Data Management
in Research Infrastructures

K. Atakan

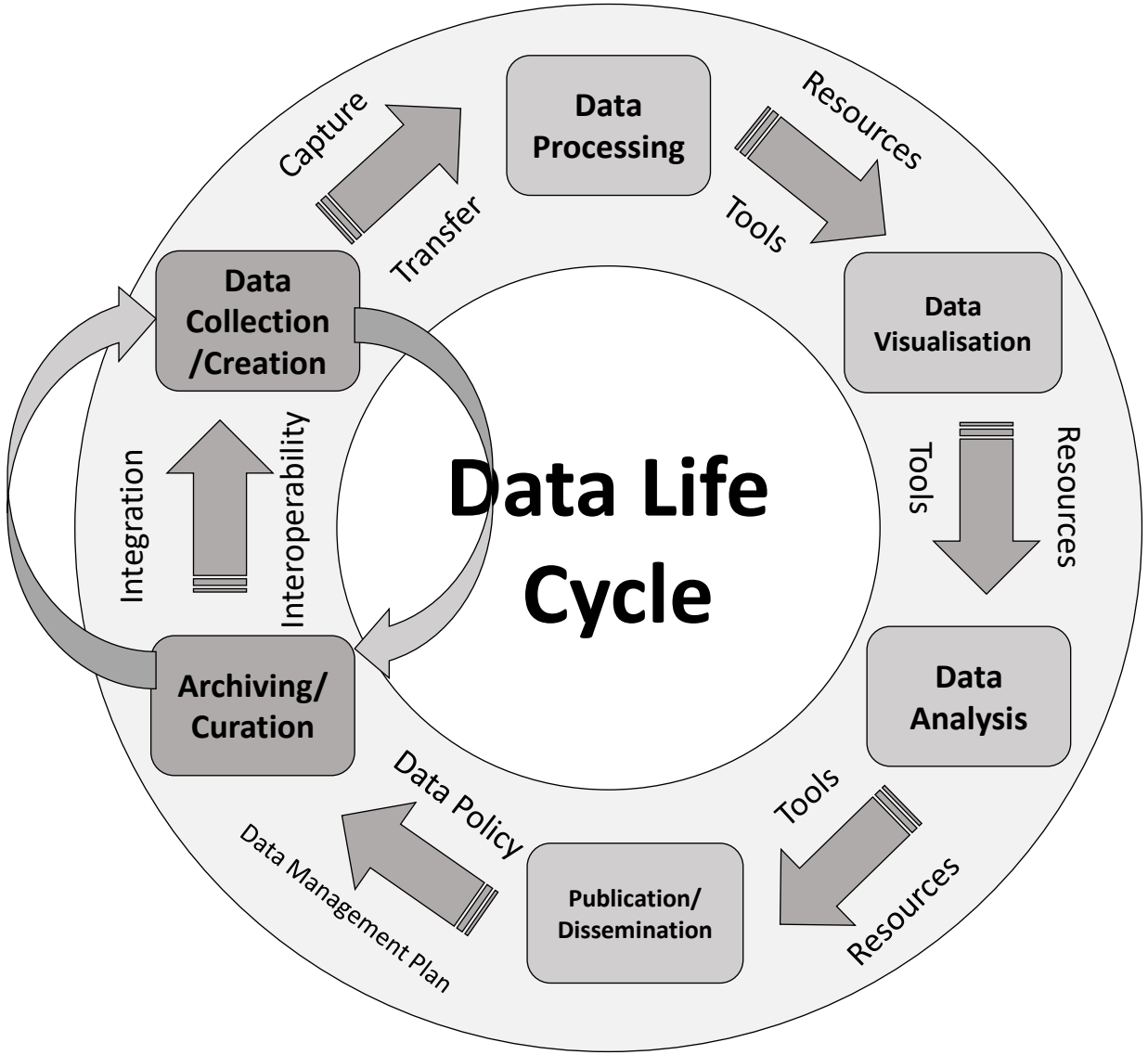
Department of Earth Science

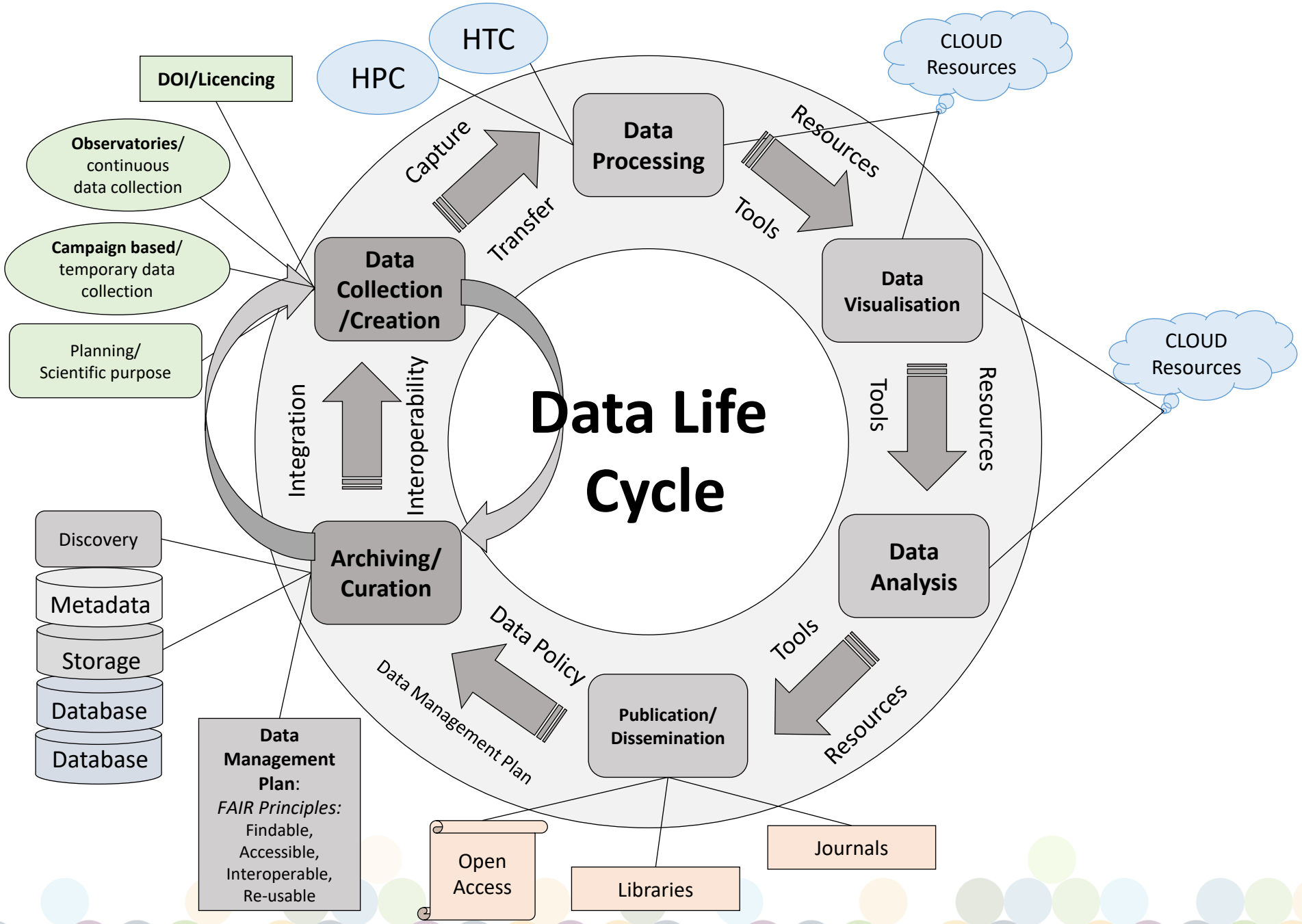
University of Bergen

Kuvvet.Atakan@uib.no

What is Data Management?

It all starts with understanding the
DATA LIFE CYCLE





Data Management has several dimensions:

- Technical
- Legal
- Governance
- Financial

- Strategic
- Policy
- Resource

- Security
- Privacy
- Sensitivity
- Ethical

Data Management - technical aspects

Technical aspects of Data Management

- **Storage, archiving and maintenance**, after data creation/generation
- **Curation** activities enable data **discovery and retrieval**, maintain quality, add value, and provide for re-use over time.

- **Access provisions** and regulations, data retrieval, AAAI
- **Availability**, discovery services, GUI

- **Web-services**, APIs
- **Metadata** and data standards

- **Data quality** (QA-systems), integrity, completeness, provenance information
- **Authoritative data**, certification (see below)

- **FAIR principles** (**F**indable, **A**ccessible, **I**nteroperable and **R**e-usable)

Data Management – Legal (L), Governance (G), Financial (F) aspects

- **Legal**
  - **Data ownership**, distinctions between data creators/collectors, data providers and service providers
  - **Identifiers, licencing**,
  - **Data handling** and data protections,
  - **Liabilities**, disclaimers, IPR

- **Governance**
  - **Data stewardship**, administration and management responsibilities
  - Implementation of data policies

- **Financial**
  - **Operational costs** of data management (cost categories, accounting)
  - Available financial resources and support
  - Conditions of financial support

Data Management - SSSPE

- **Strategic**
  - **Data policies**, **access rules/restrictions** (Open access: anonymous, authenticated, authorized, embargoed, restricted)
  - **Certification** (e.g. CoreTrustSeal, RDA-Trust)
  - Sustainability aspects in TLGF
- **Security**
  - Security certificates, Public key certificate
- **Sensitivity**
  - Sensitive data
  - Confidential data
- **Privacy**
  - GDPR
  - National regulations
- **Ethical**

Data processing may involve various processes, including:

- **Validation**
  - Ensuring that supplied data is correct and relevant.
- **Sorting**
  - arranging items in some sequence and/or in different sets."
- **Summarization**
  - reducing detailed data to its main points.
- **Aggregation**
  - combining multiple pieces of data.
- **Analysis**
  - the "collection, organization, analysis, interpretation and presentation of data."
- **Reporting**
  - list detail or summary data or computed information.
- **Classification**
  - separation of data into various categories.

Creating a **Data Taxonomy** may be needed

Useful Links for Data Management

OpenAIRE - Open Science training

http://catalogue.openaire.eu/service/openaire.open_science_training

OpenAIRE - Research Data Management (RDM) Handbook

https://www.openaire.eu/rdm-handbook

University of Bergen (UiB) – DMP Course 21 Oct.2020

https://www.uib.no/en/ub/137940/introduction-data-management-plan-dmp

SCIENCE EUROPE - PRACTICAL GUIDE TO THE INTERNATIONAL ALIGNMENT OF RESEARCH DATA MANAGEMENT

Research Council of Norway (RCN) – Data Management

https://www.forskningsradet.no/en/Adviser-research-policy/open-science/open-access-to-research-data/

Online links for DMP:

Digital Curation Centre – DMP Online help

https://dmponline.dcc.ac.uk/help

easy.DMP – EUDAT – online service for creating DMP

https://eudat.eu/catalogue/easyDMP

https://easydmp.eudat.eu/

https://www.sigma2.no/easydmp/how-to

easy.DMP - Uninett Sigma 2 – online service for DMP

https://www.sigma2.no/data-planning

Useful links for FAIR:

- FORCE11 – FAIR Principles
- https://www.force11.org/group/fairgroup/fairprinciples
- GO FAIR: The Internet of FAIR Data and Services
- https://www.go-fair.org/resources/internet-fair-data-services/

Data Management in EPOS

EPOS Data Management Plan

- In order to provide a meaningful **Data Management Plan (DMP)**, it is necessary to take into account the **distributed nature of the EPOS** Research Infrastructure.
- The **EPOS architecture** encompasses research infrastructures at different levels: **local, regional, national, transnational and Pan-European**.
- **EPOS goal is to harmonize** them into a single framework and make data accessible through a unique European platform **following the FAIR (findable, accessible, interoperable, reusable) principles**.
- A **first release of the EPOS DMP**, in the form of **guidelines**, is provided to the communities.

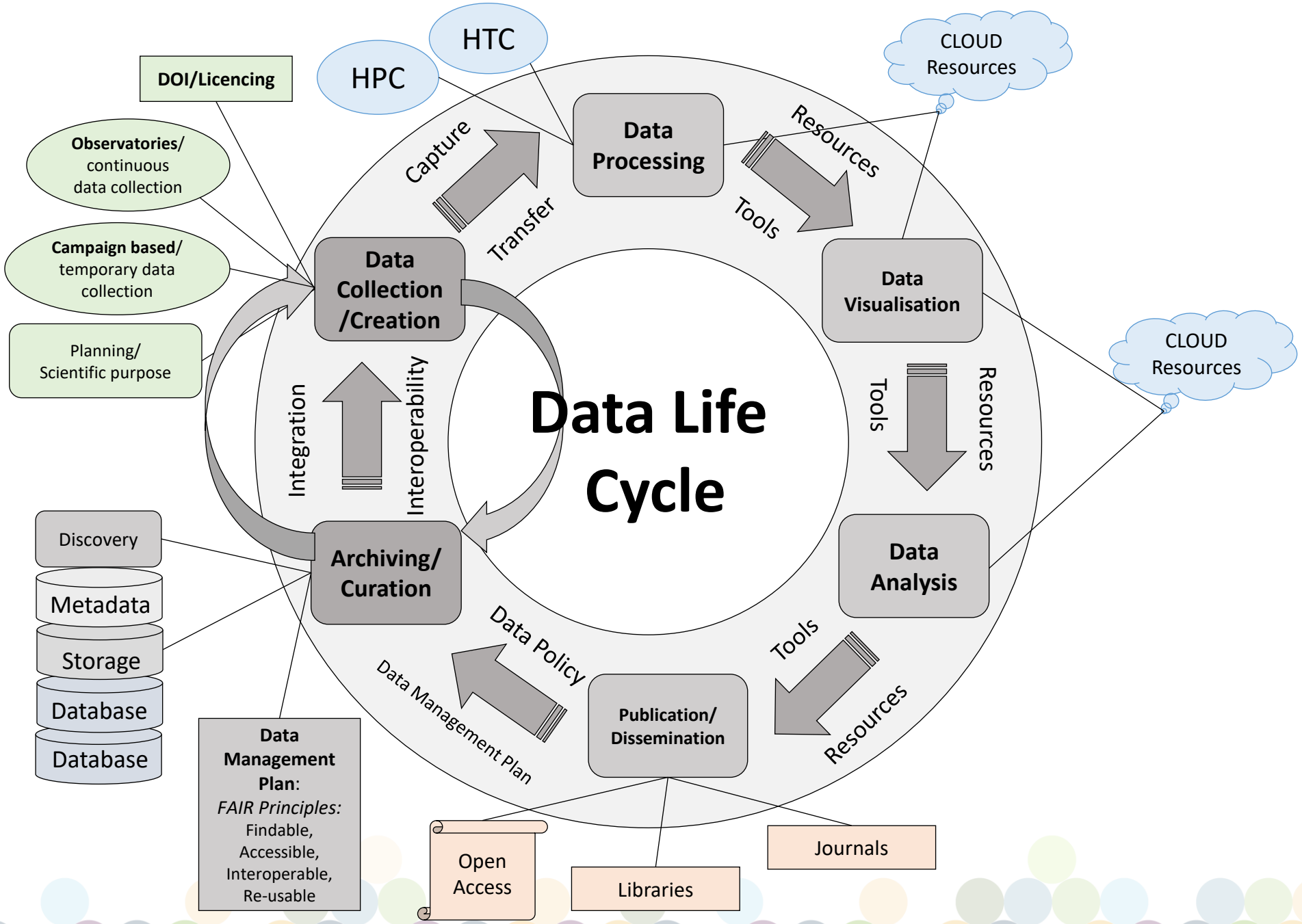
EPOS Data Management Plan

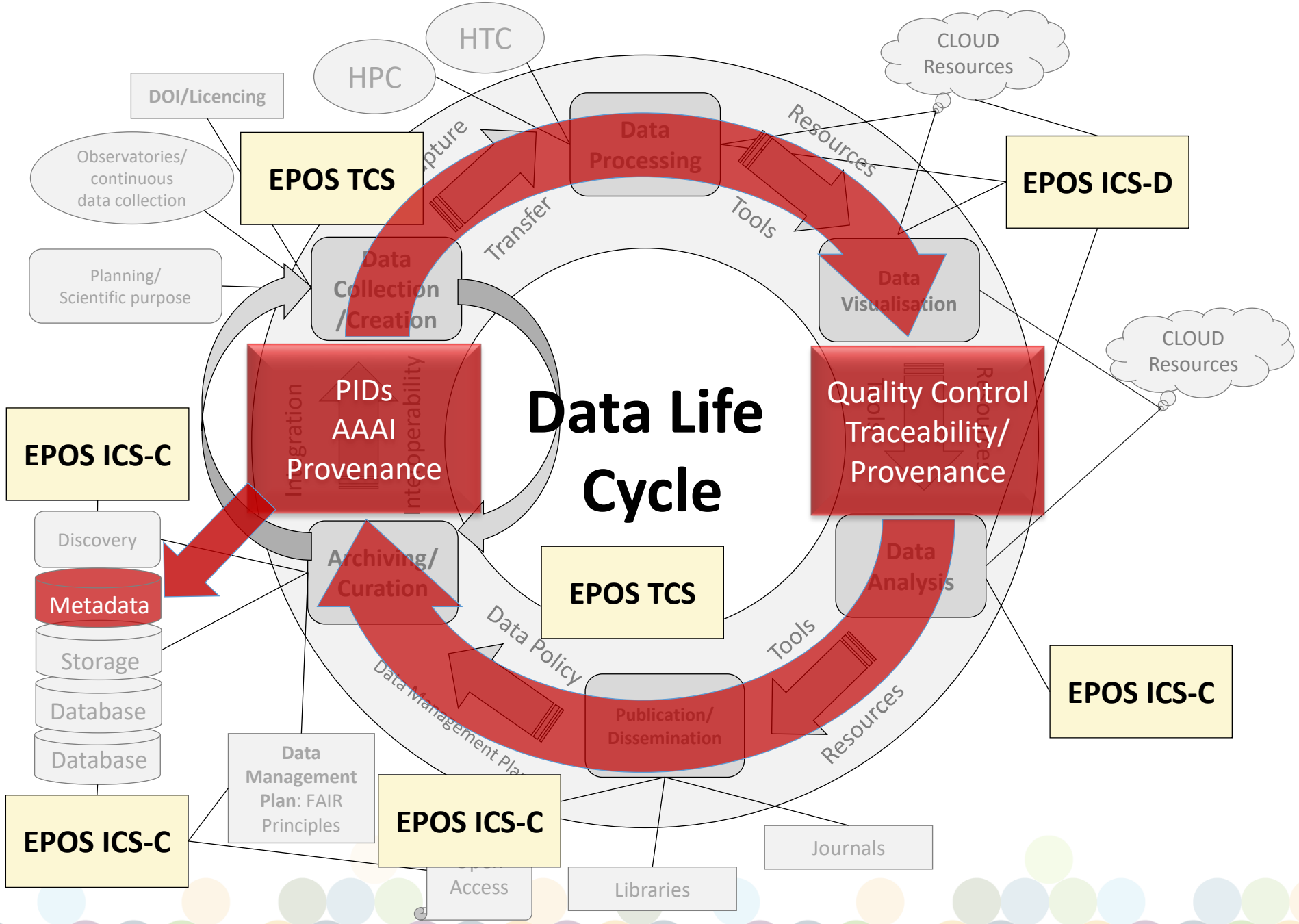
- Each of the TCS community contains **data, data products, software and services (DDSS)** that vary significantly
- DDSS are provided by a **large number of institutions** that constitute the underlying **National Research Infrastructures** (NRIs)
- **Each NRI** have its **own DMP**
- **Not all TCSs** on the other hand, have yet clearly **developed DMPs**
- There is indeed a need for harmonizing DMPs at TCS level
- There is a need for **creation** of the **EPOS central hub DMP**
- Finally, a **higher level DMP** will be elaborated, taking into account both the **harmonized view of the data principles** from communities and the **global EPOS DMP**

EPOS Data Management Plan

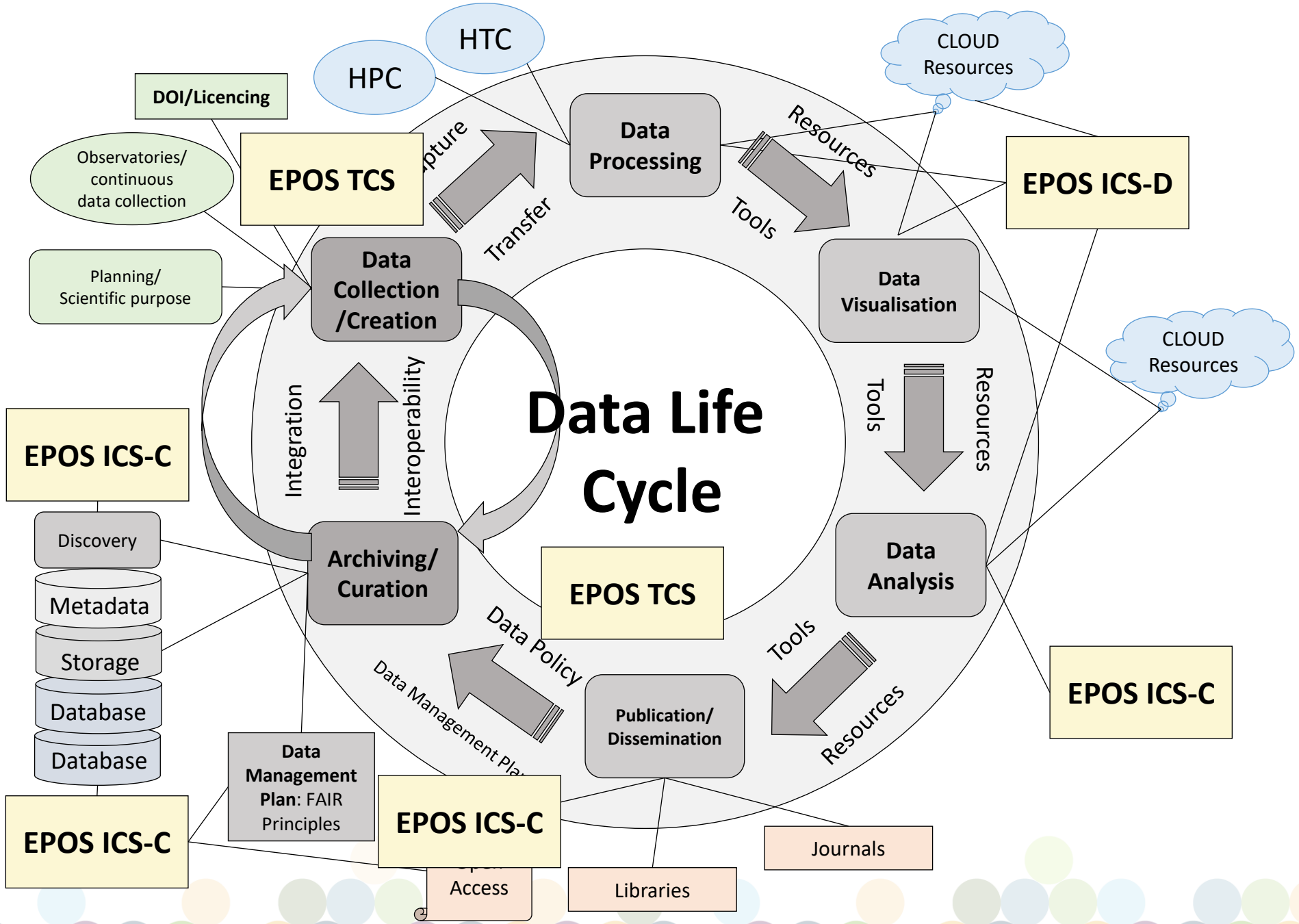
EPOS-DMP based on FAIR Principles have six sections

- Data Generalities
- **FAIR** data
  - Making data **findable**, including provisions for metadata
  - Making data openly **accessible**
  - Making data **interoperable**
  - Increase data **re-use** through clarifying licenses
- Allocation of **Resources**
- Data **Security**
- **Ethical** Aspects
- Other Issues

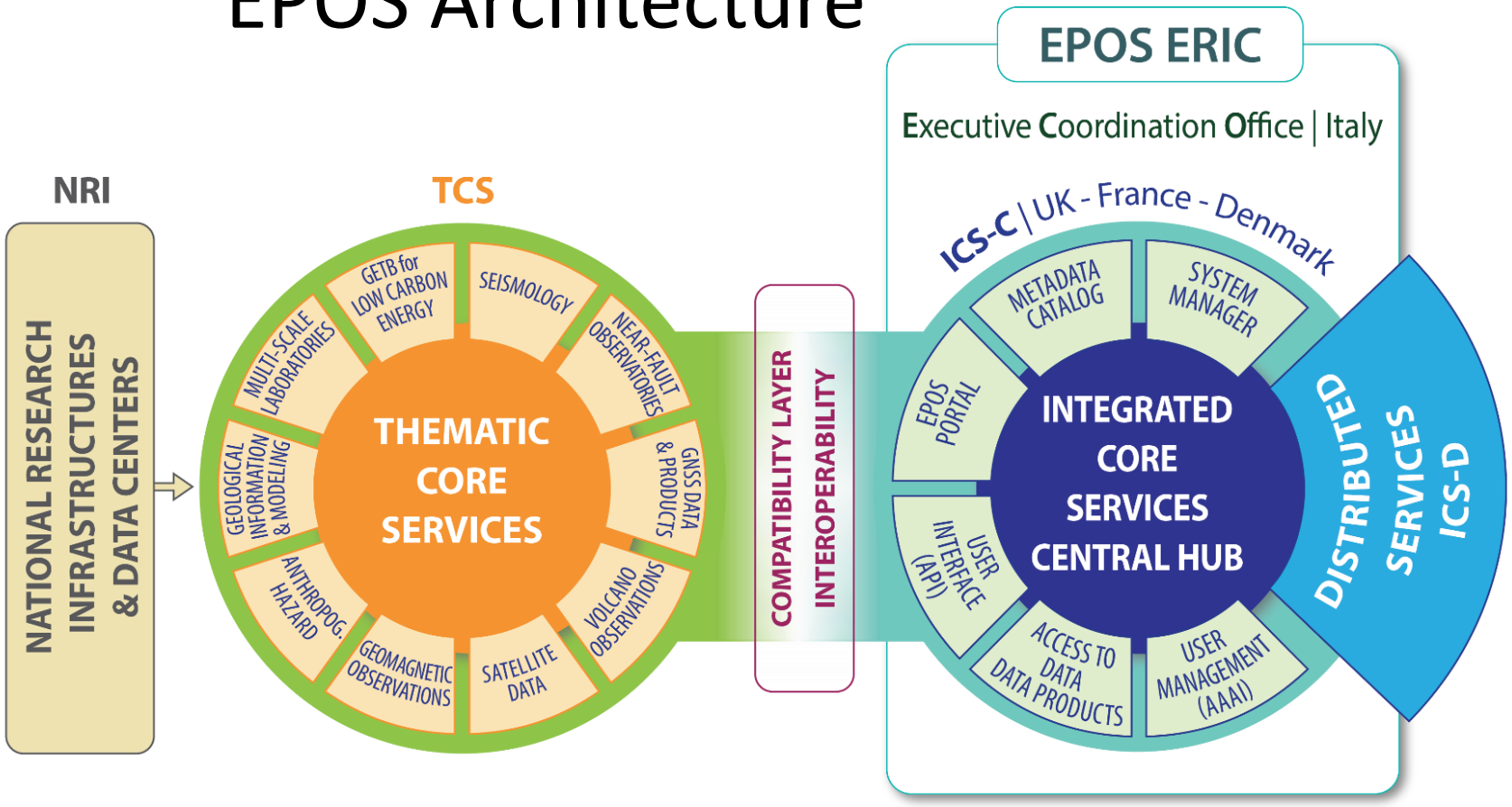




ERIC ESTABLISHED BY THE EUROPEAN COMMISSION IMPLEMENTING DECISION 2018/1732



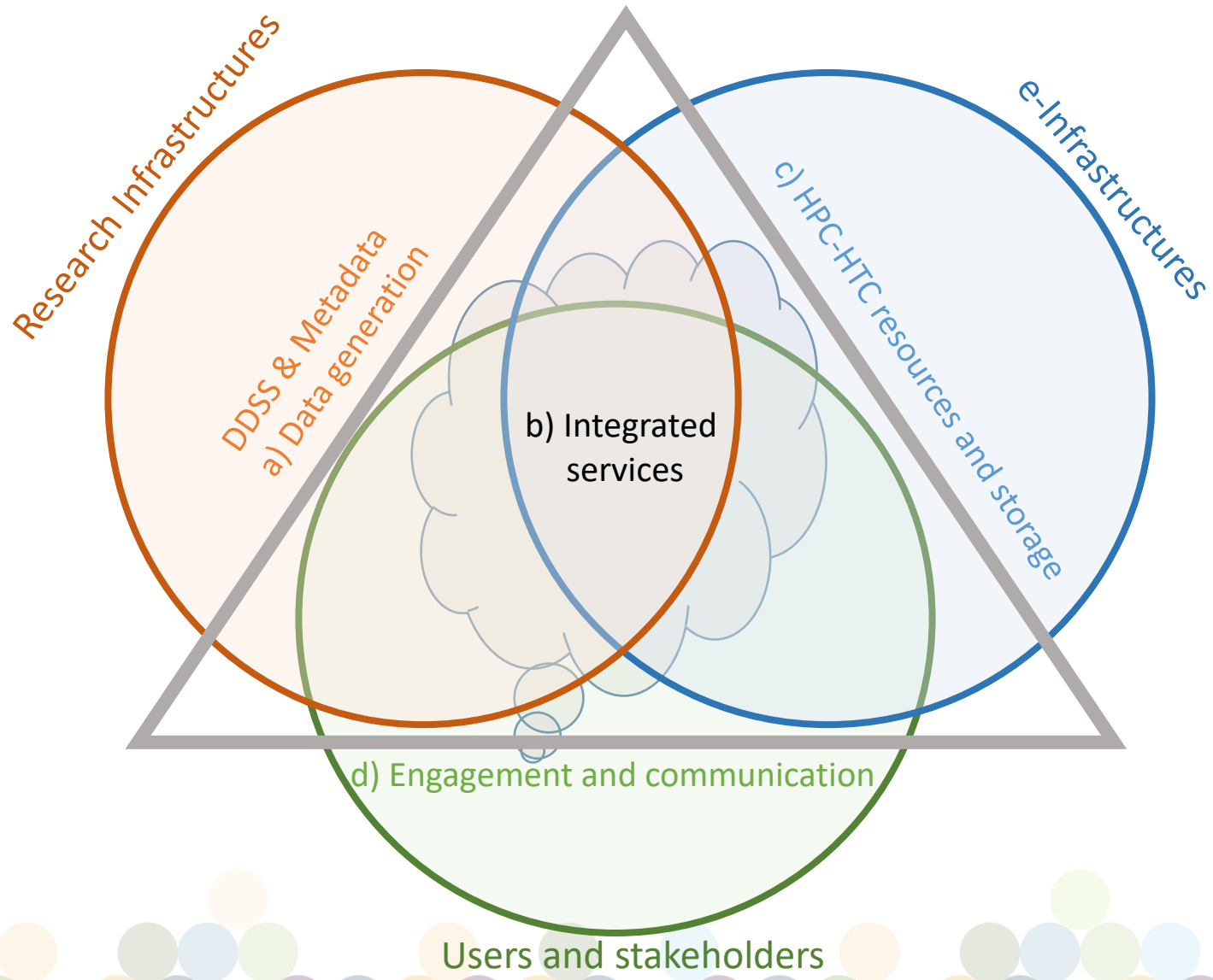
EPOS Architecture

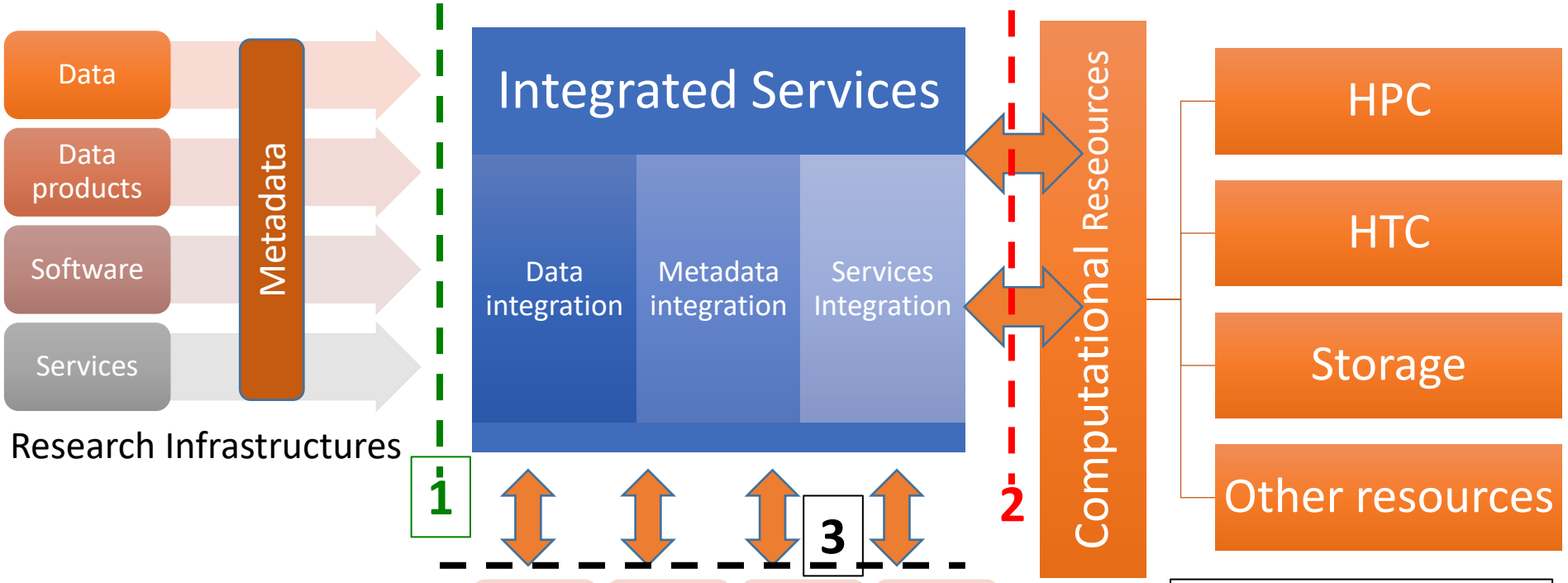


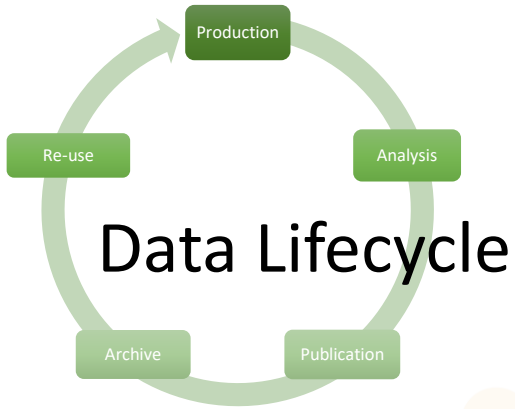
Main elements of the EPOS Architecture, the Integrated Core Services Central Hub (ICS-C) and the Executive and Coordination Office (ECO) belong to the EPOS-ERIC legal subject.

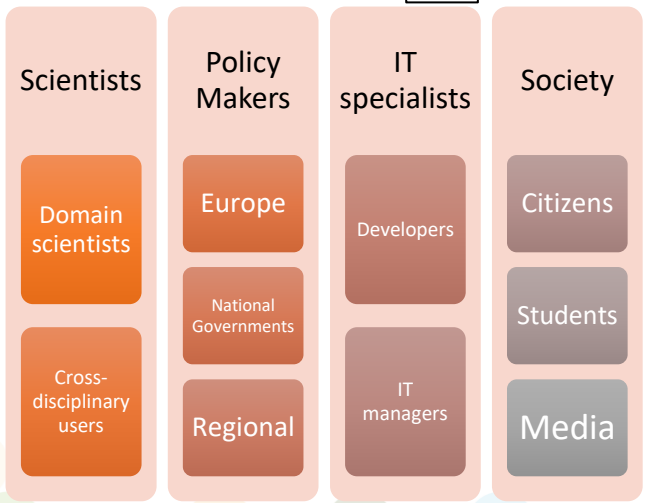
ERIC ESTABLISHED BY THE EUROPEAN COMMISSION IMPLEMENTING DECISION 2018/1732

EPOS represents the **Integrated Services** (b) and is defined by the cross-section of the three elements a,c and d



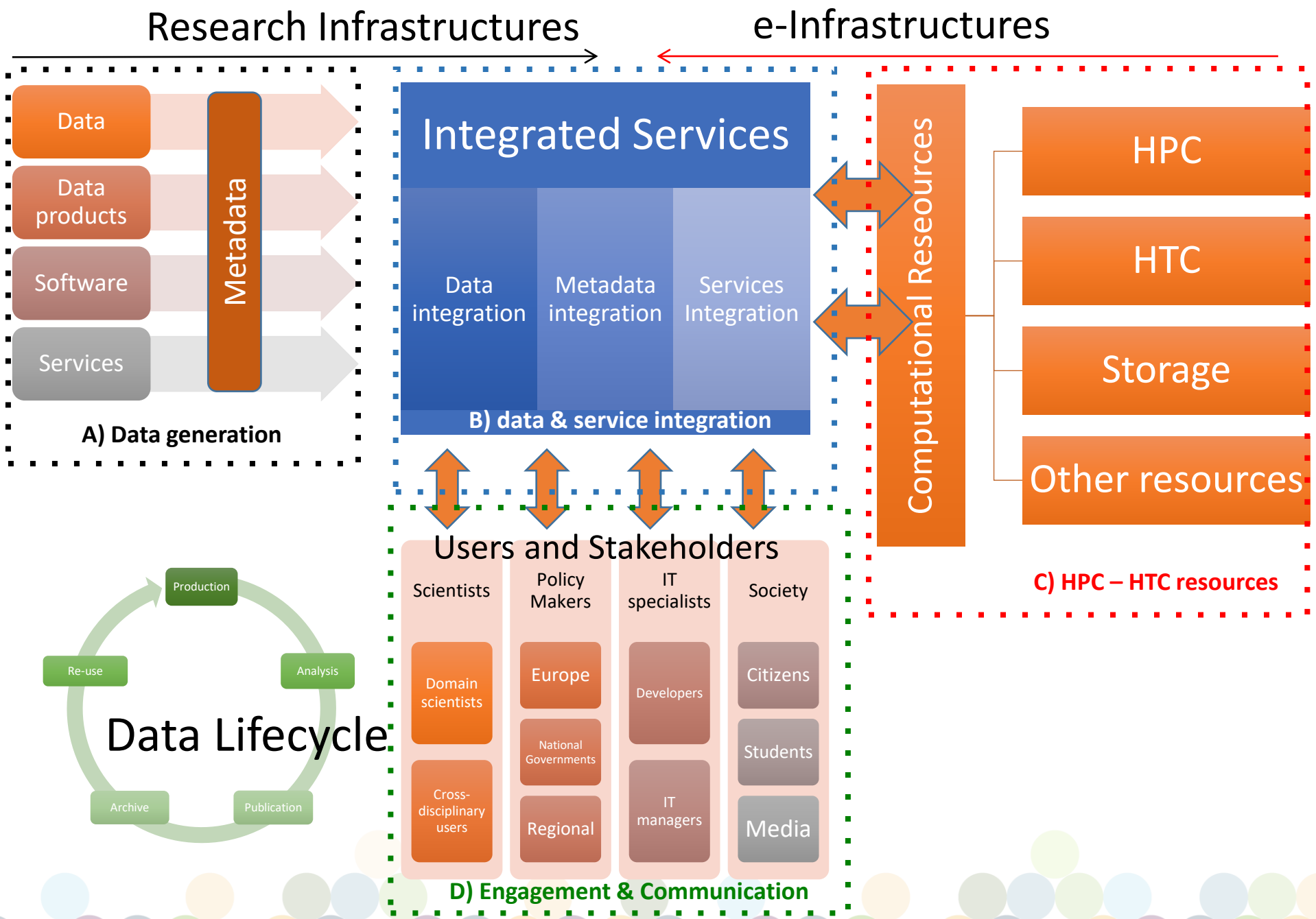






Users of Research Infrastructures

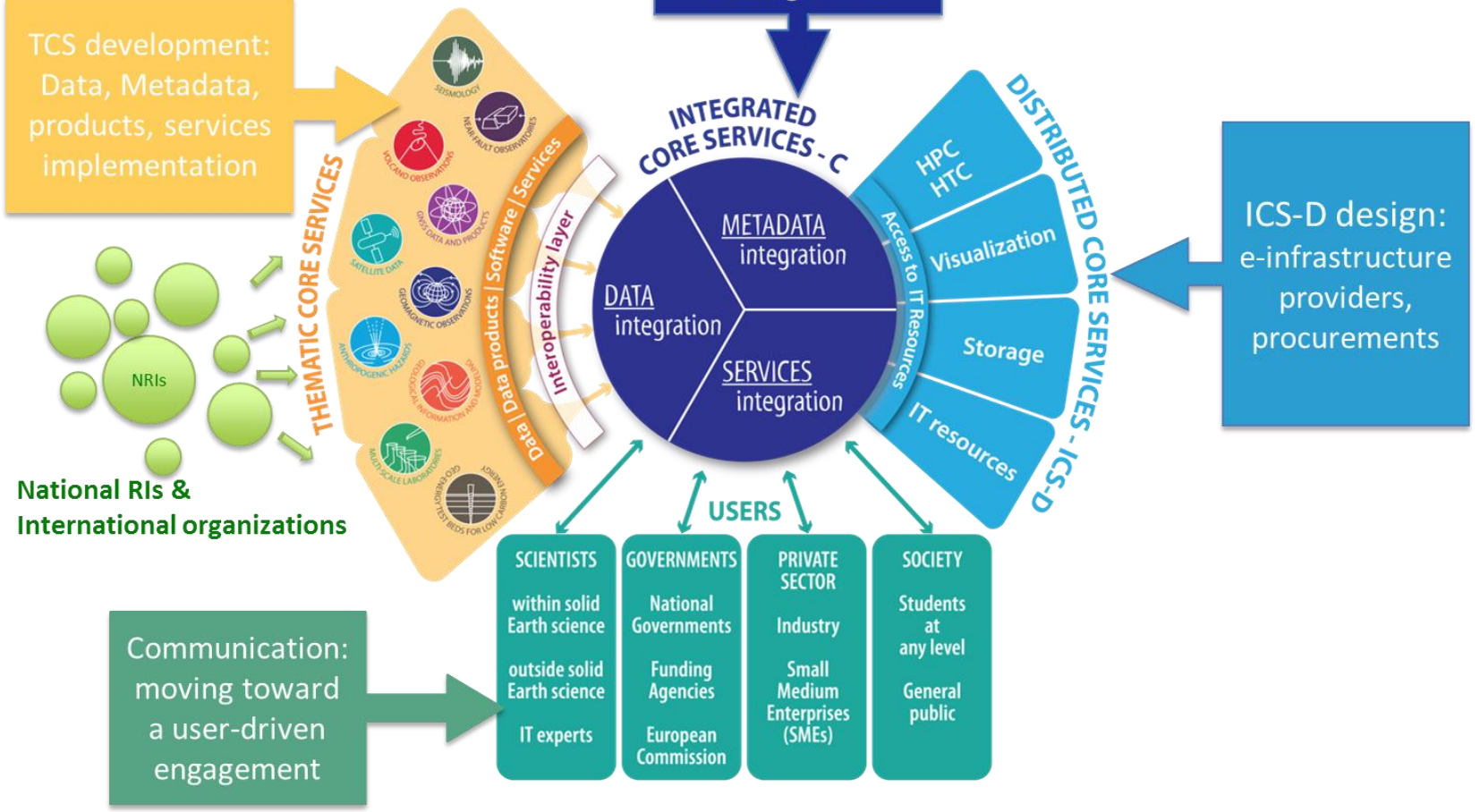
- **INTERFACE 1** is clearly a responsibility of RIs since they have engaged the communities
- **INTERFACE 2** is a challenge and relies on both e-Is and user-requirements also represented by RIs.
- **INTERFACE 3** is of common interests for RIs & e-Is and the design should be done coherently with existing strategies, maximizing the impact



Status of technical implementation

ICS-C development: interoperability demonstrator, user management

TCS development: Data, Metadata, products, services implementation

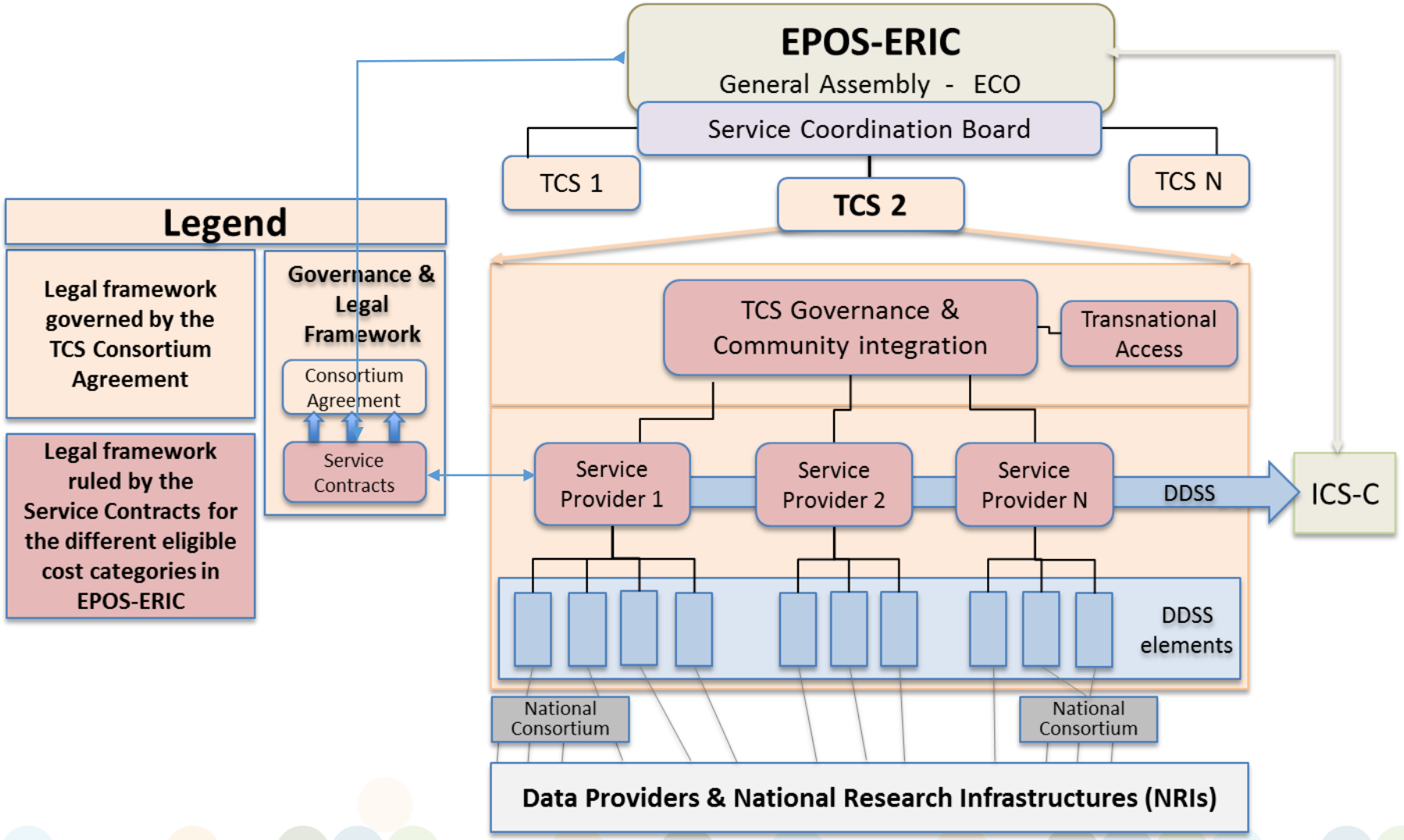


ICS-D design: e-infrastructure providers, procurements

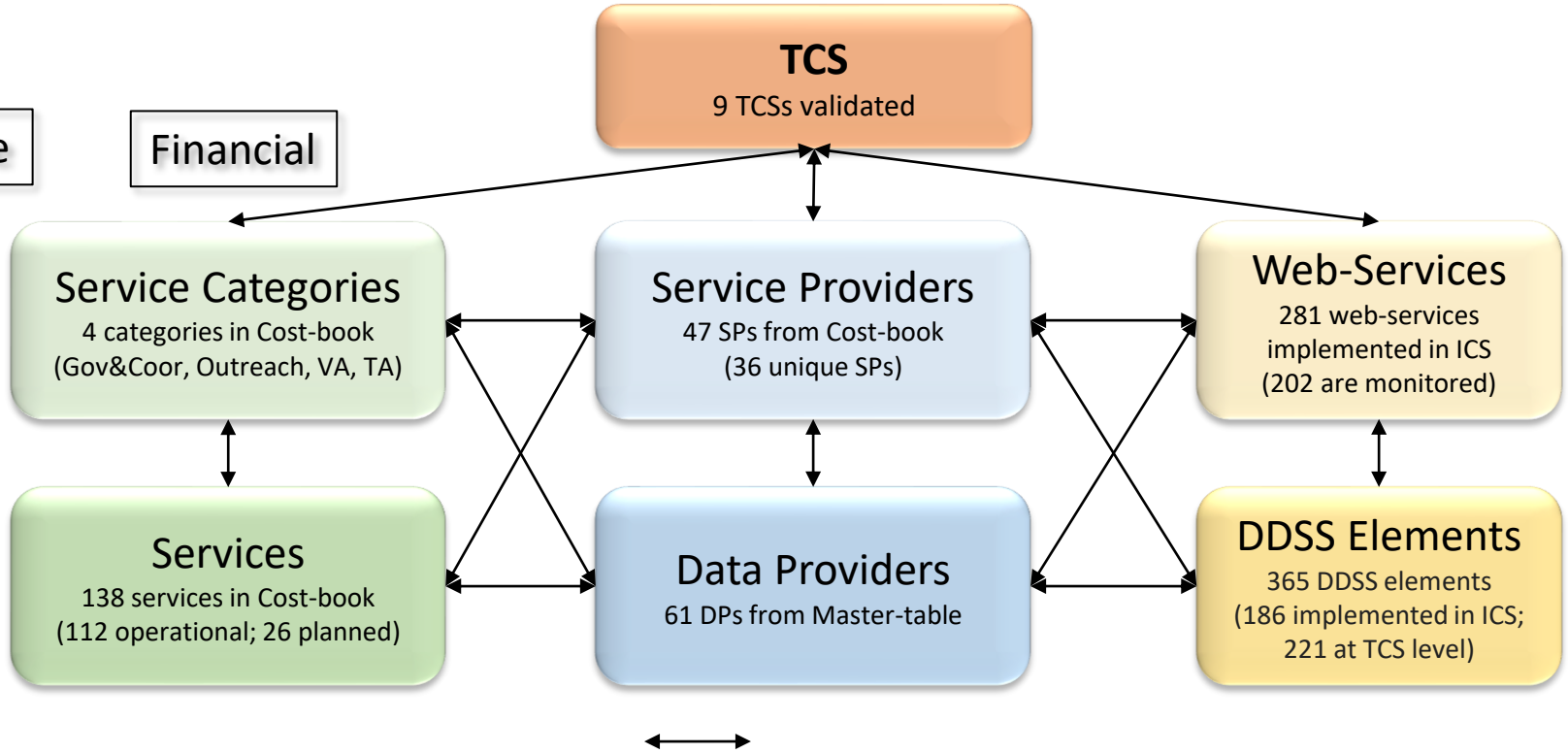
Communication: moving toward a user-driven engagement

ERIC ESTABLISHED BY THE EUROPEAN COMMISSION IMPLEMENTING DECISION 2018/1732

EPOS Data Management Structure



Terminology in EPOS Delivery Framework



Service Provider: A legal entity (an organization) providing various DDSS elements from one or several data providers

Data Provider: An organization providing one or more DDSS elements

TCS: Thematic Core Services in EPOS

Services: An **EPOS service** from a specific **TCS**, under a specific **service category**, given by a **service provider,** on behalf of various **data providers**, through a **web-service**, that give access to everal **DDSS**.

Web-service: a server running on a computer device, listening for requests at a particular port over a network, serving web documents (HTML, JSON, XML, images), and creating web applications services, which serve in solving specific domain problems over the Web (WWW, Internet, HTTP)

DDSS elements: data, data-products, software and services

There are several combinations that are possible in this matrix and hence the question of **number of "services" in EPOS is complex** and depends on where the question is initiated among the five boxes above. Numbers are as of Oct.2019.

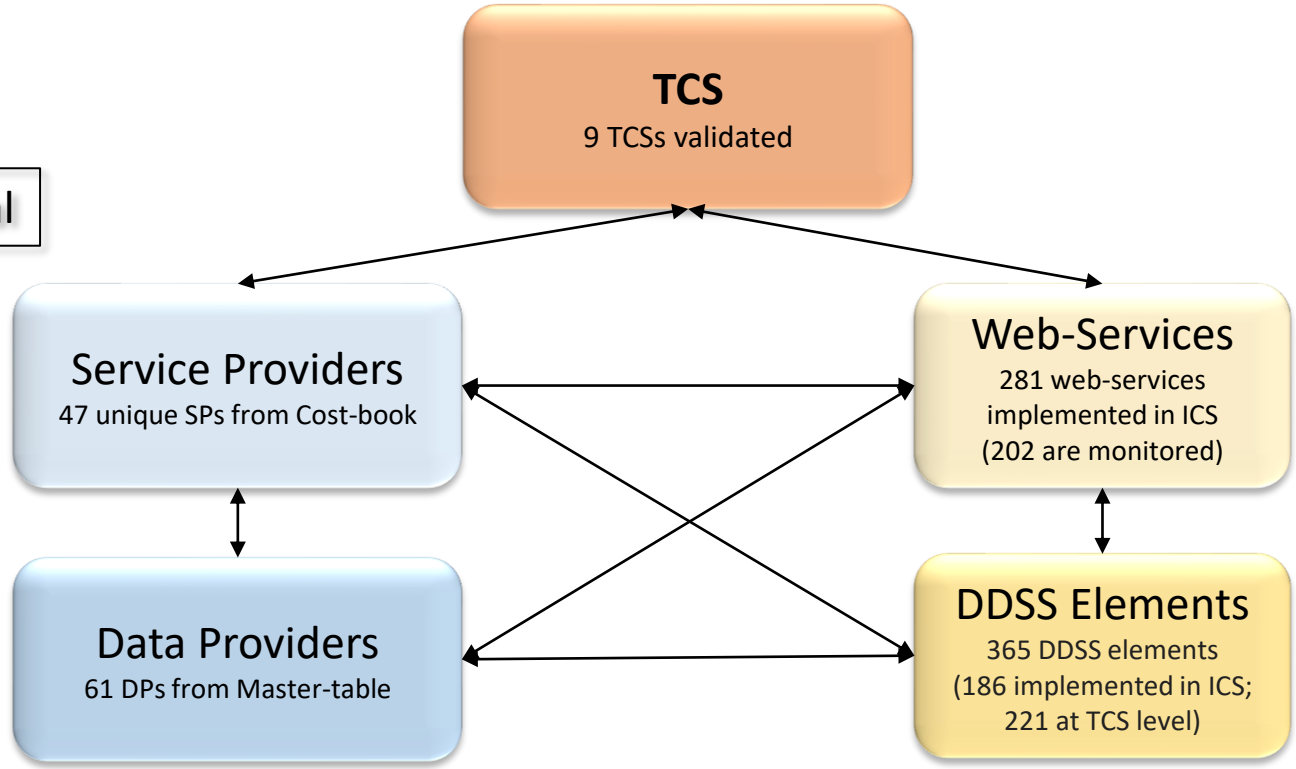
Terminology in EPOS Delivery Framework

Legal, governance, financial

Service Provider: A legal entity (an organization) providing various DDSS elements from one or several data providers

Data Provider: An organization providing one or more DDSS elements

TCS: Thematic Core Services in EPOS



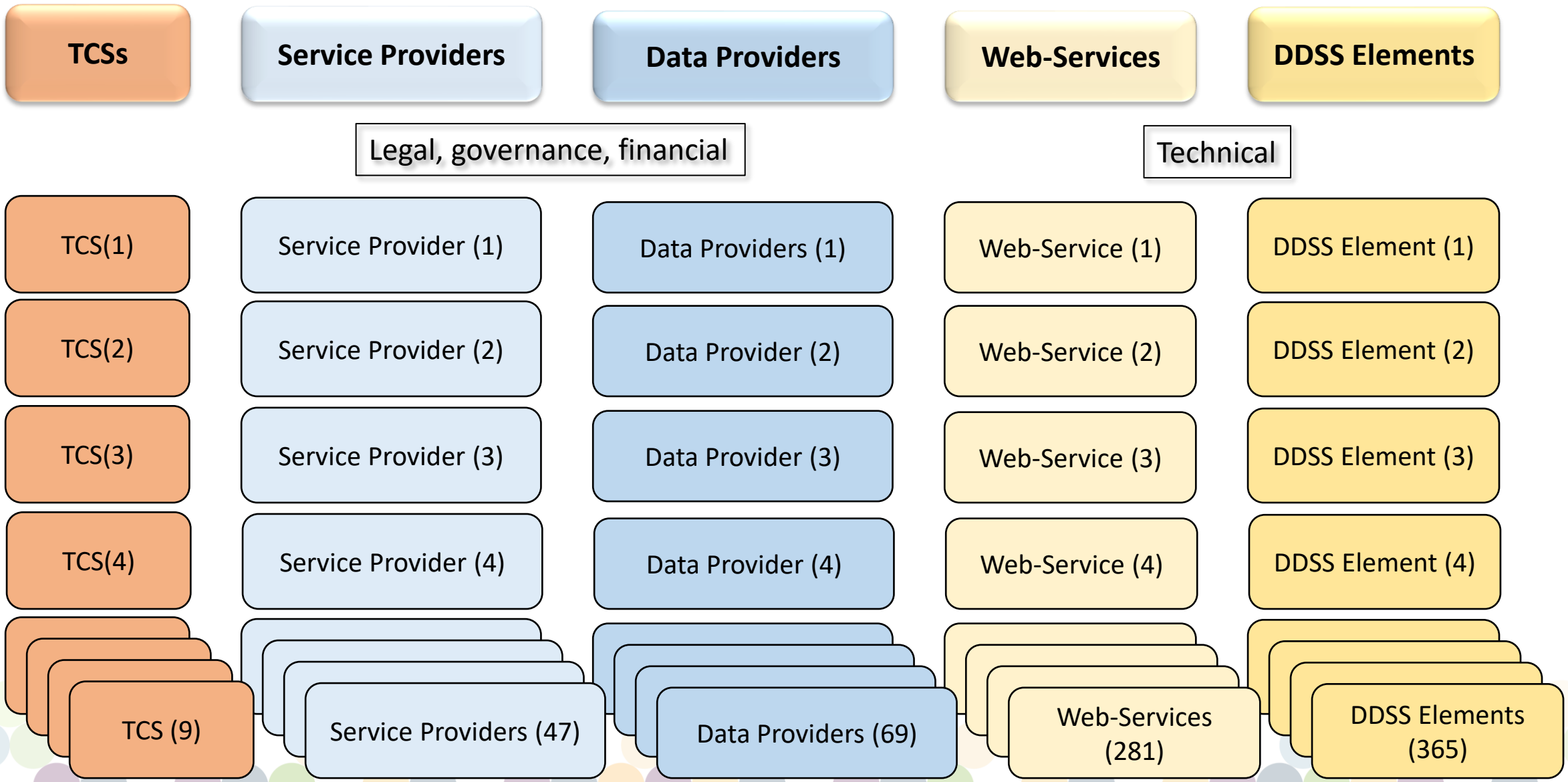
Technical

Web-service: a server running on a computer device, listening for requests at a particular port over a network, serving web documents (HTML, JSON, XML, images), and creating web applications services, which serve in solving specific domain problems over the Web (WWW, Internet, HTTP)

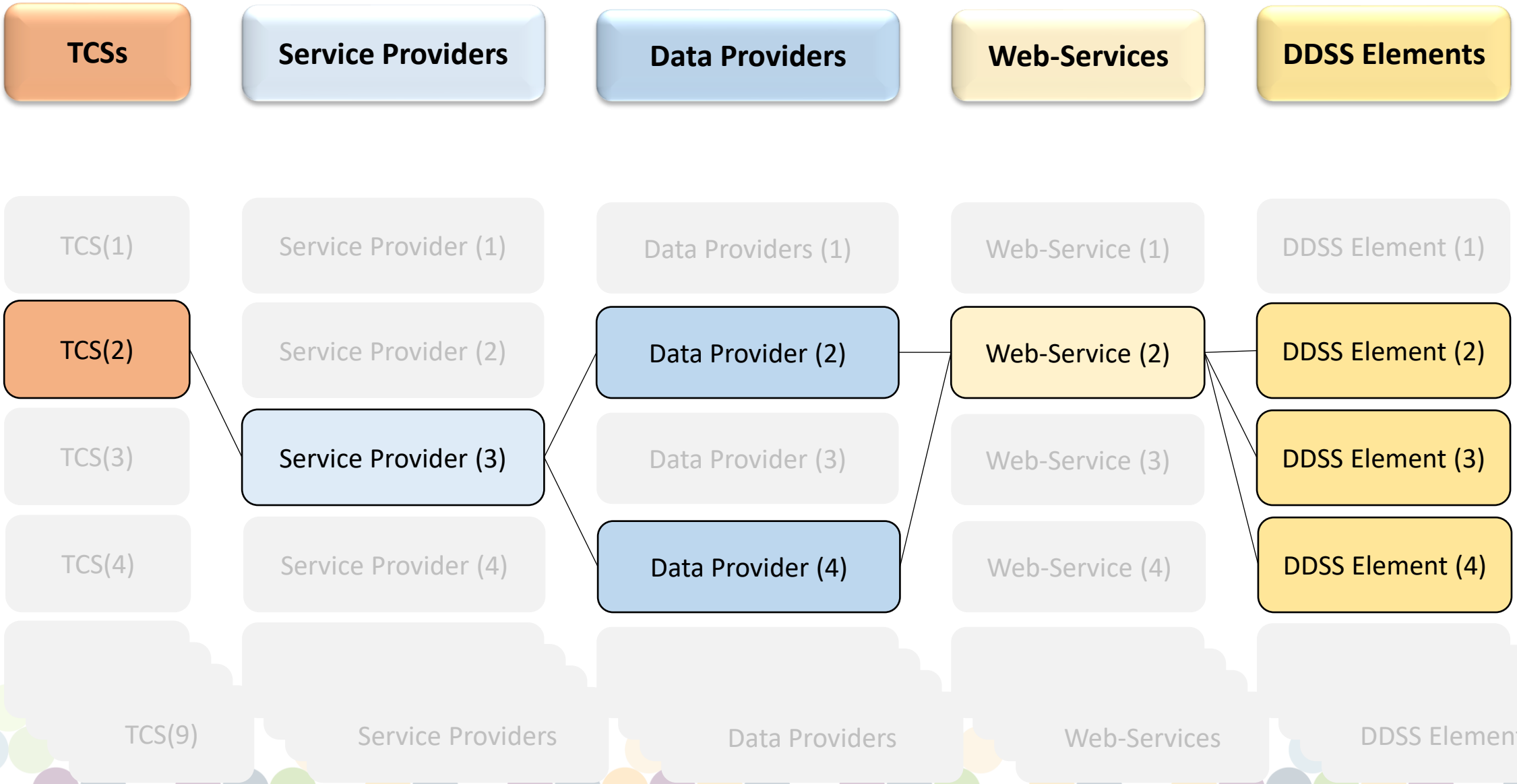
DDSS elements: data, data-products, software and services

There are several combinations that are possible in this matrix and hence the question of **number of "services" in EPOS is complex** and depends on where the question is initiated among the five boxes above. Numbers are as of Oct.2019.

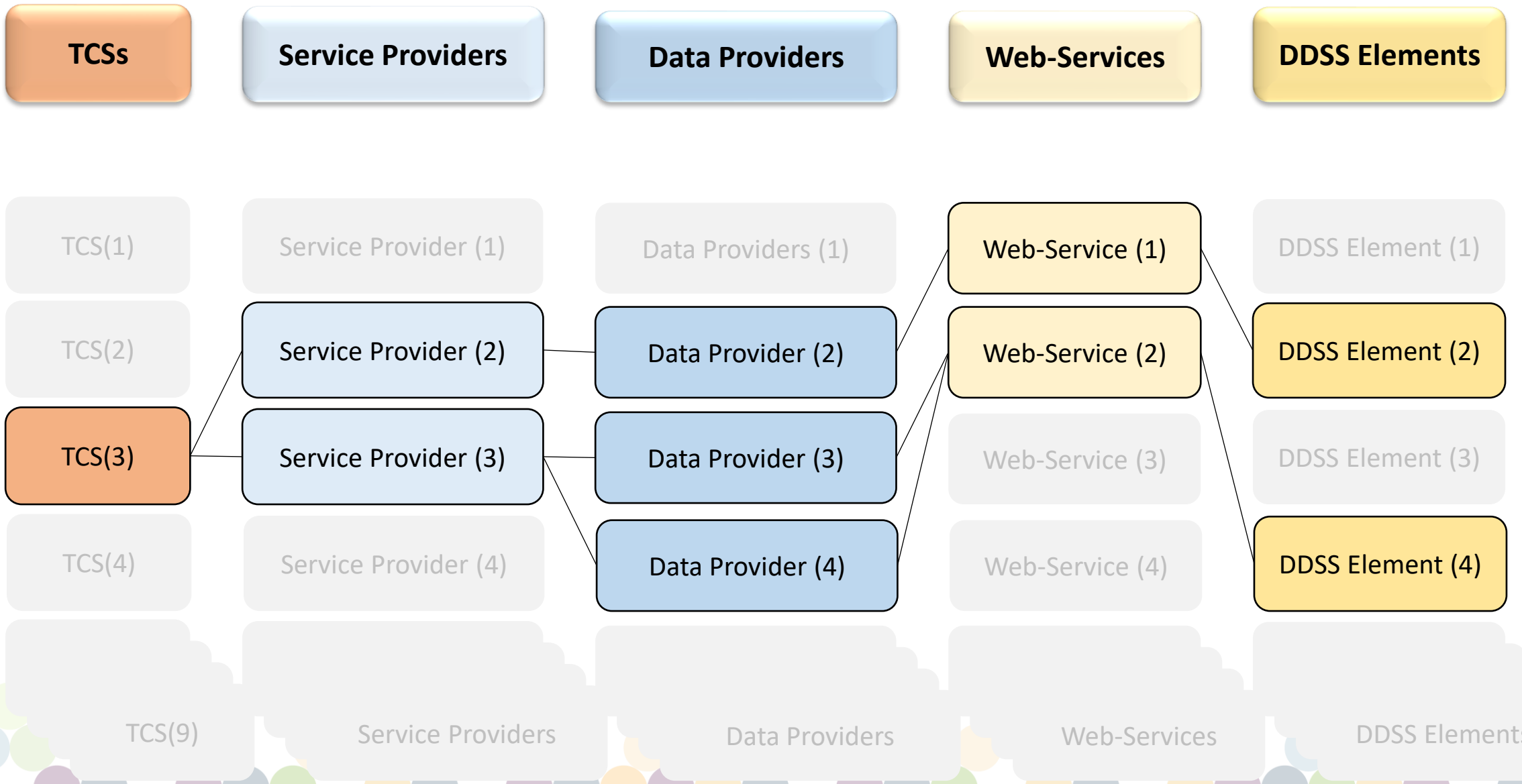
TCS-SP-DP-WS-DDSS-Matrix



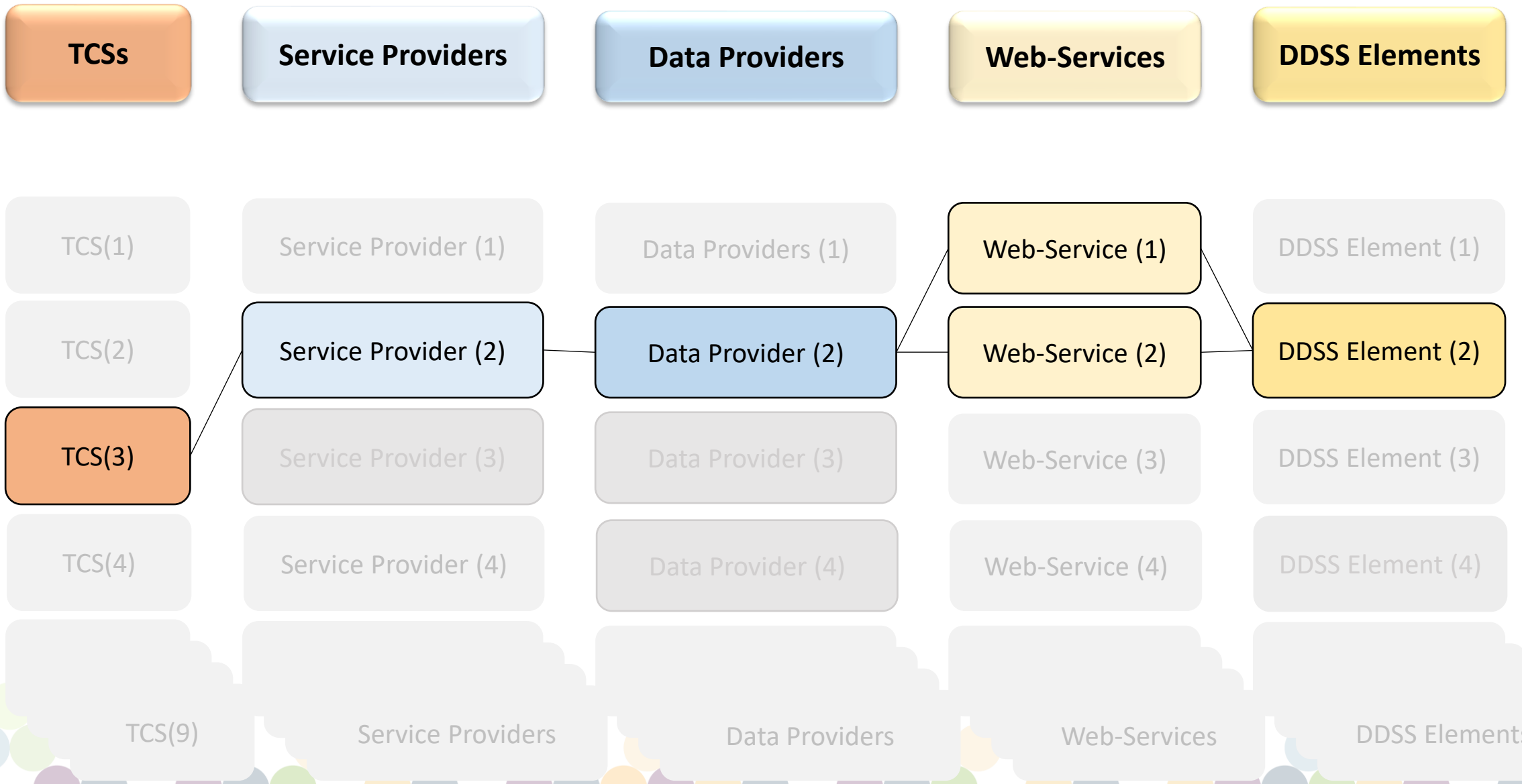
TCS-SP-DP-WS-DDSS-Combinations: Example 1



TCS-SP-DP-WS-DDSS-Combinations: Example 2



TCS-SP-DP-WS-DDSS-Combinations: Example 3



Thank you for your attention!

WebSite

www.epos-eu.org

Newsletter

www.epos-eu.org/newsletter

Social