A European Strategy for Data High-value datasets
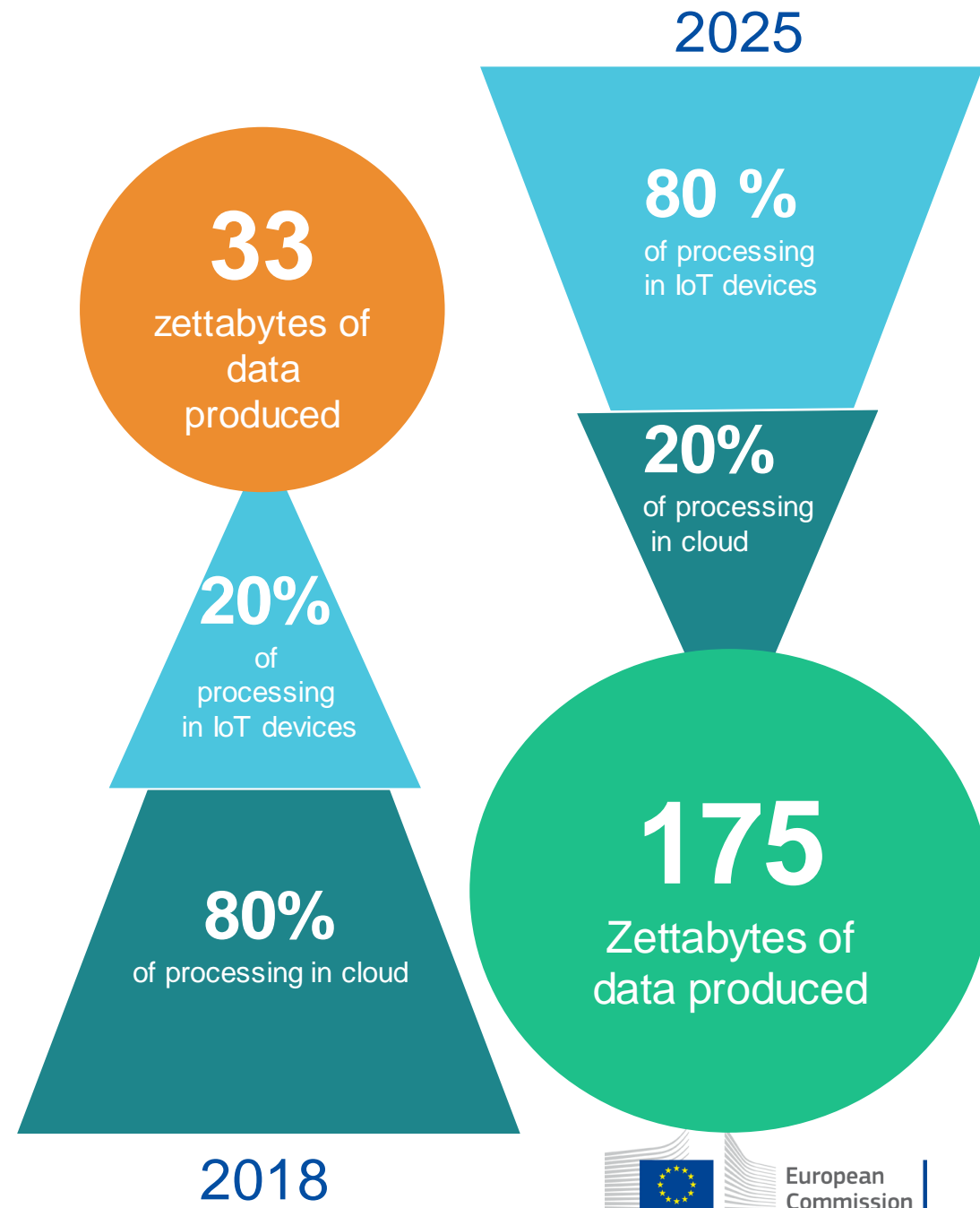
Meeting of the INSPIRE MIG Expert Group

27 November 2020
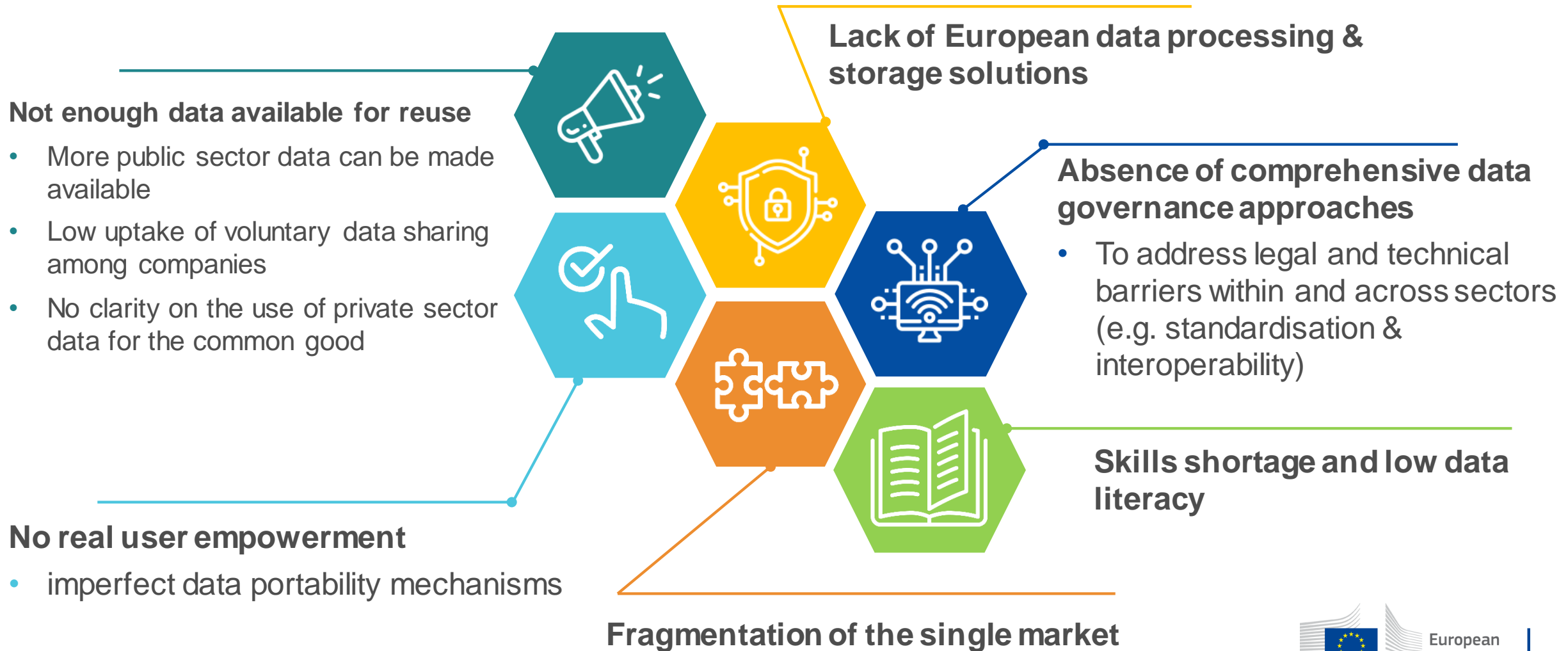
Jiri PILAR
Unit G1 - Data Policy and Innovation
DG CNECT, European Commission
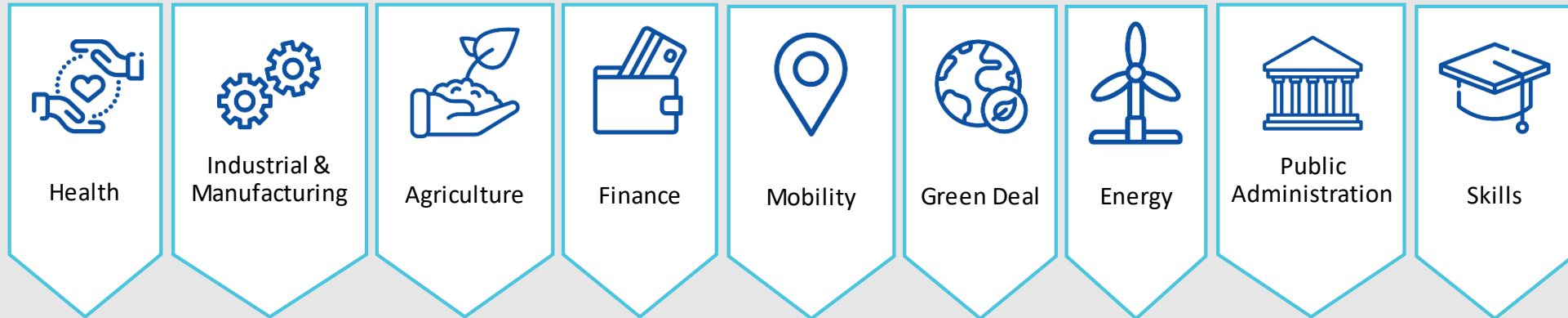
Europe has everything to play for

- Data can transform all sectors of the economy and is crucial for AI

- Personal and non-personal data can be a source of innovation for new products and services

- Data can contribute to tackle societal challenges such as climate change, health, mobility, etc.

- Data can make our lives and work easier and better

2025

80 %
of processing in IoT devices

20%
of processing in cloud

175
Zettabytes of data produced

33
zettabytes of data produced

20%
of processing in IoT devices

80%
of processing in cloud

2018

European Commission

European Strategy for Data: What are the problems?

Lack of European data processing & storage solutions

Not enough data available for reuse

- More public sector data can be made available
- Low uptake of voluntary data sharing among companies
- No clarity on the use of private sector data for the common good

Absence of comprehensive data governance approaches

- To address legal and technical barriers within and across sectors (e.g. standardisation & interoperability)

Skills shortage and low data literacy

No real user empowerment

- imperfect data portability mechanisms

Fragmentation of the single market

European Commission

Closing in on common European data spaces

Health | Industrial & Manufacturing | Agriculture | Finance | Mobility | Green Deal | Energy | Public Administration | Skills

- Driven by stakeholders
- Rich pool of data of varying degree of openness
- Technical tools for data pooling and sharing

- Sectoral data governance (contracts, licenses, access rights, usage rights)
- IT capacity, including cloud storage, processing and services

Personal data spaces

Public sector data (high value data, 'sensitive' data)

Horizontal framework for data spaces:

- Trust in novel data intermediaries that respect 'data sovereignty'
- Data sharing schema
- Governance of standards for cross-sector interoperability

Commission

Overview of data actions

[D] What data are we talking about?

[H] Who holds such data?

[A] What policy intervention?

Q4 20 — Data governance Act

Good governance of data cannot wait

[D] Data voluntarily made available by data holders

[H] Public sector, business, individuals, researchers

[A] Make such data easier to share in a controlled manner (technical, legal and with organisational support); Build trust in data sharing; Ensure data interoperability across sectors
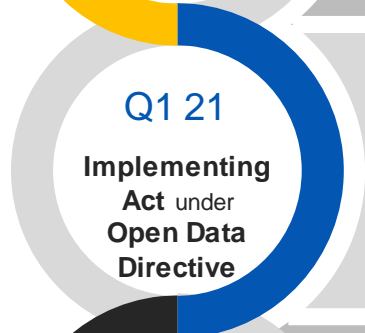
Q4 20 — Digital Market Act

Data: a key element of Big Tech's market power

[D] Data held by online platforms originating from the users (both businesses and individuals)

[H] Online platforms

[A] Among other policy options, identify appropriate data access and data portability remedies
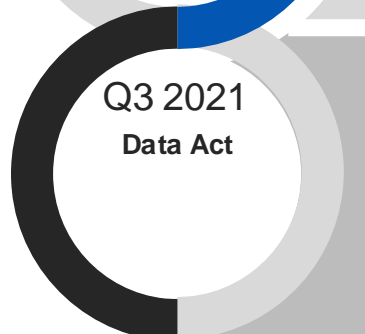
Q1 21 — Implementing Act under Open Data Directive

High quality government data for SMEs & innovation

[D] 'High value' Open Government Data (core reference data)

[H] Public sector

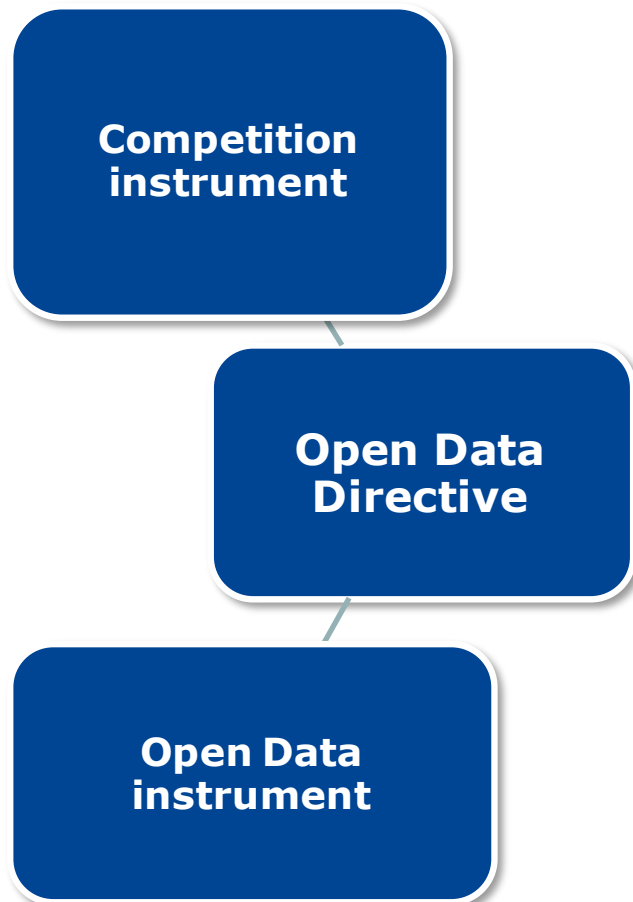**[A]** Make such data available for re-use free of charge

Q3 2021 — Data Act

Better access to and control over data for a fair data economy

[D] Co-generated, IoT data from industry and individuals, Big Data sources held by business

[H] Business

[A] Ensure flexible use of Big Data sources by government for the common good; Establish fairness in use of co-generated, IoT data; Make sure that Europeans stay in control over their data vis-à-vis third country jurisdictions; Examine IPR legislation for possible obstacles

European Commission

Open Data Directive – basic facts

Competition instrument

Open Data Directive

Open Data instrument

Introduces a <u>minimal set of rules</u> to make more data from the public sector easier and cheaper to use for innovation

The Directive does not apply to:
- data not publicly accessible under access to information regimes
- information protected by third party intellectual property rights

Application of the Directive must fully respect the Union and the national data protection rules

Latest revision in 2019: now entitled '**<u>Directive on Open Data and the re-use of public sector information</u>**'

Progress so far and next steps

➢ The Directive was published in the Official Journal of the EU on 26 June 2019

➢ 2 years of transposition in Member States

➢ Work towards the definition of the List of HVDs (2019-21)

　➢ Public Sector Information Group (expert group twice a year)

　➢ Open Data Committee (comitology: for the Implementing Regulation)

　➢ Public consultation – closed on 31 May 2020

　➢ Impact assessment – supported by the ongoing study till September 2020

　➢ Dedicated actions (workshops, focus groups)

　➢ Proposal of the list of HVDs to be submitted to the Open Data Committee in Q1 2021

European Commission

High value datasets – key points

Datasets listed in the implementing act to be made available for free, in machine-readable formats, via APIs and (where relevant) as bulk downloads

Exceptions:
- Free availability requirement shall not apply to public undertakings if there is a risk of competition distortion
- In case of a substantial impact on the budget of the public bodies involved, free availability can be delayed by up to 2 years

Thematic categories of high value datasets

Geospatial

Earth observation and environment

Meteorological

Statistics

Companies and company ownership

Mobility

An Implementing Regulation planned for 2021 will define the list of specific high-value data sets within the 6 thematic categories set out in Annex I among the documents to which the Directive applies

Examples in recital 66:

"the thematic categories listed in the Annex could inter alia cover postcodes, national and local maps (Geospatial), energy consumption and satellite images (Earth observation and environment), in situ data from instruments and weather forecasts (Meteorological), demographic and economic indicators (Statistics), business registers and registration identifiers (Companies and company ownership), road signs and inland waterways (Mobility)."

Categories can be extended (delegated act).

Results of the Impact Assessment

- Two policy options of varying intensity were assessed per each data category

- The intensity refers to the scope of data and the 're-use friendliness' of publishing arrangements (PO1 – 'low hanging fruit'/PO2 – 're-users wishlist)

- As a result of an MCA, an optimised (mixed) intervention option is proposed:

	Preferred PO per theme (Equal Weight 0.20)
Company & company ownership	PO1
Geospatial data	PO1
Meteorological data	PO2
Earth observation & Environment	PO2
Statistics	PO2
Mobility	PO1

European Commission

HVDs analysed: Geospatial data

Policy option 1 (low hanging fruit)	Policy option 2 (re-users' wish list)
Scope Administrative units, Place Names, Addresses, Buildings, Geometry, identification code and location of cadastral parcels **Publication modalities** • CC-BY 4.0 licence • Format: GeoJSON, as per INSPIRE requirements • Metadata to match the INSPIRE requirements allowing transformation to geoDCAT-AP • Partial national or national coverage in terms of granularity • Key attributes as per INSPIRE	**Scope** Administrative units, Place Names, Addresses, Buildings and Full cadastral Parcels **Publication modalities** • CC0 licence • Traceability: both national open data catalog and geodata catalog • Higher granularity of data (scale 1:5000 and beyond) and high update frequency • More stringent requirements regarding key attributes (e.g. height of buildings)

HVDs analysed: Earth observation and environment

- Higher intensity intervention: includes Environmental e-reporting priority data **(in bold: the lower intensity)**, and adds the listed INSPIRE themes in full. Adds open data requirements to INSPIRE (download) services and to e-reporting priority data. Allows the combinations found across varied use cases.

- Removes restrictive terms of use and fees. INSPIRE's data harmonisation efforts extend to open data. Follows current/future INSPIRE standards. API and bulk download.

Earth observation	Environmental data	
Hydrography (I)	**Air quality (prio)**	Protected sites (I)
Land parcels (I), cover (II), use (III)	**Biodiversity (prio)**	Bio-geographical regions (III)
Elevation (II)	**Emissions (prio)**	**Environmental monitoring facilities (III)**
Geology (II)	**Nature preservation (prio)**	Habitats/Biotopes (III)
Ortho-imagery (II)	**Noise (prio)**	Natural Hazards (III)
Oceanography (III)	**Waste (prio)**	Soil (III)
Sea Regions (III)	**Water (prio)**	Species distribution (III)

HVDs analysed: Mobility data

Policy option 1 (low hanging fruit)	Policy option 2 (re-users' wish list)
Scope Inland waterway and river infrastructure data based on the RIS Directive (static data, dynamic data and navigational charts - Inland ECDIS). Applies only to waterways within the scope of the RIS Directive. Also: datasets from the transport network theme under INSPIRE (road transport, rail transport, water transport, air transport and cableways data aligned with INSPIRE specifications). **Publication modalities** • CC-BY 4.0 licence • CSV, XML or geoJSON formats • As a rule: Formats prescribed by the RIS Directive and its implementing acts; publicly documented widely used standard format. • Bulk download and API (web service API for static data) • Clear complete documentation and use of shared vocabularies • Granularity: individual waterway, National waterway network and cross-border nodes (for charts) within the scope of RIS Directive • ISRS Location Code as the key attribute • Provision in (near) real-time in order to guarantee the accuracy of RIS services • For INSPIRE data: INSPIRE requirements regarding data and metadata quality and data provision should be applied	**Scope** The same as the lower intensity scenario plus all national transport network datasets i.e. not only those covered by the INSPIRE Directive. **Publication modalities** As in the low intensity scenario but for the datasets not covered by EU legislation (RIS or INSPIRE Directives), the INSPIRE data quality and publication standard would only be a recommendation (non-binding).

Non-legislative activities to promote Open Data

- **Open Data digital infrastructure: European Data Portal and EU Open Data Portal + CEF (Connecting Europe Facility) building blocks**

- **Green Deal 2019 + EU Recovery Plan 2020: Green and digital go hand in hand**

- **Digital Europe Programme (DEP)**: Specific Objective 2 'Data for Artificial Intelligence (AI)' will strengthen core AI capacities in Europe, including data resources. Calls will focus on, inter alia, making specific datasets interoperable and fit for AI applications. Activities could cover, for example:
  - **curation;**
  - **semantic annotation;**
  - **harmonisation of metadata;**
  - **facilitating publication in machine-readable formats and accessibility through APIs.**

Thank you very much for your attention

For further questions:

email: CNECT-G1@ec.europa.eu
Unit G1 of DG CONNECT

Websites with more information:
https://ec.europa.eu/digital-single-market/en/open-data
https://eur-lex.europa.eu/legal-content/EN/LSU/?uri=celex:32019L1024
https://ec.europa.eu/info/strategy/priorities-2019-2024/europe-fit-digital-age/european-data-strategy_en