Data Spaces made Simpl

EOSC Symposium – 21 September 2023

Manuel Mateo Goyet

Cloud & Software, DG Connect

European Data Strategy: 4 pillars

A governance framework for data access and use

Investments in capabilities and infrastructures

Competences

User empowerment
Data literacy
Skills
Capacity building for SMEs

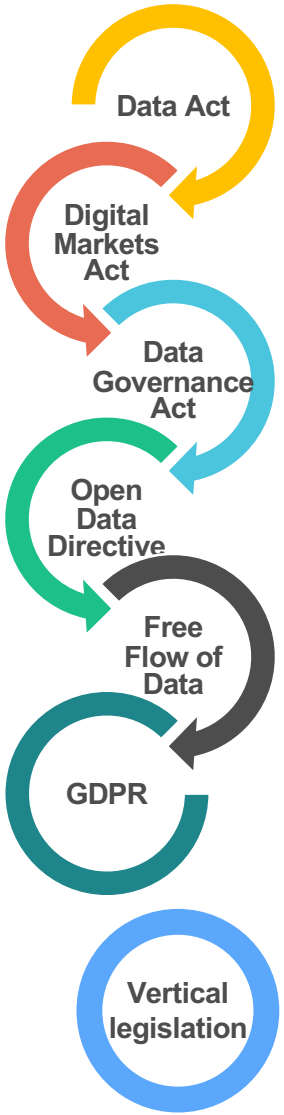
Common European data spaces

in crucial economic sectors and domains of public interest

International Aspects: Analytical framework for measuring data flows

European Commission

Legal instruments

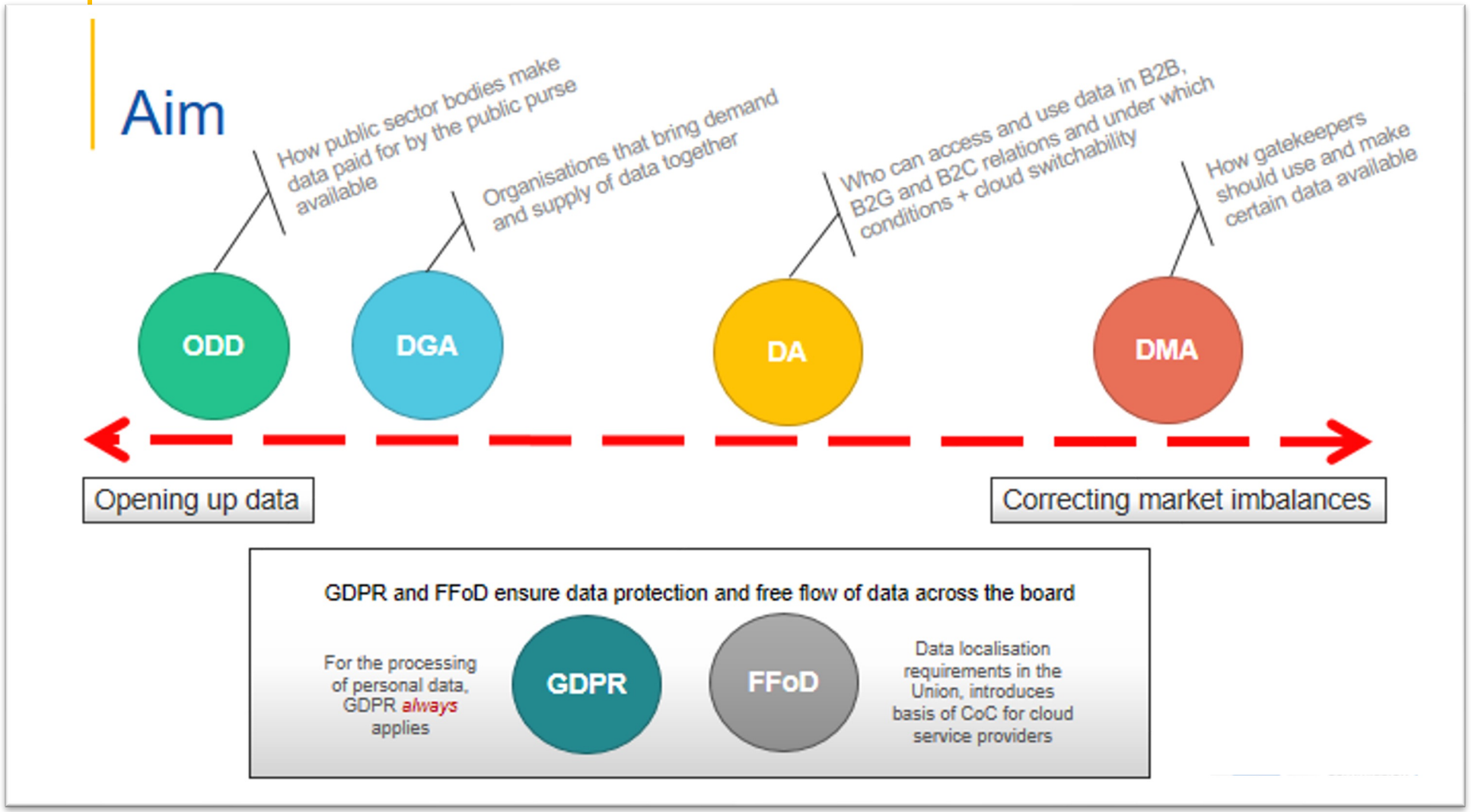


Aim	Data Covered	Regulated Actors
Ensure **FAIRNESS** in the allocation of data value among the actors of the data economy	Private sector data, personal and non-personal data, and co-generated (IoT) data	Businesses, public sector bodies, cloud and other data processing service providers
Tackle imbalances caused by the **MARKET POWER** of gatekeepers	Personal data and private sector data held by online platforms and originating from the users	Cloud and other data processing service providers, large data platforms
Ensure **TRUST** in data transactions	Public and private non-personal data, and personal data voluntarily made available by data holders	Data intermediation service providers, public sector bodies, (Recognised) Data Altruism Organisations
Promote use of **OPEN DATA**	Data in an open format that can be freely used, re-used and shared by anyone for any purpose	Public sector bodies, bodies governed by public law, public undertakings, universities
Ensure **FREE FLOW OF DATA** other than personal data within the Union	Non-personal data	Member States, competent authorities, professional users
Ensures a high-level of **DATA PROTECTION** and free flow of personal data in the Union	Personal data	Data controller, data processor, data subject, DPO, supervisory authorities, EDPB

Promote a competitive market according to **SECTOR-SPECIFIC** rules where necessary, e.g. automotive	Personal and non-personal data	Individuals and private and public sector bodies

European Commission

Data instruments

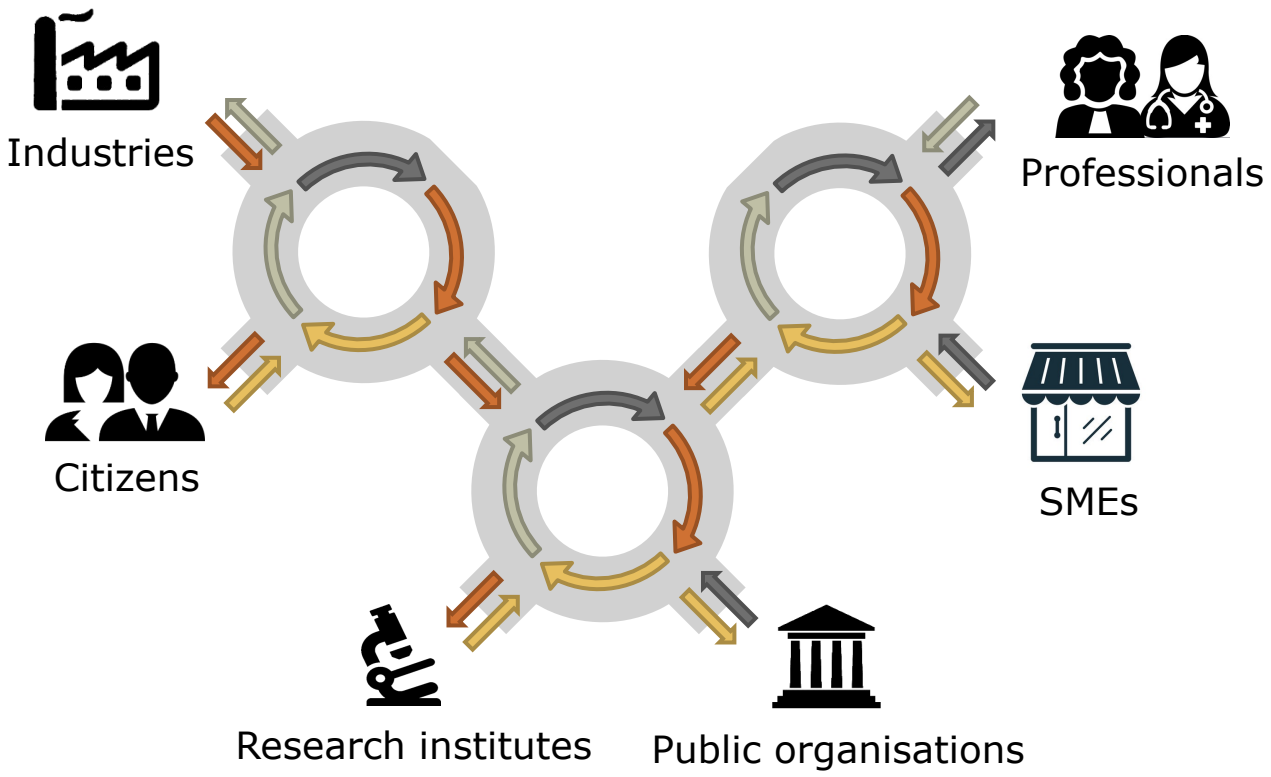


Legend

DA	Data Act
DMA	Digital Markets Act
DGA	Data Governance Act
ODD	Open Data Directive
FFoD	Free Flow of Data Regulation
GDPR	General Data Protection Regulation

European Commission

Data Spaces



Defined as a federated data ecosystem based on shared policies and rules. The participants of such data spaces are enabled to access data in a secure, transparent, trusted, easy and unified fashion.

Data holders are in control of who can have access to their data, for which purpose and under which conditions it can be used.

From a technical perspective, a data space can be seen as a data integration concept which does not require common database schemas and physical data integration. It is rather based on distributed data stores and integration on an “as needed” basis.

Key characteristics of a data space

- A secure and privacy-preserving IT infrastructure to pool, access, process, use and share data.
- A data governance mechanism, comprising a set of rules of legislative, administrative and contractual nature that determine the rights to access, process, use and share data in a trustful and transparent manner.
- Data holders are in control of who can have access to their data, for which purpose and under which conditions it can be used.
- Presence of vast amounts of data that are made available on a voluntary basis and can be reused against remuneration or for free, depending on the data holder's decision.
- Participation by an open number of organisations/individuals.

Data Spaces – Commission intervention

The Commission will fund the creation of common European data spaces in specific sectors where the EU financial contribution will have an impact.

In strategic sectors, the Commission will facilitate the development of EU-wide common data spaces:

- contributing to the definition of their objectives through sectoral policies.
- offering a technical solution (a smart middleware) that will pool existing data infrastructures (cloud federation)

Stakeholders are already organising data spaces in different sectors. The Data Governance Act lays down a number principles to increase trust in neutral data intermediaries that will help match data demand and supply.

Design principles for common European data spaces

Promote tools to control, pool, access, use and share all types of data favouring FRAND principles

Data access and control

Governance

Put in place an *appropriate governance structure* (mechanisms) to ensure fair, transparent, proportionate and non-discriminatory access to, sharing and use of data.

Comply with applicable *EU legal framework*, and put in place adequate measures to prevent unauthorised access to data

Respect of EU rules and values

Interconnection and interoperability

Develop on international *standards*, INSPIRE, FAIR principles to progressively lead to single EU data space

Open to all actors (organisations/individuals) that respect EU rules and values and comply with data space rules

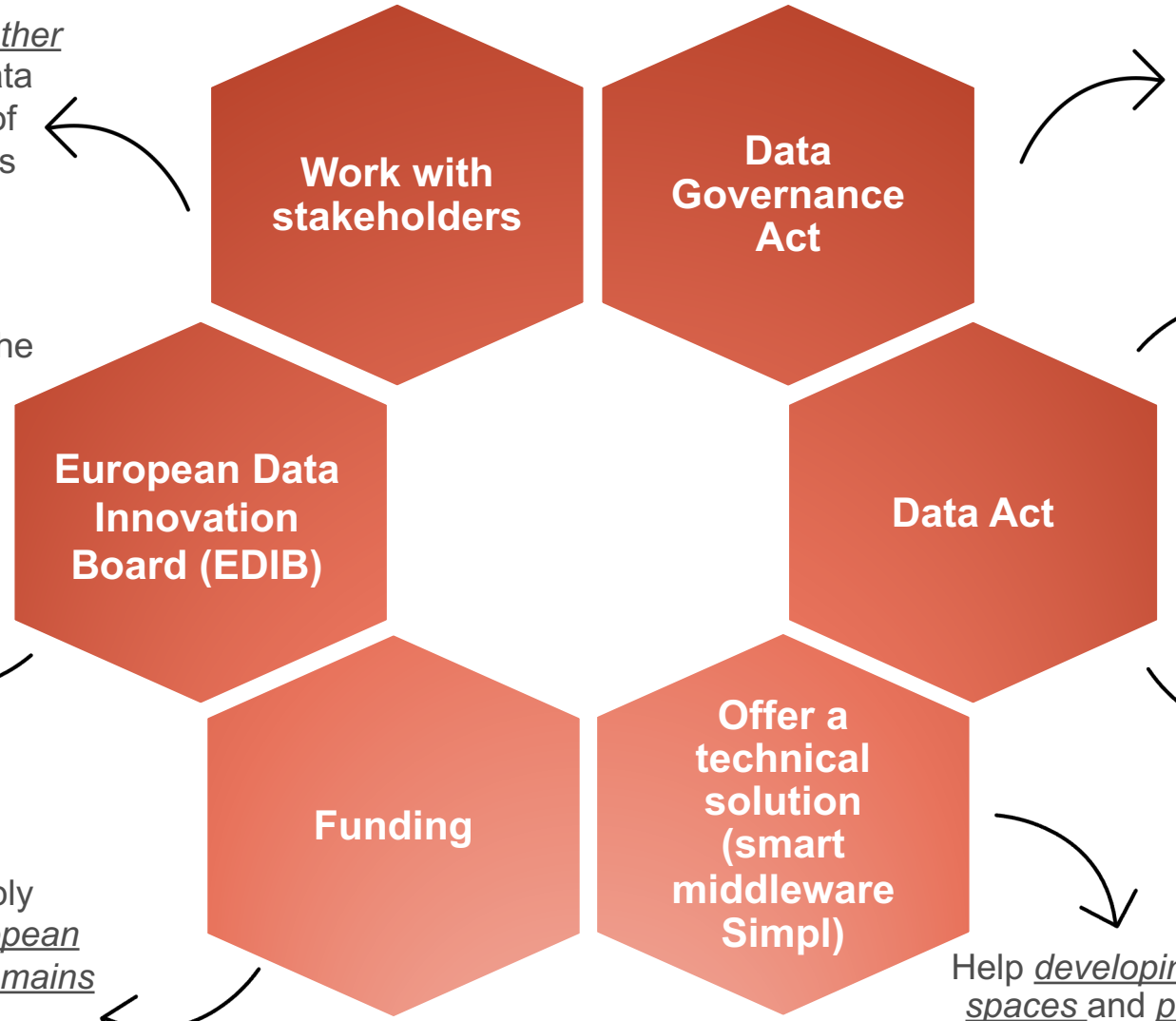
Openness

Technical data infrastructure

Encourage actors to use *common technical data infrastructure* (including sector-specific framework and cybersecurity-, and data protection-by-design principles)

European Commission

The EC's role in supporting the creation of data spaces



Bring stakeholders together with a common EU data policy, promote use of data space guidelines

Increase trust in data sharing (neutral data intermediaries, i.e., the orchestrators of the data spaces

Support the Commission in:

- *Issuing guidelines* to facilitate the development of the common European data spaces
- Identifying the relevant standards and interoperability requirements for *cross-sector data sharing*

Unlock the potential of the EU's *industrial data* (IoT data), ensure fairness in the allocation of data value among the actors of the data economy

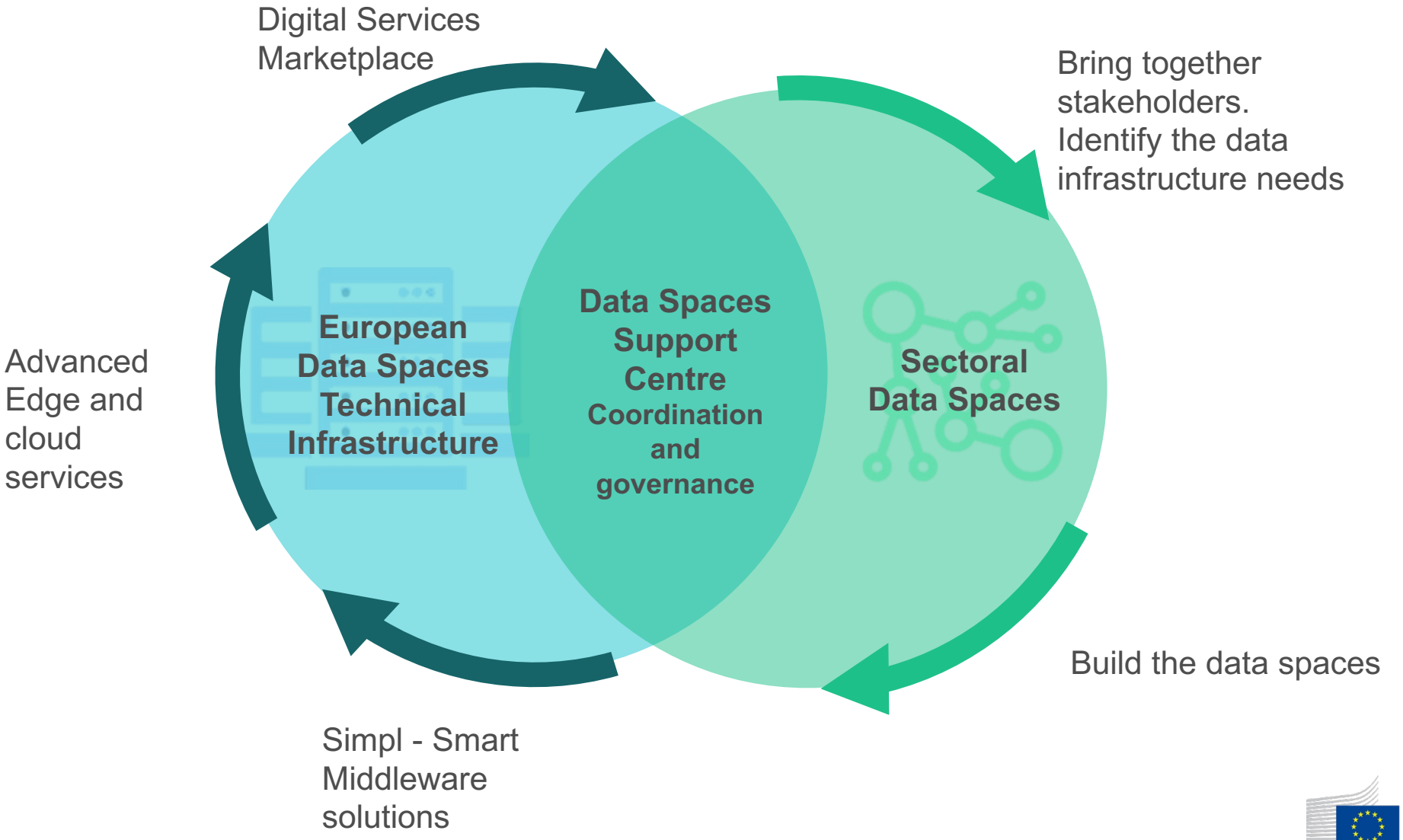
Facilitate *interoperability* of data, data sharing mechanisms and services (incl. data spaces) between sectors

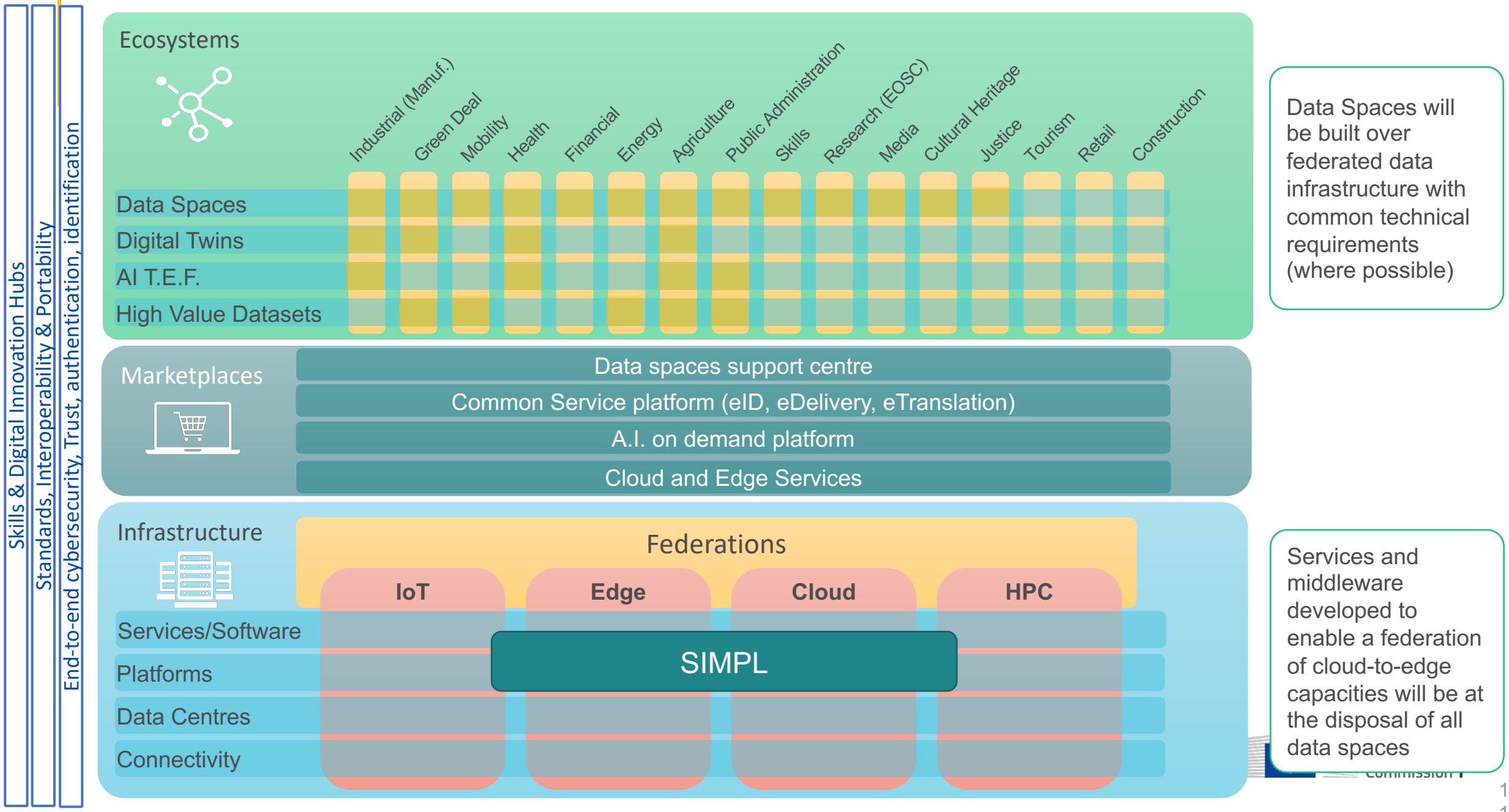
EU financial contribution (notably DIGITAL) to *deploy common European data spaces in strategic sectors/domains*

Help *developing data spaces* and *pooling existing data infrastructures* (cloud federation)

European Commission

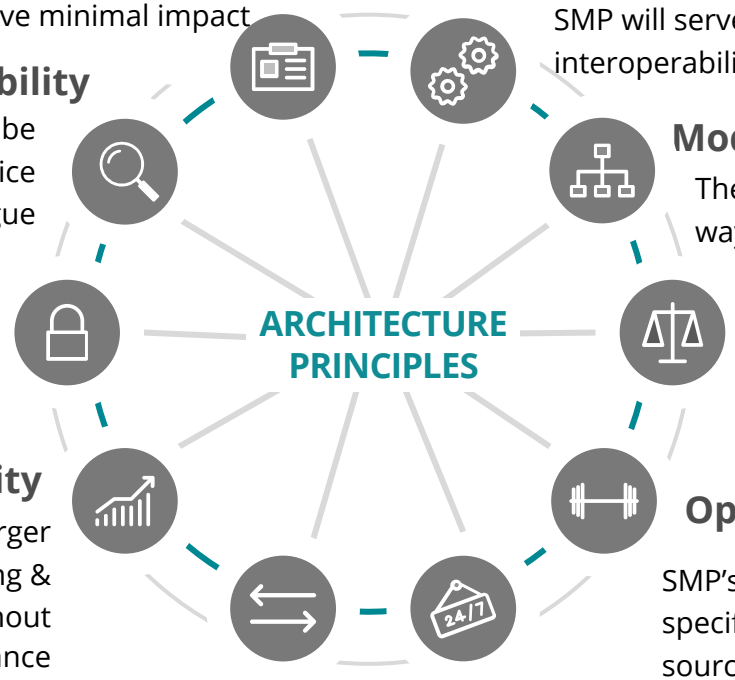
Data Spaces deployment (Digital Europe Programme)




Skills & Digital Innovation Hubs
Standards, Interoperability & Portability
End-to-end cybersecurity, Trust, authentication, identification
Ecosystems
Industrial (Manuf.)
Green Deal
Mobility
Health
Financial
Energy
Agriculture
Public Administration
Skills
Research (EOSC)
Media
Cultural Heritage
Justice
Tourism
Retail
Construction
Data Spaces
Digital Twins
AI T.E.F.
High Value Datasets
Marketplaces
Data spaces support centre
Common Service platform (eID, eDelivery, eTranslation)
A.I. on demand platform
Cloud and Edge Services
Infrastructure
Federations
IoT
Edge
Cloud
HPC
Services/Software
Platforms
Data Centres
Connectivity
SIMPL
Data Spaces will be built over federated data infrastructure with common technical requirements (where possible)
Services and middleware developed to enable a federation of cloud-to-edge capacities will be at the disposal of all data spaces

Architecture principles

Ten guiding principles for designing the architecture of the open-source smart middleware platform



Resilience

SMP's architecture components must be fault tolerant, so that failures have minimal impact

Federation

SMP will serve as a federation of both data and infrastructure, enabling interoperability and decentralization

Discoverability

All services provided by the SMP must be publicly exposed and discoverable in a service registry or catalogue

Modularity

The SMP architecture must be defined in a modular way and with open-source technologies

Security, privacy & trust

The SMP must provide a secure and trustworthy environment while complying with relevant regulations

Loose coupling

The implemented components should be loosely coupled to minimize dependencies between them

Scalability & elasticity

The SMP must provide means to accommodate larger workloads and assure correct resource pooling & allocation while allowing new entities and users without impacting performance

Openness and agnosticism

SMP's components must be designed following open specifications and standards and implemented using open-source technologies

Interoperability

The SMP must ensure that all participants will be able to interact with each other and exchange data and resources in a well-specified way.

Composability & extensibility

The SMP architecture must allow for open to iterative growth while allowing the addition of new services and capabilities

More details al Simpl website, The architecture vision;

European Commission

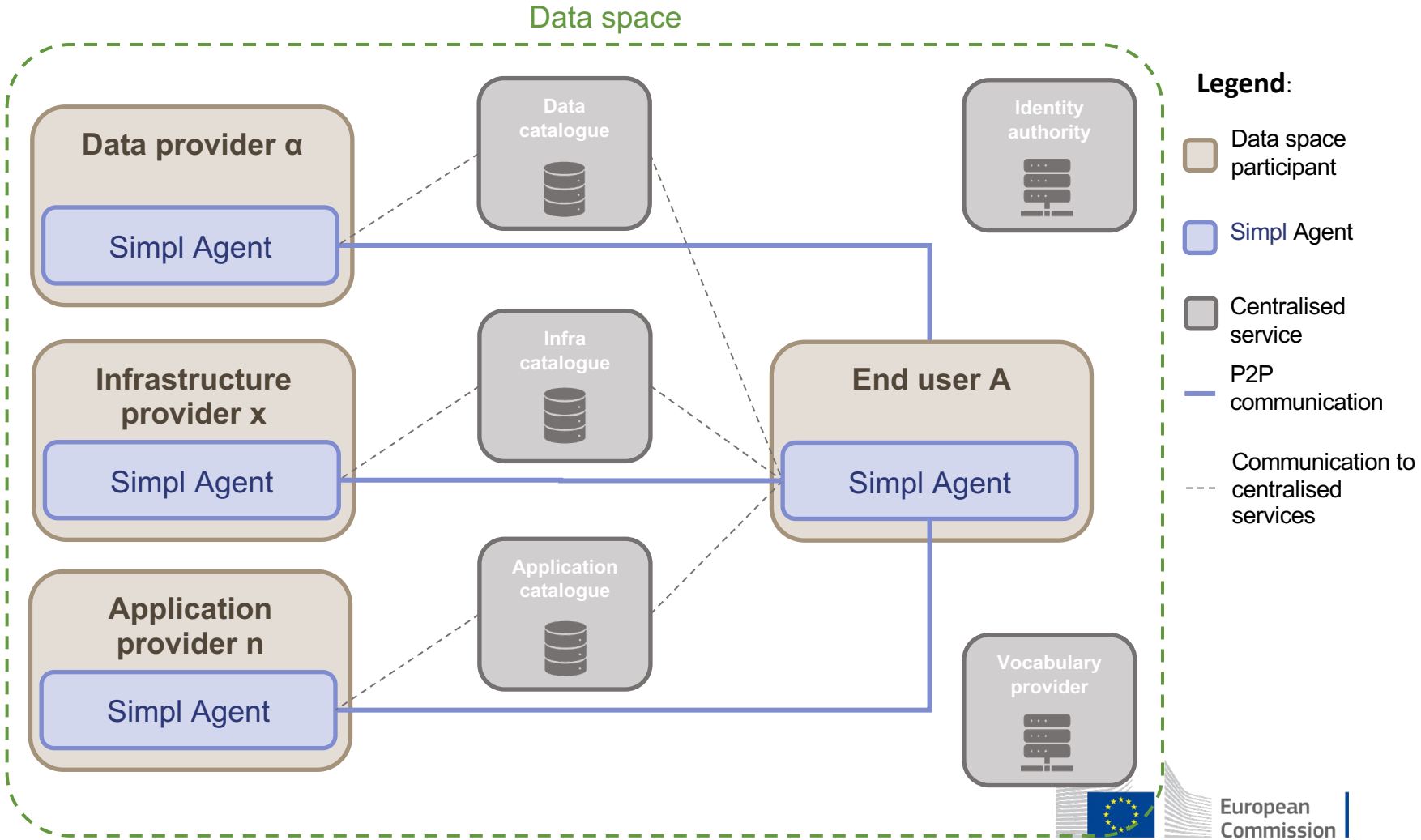
Simpl system architecture (1/2)

The platform capabilities are mapped onto centralised and decentralised system components

Key takeaways

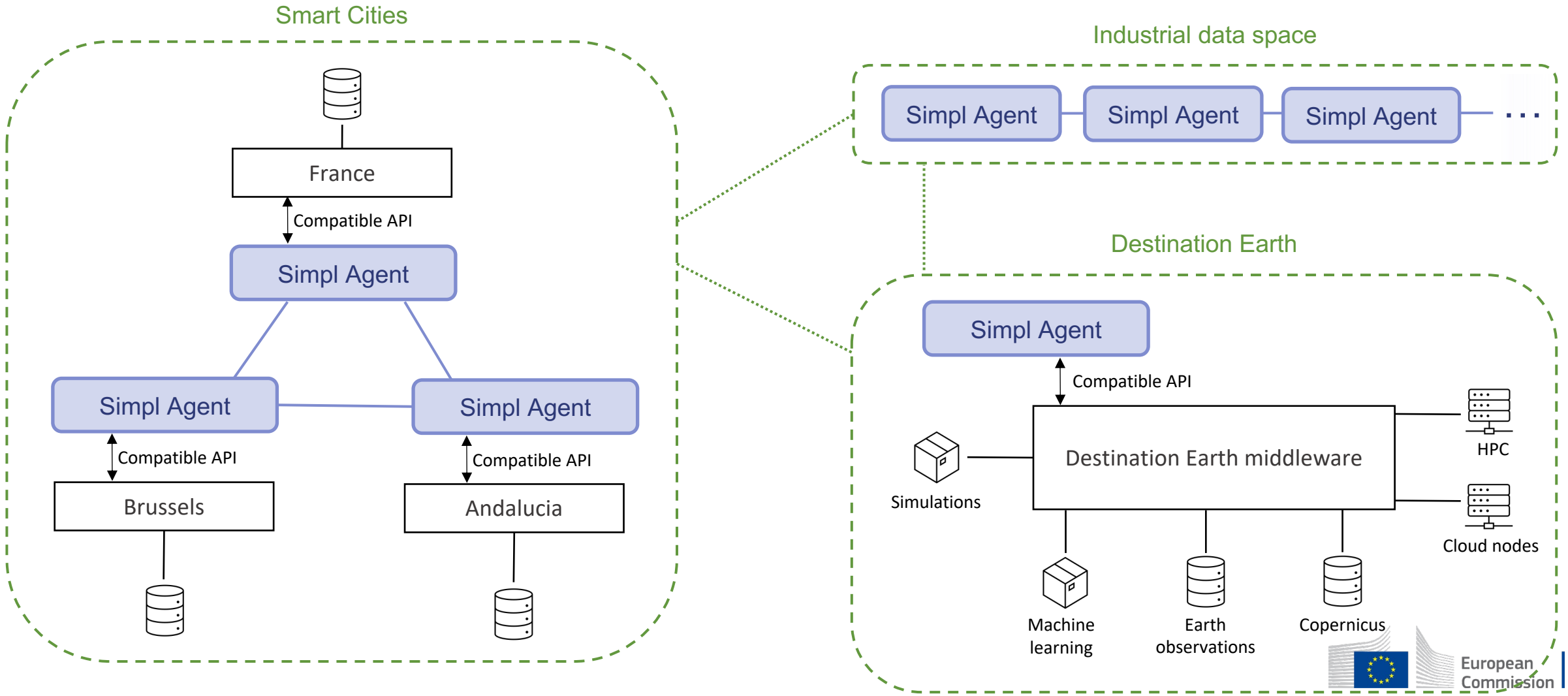
- Centralised services = services that provide capabilities through centralised system components.

- *Data, Infrastructure and application catalogues* provide the cataloging service for end users to discover shared services in the data space
- *Vocabulary providers* provide the definition of metadata representation, vocabularies, and ontologies
- *Identity authorities* manage the identities of the data space participants and provides proofs that other participants can use for authentication and authorization



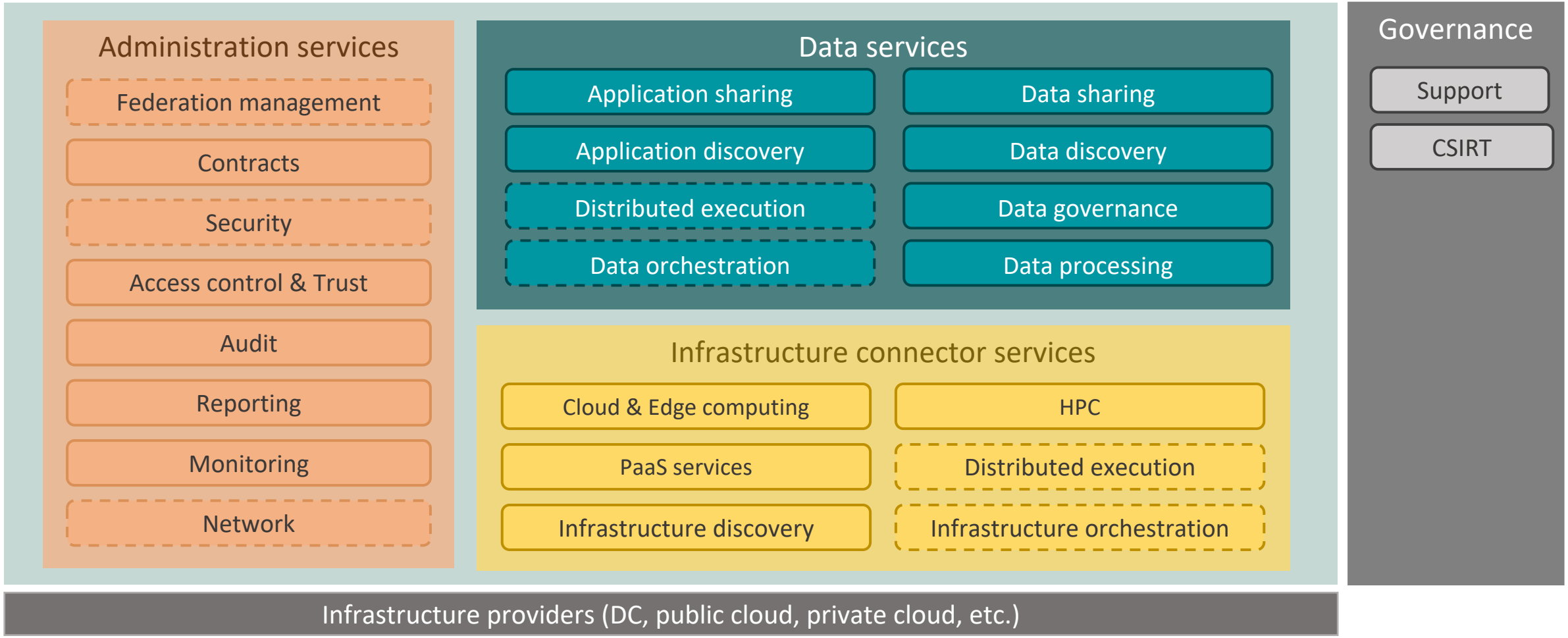
Simpl system architecture (2/2)

An example deployment of different initiatives and data spaces using Simpl



Simpl conceptual architecture (1/5)

Four architectural layers describe the capabilities of the Smart Middleware Platform



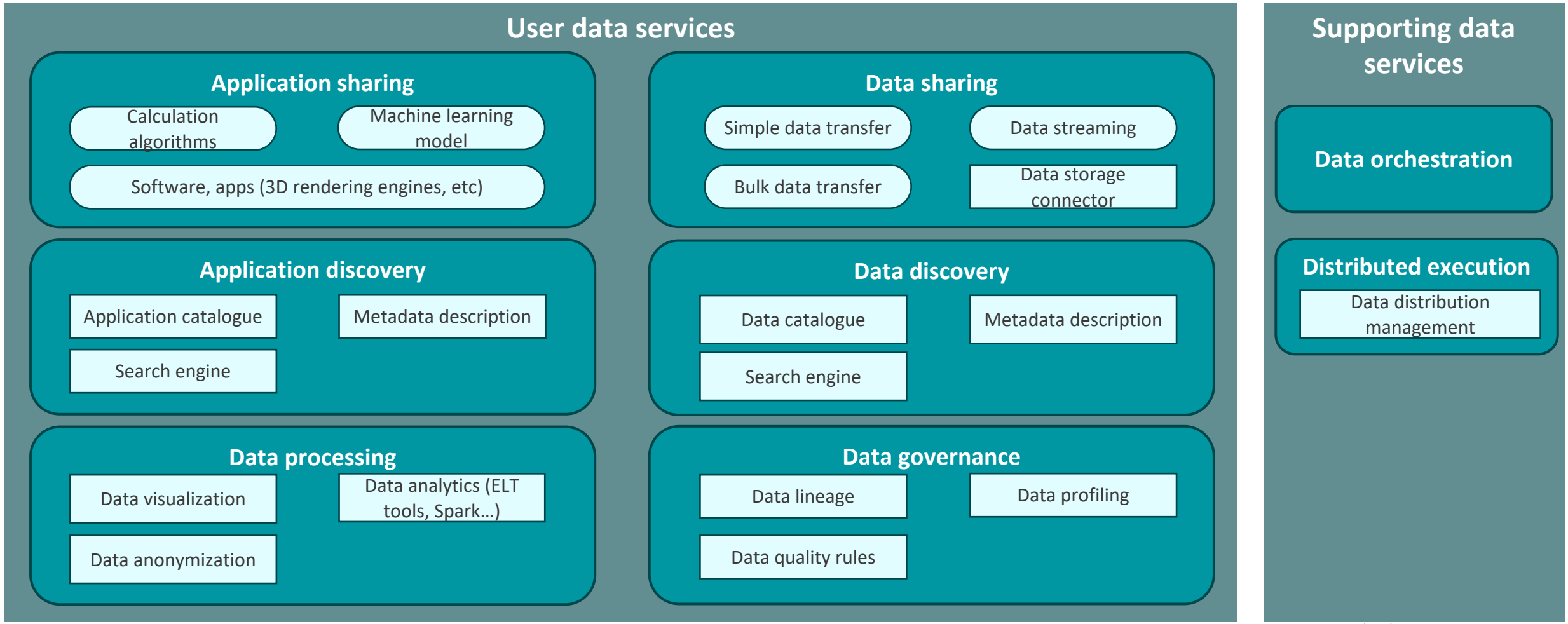
Legend: User services | Supporting services

More details al Simpl website, The architecture vision;

European Commission

Simpl conceptual architecture (2/5)

Capabilities of the data layer services are subdivided into several building blocks



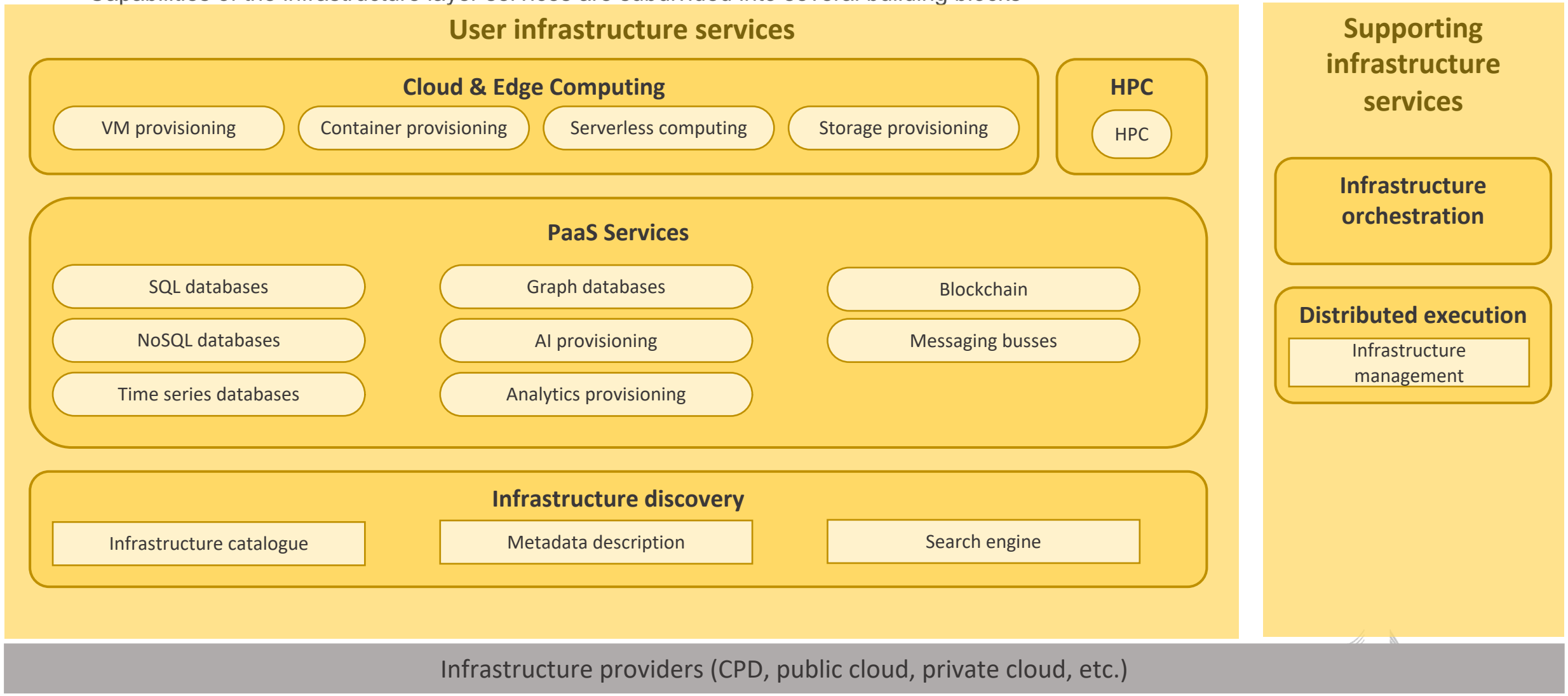
Legend: Access-through | Built-in

More details al Simpl website, The architecture vision;

European Commission

Simpl conceptual architecture (3/5)

Capabilities of the infrastructure layer services are subdivided into several building blocks

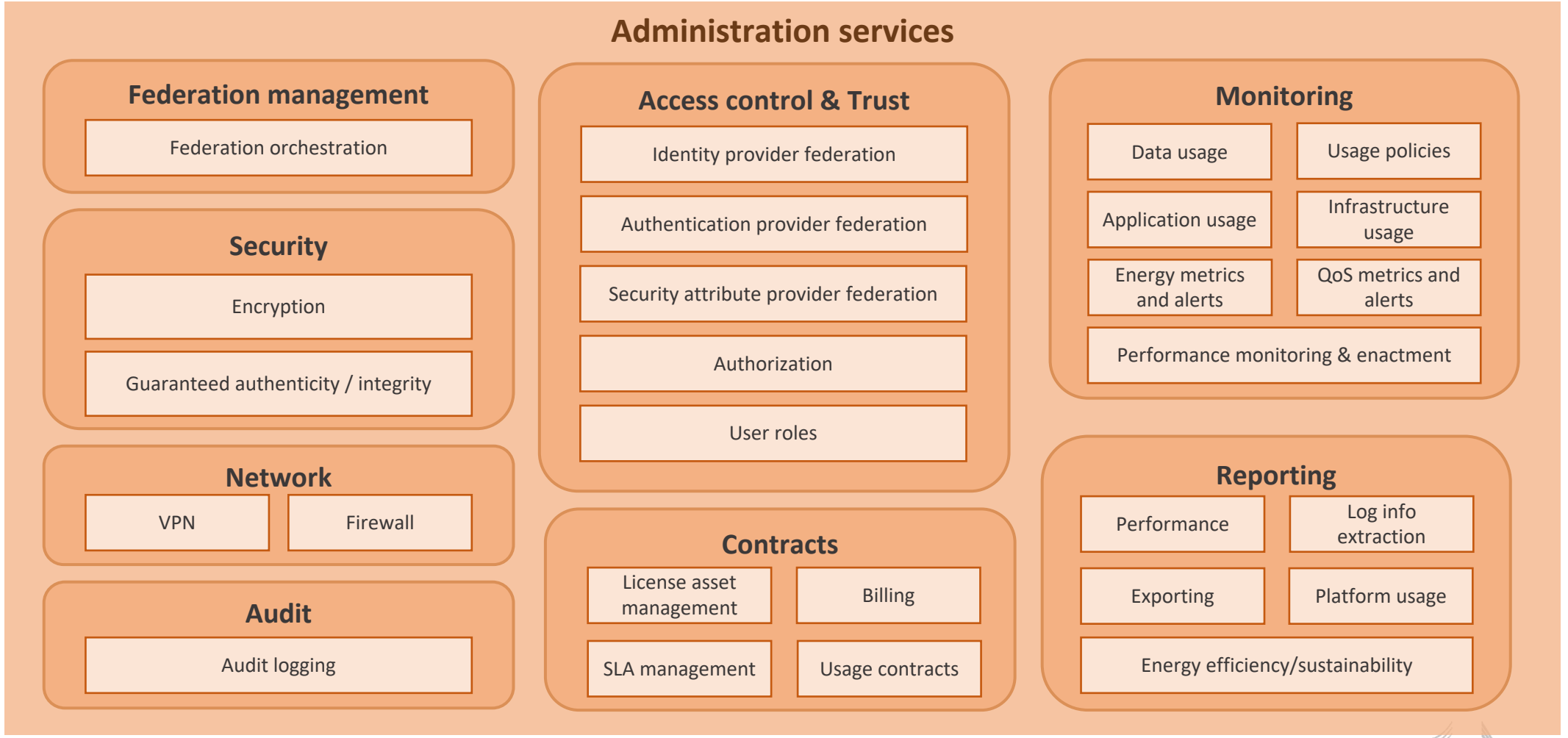


Legend: Access-through | Built-in

More details al Simpl website, The architecture vision;

Simpl conceptual architecture (4/5)

Capabilities of the administration layer services are subdivided into several building blocks



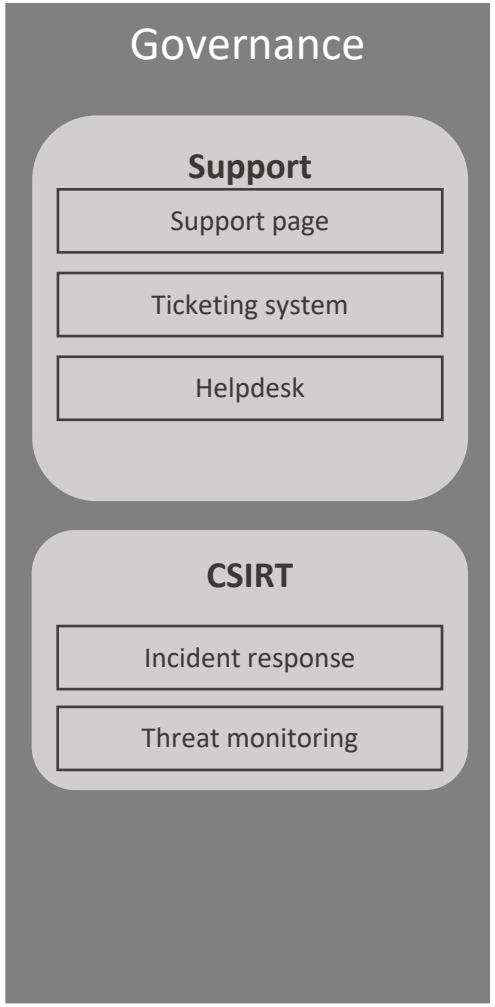
Legend: Built-in

More details al Simpl website, The architecture vision;

European Commission

Simpl conceptual architecture (5/5)

Capabilities of the governance layer services are subdivided into several building blocks



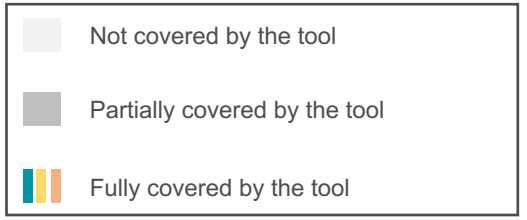
Legend: Built in

More details al Simpl website, The architecture vision;

European Commission

Product coverage of SMP building blocks

Selection made according to building block mapping within the Open-Source market possibilities as well as EU community usage



	Infra. Layer (10 blocks)	Data layer (21 blocks)	Admin. Layer (23 blocks)
EGI	**7/10** fully covered	**13/21** fully covered	**10/23** fully covered
OpenStack	**8/10** fully covered	**4/21** fully covered	**18/23** fully covered
VanillaStack	**8/10** fully covered	**Doesn´t apply to this layer**	**19/23** fully covered
Linux	**10/10** fully covered	**16/21** fully covered	**15/23** fully covered
X - Road	**Does not apply to this layer**	**11/21** fully covered	**10/23** fully covered
Apache	**Does not apply to this layer**	**18/21** fully covered	**Does not apply to this layer**

Blocks not entirely covered by products

- Infrastructure layer weak building block (BB) coverage:
  - Serverless computing
  - PaaS services
- Data layer weak BB coverage:
  - Software, apps
  - Anonymisation
  - Streaming
  - Quality rules
- Administration layer weak BB coverage:
  - License asset management
  - Usage contracts
  - SLA management

More details at: Simpl website, The market analysis report

European Commission

Product assessment results

Product mapping against the established set of criteria

Criteria	EGI	OpenStack*	X - Road	Linux Foun.	Apache
Scalability	🟢	🟢	🟡	🟢	🟢
Interoperability	🟢	🟡	🟡	🟡	🟡
Reusability	🟢	🟡	🟡	🟢	🟢
Time to market	🟡	🟡	🟡	🟡	🟡
Vendor lock-in	🟡	🟡	🟢	🟢	🟢
License	🟢	🟢	🟢	🟢	🟢
European market penetration	🟢	🟢	🟢	🟢	🟢
Politics of the community	🟡	🟡	🟡	🟡	🟢
Community contributions mechanism and status	🟡	🟡	🟡	🟢	🟢
Coverage with SMP requirements	🟡	🟢	🟡	🟢	🟡
Effort to add new functionalities	🟡	🟡	🟢	🟢	🔴
Security & privacy by design	🟡	🟡	🟡	🟡	🟡
Technical integration complexity	🟡	🔴	🟡	🔴	🟢

Conclusions

- No single product covers the majority of SMP requirements
- EGI is promising regarding infrastructure and data layers. The main downside is its specific research and scientific orientation
- Highly active OpenStack tools could cover administration services alone or integrated with EGI
- Linux foundation has a very active community with a broad range of tools, but integration is complex
- Apache provides multiplatform tools and a wide community, but integration is complex
- X – Road offers a trust federation for data services to exchange, manage and set access control features for the SMP, but does not cover some key capabilities such as Contracts or Reporting and Monitoring

* VanillaStack is not displayed as it encompasses OpenStack and all its other solutions are part of CNCF. VanillaStack has the same evaluations as OpenStack.

🟢 Complies 🟡 Partially complies 🔴 Doesn´t comply

More details at: Simpl website, The market analysis report

Complementary components

In addition to the main products mapping, a variety of open-source initiatives provide potentially useful tooling for specific SMP capabilities

	Building Blocks	FIWARE	ECLIPSE FOUNDATION	HARVESTER / RANCHER	truedat
Description	• Digital Europe Program initiative consisting in 5 main blocks • Reduced evolution of the former group of ten building blocks established in the CEF concluded programme	• Fiware components can be assembled together and with third-party components. • A Context Broker Generic Enabler is required based on Context Broker Generic enablers built around the Context Broker	• Long list of open-source products • Successful community • Member of GAIA-X • Relevant projects: -Dataspace Connector -Eclipse IoT	• Harvester is designed for users looking for cloud native HCI solutions • Rancher allows Container-as-a-service delivery • Rancher includes RKE, K3S and Kubernetes operations	• Open-source data governance business solution • Enables to organize & enrich information through configurable workflows and monitor data governance activity
Specific Tooling	eDelivery eID eSignature eInvoicing Context broker	Orion context broker Scorpio Broker OpenVudu Fogfow Keyrock Wilma		RKE K3S	

More details at: Simpl website, The market analysis report

Complementary components

In addition to the main products mapping, a variety of open-source initiatives provide potentially useful tooling for specific SMP capabilities

Building Blocks

- Data orchestration
- Simple data transfer
- Authentication provider federation
- Identity provider federation
- Guaranteed authenticity / integrity
- Billing

FIWARE

- Data distribution management
- Federation orchestration
- Identity provider federation
- Guaranteed authenticity / integrity

ECLIPSE FOUNDATION

- Search engine
- Bulk data transfer
- Simple data transfer
- Machine learning models
- Cloud & Edge Computing
- Audit logging
- Usage contracts
- Data distribution management
- Data catalogue
- Calculation apps
- Data visualization
- Software & apps
- Data and app usage

HARVESTER / RANCHER

- VM provisioning
- Storage provisioning
- Container provisioning
- Audit logging
- Authentication provider federation
- Application usage
- Guaranteed authenticity / integrity

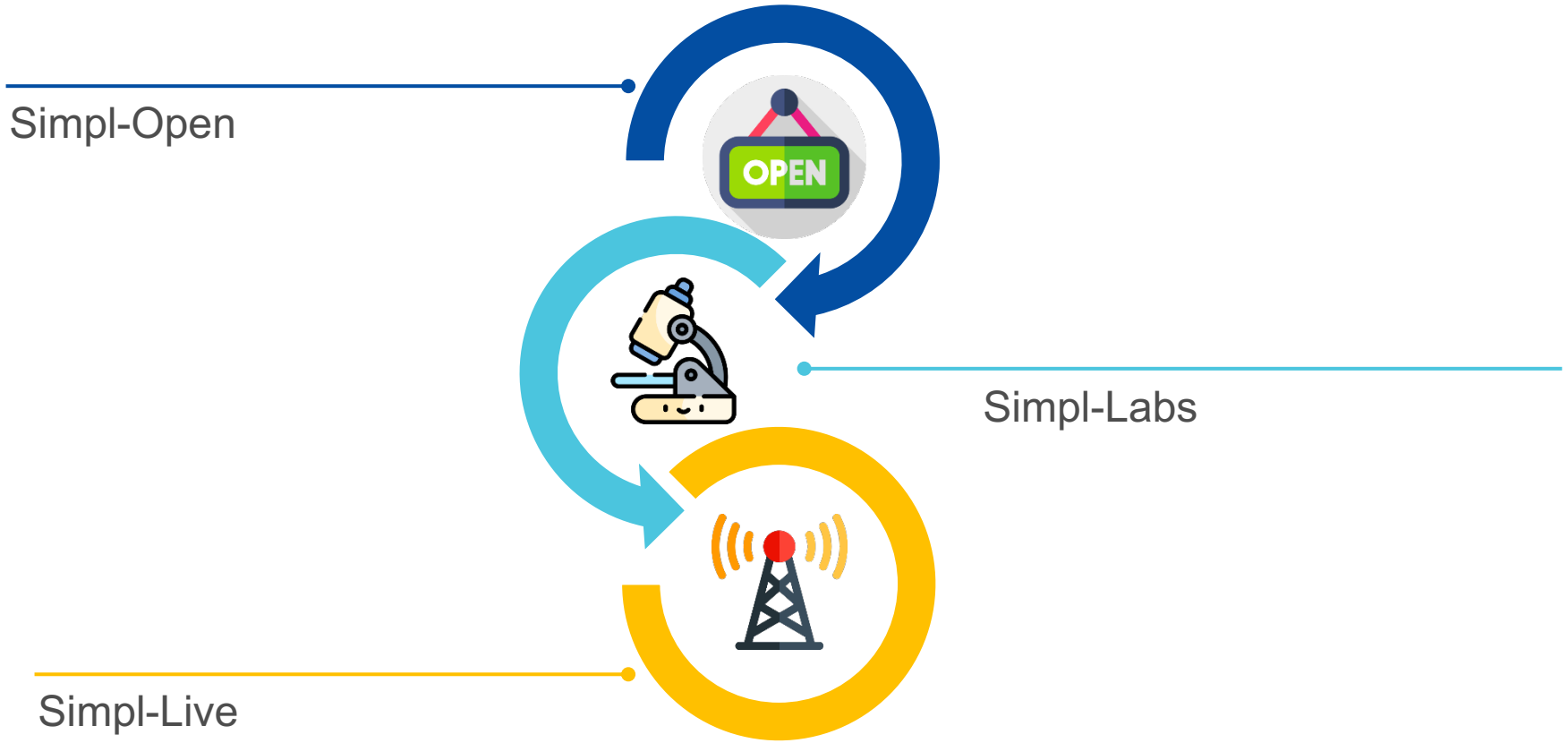
truedat

- Data quality rules
- Data lineage
- Data profiling
- Metadata description

Data layer BB
Infrastructure layer BB
Administration layer BB

More details at: Simpl website, The market analysis report

A tender for three products



European Commission

Simpl-Open

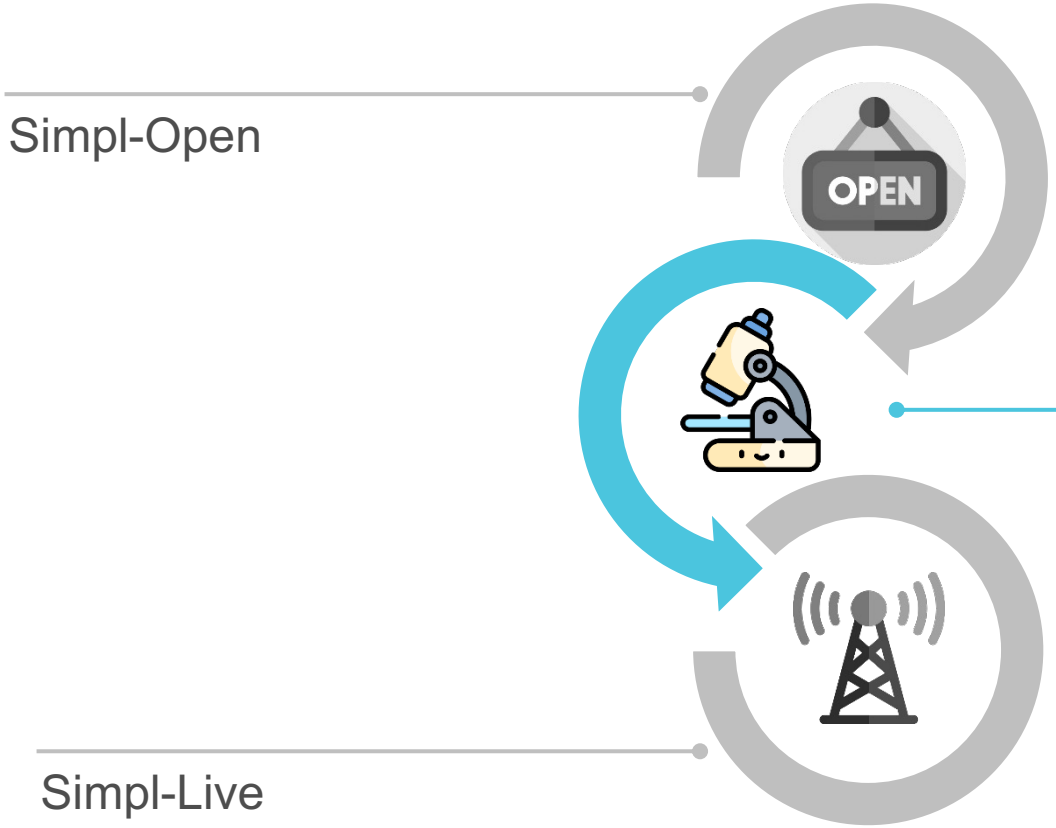
Simpl-Open

The core product of Simpl
Anchored in the prep study

Simpl-Labs

Simpl-Live

Simpl-Labs



= playground and demonstration environment for Simpl-Open

Simpl-Live

Simpl-Open

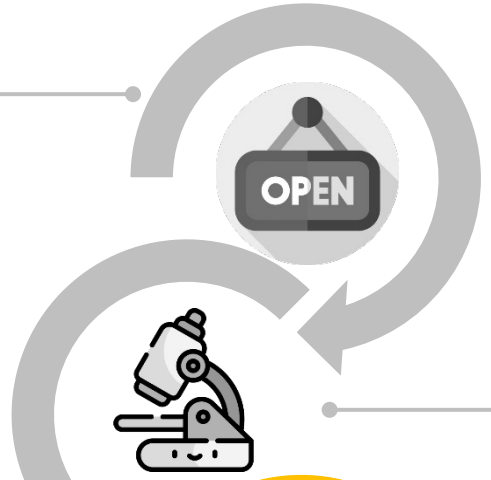
Simpl-Labs

= instances of Simpl-Open for sectoral data spaces

Simpl-Live

European Commission

Simpl-Live



Simpl-Open

Simpl-Labs

= instances of Simpl-Open for sectoral data spaces

Simpl-Live

Secondary eHealth Data Space

Public Procurement Data Space

Language Data Space

European Commission

Content highlights

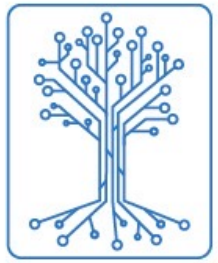
Deloitte Study

Open-Source

Community building

Agile / DevSecOps

Cloud Native Computing Foundation



Gaia-X

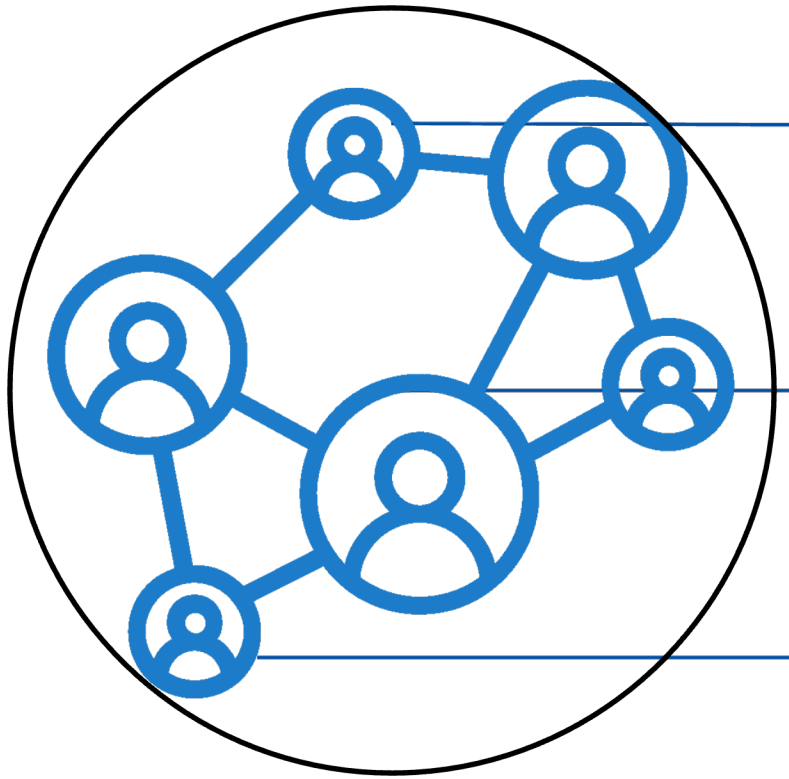
Building blocks

eosc

EOSC

European Commission

Data Spaces Support Centre



Create a network of stakeholders

- Closely work with CSAs and projects funded under DIGITAL

Identify common requirements:

- Support the work of the European Data Innovation Board (DGA)

Create a platform for knowledge exchange:

- Support the deployment of data spaces