



# WeTransform!

## High-Value Datasets: Why, What, Who, How?

Thorsten Reitz, Co-Founder/CEO  
June 6th, 2023

Webinar

# wetransform IN A NUTSHELL

## Who are we and for whom do we work?

- Based in Darmstadt, Germany
- Mission: Building Green Data Ecosystems
- More than 110 customers, over 4,500 users
- hale»studio: OS software for data transformation
- hale»connect: Integrated & automated (INSPIRE) data platform



Simon



Thorsten



Flaminia



Claudia



Livia



Kapil



John



Emanuela



Akshat



Johanna



Kate



Florian



Franziska



Christopher



Anna



Somakanthan



Arbeitsgemeinschaft der Vermessungsverwaltungen der Länder der Bundesrepublik Deutschland (AdV)



Rijkswaterstaat  
Ministerie van Verkeer en Waterstaat

Europäische Umweltagentur



Miljø- og Fødevareministeriet  
Miljøstyrelsen



BEUTH HOCHSCHULE FÜR TECHNIK BERLIN  
University of Applied Sciences



LANDKREIS NIENBURG / WESER  
M.O.S.S.  
Computer Grafik Systeme Geoinformationssysteme



Rheinland-Pfalz  
LANDESAMT FÜR VERMESSUNG UND GEOBASISINFORMATION



ForstBW



irCELine



swisstopo



PRISMA solutions



LGL umweltbundesamt



JRC  
EUROPEAN COMMISSION



Komm.ONE



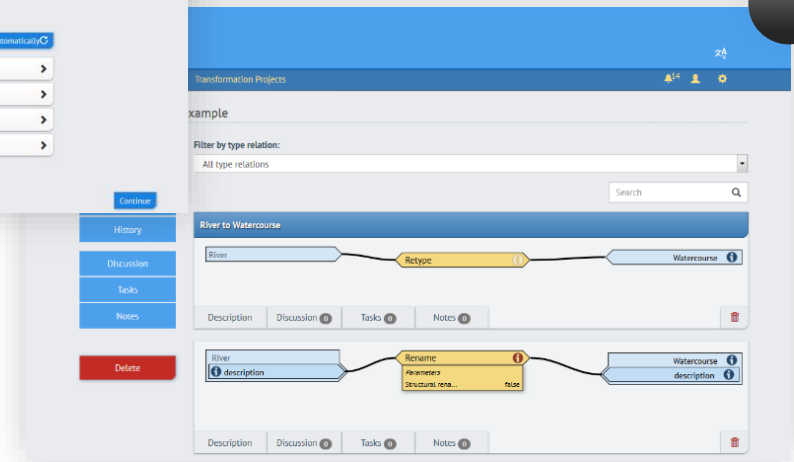
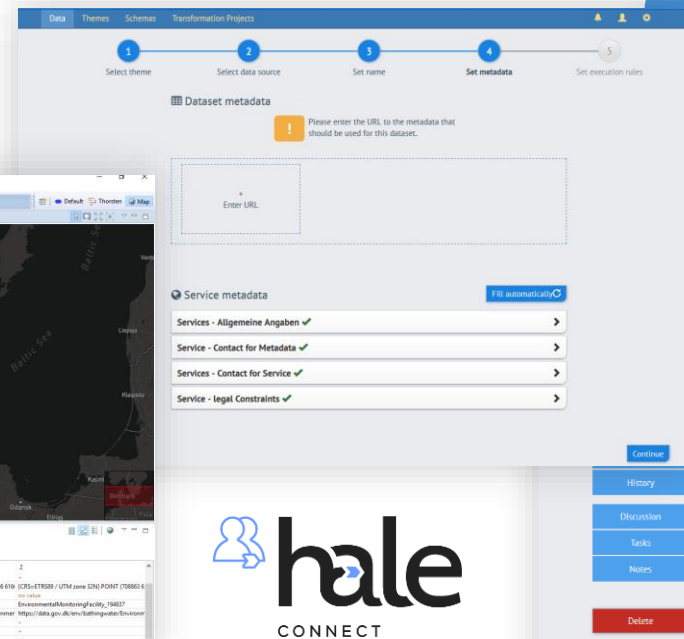
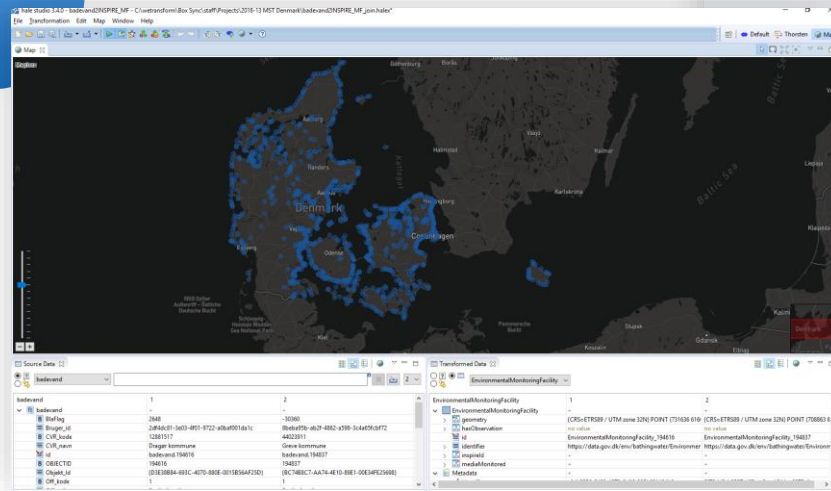
EPSILON  
Italia



MINERVA INTELLIGENCE

# INSPIRE AS A SERVICE – SINCE 2015

## 1.300+ organisations, 75.000+ INSPIRE Network Services



### Key Benefits



Fully Automated Publishing and Simple Harmonisation reduce effort by 80 to 95%



Continuous Maintenance and High Availability Guarantee Compliance

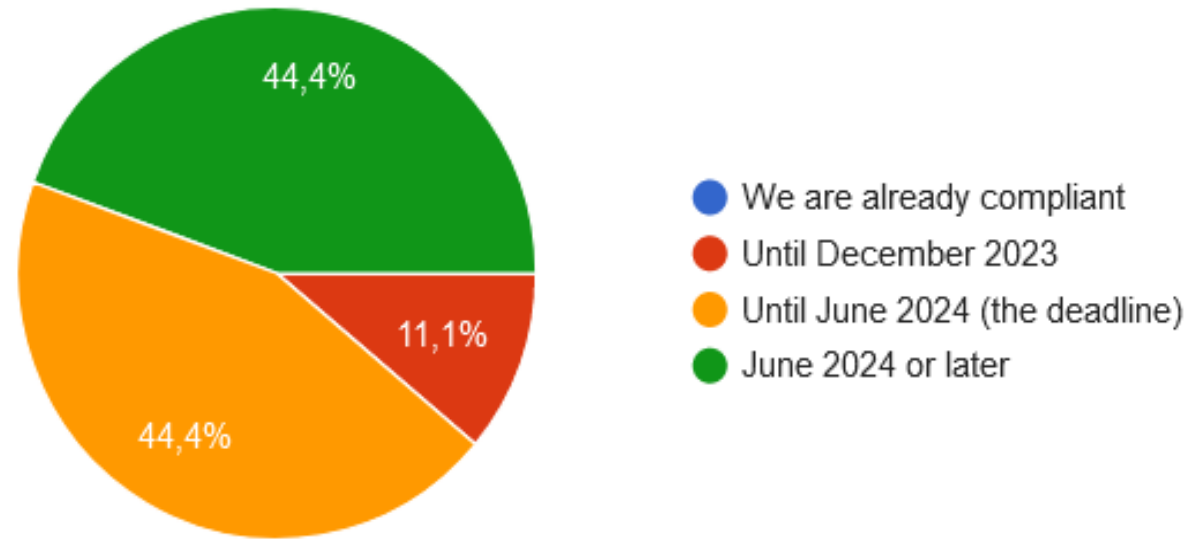


New APIs, Models, Sharing Rules and Formats can be automatically added

# RESULTS FROM THE SURVEY

When do you expect to be fully compliant with the High-value Datasets act?

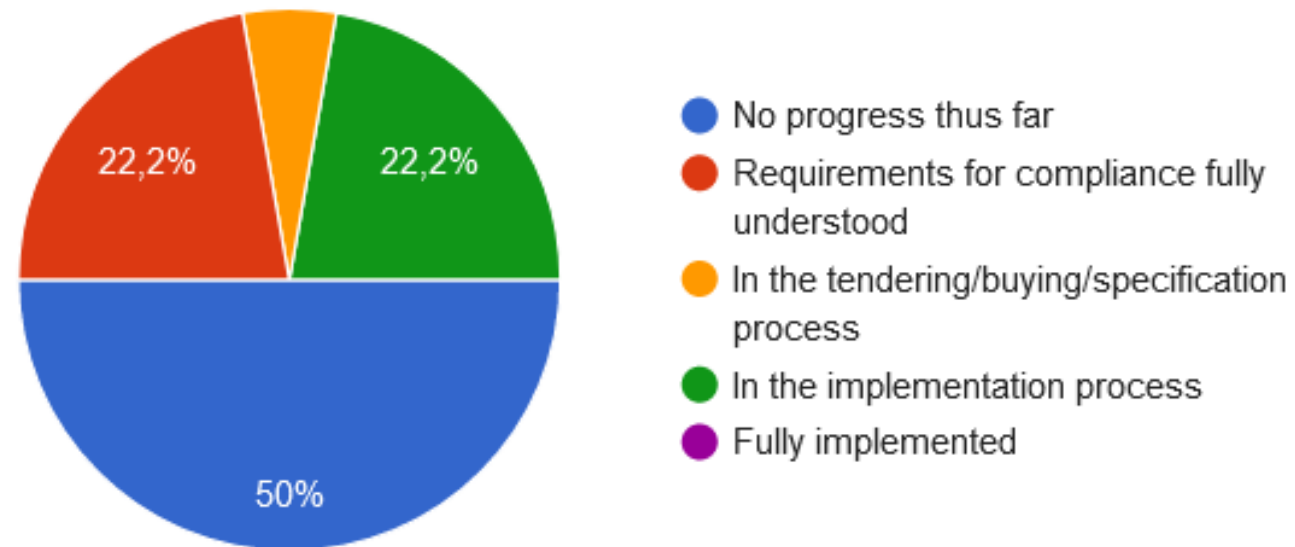
18 Antworten



# RESULTS FROM THE SURVEY

Where do you currently stand in your implementation of High-value Datasets?

18 Antworten



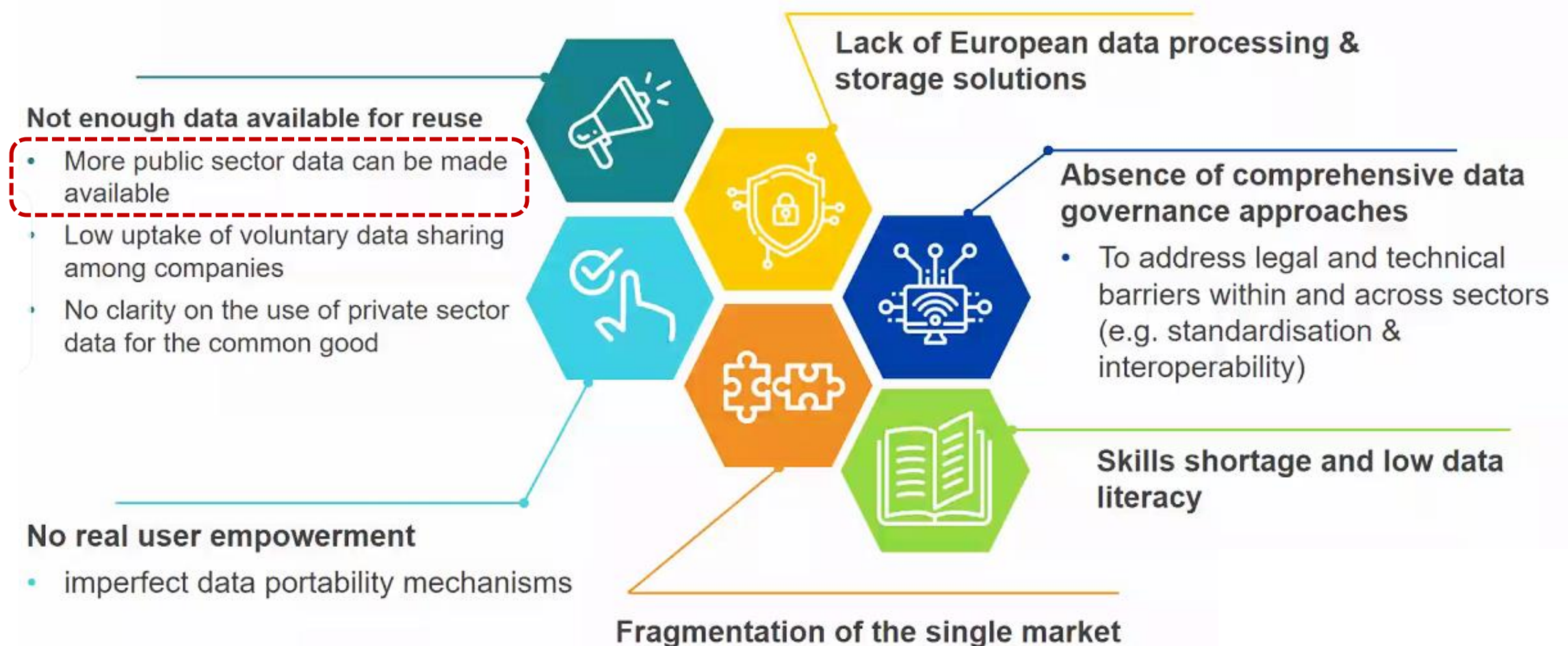
# CHALLENGES FROM THE SURVEY

- Data security, data analysis, and data interpretation
- HVD services for big datasets
- Compliance with and use of APIs
- Identifying the right data for use in HVD
- Data models (data structure, data specifications) for HVD are not fully specified, yet. They are not 100 per cent identical with INSPIRE data models. These data specifications are needed for hassle-free implementation
- Lack of recommendations e.g. related to metadata keyword
- Lack of enough workforce and expertise within the organisation
- Non-GIS-Data / not included in INSPIRE
- To understand exactly what implies
- Applying HVD rules to data provided by [...] but owned by partners, some of them willing to protect data rather than to open it (e.g. landscape features on agricultural parcels)
- the specification process for the datasets, workflow of transformation
- "Machine readable" means new infrastructure (API's,...)
- To understand what is needed to implement it
- Purpose of HVD at all, when INSPIRE datasets are covering same area of sources
- Availability of resources to implement the HVD
- We provide all geodata of [...] (without sensor data till now) incl. INSPIRE datasets. Biggest problem seems to be the update cycle of cadastral parcels and buildings
- cost coverage
- Adapting all the data according to the regulation
- Compliance and related costs

# AGENDA

- **The Purpose and Context of HVD**
  - Which data is in scope?
  - Update cycles, historic data
  
- **Making Data Findable and Accessible**
  
- **Making Data Interoperable and Re-usable**
  
- **Compliance reporting**
  
- **Methods and solutions for implementation**
  
- **Your Next Steps, Feedback & Discussion**

# CONTEXT: CHALLENGES IN DATA RE-USE





# PURPOSE OF HIGH-VALUE DATA SETS IR

## → Make More (Public Sector) Data FAIR

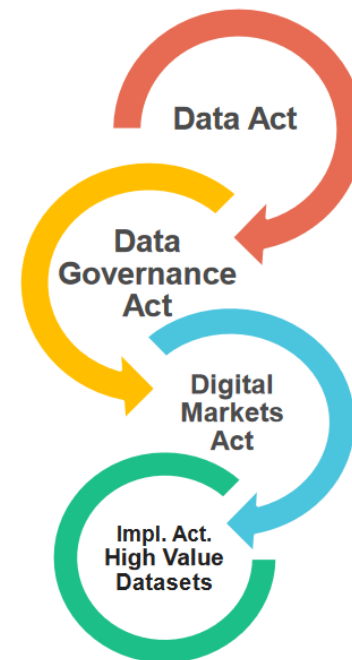
- Improve data value allocation
- Ensure Trust
- Remove Gatekeepers
- Push Innovation
- Cross-border data

## → Push Implementation of Related Policy

This implementing regulation is in support of other Policy such as PSI, Open Data, INSPIRE and the wider **EU Data Strategy**.

## → Invigorate Data Re-Use

Focus activities on data sets with particularly high value to other users and usages and create conditions that allow for maximum re-use



## FAIR Principles



### Findability

Resource and its metadata are easy to find by both, humans and computer systems. Basic machine readable descriptive metadata allows the discovery of interesting data sets and services.



### Accessibility

Resource and metadata are stored for the long term such that they can be easily accessed and downloaded or locally used by humans and ideally also machines using standard communication protocols.



### Interoperability

Metadata should be ready to be exchanged, interpreted and combined in a (semi)automated way with other data sets by humans as well as computer systems.



### Reusability

Data and metadata are sufficiently well-described to allow data to be reused in future research, allowing for integration with other compatible data sources. Proper citation must be facilitated, and the conditions under which the data can be used should be clear to machines and humans.

Source: <https://ccafs.cgiar.org/open-access-and-fair-principles>

# PRIORITY DATA vs. HIGH-VALUE DATA

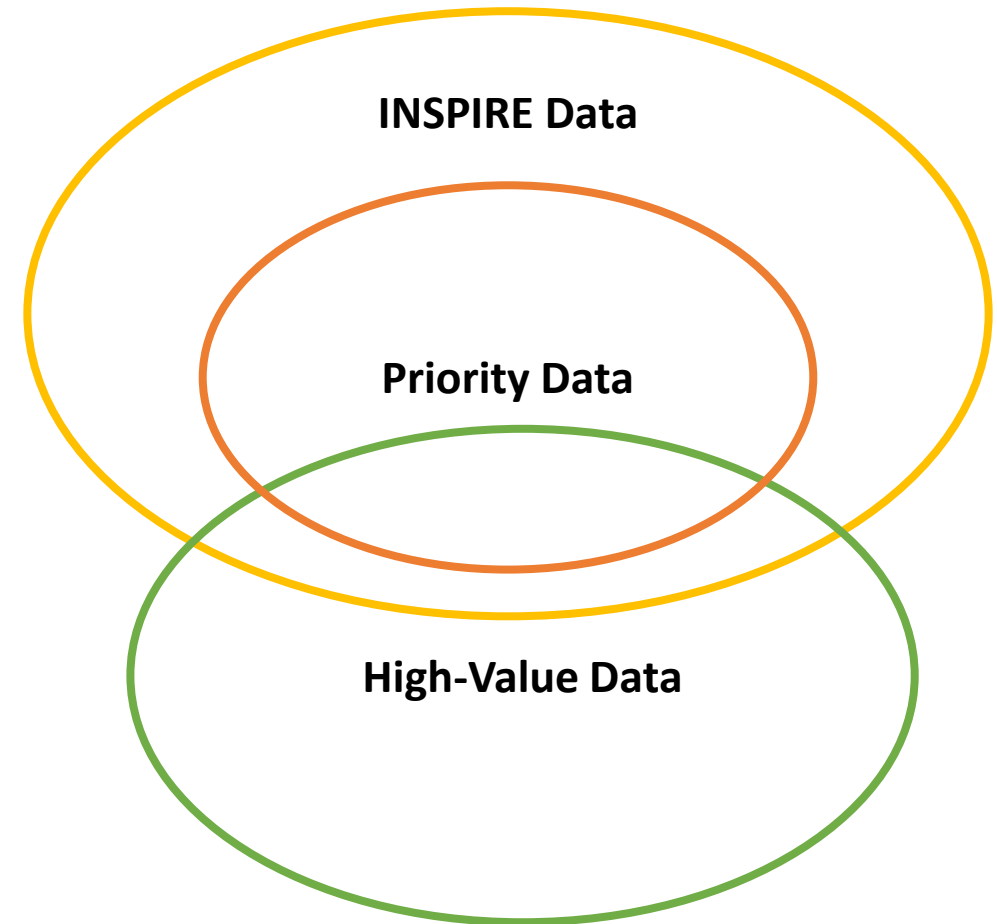
## → Scope

- All priority data is in the INSPIRE Scope
- Some High-Value data is outside of the INSPIRE Scope
- Some are in scope of both, e.g. Noise

## → Conditions for Re-Use:

- High Value data sets must be licensed according to CC-BY or less restrictive
- No comparable rule for Priority Datasets

Label	Governance level	Status
Directive 1999/31/EC	Technical (EU)	Valid
Directive 2000/60/EC	Technical (EU)	Valid
Directive 2002/49/EC	Technical (EU)	Valid
Directive 2006/21/EC	Technical (EU)	Valid
Directive 2006/7/EC	Technical (EU)	Valid
Directive 2007/60/EC	Technical (EU)	Valid
Directive 2008/50/EC	Technical (EU)	Valid
Directive 2008/56/EC	Technical (EU)	Valid
Directive 2009/147/EC	Technical (EU)	Valid
Directive 2010/75/EU	Technical (EU)	Valid



# DATA IN SCOPE

## HVD

*Geospatial*

*Earth observation  
and environment*

*Meteorological*

*Statistics*

*Companies and  
company ownership*

*Mobility*

## INSPIRE

Administrative Units

Buildings

Hydrography

Geology

Area Management

Env. Monit. Facilities

Mineral Resources

Climate

Transport Networks

Geographical Names

Cadastral Parcels

Protected Sites

Land Cover

Bio-Geogr. Regions

Habitats & Biotopes

Natural Risk Zones

Emissions/Noise

Addresses

Ref./Agricult. Parcels

Elevation

Orthoimagery

Energy Resources

Land Use

Air

Waste/Water/...

*Total: 25 INSPIRE themes!*

# UPDATE CYCLES & SCALE

## 1. Up-to-date data sets

1. Required Update Frequency requirements depend on category



1. Weather stations, weather radar: 5-10min
2. Alerts: hourly or faster
3. Weather models: every 6 hours
4. Climate data: daily
5. Other Geospatial and Environmental data that also falls under INSPIRE: At least every 6 months

## 2. Historical Data Sets

1. Regulation also applies to historical data for earth observation and the environment (see Article 4.2)
2. Can optionally be offered through Bulk Download only (See Annex 2.2.a) III)

## 3. Scale

1. All scales down to 1:5.000 or better when available are in scope

# AGENDA

- The Purpose and Context of HVD
- **Making Data Findable and Accessible**
  - API options for geodata and tabular data
  - Bulk Download options for geodata and tabular data
  - Metadata, specific keywords
- Making Data Interoperable and Re-usable
- Compliance reporting
- Methods and solutions for implementation
- Your Next Steps, Feedback & Discussion

## Requirements

1. *“The metadata describing the datasets within the scope of the INSPIRE data themes shall contain at least the metadata elements set out in Commission Regulation.” (Annex 1.2, 2.2)*
2. *“[...] set out and publish the terms of use of the API and the quality-of-service criteria on its performance, capacity and availability. The terms of use shall be available in a human-readable and machine-readable format.” (Article 3 (2))*
3. *“[...]persistent link to the licensing conditions applicable to the re-use of high-value datasets”*
4. *“Public sector bodies holding high-value datasets [...] shall ensure that the datasets are denoted as high-value datasets in their metadata description.” (Article 3 (5))*

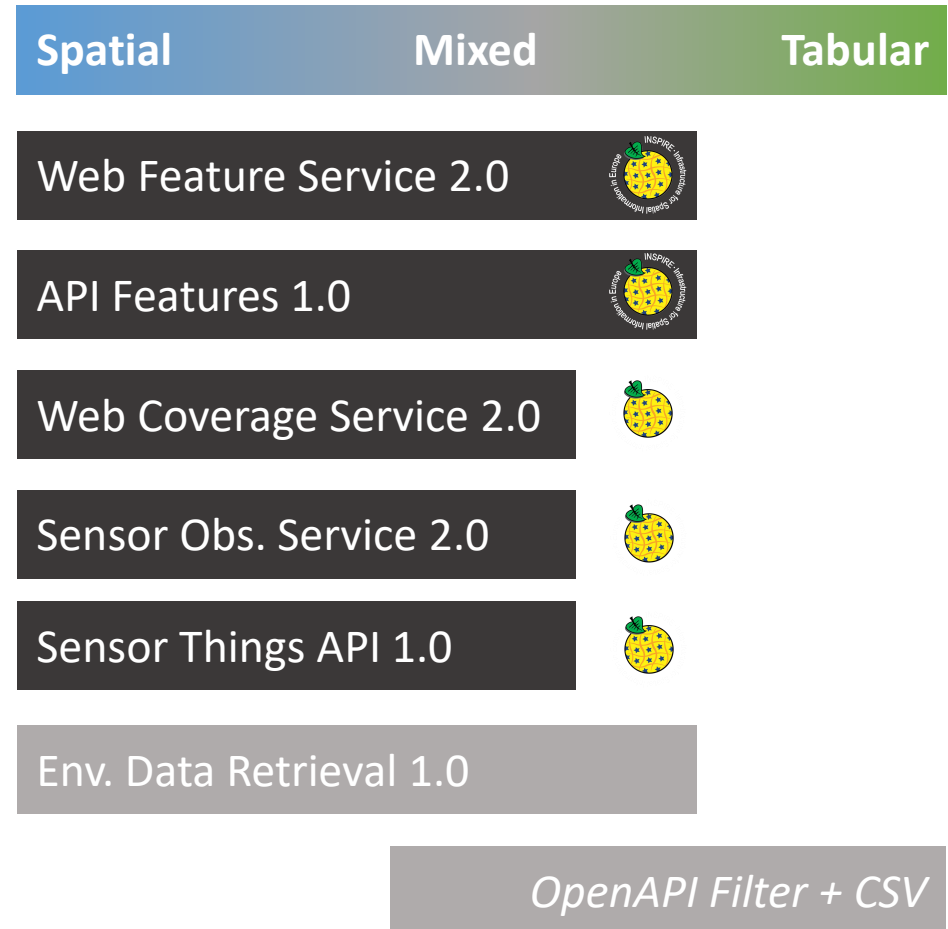
## Implementation

1. Use INSPIRE Metadata
  1. INSPIRE theme
  2. PDS/InVeKoS keywords
2. There are various proposals to add specific keywords, e.g. in GDI-DE
3. DCAT-AP Discussion:  
<https://github.com/SEMICeu/DCAT-AP/issues/231>

# API REQUIREMENTS & OPTIONS

## Requirements

1. “[...] *the datasets described or referenced in the Annex are made available in machine-readable formats via APIs corresponding to the reasonable needs of re-users.*” (Article 3 (1))
2. “[...] *set out and publish the terms of use of the API and the quality-of-service criteria on its performance, capacity and availability. The terms of use shall be available in a human-readable and machine-readable format.*” (Article 3 (2))
3. “*API terms of use shall be accompanied by API documentation in a Union or internationally recognised open, human-readable and machine-readable format*” (Article 3 (3))
4. “[...] *designate a point of contact for questions and issues related to the API [...] to ensure the availability and maintenance of the API [...] the high-value datasets*” (Article 3 (4))



# BULK DOWNLOAD REQUIREMENTS & OPTIONS

## Requirements

1. “[...] *the datasets described or referenced in the Annex are made available [...]. Where indicated in the Annex, the datasets shall also be made available as a bulk download.*” (Article 3 (1))
2. “[...] *set out and publish the terms of use of the API and the quality-of-service criteria on its performance, capacity and availability. The terms of use shall be available in a human-readable and machine-readable format.*” (Article 3 (2))
3. “*API terms of use shall be accompanied by API documentation in a Union or internationally recognised open, human-readable and machine-readable format*” (Article 3 (3))
4. “[...] *designate a point of contact for questions and issues related to the API [...] to ensure the availability and maintenance of the API [...] the high-value datasets*” (Article 3 (4))

Spatial

Mixed

Tabular

The APIs from before 😊

Atom Feed



STAC

Direct Link to Archive



DCAT-based Open Data Catalogues

OpenAPI Filter + CSV\*

\* See <https://www.w3.org/TR/tabular-data-primer/>



# AGENDA

- The Purpose and Context of HVD
- Making Data Findable and Accessible
- **Making Data Interoperable and Re-usable**
  - Data Licensing
  - To harmonise or not to harmonise?
- Compliance reporting
- Methods and solutions for implementation
- Your Next Steps, Feedback & Discussion

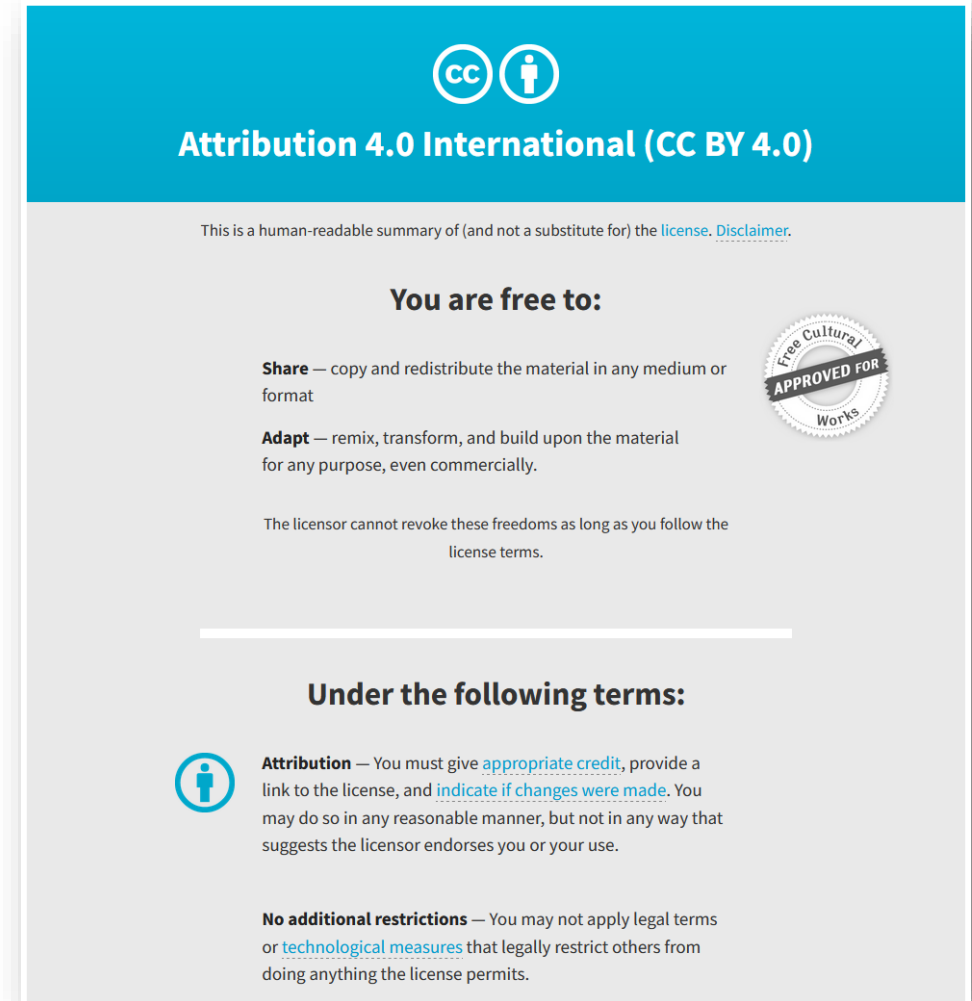
# DATA LICENSING & CAVEATS

## 1. Conditions for the data license

1. *“High-value datasets shall be made available for re-use under the conditions of the Creative Commons Public Domain Dedication (CC0) or, alternatively, the Creative Commons BY 4.0 licence, or any equivalent or less restrictive open licence, as set out in the Annex, allowing for unrestricted re-use.”*

## 2. Exemptions

1. Grace period of up to 24 months, e.g. for data sets with which significant revenue is currently generated
2. Data sets where a GDPR assessment has ruled out publication
3. Exemptions shall be published.



The image shows a summary card for the Creative Commons Attribution 4.0 International (CC BY 4.0) license. At the top, there is a blue header with the CC logo and a person icon, followed by the text "Attribution 4.0 International (CC BY 4.0)". Below this, a disclaimer states: "This is a human-readable summary of (and not a substitute for) the license. [Disclaimer.](#)". The main content is divided into two sections: "You are free to:" and "Under the following terms:". Under "You are free to:", there are two bullet points: "Share" (copy and redistribute) and "Adapt" (remix, transform, and build upon). A circular seal on the right says "Free Cultural Works APPROVED FOR". Below this, it states: "The licensor cannot revoke these freedoms as long as you follow the license terms." Under "Under the following terms:", there is a person icon and a bullet point for "Attribution" (give appropriate credit, provide a link, and indicate changes). Below that, it states: "No additional restrictions" (do not apply legal terms or technological measures).

# DATA HARMONISATION

## → Heterogeneous rules by data category

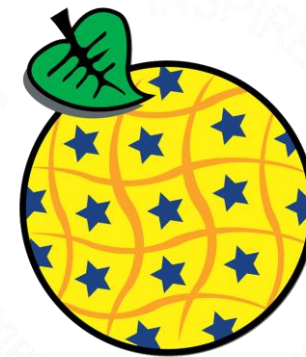
- Meteorological and Statistical data needs to be provided in defined formats (e.g. SDMX)
- Transport Networks, Reference & Agricultural Parcels need to be provided compliant to the INSPIRE IR/TG
- For the other geospatial and environmental data sets, there are *no* formats mentioned, but...



But it still needs to be an Open Format!

## → Semantic harmonisation is a minimum

- “The datasets shall be described in a **complete and publicly available online documentation describing at least the data structure and semantics.**”
- “The datasets shall use **Union or internationally recognised and publicly documented controlled vocabularies and taxonomies, where available.**”



Semantics

GML

# A POSSIBLE “MINIMAL” APPROACH

## → Creating Documentation

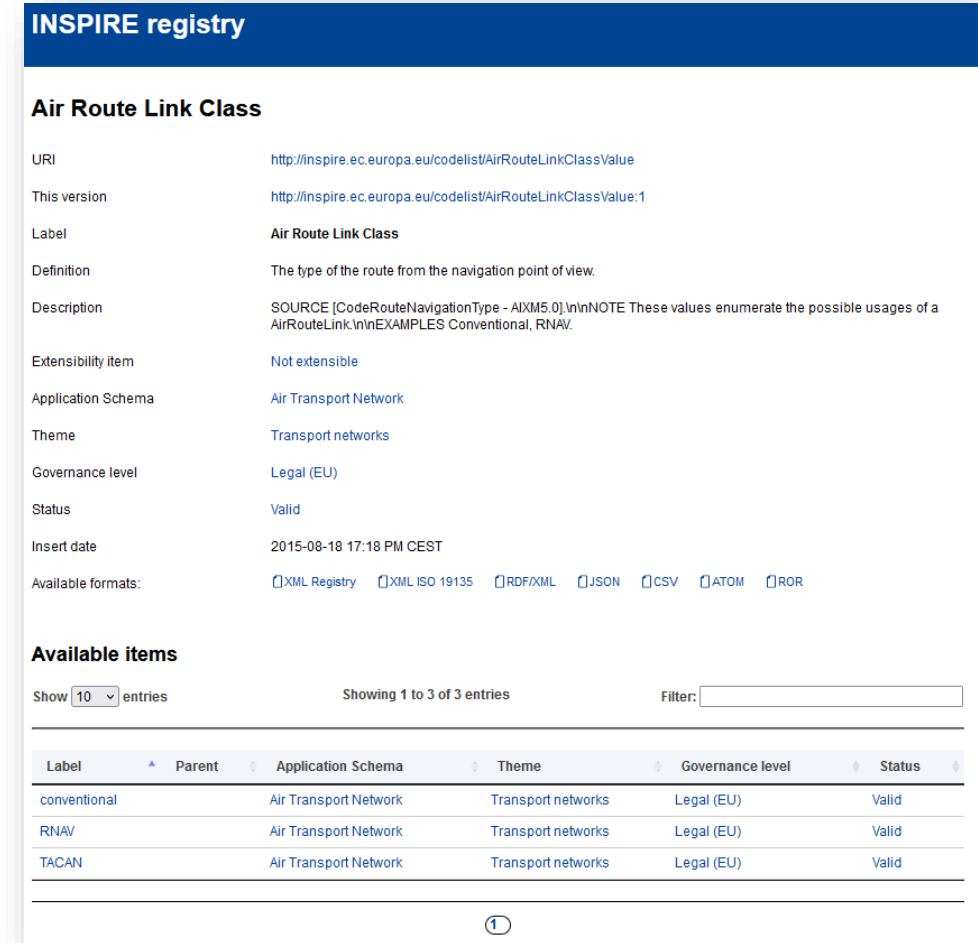
- About 75% of all data sets we have encountered are completely undocumented
- Recommendation: Directly create documentation as code and publish on Github or similar platform
- Typical effort: 2 to 4 hours per Table/Type

## → Semantic Harmonisation / Codelists

- In themes such as Geology, this can be difficult and require a significant part of the overall effort

## → Implementation

- Use a schema close to the source schema, but with added fields for INSPIRE IDs and INSPIRE/EU code lists/vocabularies
- Encode Spatial data as **GeoPackage** or **GeoJSON**
- Encode Tabular data as **CSV RFC 4180** or **Parquet**



The screenshot shows the INSPIRE registry page for the 'Air Route Link Class'. The page is titled 'INSPIRE registry' and 'Air Route Link Class'. It displays various metadata fields for this class, including its URI, version, label, definition, description, extensibility, application schema, theme, governance level, status, and insert date. Below the metadata, there is a section for 'Available items' which shows a table of three entries: 'conventional', 'RNAV', and 'TACAN'. Each entry is associated with the 'Air Transport Network' application schema, 'Transport networks' theme, 'Legal (EU)' governance level, and 'Valid' status.

Label	Parent	Application Schema	Theme	Governance level	Status
conventional		Air Transport Network	Transport networks	Legal (EU)	Valid
RNAV		Air Transport Network	Transport networks	Legal (EU)	Valid
TACAN		Air Transport Network	Transport networks	Legal (EU)	Valid

# AGENDA

- The Purpose and Context of HVD
- Making Data Findable and Accessible
- Making Data Interoperable and Re-usable
- **Compliance reporting**
- Methods and solutions for implementation
- Your Next Steps, Feedback & Discussion

# REPORTING & MONITORING

## → 1<sup>st</sup> Reporting due: December 2024

## → Reporting Content

- a) a list of specific datasets at Member State level (and, where relevant, subnational level) corresponding to the description of *each high-value dataset* in the Annex to this Regulation and **with online reference to metadata** that *follow existing [INSPIRE] standards*, such as a single register or open data catalogue;
- b) persistent **link to the licensing conditions** applicable to the re-use of high-value datasets listed in the Annex to this Regulation, per dataset referred to in point a)
- c) persistent **link to the APIs** ensuring access to the high-value datasets listed in the Annex to this Regulation, per dataset referred to in point a);
- d) where available, guidance documents issued by the Member State on publishing and reusing their high-value datasets;
- e) where available, the existence of data protection impact assessments [...];
- f) the number of public sector bodies exempted in accordance [...].

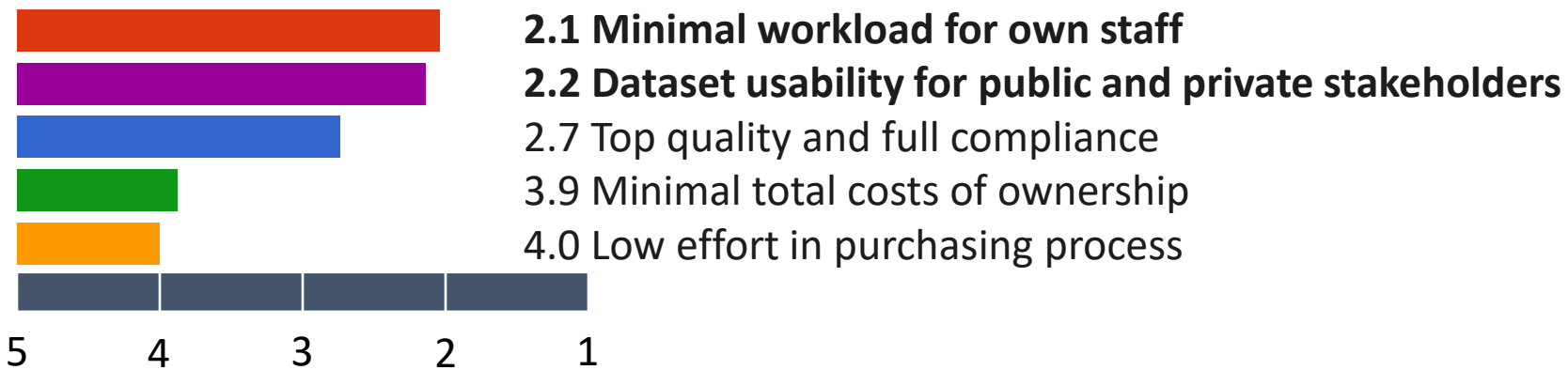
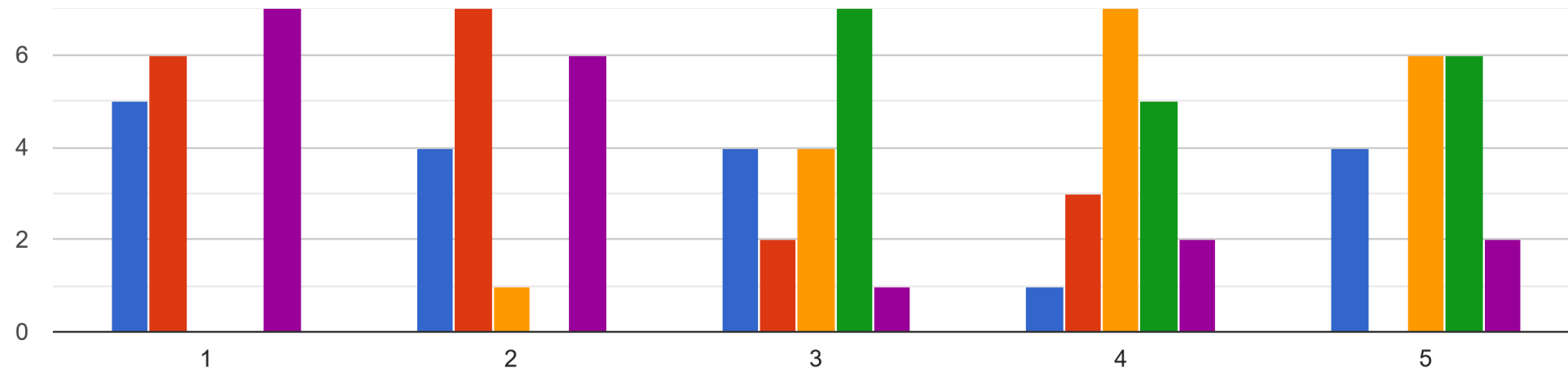
## → Follow-Ups every two years

# AGENDA

- The Purpose and Context of HVD
- Making Data Findable and Accessible
- Making Data Interoperable and Re-usable
- Compliance reporting
- **Methods and solutions for implementation**
- Your Next Steps, Feedback & Discussion

# FINDING AN APPROACH

What is most important to you in your High-value dataset implementation? (Rank)

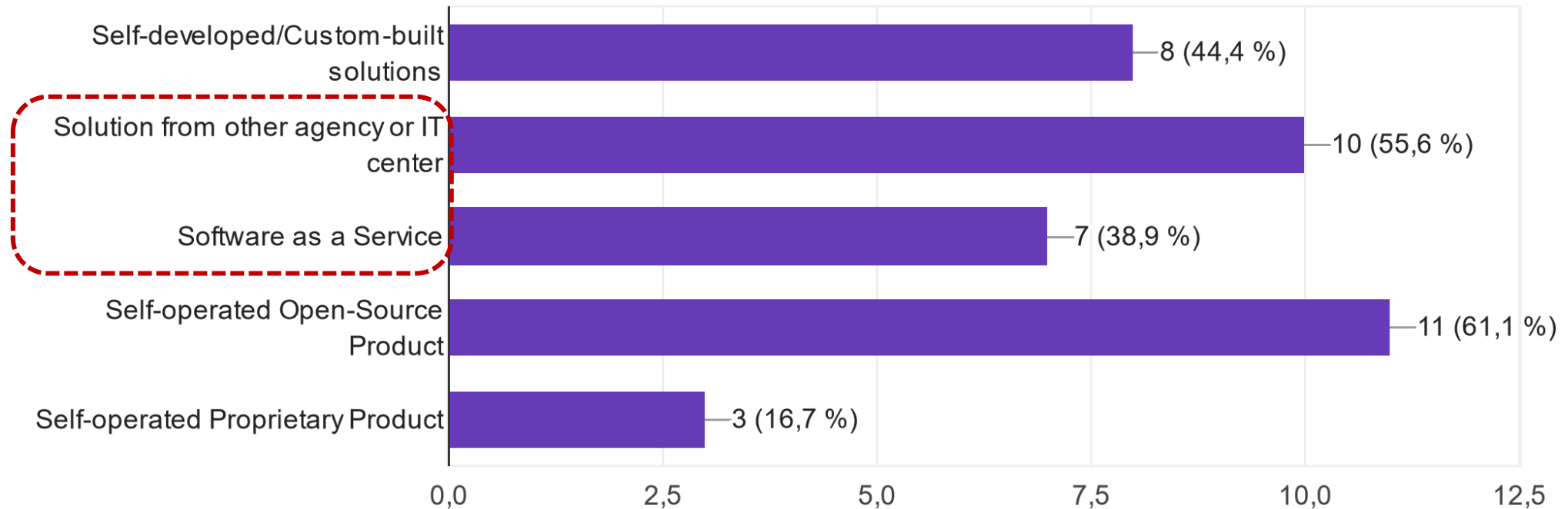




# FINDING AN APPROACH

What type of solutions are you considering? (Check all that apply)

18 Antworten



# WETRANSFORM ♥ HIGH VALUE DATASETS

Literally our Mission: Making Data More Accessible and Useable

1. Simple upload of vector and raster data in many formats
2. Simple templates for semantically pre-harmonised data and automated transformation to INSPIRE GML/GeoPackage
3. Publication of data as API (API Features, WFS, WCS) and as bulk download (Atom Feed, STAC, Direct Link)
4. Automated generation/update of metadata and publication of services
5. Automated processing of dataset updates from their respective sources
6. Hosting and operation for data volumes from 1GB to 100TB
7. Regular updates to support the latest HVD requirements
8. Automated generation of compliance report required by the Implementing Act
9. *Optional: Provision of datasets and services according to INSPIRE Technical Guidance 2.0 with the INSPIRE SLA (cross compliance)*

The screenshot displays the 'Quality' tab for the dataset 'DFS — INSPIRE Air Transport Network'. It features a navigation bar with 'Dataset', 'Quality', and 'Similar datasets' tabs. The 'Quality' section is titled 'Metadata quality' and includes a brief explanation of the Metadata Quality Assurance process. Below this, five categories of indicators are shown with their respective values:

- Accessibility:** Download URL (100%), Most frequent accessU... (200), Most frequent download... (200)
- Reusability:** Access restrictions (true), License information (100%), Access restrictions voc... (false), Contact point (true), Publisher (true)
- Contextuality:** File size (0%), Rights (0%), Modification date (true), Modification date (100%)
- Findability:** Keyword usage (true), Categories (true), Geo search (true), Time based search (false)
- Interoperability:** DCAT-AP compliance (false), Format (100%), Media type (0%), Format / Media type fro... (0%)

# AGENDA

- The Purpose and Context of HVD
- Making Data Findable and Accessible
- Making Data Interoperable and Re-usable
- Compliance reporting
- Methods and solutions for implementation
- **Your Next Steps, Feedback & Discussion**

# YOUR NEXT STEPS

→ Q&A

→ Questionnaire

→ Discussion

→ Go for it 😊!



**Any questions?  
Reach out to us!**

+49 6151 6290 890

[info@wetransform.to](mailto:info@wetransform.to)

[www.wetransform.to](http://www.wetransform.to)

**The Path to High-Value Data!**

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