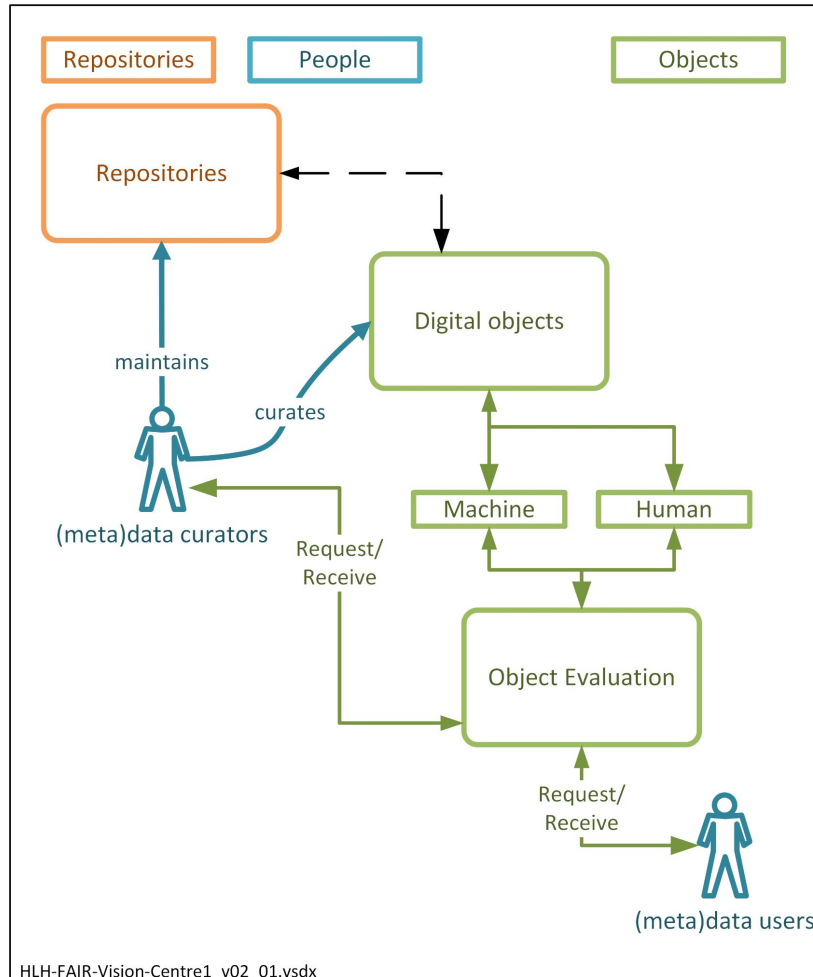
## **R7. Data integrity and authenticity**

## **R14. ReUse**

# Repository...





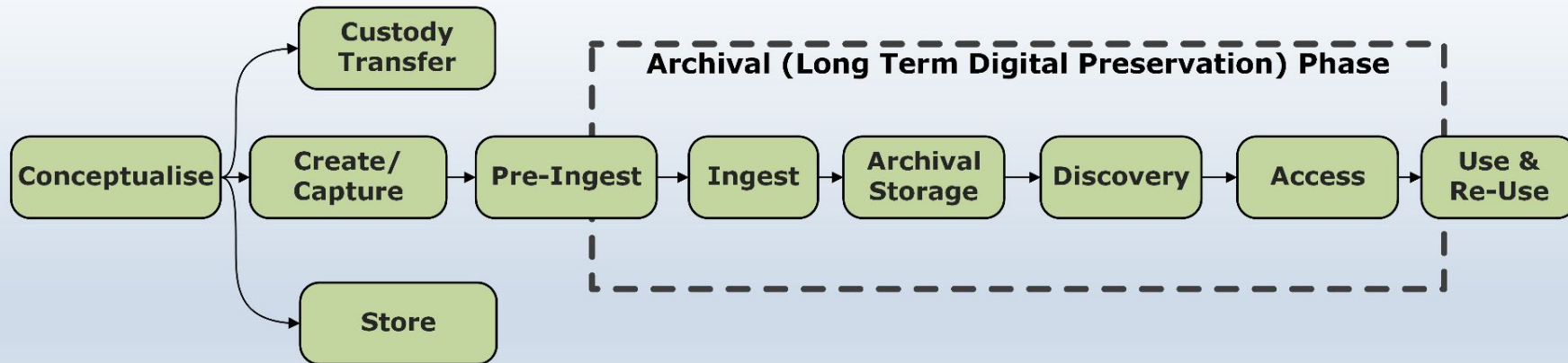


# Attention, Journeys and Scope

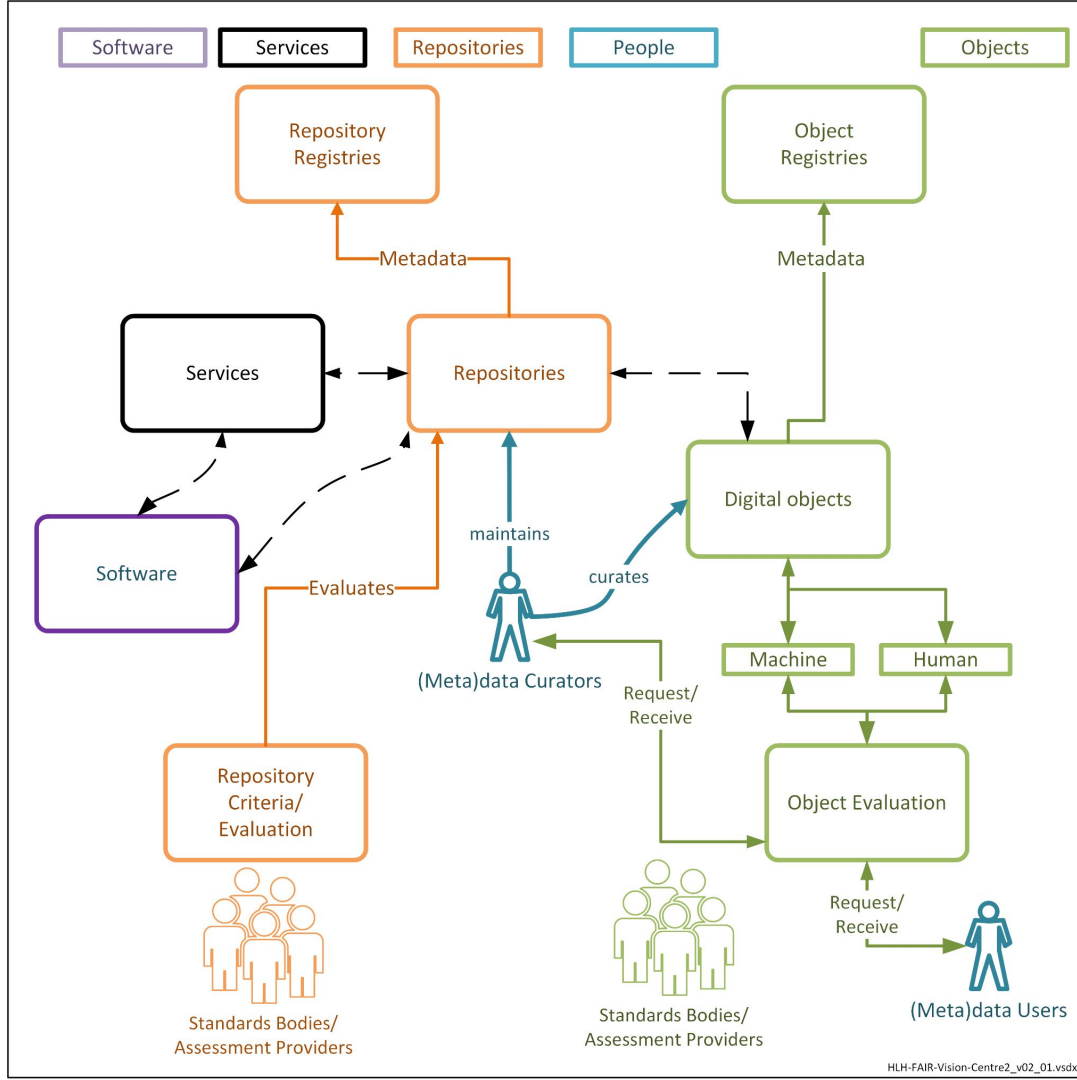
- Lots of attention on repositories, practices, standards and certification
- Recognition that FAIR and Trust are an ongoing journey
- But the lifecycle and ecosystem are made of other services too

**Different functions and services are offered by different organisations through complex data flows and partnerships.**

Sequential Actions across curation and archiving systems



Conception to Use/Re-Use including curation/preservation actions, versioning, identification, publication and citation



# Service assessment

Maintaining Trustworthiness and FAIRness/FAIR-Enabling?

- Technical: FAIRenabling, Quality, Open & Connected
- Socially: Use-centric, Transparent, Longevity, Ethical & Legal

To apply **standards, evaluate** and **assess compliance**

(from “self-assess” to “certify”)

depends on knowing what we mean by “This Service”

# UK Data Service, UK Data Archive

UK Data Archive. CoreTrustSeal Certified Trustworthy Digital Repository (TDR)

“The UK Data Archive provides the CoreTrustSeal certified Trusted Digital Repository component to the UK Data Service with a mission to provide **reliable, long-term access to managed digital resources.**”

Lead partner in **UK Data Service**. UKDS is the UK Service Providers to **CESSDA**.

## Vision

“To continue to be a critical part of the UK’s research infrastructure where the exceptional economic and social data we **make available** are central to the achievement of excellence in research, teaching and in the realisation of public benefit.”

## Mission

“To **support** high quality social and economic research, teaching and learning through **assuring** long term **access** to quality economic and social data, supporting and **promoting** their use, value and impact.”

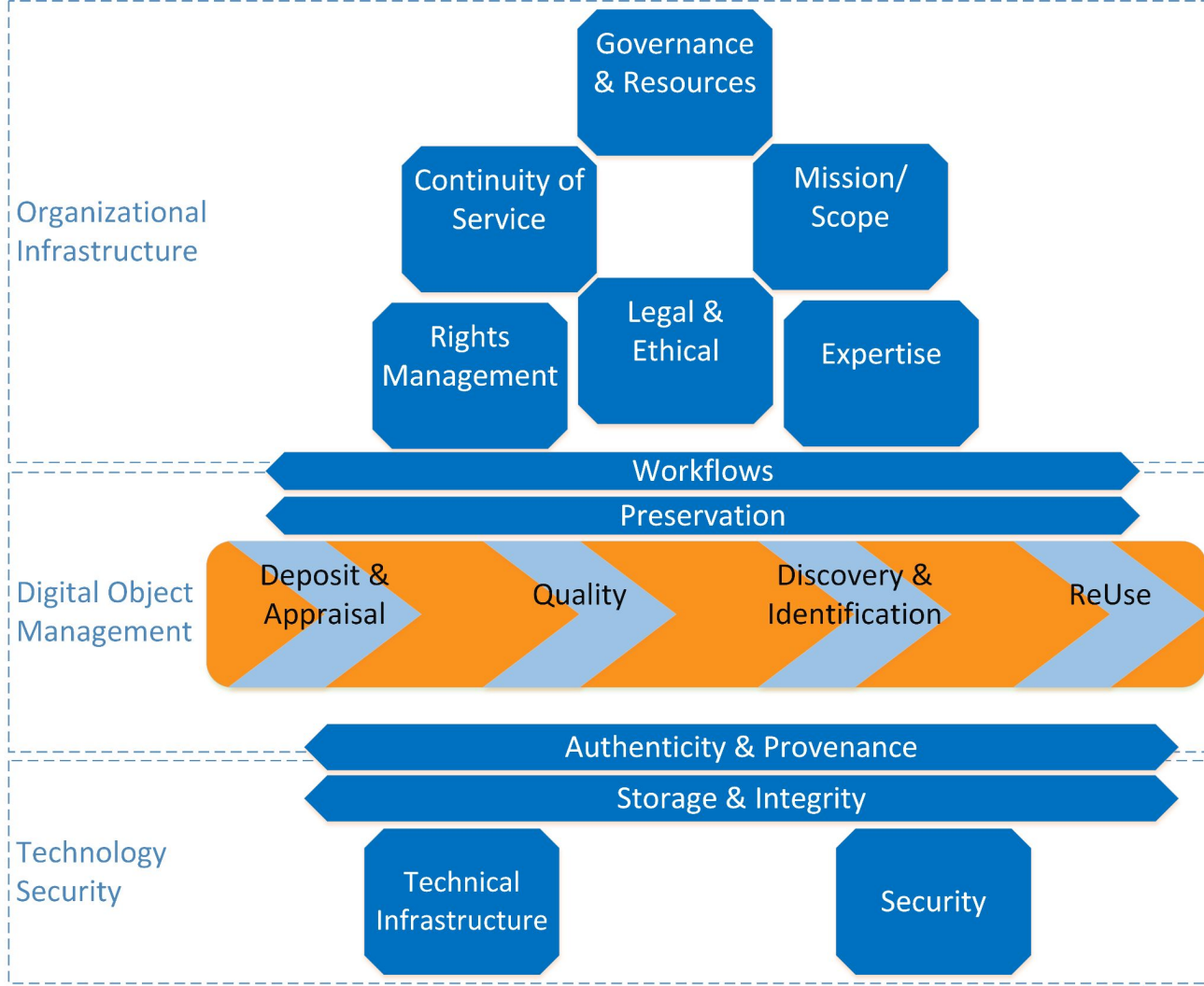
# The UK Data Service Services are at your Service

Internal services, deposit, access and support services, products vs services, technical (IT Service Management) services.

Services as groups of:

- Inputs
- Functions
- Processes
- Activities
- Workflows
- Outcomes

# CoreTrustSeal Requirements



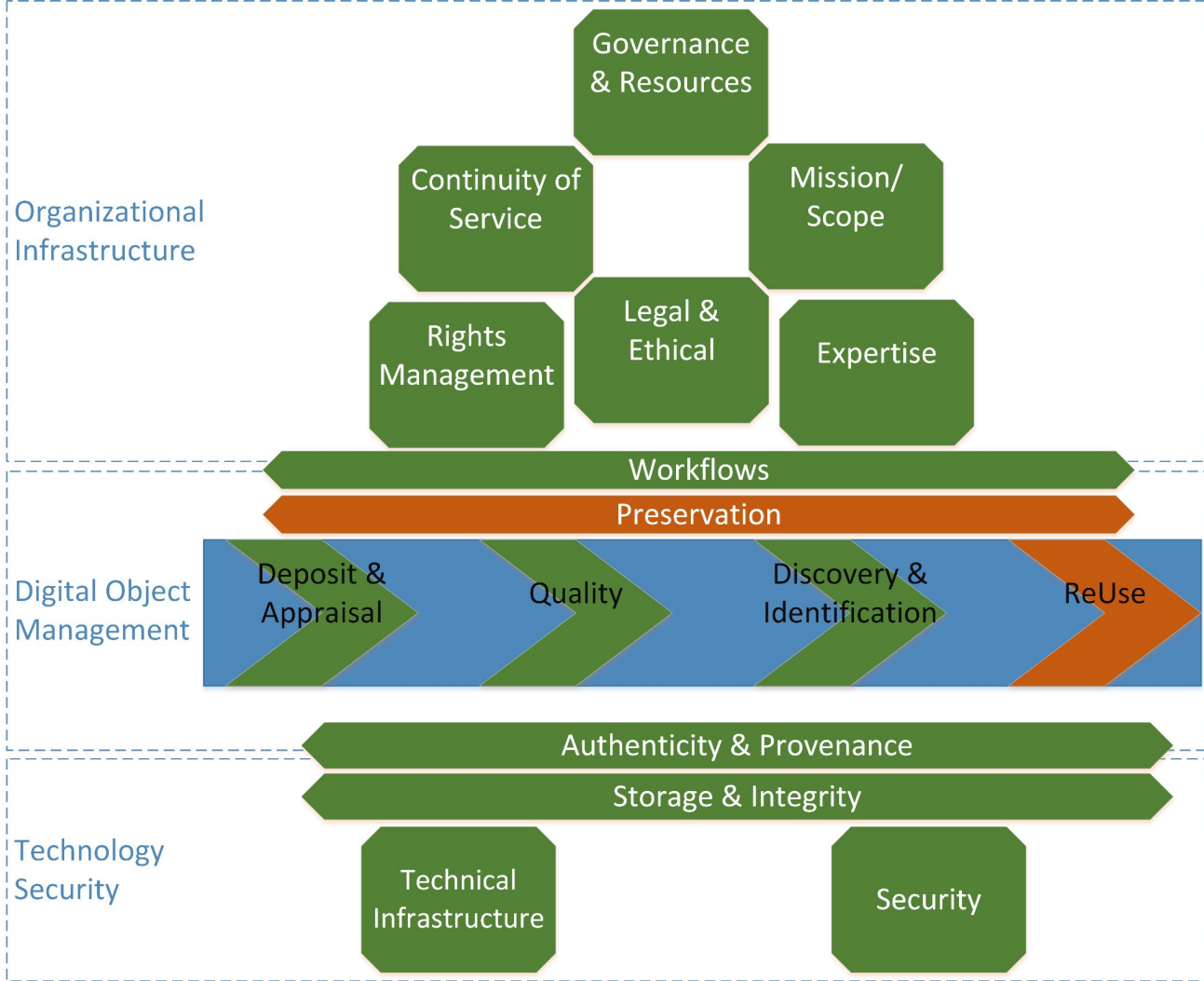


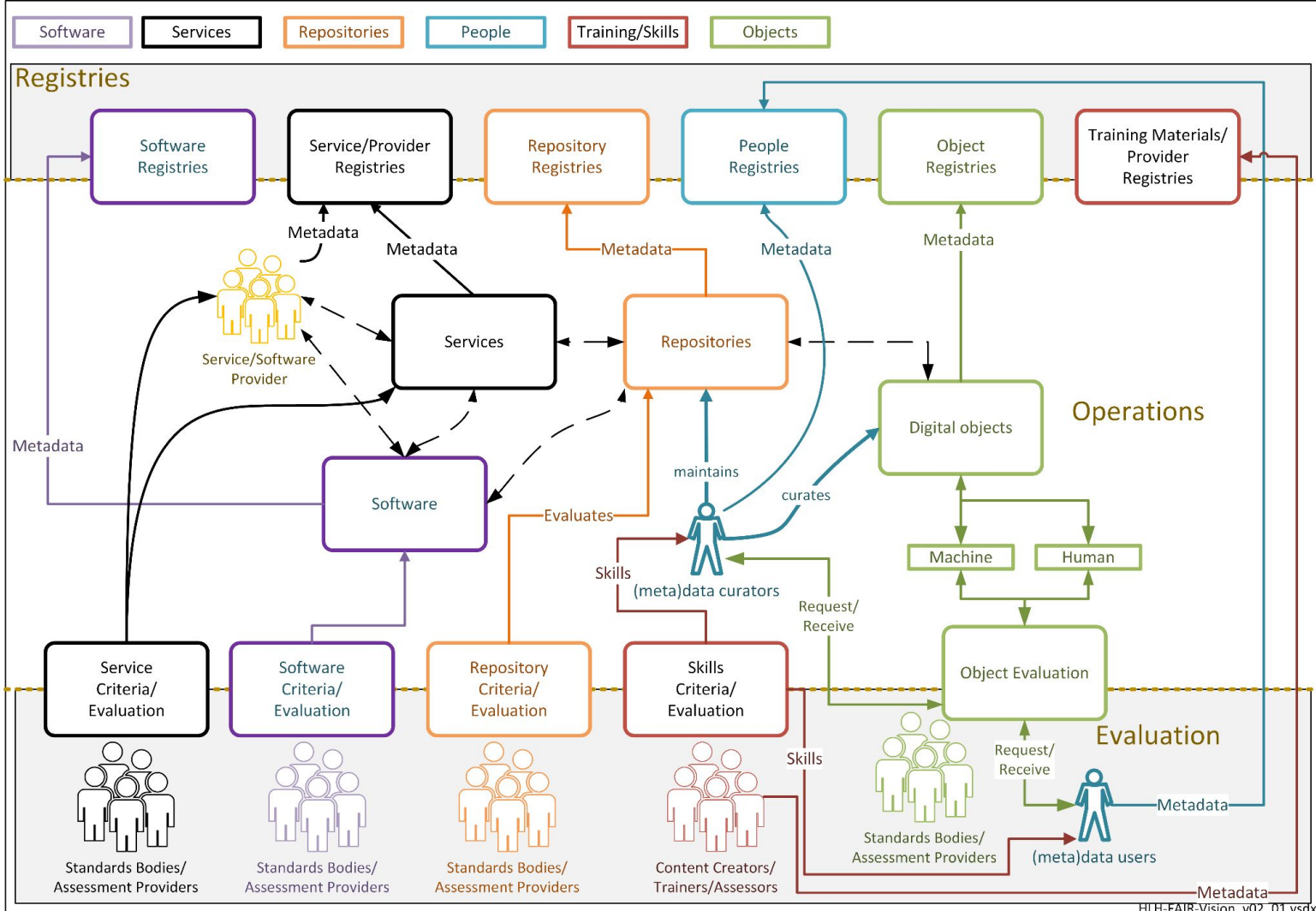
# CoreTrustSeal for Data Services?

- No Long Term Preservation (LTDP) Services (x Preservation Plan)
- Non-Curation Services (x ReUse) Snapshot
- 14 Requirements applicable to any deposit, storage and access system that 'touches' data or metadata.

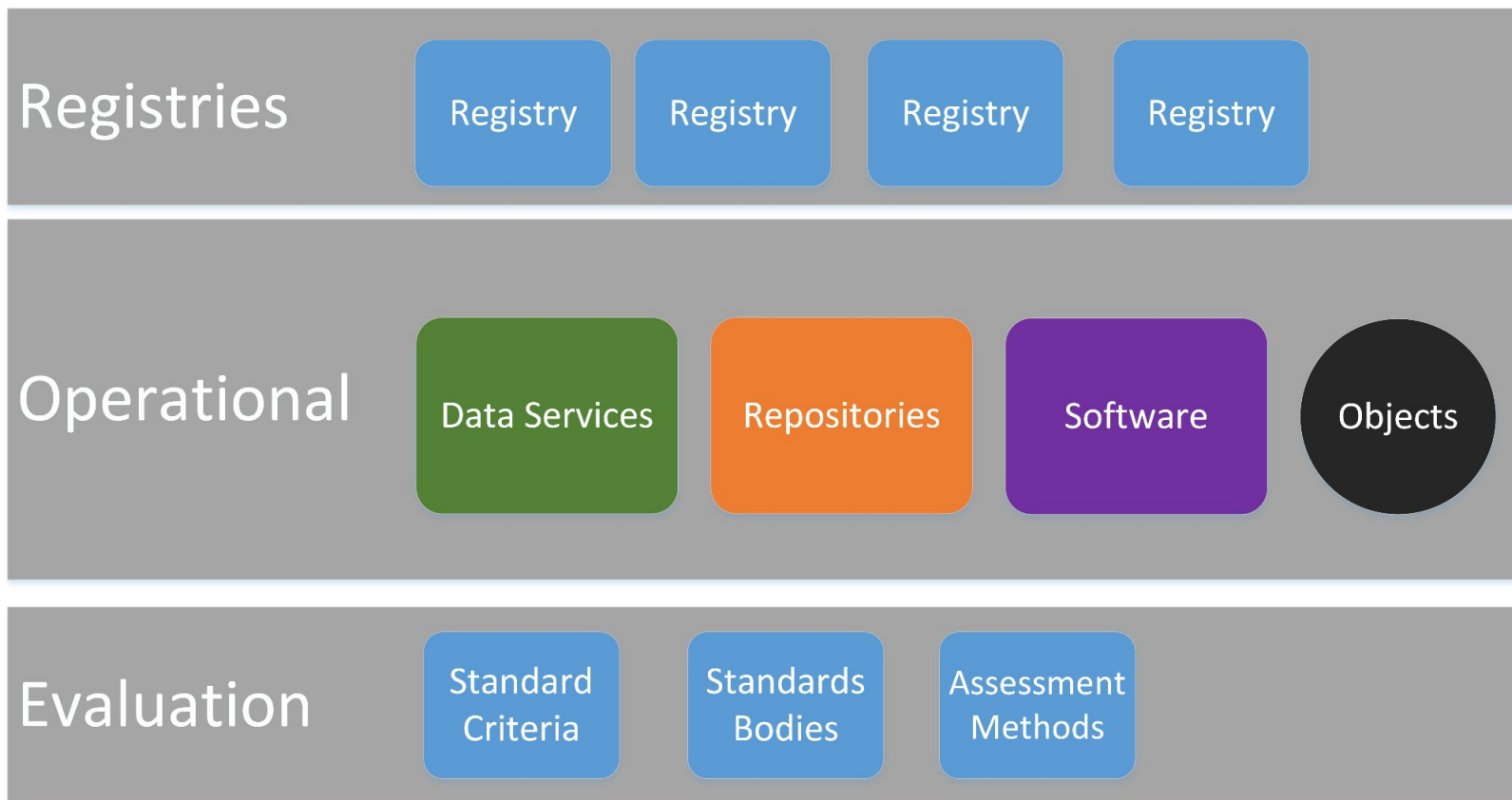
# CoreTrustSeal Requirements\*

\*Everything in green applies to wider data services





# Operations, Standards & Assessment, Registries



# Beyond the “core” and beyond the CoreTrustSeal?

## **Beyond the “core”**

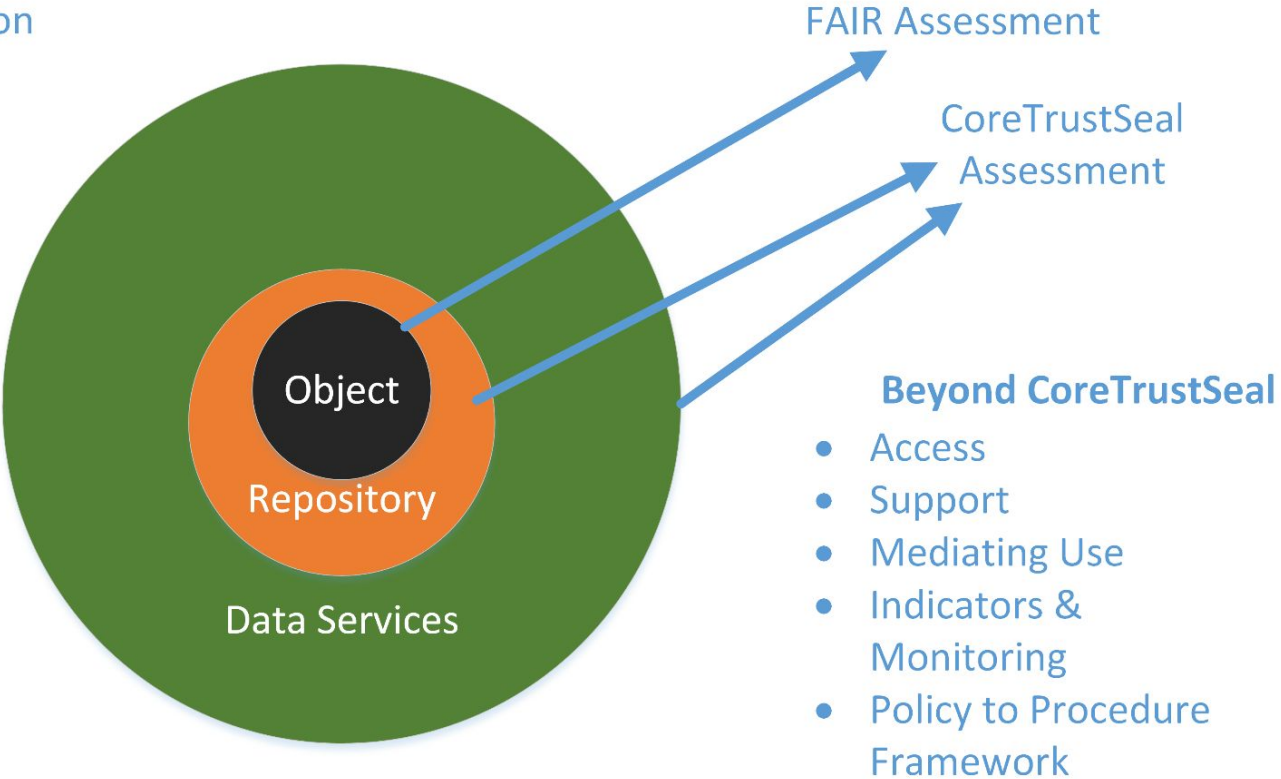
- Sensitive data handling Information Security (ISO27001),
- IT Service Management (FitSM)

## **Beyond the CoreTrustSeal**

- Access (implied)
- Support (& Training)
- Mediating Use (Secure Access & Safe Room)
- Indicators & Monitoring (data driven operations & Impact)
- Policy to Procedure Framework (Implied) for practice and evidence

# Transparency, Efficiency, Interoperability:

- Registries
- Standards and Evaluation



## Beyond the “Core”

- IT Service Management
- Information Security

## Beyond CoreTrustSeal

- Access
- Support
- Mediating Use
- Indicators & Monitoring
- Policy to Procedure Framework

# Quality Culture for Trustworthy Data Services Information Architecture?

- Respond to change, predict the future, prepare future services
- Technical Watch, Community Watch, Landscape monitoring
- Transparency and expertise
- Shared resources and responsibilities to ensure best practice
- Exemplar metadata registries and emerging standards and assessment

Services offer Services to Services

# Services offer Services to Services

- Agreement on functions/processes that make up “services”
- Ensure Trust (and FAIR) beyond the TDRs
- Help define shared resources (registries)
- Help define shared standards and assessment for consistency and transparency of practice

Findable and Accessible... but

- **Interoperable and ReUsable digital objects depend on interoperable and reusable services, cooperating efficiently and at scale.**



Utgång, förföljd av björn



Sortie poursuivie par l'ours...