# Semantic Mapping in ENVRI and EPOS

Keith G Jeffery
keith.jeffery@keithgjefferyconsultants.co.uk

ENVRI and EPOS

- **ENVRI: 27 RIs variously ERICs, Landmarks...**
- **Atmosphere, Hydrosphere, Geosphere, Biosphere**

- **EPOS: 1 RI (ERIC)**
  - **~250 asset suppliers**
  - **Central homogeneous rich metadata catalog**
  - **Ingestion through conversion of ~17 heterogeneous metadata formats to EPOS-DCAT-AP and hence to CERIF**

- **ENVRI now building a catalog for all ENVRI RIs based on EPOS technology**
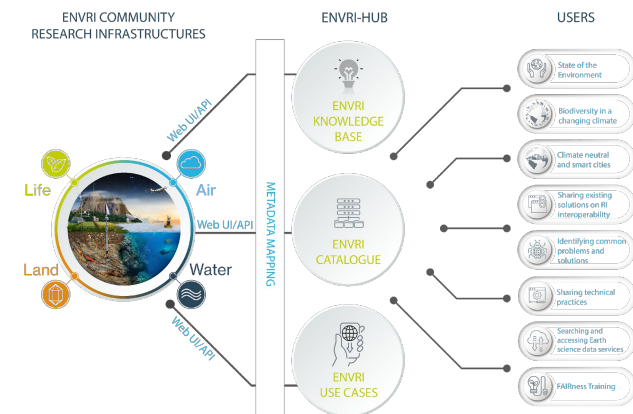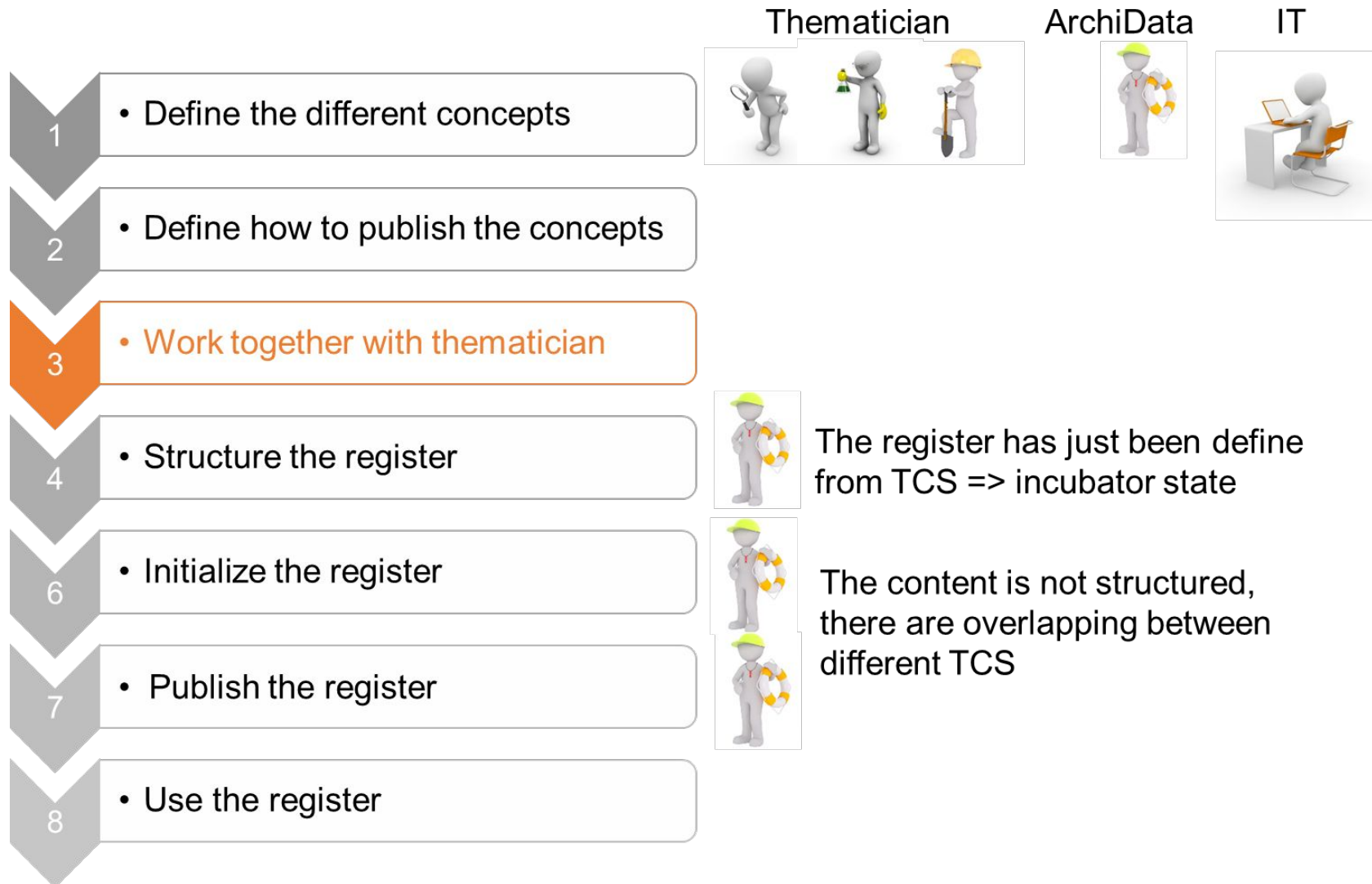
Aspects of the work

- **Vocabularies**
- **Harmonisation (as much as possible)**
- **Subset of common terms**
- **Catalog**

EPOS
EUROPEAN PLATE OBSERVING SYSTEM

Vocabularies

What has been done

Thematician   ArchiData   IT

1. • Define the different concepts
2. • Define how to publish the concepts
3. • Work together with thematician
4. • Structure the register
6. • Initialize the register
7. • Publish the register
8. • Use the register

The register has just been define from TCS => incubator state

The content is not structured, there are overlapping between different TCS

ENVRI
FAIR

Incubator register

Why an intermediate stage?

The entities are defined as concepts with label and description

Relations broader and narrower between entities are done

But everything is mixed up together

Harmonisation

Harmonisation

- Group working on this in EPOS
- And also in other RIs in ENVRI
  - Especially Oceanography (Alexandra Kokkinaki, BODC)
- Very slow and human resource-heavy work
- Is it worth it – should we not do mappings?

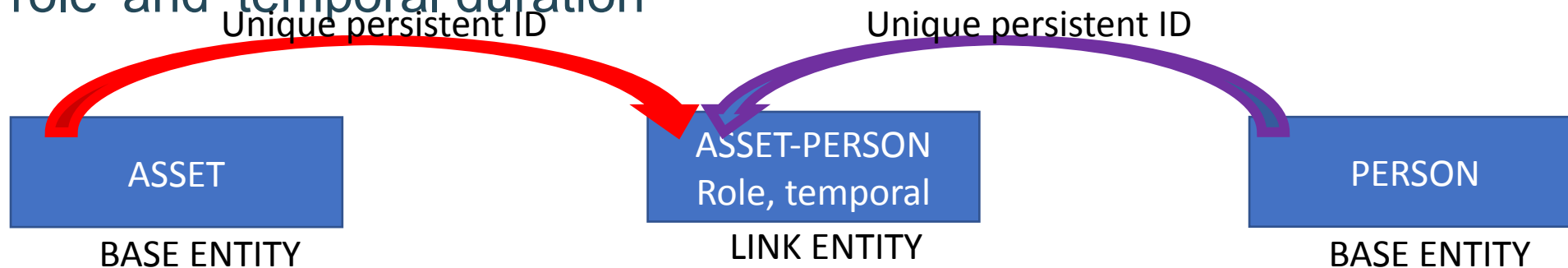# Common Subset

Common SUbset

- In ENVRI proposed by oceanography domain (Katrina Exter, VLIZ) for use in the ENVRI catalog

- A very small number of agreed terms common to all subdomains (atmosphere, hydrosphere, geosphere, biosphere)

- Ongoing work, almost complete

- Provides some homogeneity in discovery of digital assets across the heterogeneous subdomains and even more heterogeneous RI digital assets

Catalog

CERIF: the basic model

- **Base entities**: objects in the real world e.g. Asset, Person which have attributes

- **Link Entities**: relationships between base entities with attributes including 'role' and 'temporal duration'

Unique persistent ID   Unique persistent ID

ASSET	ASSET-PERSON Role, temporal	PERSON
BASE ENTITY	LINK ENTITY	BASE ENTITY

- Example Link entity instance:

-   ASSET   ROLE   TEMPORAL INTERVAL
  PERSON

- <Paper X>[<authored by> <19970128 00:00:00><19980812 00:00:00>]<Person P>
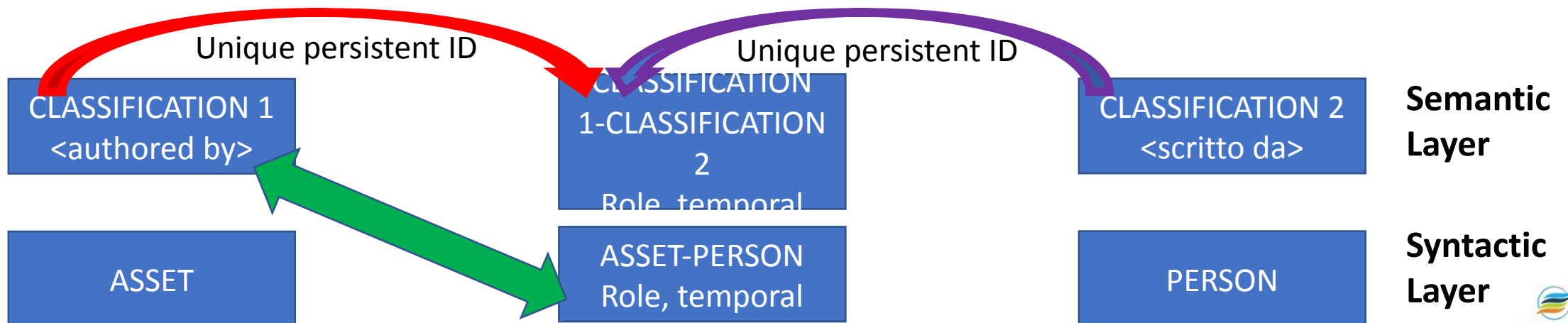
- (optionally, also add probability as an attribute – modal logic)

ENVRI
FAIR

CERIF: syntactic and semantic layers
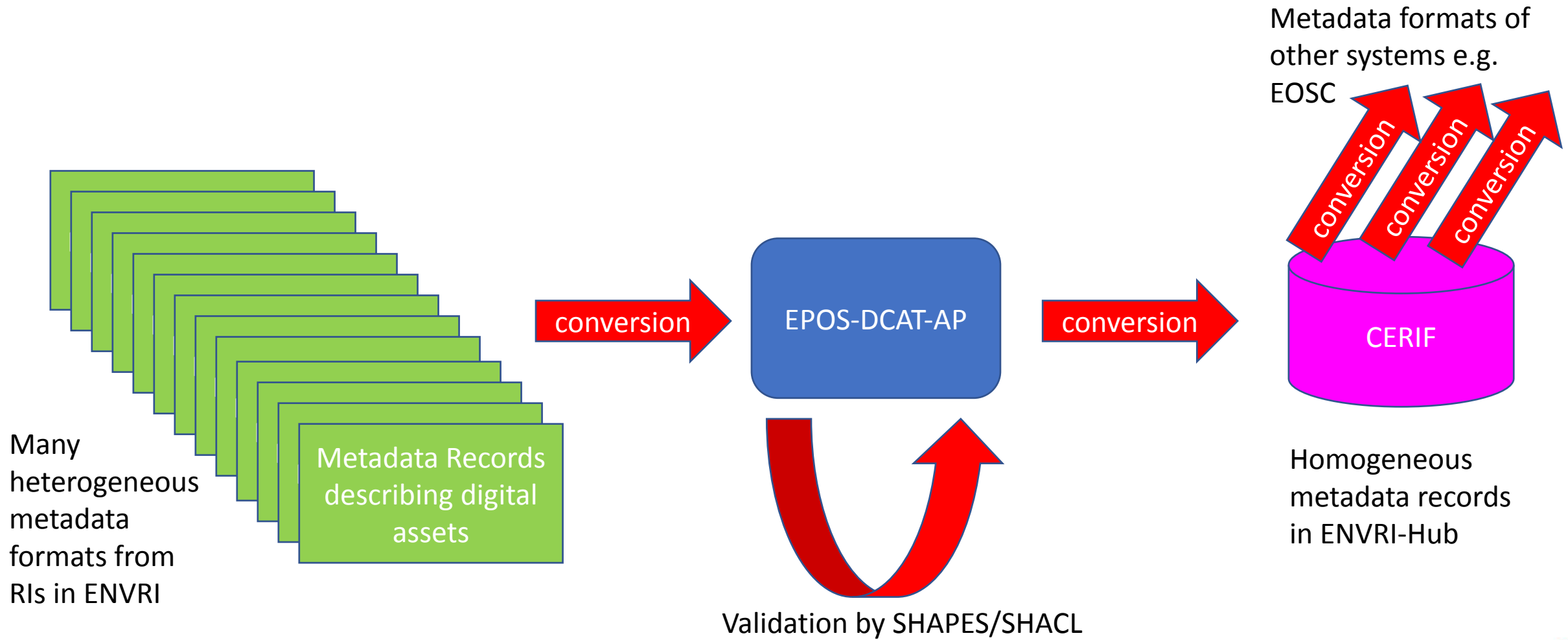
- Syntactic layer: referential and functional integrity
- Semantic Layer: semantic consistency and interoperability

- In fact 'authored by' is not represented as a text string in the role attribute of the link entity but as a UUID which points to the semantic layer where a role 'authored by' is stored in a base entity and can be related to textual terms in the same or other base entities by link entities thus allowing full ontological capability including multilinguality (see "scritto da")
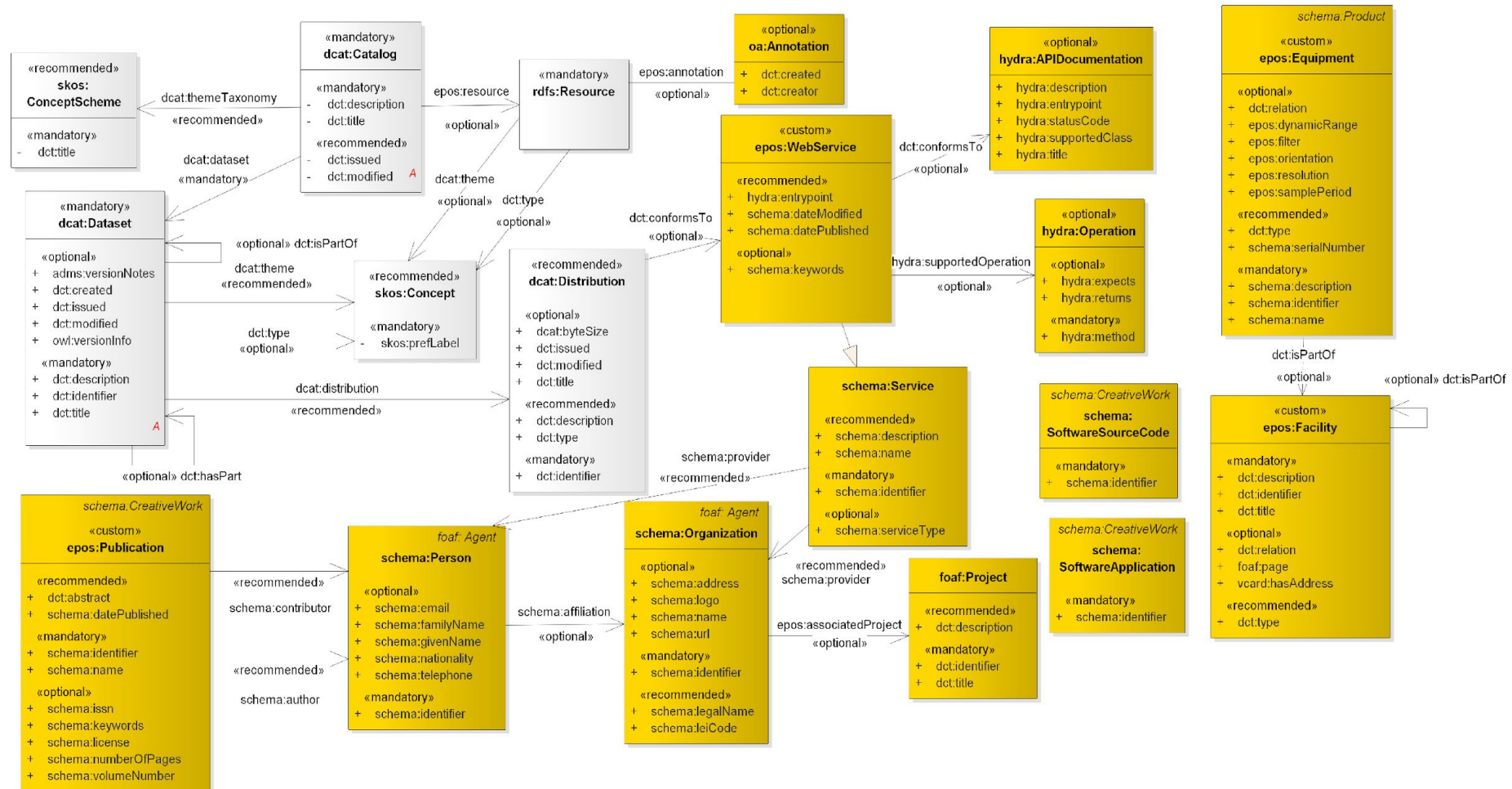
Unique persistent ID   Unique persistent ID

| CLASSIFICATION 1 <authored by> | CLASSIFICATION 1-CLASSIFICATION 2 Role, temporal | CLASSIFICATION 2 <scritto da> | **Semantic Layer** |
| ASSET | ASSET-PERSON Role, temporal | PERSON | **Syntactic Layer** |

ENVRI
FAIR

Metadata Ingestion Pipeline

Catalog In...

This is still very much work in progress....

Keith G Jeffery
keith.jeffery@keithgjefferyconsultants.co.uk