



Data Spaces & Research Data

SAGE Overview, Use Cases...

How Research Data might be included in SAGE

Presentation to eIRG, Poznan 25 June 2025

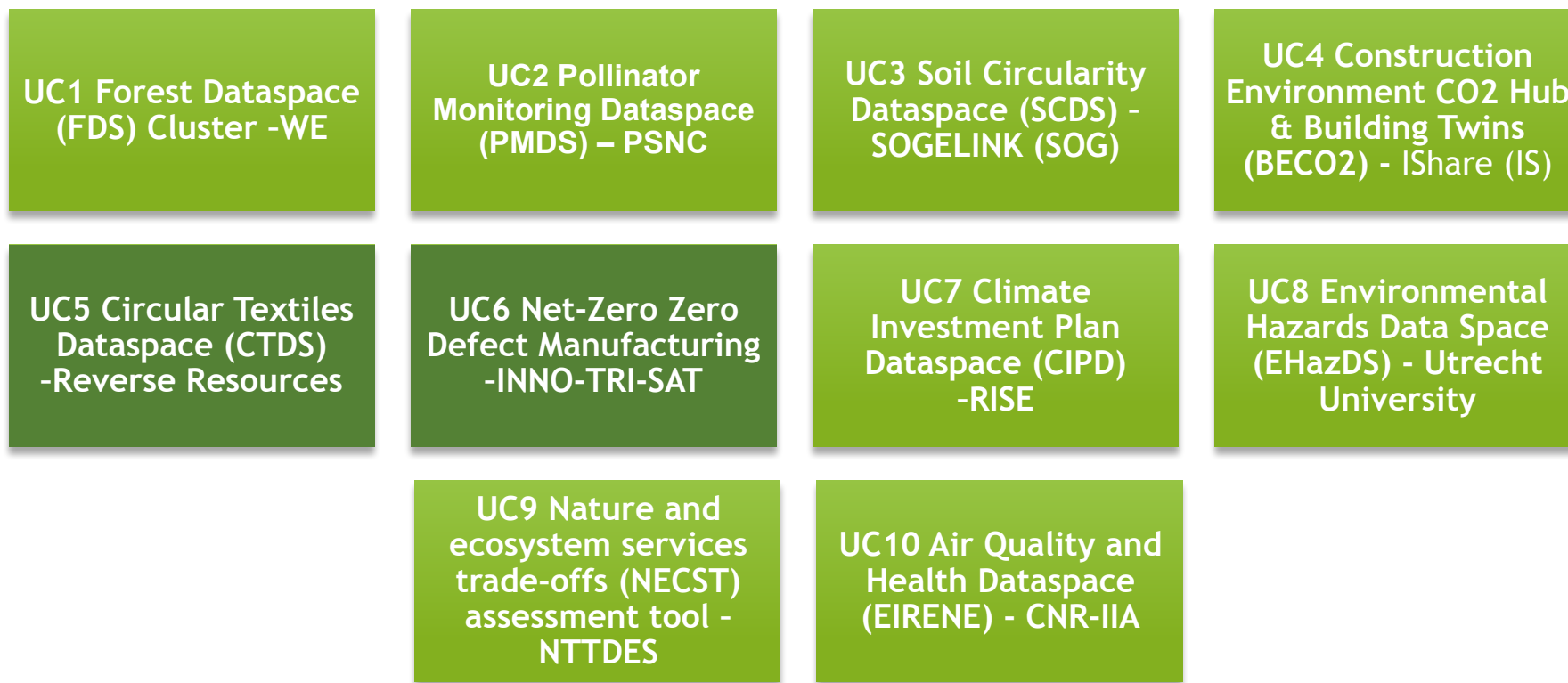




Project Overview

36 months, 40+ partners, €16 million (only 50% funded by EC), 1 March 2025 – 28 February 2028

Project lead: International Data Corporation (IDC)





What do* Data Spaces do?

Enable Data Sovereignty by Data Providers and Data Rights Holders

Enable Data Consumers to Find the Data They Need

Support Interoperability (all layers)

- Within Use Cases
- Across Use Cases
- Between Data Spaces

Maintain Data Sovereignty Even after Data is Combined

Enable Value to Be Created from Data

Enable Trust Between Data Providers/Rights Holders and Data Consumers

- Even when data moves out of the control of the Data Provider/Rights Holders

Data Spaces do NOT store or process data

- They DO store and process metadata/self-descriptions

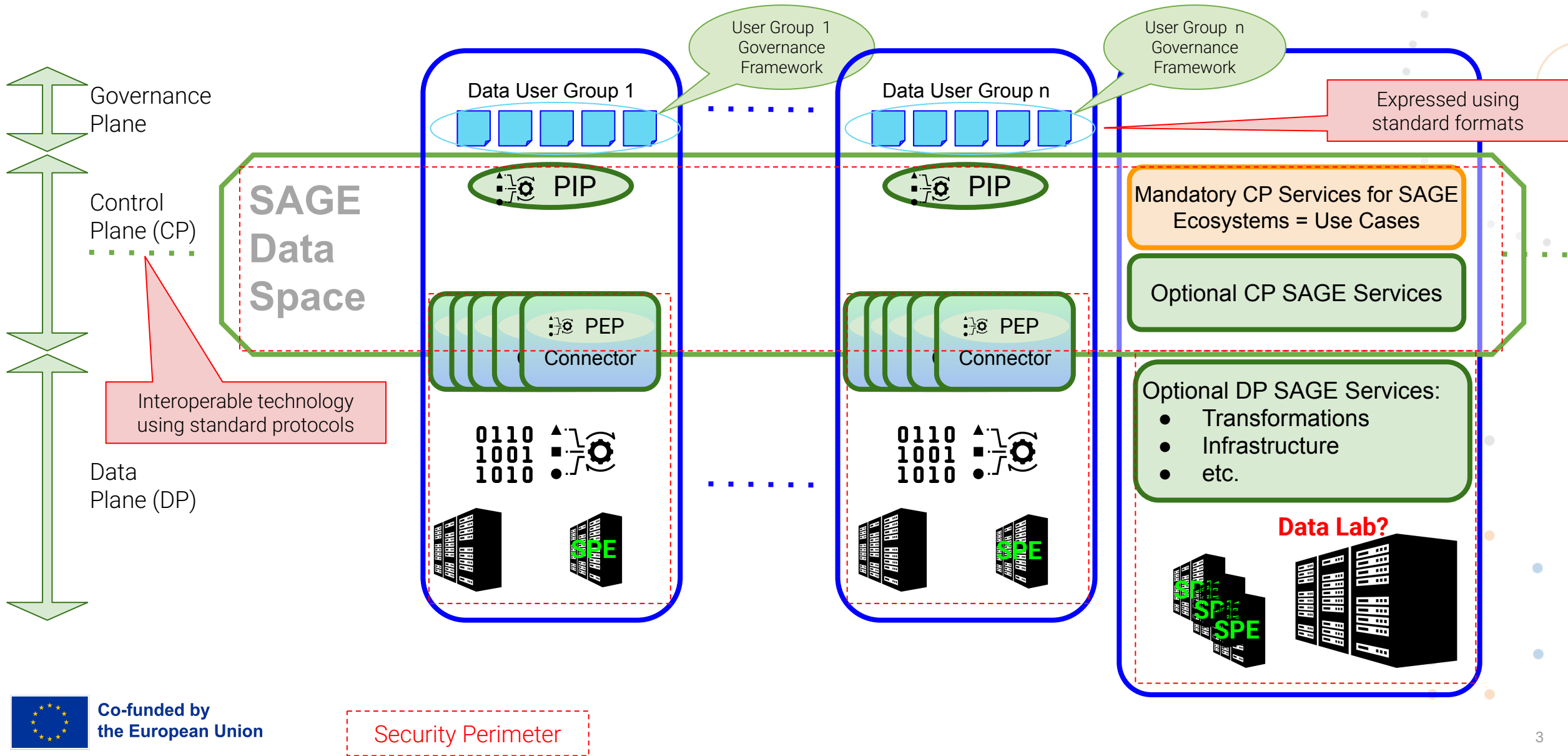
* perhaps “should” since data spaces have yet to fulfill the promise completely.





SAGE | The Data Space for a Sustainable Green Europe

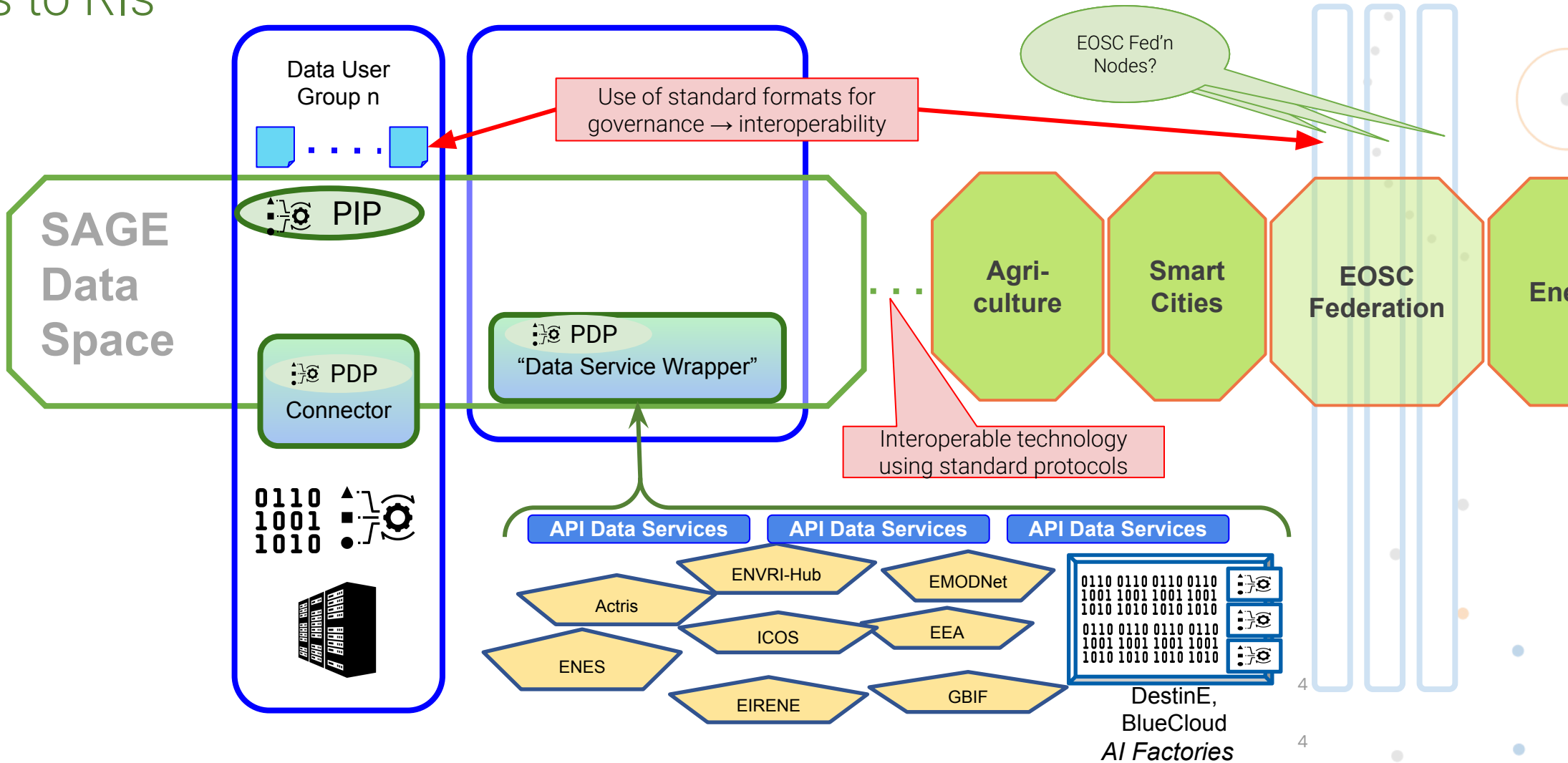
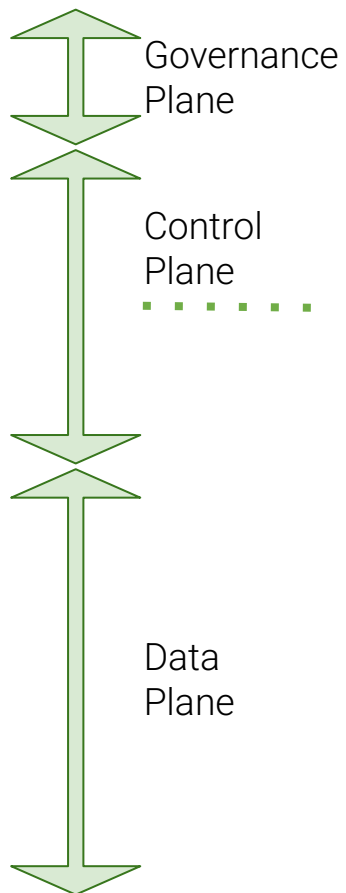
Architecture Overview





SAGE | The Data Space for a Sustainable Green Europe

Links to RIs





Links to Research Data by Use Case (preliminary)

Use Case	Research data	Public Data	Non-Public Data	Use Case “Pattern”
UC1: Forest Transformation		Earth Observation, Climate, Weather	National Forest Inventories, private forest owner data	Add Value to Private Data from “Big Pile of Data” (BPOD)
UC2: Pollinator Monitoring*	Project-based	Earth Observation, Climate, Weather, Land Use/ Agricultural	National, local agricultural info, private farm data (incl. FMIS), apiary business data	Add Value to Private Data from BPOD
UC3: Soil Circularity		Earth Observation, Land Use	National, local soil data	Create Private Data from BPOD, Enable Market and Circular Data Flows
UC4: Built Environment CO2 Monitoring	Life Cycle Analyses, CO2 intensity results	Local CO2 intensity (e.g. by energy source)	Building owner data, supplier, supply chain, supplier reporting	Enable Circular Data Flows in Supply Chain
UC5: Circular Textiles		Local CO2 intensity (e.g. by energy source)	Textile manufacturer/recycler data, supplier, supply chain, supplier reporting	Enable Circular Data Flows in Supply Chain

** More Detail from next speaker!*





Links to Research Data by Use Case (preliminary)

Use Case	Research data	Public Data	Non-Public Data	Use Case “Pattern”
UC6: Net Zero Defect Manufacturing	Life Cycle Analyses, CO2 intensity results	Local CO2 intensity (e.g. by energy source)	Supply Chain data	Optimize supply chain and manufacturing performance
UC7: Climate Investment Plan	Project-based, sensitivity analyses	Earth Observation, Climate, Weather, in situ obs.	Municipal data (many sources/types)	Add Value to Municipal Data from “Big Pile of Data” (BPOD) for scenario comparison, planning
UC8: Environmental Hazards	Environmental RIs	Earth Observation, Climate, Weather, Land Use/ Agricultural	Future: Location data, personal travel history	Create Big Pile of Data, Then Add Value to Private Data from BPOD
UC9: Nature and ecosystem services trade-offs	Environmental RIs	Earth Observation, Climate, Weather, Land Use/ Agricultural	Private Landholder scenario planning	Add Value to Private Data from BPOD for scenario comparison, planning
UC10: Air Quality and Health	Environmental RIs	Earth Observation, Climate, Weather, Land Use/ Agricultural	Clinical health data Future: Location data, personal travel history	Create Big Pile of Data, Then Add Value to Private Data from BPOD





Challenges Ahead – a few to think about...

Interoperability

- Federated Discovery: no good solutions in data spaces, RIs, or EOSC
- Common Data (and Services) Visibility, Access and Use (VAU) vocabularies – if used across distinct user groups and domains
 - Should cover both organisational and legal interoperability – important for AI
- Consistent logic for combining VAU policies when data with date with different policies are combined
 - Important for AI!
- Semantic Interoperability – without mandating specific data models
 - ODI's "Framework for AI Ready Data" is only a framework – with a few well-known additions to FAIR

Cybersecurity and Secure Processing Environments

- Data Planes of different data providers in Data Spaces (INCLUDING RIs in EOSC Federation!) need to agree on common frameworks
- Also agree to work together, to achieve a secure end-to-end system (infrastructure commons)
- Secure Processing Environments (SPEs) – important for data sovereignty. ideally robust frameworks for defining and deploying these across many infrastructures.
 - These also require consistent logic for combining data

Data Labs?

- Ironically, Data Spaces don't store or process data – this is the responsibility of Participants, in the Data Plane
 - This IS a problem for SMEs, start-ups, companies that normally use US Hyperscalers, and almost all Research Infrastructures!
- Wouldn't it be nice to have a good place to store good data that respects data sovereignty, enforces trust and enables interoperability?





SAGE | The Data Space for a Sustainable Green Europe

Thank you!

email: mjdietch@bloodstone-consulting.com



Co-funded by
the European Union