



WeTransform!

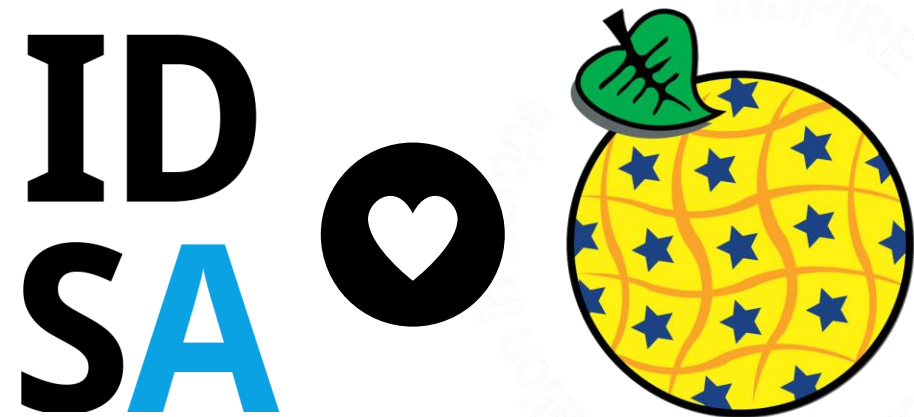
Interoperability Provisions for the Green Deal Data Space

Thorsten Reitz, Co-Founder/CEO
May 17th, 2024

78th INSPIRE MIG-T Meeting

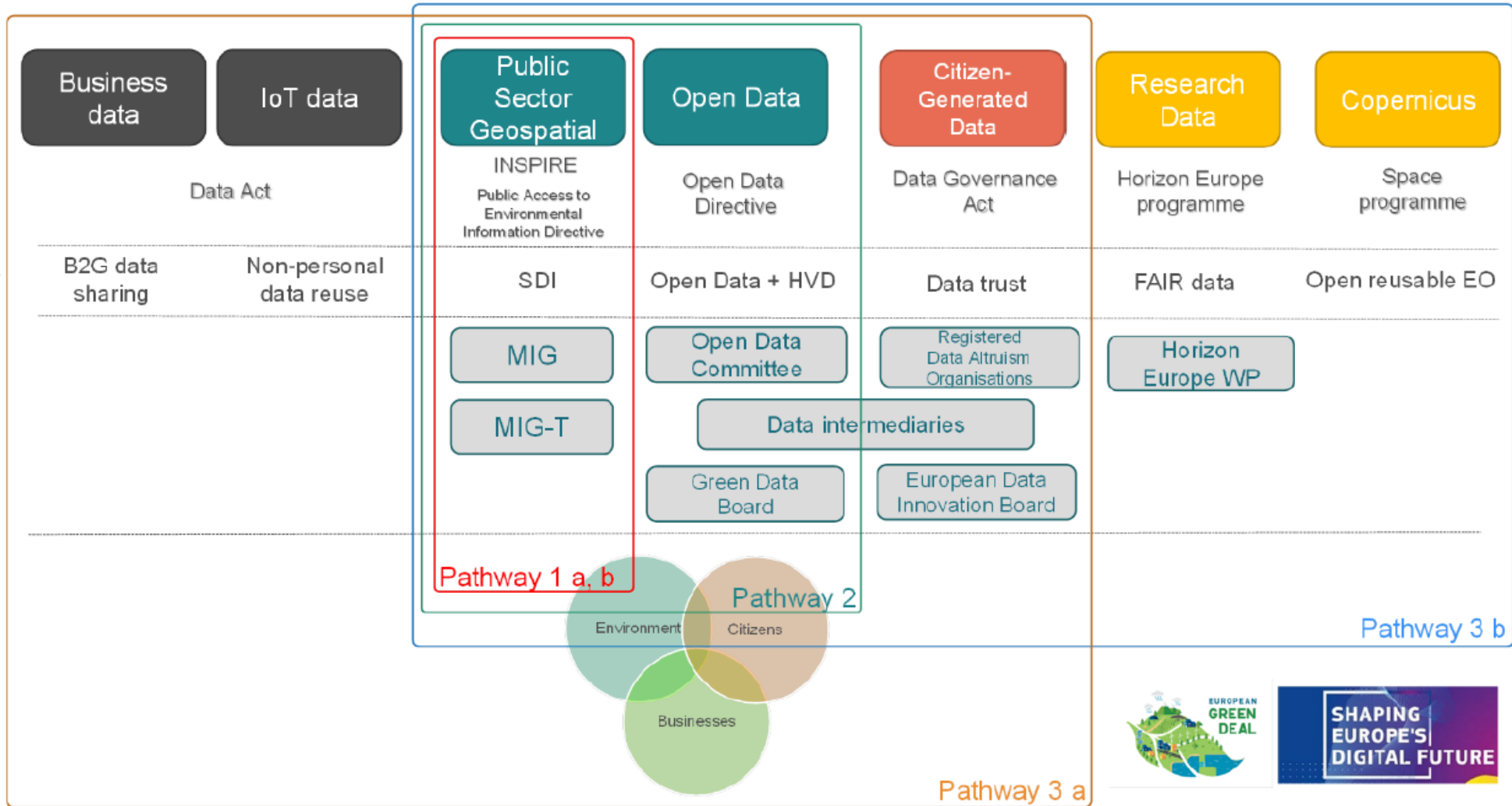
SCOPE & OBJECTIVES

- **JRC Expert Contract 06/2023 to 02/2024**
- **Objectives of the Study:**
 - Analyse existing provisions and their impact *qualitatively* and *quantitatively*
 - Propose new provisions for interoperability in the EGDDS
 - Take into account different policy pathways for the transition from INSPIRE to the EGGDS
 - Discuss and validate provisions
- **Present Conclusions**
- **Steps to Implement and Evaluate Provisions**



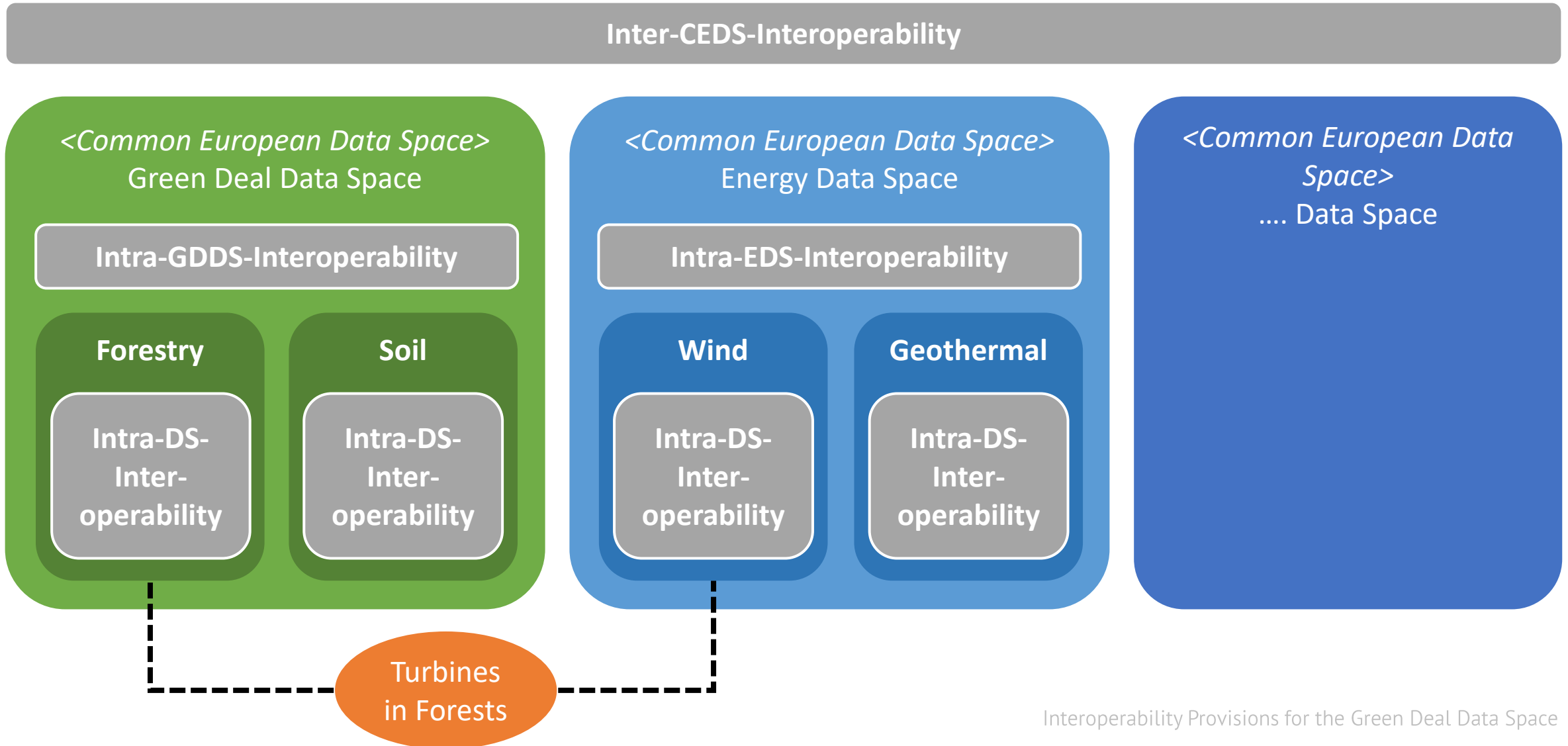
POLICY PATHWAYS

Where do we go? Defining the Data Scope of the Green Deal Data Space



FRAMEWORK: SCOPE OF INTEROPERABILITY

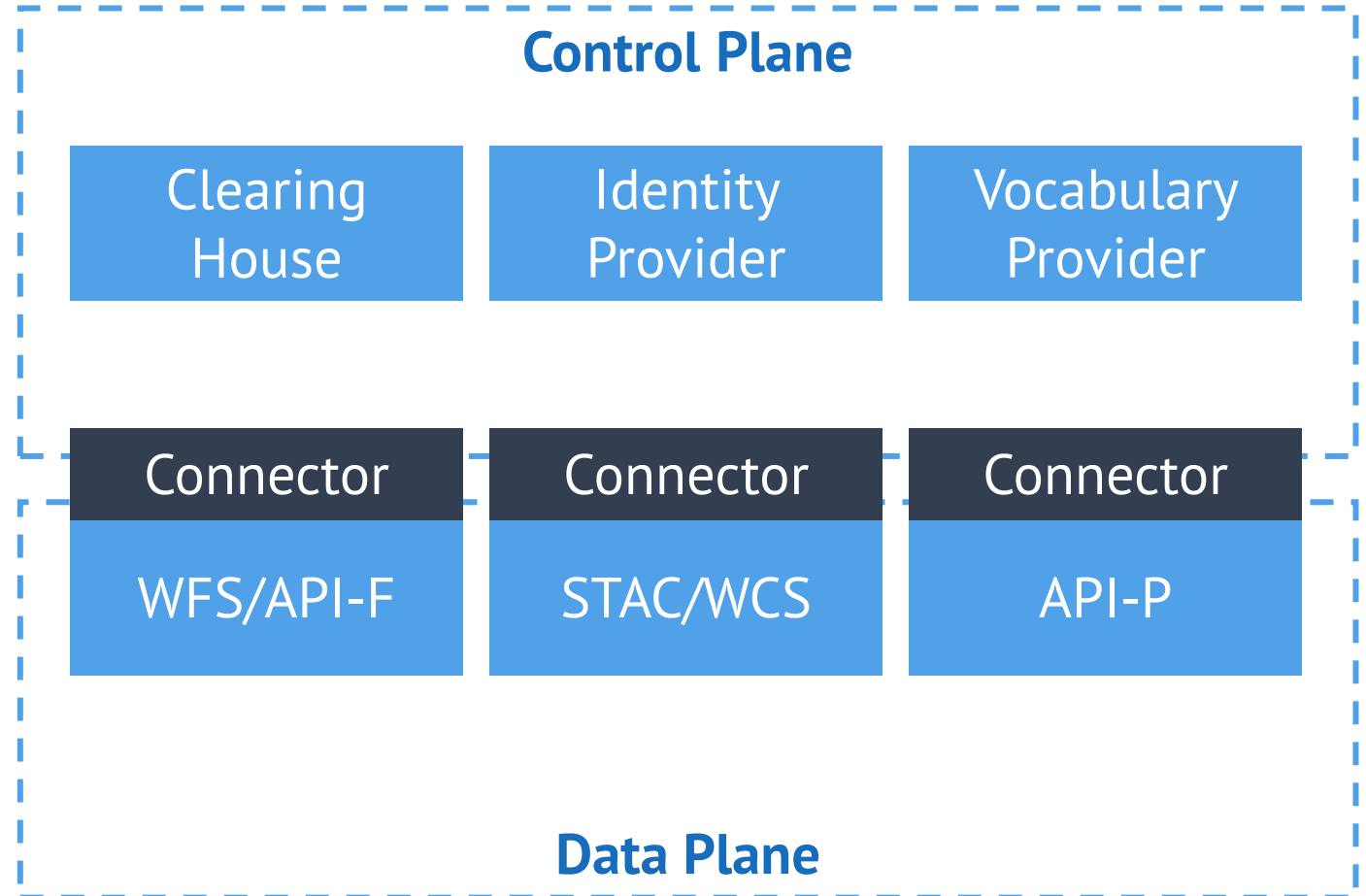
From Intra-DS to Inter-CEDS Collaboration



FRAMEWORK: DATA AND CONTROL PLANE

Interoperability as part of the Architecture

- **Control Plane**
 - Data Spaces Protocol implementation
 - Contracts/Policies
 - Identification
 - Metadata
- **Shared**
 - Vocabulary/Semantics
 - Quality, Provenance, Traceability
- **Data Plane**
 - Protocol
 - Encoding
 - Data Structures/Models



FRAMEWORK: ASPECTS

What contributes significantly to data interoperability?

- **Semantic Interoperability**

- ➔ a) Dictionaries, Vocabularies, Taxonomies
- b) Standards-based conceptual models
- ➔ c) Well-defined units of measure and reference systems
- d) Semantic Mappings and documented mismatches

- **Technical Interoperability**

- ➔ a) Usage of open data formats
- b) Usage of open APIs and protocols
- ➔ c) Formally specified logical data models
- d) Validation and Quality Assurance Infrastructure

- **Process Interoperability**

- a. Common Quality Standards
- b. Lifecycle rules
- c. Monitoring and Reporting Mechanisms
- d. Governance and Maintenance Mechanisms

What 3 aspects
have the best
cost/benefit ratio
at integration
time?

What aspect helps
most to achieve
integration at
scale?

CONCLUSIONS: IMPACT OF INTEROPERABILITY PROVISIONS

What can we learn from history?

- **Analysis of the *Impact* of EC and national regulations on the 12 aspects**

1. Semantic Interoperability
2. Technical Interoperability
3. Process Interoperability

- **Legend:**

- *None*: No discernible impact
- *Minor*: Partial impact on some (10-50%) resources
- *Major*: Substantial impact on some resources
- *Full*: Substantial impact on most (50%+) resources

Aspect	PSIE	INSPIRE	END	HVD	TN-ITS	XPlanung
1a	None	Major	Full	Minor	Full	Full
1b	None	Full	Full	Minor	Full	Full
1c	None	Full	Full	None	Full	Full
1d	None	None	Major	None	None	Minor
2a	Minor	Major	Full	Minor	Full	Full
2b	None	Full	None	Minor	Major	None
2c	None	Major	Full	None	Full	Full
2d	None	Minor	Full	None	Major	Major
3a	None	Minor	Major	None	Major	Minor
3b	None	Minor	Full	Minor	Full	Major
3c	None	Major	Major	None	Minor	None
3d	None	Minor	None	None	Major	Major

CONCLUSIONS: DATA INTEROPERABILITY IN INSPIRE

What to retain, what to change



- Standards that define concepts, vocabularies and structure have been established
- Conceptual and logical models have been built
- Most available data uses open formats and open APIs
- Harmonised data is well-documented
- Monitoring, Validation and Compliance Mechanisms and Infrastructure



- Early implementation testing
- All-or-nothing compliance
- Gaps in Vocabularies / Codelists
- Accidental complexity in logical data models
- Semantic and logical mappings to local and related standards missing
- Inflexible legal framework and slow fixes until ~2021
- Little consequence to not being interoperable

CONCLUSIONS: LOW-HANGING FRUITS

Aspects with Best Impact to Difficulty Ratio

- **1a Units and Reference Systems (Ratio 1:2.3)**
- **2c Open APIs (Ratio 1:2.2)**
- **2b Open Formats (Ratio 1:2.0)**
- **3b Lifecycle Rules (Ratio 1:1.8)**
- **2a Logical Models (Ratio 1:1.7)**
- **3c Monitoring Reporting (Ratio 1:1.7)**

Aspect	Maturity	Difficulty	Impact
1a Units/Reference Systems	High	Overall Difficulty: Low (2.0), individual factors: <ul style="list-style-type: none"> - Specialist domain knowledge: 2 - Specialist technical knowledge: 2 - Broad Agreement: 1 (SI units, base standards) - Complexity of the domain: 3 - Tooling/Methods maturity: 2 	Overall Impact: High (4.5), individual factors: <ul style="list-style-type: none"> - Understanding: 4 - Automation: 5 - Combination: 5 - Usage: 4
1b Dictionaries	Moderate	Overall Difficulty: Medium (3.4), individual factors: <ul style="list-style-type: none"> - Specialist domain knowledge: 4 (vocabularies need to be added or extended) - Specialist technical knowledge: 3 (for more complex taxonomies) - Broad Agreement: 4 (in the remaining domains, this is difficult) - Complexity of the domain: 4 (the domains missing such lists are complex) - Tooling/Methods maturity: 2 	Overall Impact: High (4.25), individual factors: <ul style="list-style-type: none"> - Understanding: 5 - Automation: 4 - Combination: 4 - Usage: 4
1c Conceptual Models	Moderate	Overall Difficulty: Medium (3.2), individual factors: <ul style="list-style-type: none"> - Specialist domain knowledge: 4 (extensions, simplifications) - Specialist technical knowledge: 3 - Broad Agreement: 3 - Complexity of the domain: 4 - Tooling/Methods maturity: 2 	Overall Impact: High (3.75), individual factors: <ul style="list-style-type: none"> - Understanding: 5 - Automation: 2 - Combination: 3 - Usage: 4
1d Mappings	Low	Overall Difficulty: Medium (3.0), individual factors: <ul style="list-style-type: none"> - Specialist domain knowledge: 4 - Specialist technical knowledge: 4 - Broad Agreement: 2 (using more intermediaries simplifies this) - Complexity of the domain: 4 - Tooling/Methods maturity: 1 	Overall Impact: High (4.25), individual factors: <ul style="list-style-type: none"> - Understanding: 4 - Automation: 4 - Combination: 5 - Usage: 4
2a Logical Models	Moderate	Overall Difficulty: Medium (2.2), individual factors: <ul style="list-style-type: none"> - Specialist domain knowledge: 3 (useful simplifications are not trivial to define) - Specialist technical knowledge: 4 (tools 	Overall Impact: High (3.75), individual factors: <ul style="list-style-type: none"> - Understanding: 3 - Automation: 4 - Combination: 3

2c Open APIs	High	Overall Difficulty: Low (1.6), individual factors: <ul style="list-style-type: none"> - Specialist domain knowledge: 2 - Specialist technical knowledge: 1 - Broad Agreement: 2 (some new APIs are coming up) - Complexity of the domain: 2 - Tooling/Methods maturity: 1 	Overall Impact: Medium (3.5), individual factors: <ul style="list-style-type: none"> - Understanding: 2 - Automation: 5 - Combination: 3 - Usage: 4
2d Q&A Infrastructure	High	Overall Difficulty: Medium (2.6), individual factors: <ul style="list-style-type: none"> - Specialist domain knowledge: 2 - Specialist technical knowledge: 4 - Broad Agreement: 2 - Complexity of the domain: 3 - Tooling/Methods maturity: 2 	Overall Impact: High (4.0), individual factors: <ul style="list-style-type: none"> - Understanding: 3 - Automation: 4 - Combination: 5 - Usage: 4
3a Quality Standards	Moderate	Overall Difficulty: Medium (2.8), individual factors: <ul style="list-style-type: none"> - Specialist domain knowledge: 4 - Specialist technical knowledge: 2 - Broad Agreement: 3 - Complexity of the domain: 3 - Tooling/Methods maturity: 2 	Overall Impact: Medium (3.25), individual factors: <ul style="list-style-type: none"> - Understanding: 3 - Automation: 2 - Combination: 4 - Usage: 4
3b Lifecycle Rules	Moderate	Overall Difficulty: Low (1.8), individual factors: <ul style="list-style-type: none"> - Specialist domain knowledge: 2 - Specialist technical knowledge: 1 - Broad Agreement: 2 - Complexity of the domain: 2 - Tooling/Methods maturity: 2 	Overall Impact: Medium (3.25), individual factors: <ul style="list-style-type: none"> - Understanding: 2 - Automation: 4 - Combination: 3 - Usage: 4
3c Monitoring/Reporting	Moderate	Overall Difficulty: Medium (2.4), individual factors: <ul style="list-style-type: none"> - Specialist domain knowledge: 2 - Specialist technical knowledge: 3 - Broad Agreement: 3 - Complexity of the domain: 2 - Tooling/Methods maturity: 2 	Overall Impact: High (4.0), individual factors: <ul style="list-style-type: none"> - Understanding: 2 - Automation: 5 - Combination: 5 - Usage: 4

DRAFTING POTENTIAL PROVISIONS

Clarify Roles & Responsibilities from the beginning

- **Dimensions**
 - Data / Control Plane
 - Interoperability Aspect
 - Interoperability level (Intra-DS, Inter-DS, Inter-CEDS)
- **Focus on Process. not on the selection of specific technologies**
- **Fallback to conventions and top-level specifications**
- **Clarify responsibilities**
 - Data Holders
 - Data Intermediaries
 - Data Users
 - Governance Bodies

(2b) Technical: Usage of open data formats

3.4.1.5 Inter-CEDS

No rule at this level, because it is very likely that the actual format will be highly dependent on the inter-CEDS use case. Generally prescribing XML or JSON would offer no substantial benefit.

3.4.1.6 Intra-CEDS

CEDS governance bodies shall select or define data formats to use for encoding data to be exchanged or processed. Such formats shall be open (i.e. freely accessible and usable without payment of fees) in specification and implementation, and they should be widely implemented and used. Wherever possible, formats standardized by industry or other standardization bodies shall be used. A CEDS governance body has to fully document any choice of formats through a detailed comparison with the alternatives.

If only proprietary formats are widely implemented and used in a given CEDS, the CEDS governance body has to prove so and has to make sure that usage of such a proprietary format does not create barriers to data access. At a minimum, a data format defined for data exchange in a CEDS should have at least one Open-Source reference implementation for I/O and not require the acquisition of Exploitation or Implementation Rights.

3.4.1.7 Intra-EGDDS

The EGDDS DIG shall select or define an open file format for tabular data, for geospatial vector and for geospatial raster data for the purpose of compliance testing. It shall ensure that this file format is able to fully cover the selected conceptual models and semantics to minimize information loss.

The EGDDS DIG may select or define additional file formats to use for unstructured data, for tabular data, for geospatial vector and for geospatial raster data. It shall prove that the respective format is able to fully cover the selected conceptual models and semantics to minimize information loss.

Data holders may provide their data assets in any (either the compliance testing format or any of the additional formats) of the selected or defined file formats.

Intermediaries may offer to automatically convert from one format to another.

To re-use the assets created by the Environmental and Spatial Data Community over the past 15 years, it is suggested to initially define GML as the compliance testing encoding, and GML, GeoJSON, GeoPackage or, where applicable, GeoTIFF as accepted open formats for the EGDDS.

AN EXAMPLE PROVISION

Clarify Roles & Responsibilities from the beginning

“ (1a | Intra-CEDS | both) CEDS governance bodies shall select or define dictionaries and controlled vocabularies for concepts, differentiating properties and classifications. If controlled vocabularies already exist and are commonly used, but the CEDS governance bodies decide not to use these, they **shall** provide a rationale and a mapping to the existing and used controlled vocabularies.

These dictionaries and controlled vocabularies **shall** be updated and released at least once per year and **shall** use clear versioning rules. A roadmap with the planned changes **shall** be published and updated in the same cycle.

Any recognized member of a CEDS community **may** contribute to the definition of these dictionaries and controlled vocabularies, as well as to their extension and maintenance.

”

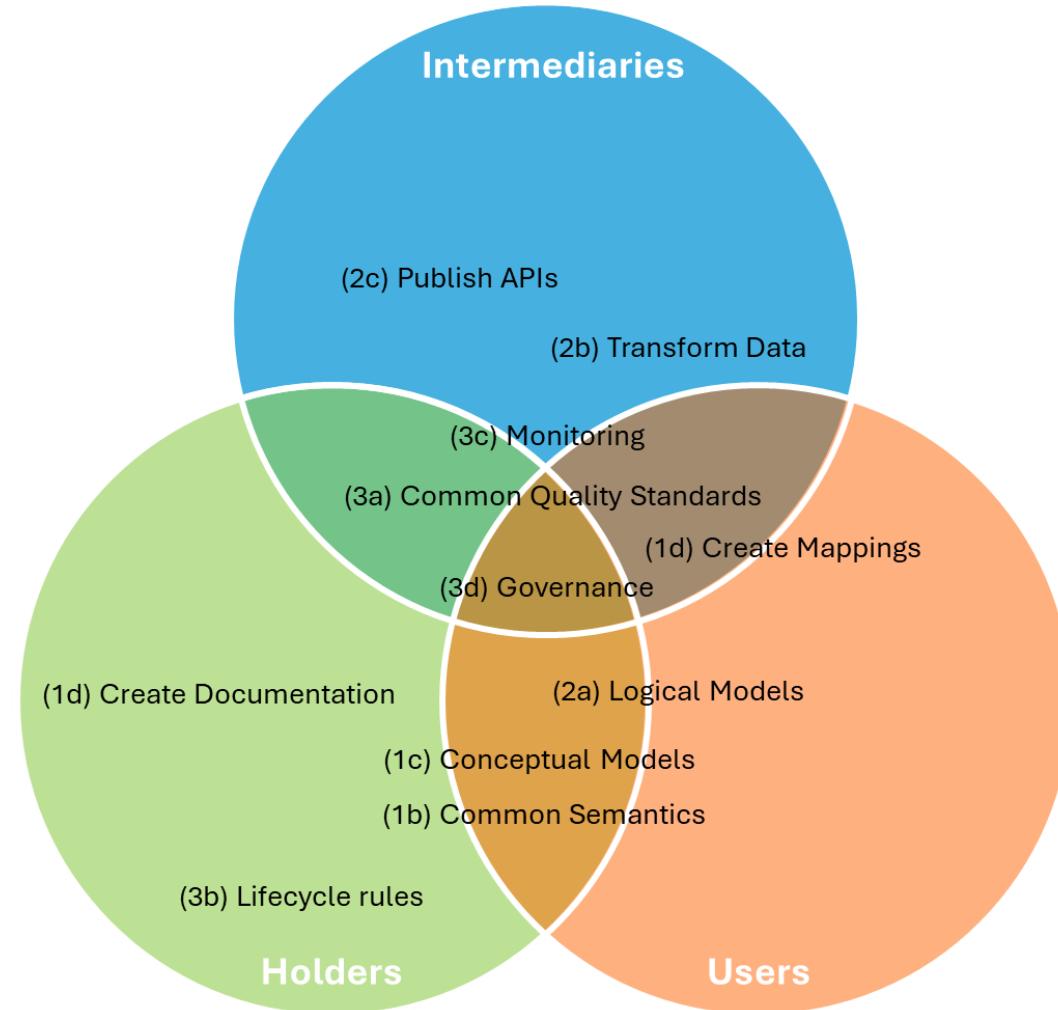
→ May include an evolution of the MIG-T

HOW WILL THESE BE IMPLEMENTED?

Tools, Projects, Activities and more

- **A Model for Iterative Implementation and Compliance**

- Each DS *iteratively* finds the best balance on who is responsible for which aspect of interoperability
- Star models, Profiles, Minimum Interoperability Mechanisms, ...
- Consider scale effects, synergies, subsidiary principle



GREEN DEAL DATASPACE DEPLOYMENT

... building on the GREAT Project

- **Vocabulary Providers**
 - Common, e.g. for Metadata, Contracts, Policies
 - Domain-specific
- **Common Harmonisation Services**
 - Data Onboarding Service
 - Data Transformation Service
- **Continuous Evaluation & Monitoring**
 - Clearing House
- **Next-Generation Tooling**
 - SIMPL Framework
 - Certified Connectors





Any questions?
Reach out to us!

+49 6151 6290 890

info@wetransform.to

www.wetransform.to

Data Spaces made easy

[www.linkedin.com/company/
wetransform-gmbh](https://www.linkedin.com/company/wetransform-gmbh)