

High Value Data

Eoin MacCuirc

emaccuirc@gmail.com

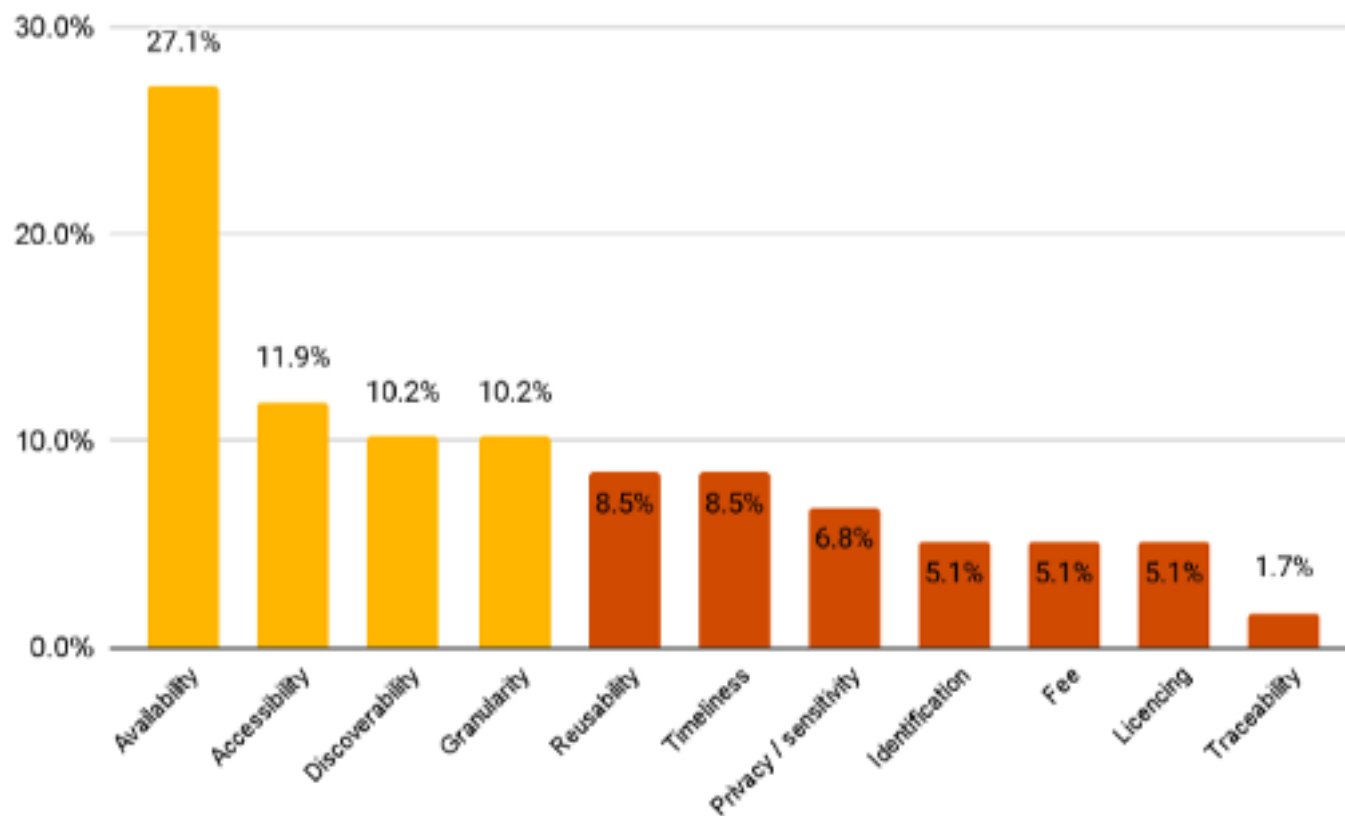
Overview

- Open Data
- The Path to the Digital Decade
- High Value Open Data
- The current list of High Value Datasets
- Data themes for the extension of high-value datasets
- Conclusions

Open Data

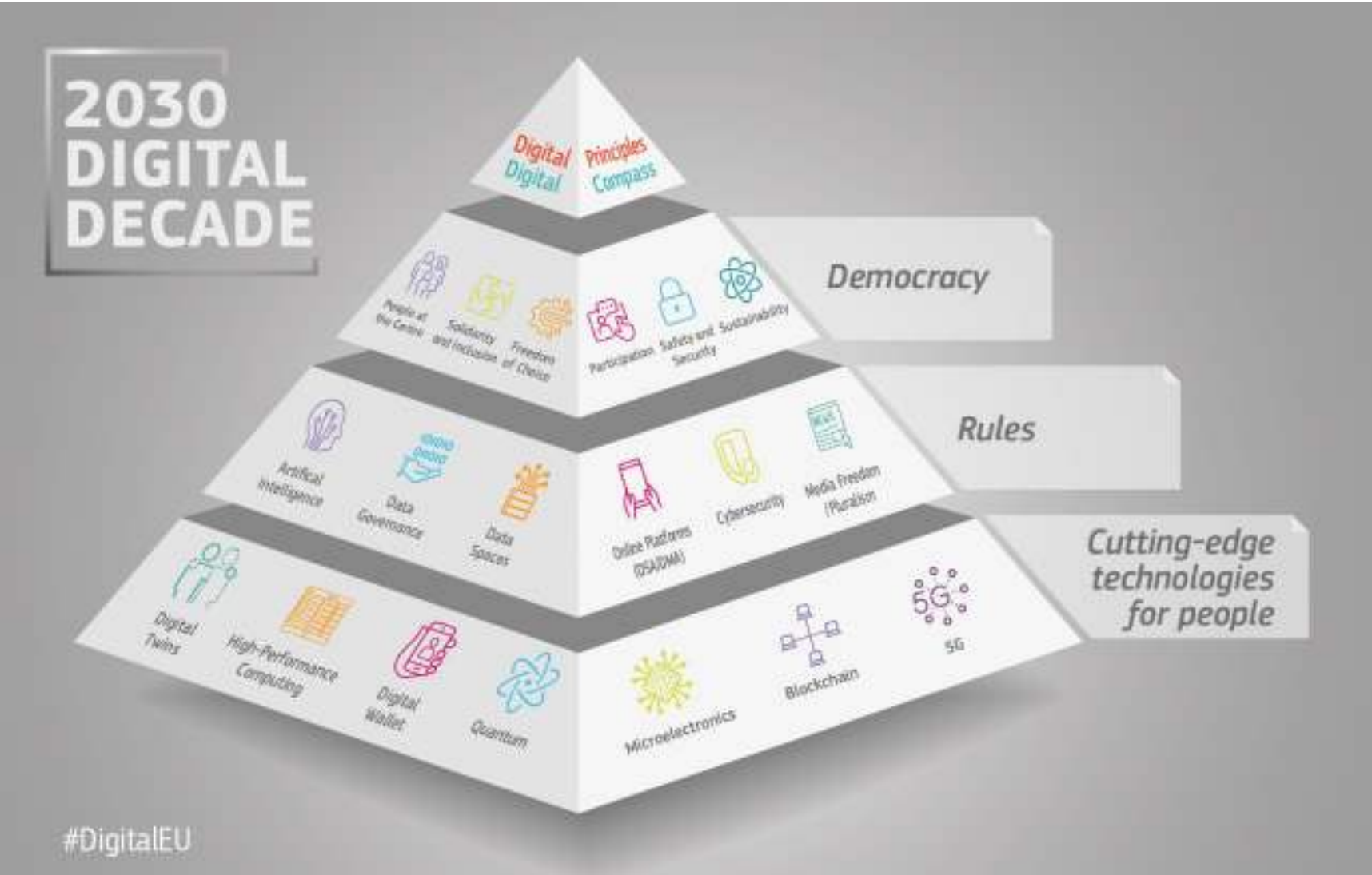
Open data refers **to data that are made available by public authorities or other entities for anyone to access, use and share, without restrictions or fees**. Open data can have various benefits for society, such as **enhancing transparency, accountability, innovation, economic growth and social inclusion**. However, open data also poses some **risks and challenges, such as ensuring data quality, interoperability, security, privacy and ethical use**.

Blocking factors for the reuse of publicly funded data according to respondents



Source: PwC EU services

The Path to the Digital Decade



The Path to the Digital Decade- Compass

Skills

20 million employed **ICT specialists**, more graduates + gender balance
80% of adults can **use tech** for everyday tasks

Government

Key Public Services - 100% online
Everyone can **access health records online**
Everyone can use **eID**



Infrastructure

Gigabit connectivity for everyone, **high-speed mobile coverage** (at least 5G) everywhere
EU produces 20% of world's **semiconductors**
10 000 **cloud edge nodes** = fast data access
EU **quantum computing** by 2025

Business

75% of companies using **Cloud, AI or Big Data**
Double the number of **unicorn startups**
90% of **SMEs taking up tech**

High Value Open Data – EU Open Data Directive

One of the key policy frameworks for open data in the EU is the Open Data Directive (EU) 2019/1024, which entered into force in July 2019 and requires member states to **make their public sector information available as open data by default, subject to certain exceptions**. The directive also **introduces the concept of high value datasets (HVDs), which are datasets that have a high potential for re-use across the EU and can generate significant social, economic and environmental benefits**. The directive mandates the European Commission to adopt implementing acts to identify the specific HVDs and the arrangements for their publication and re-use.

High Value Open Data - Timeline

In December 2022, the Commission adopted the Implementing Regulation (EU) 2023/138, which **lays down a list of specific HVDs in six thematic areas: geospatial, earth observation and environment, meteorological, statistics, companies and company ownership, and mobility**. The regulation also **sets out the technical specifications and modalities for the publication and re-use of HVDs, such as formats, standards, licences, frequency of updates and access conditions**. The regulation requires member states to **publish HVDs within 16 months from its entry into force (i.e. by April 2024)** and to ensure their full compliance with the requirements by December 2025.

High Value Open Data - Timeline

PRESS RELEASE | Publication 20 January 2023

Commission defines high-value datasets to be made available for re-use

Today, the Commission has published a list of high-value datasets that public sector bodies will have to make available for re-use, free of charge, within 16 months.

Certain public sector data, such as meteorological or air quality data are particularly interesting for creators of value-added services and applications and have **important benefits** for society, the environment and the economy – which is why they should be made available to the public.

Margrethe **Vestager**, Executive Vice-President for a Europe Fit for the Digital Age, said:

Making high-value datasets available will benefit both the economy and society, for example by helping to combat climate change, reducing urban air pollution and improving transport infrastructure. This is a practical step towards achieving a successful Digital Decade and building a more prosperous digital future.

Thierry **Breton**, Commissioner for Internal Market, said:

Data is a cornerstone of our EU industrial competitiveness. With the new list of high-value datasets, published today, we are unlocking a large amount of public data for the benefit of all. Start-ups and SMEs will be able to use this data to develop new products and innovative solutions that improve the lives of citizens in the EU and around the world.



Stock photo Getty Images plus

Related topics

[Big data](#)

[Data](#)

[Data value chain](#)

[Open data](#)

High Value Open Data – Timeline -Now

Data themes	Short description (as per Annex to the EC Implementing Regulation)	High-value datasets
Company and company ownership	Datasets containing basic company information and company documents and accounts at individual company level - including key attributes provided in the Annex.	Basic company information (key attributes listed in Annex) Company documents and accounts: - Financial statements - Non-financial statements, management reports and other statements or reports - Annual financial reports
Geospatial	The geospatial data theme includes datasets within the scope of the INSPIRE data themes, as well as Reference parcels and Agricultural parcels defined in Regulation 1306/2013	Administrative units Geographical names Addresses Buildings Cadastral parcels Reference parcels Agricultural parcels
Meteorological data	Datasets on observations data measured by weather stations, validated observations (climate data), weather alerts, radar data and numerical weather prediction model data	Observations data measured by weather stations Climate data: validated observations Weather alerts Radar data Numerical Weather Prediction (NWP) model data

High Value Open Data – Timeline -Now

Data themes	Short description (as per Annex to the EC Implementing Regulation)	High-value datasets
Earth observation and environment	Space-based or remotely-sensed data, as well as ground-based or in situ data, environmental and climate datasets within the scope of the INSPIRE data themes. In addition, it includes datasets included in relevant legal acts related to several environmental domains	INSPIRE data themes: Hydrography Protected sites Elevation Geology Land cover Orthoimagery Area management / restriction / regulation zones & reporting units Bio-geographical regions Energy Resources Environmental monitoring Facilities Habitats and biotopes Land Use Mineral Resources Natural risk zones Oceanographic geographical features Production and industrial facilities Sea regions Soil Species distribution Air Climate Emissions Nature preservation and biodiversity Noise Waste Water Environmental reporting related to environmental damage and liabilities

High Value Open Data – Timeline -Now

Data themes	Short description (as per Annex to the EC Implementing Regulation)	High-value datasets
Statistics	Statistical datasets related to reporting obligations as defined in the relevant legal acts listed in the Annex and pertaining to multiple domains (e.g. Industrial production, Tourism, national accounts, government expenditure - overall and specific sector expenditure, e.g. Health) - population (fertility, mortality), social conditions (poverty, inequalities) and employment.	Industry and sales: <ul style="list-style-type: none"> - Industrial production - Industrial producer price index breakdowns by activity - Volume of sales by activity - EU International trade in goods statistics – exports and imports breakdowns simultaneously by partner, product and flow Tourism flows in Europe Harmonised Indices of consumer prices National accounts: <ul style="list-style-type: none"> - GDP main aggregates - key indicators on corporations - key indicators on households Government <ul style="list-style-type: none"> - Government expenditure and revenue - Consolidated government gross debt Environmental accounts and statistics Social: <ul style="list-style-type: none"> - Population - Fertility - Mortality - Current healthcare expenditure - Poverty - Inequality - Employment - Unemployment - Potential labour force

High Value Open Data – Timeline -Now

Data themes	Short description (as per Annex to the EC Implementing Regulation)	High-value datasets
Mobility	Datasets within the scope of the INSPIRE data theme 'Transport networks' - and related key attributes provided in Annex I - as well as additional Inland waterways datasets (for specific MS to which Directive 2005/44/EC applies)	INSPIRE data themes: Road transport network Railway transport network Air transport network Water transport network Cable transport network Common transport elements Inland waterways datasets Static data on waterways and navigation (8 datasets) Dynamic data on waterways and navigation (7 datasets) Inland electronic and navigational charts (10 datasets)

High Value Open Data & Ireland

Ireland is one of the countries that has been actively involved in both the EU and the OECD frameworks for open data. **Ireland has transposed the Open Data Directive into national law through the European Union (Publication And Re-Use Of Public Sector Information) Regulations 2021 (S.I. No. 254 of 2021), which came into effect on 17 June 2021.** Ireland has also established an **Open Data Governance Board (ODGB) to oversee the implementation of its national Open Data Strategy.** The ODGB is composed of representatives from various public sector bodies, academia, civil society and industry.

High Value Open Data & Ireland – The Future

Ireland has also participated in the identification of HVDs at the EU level through its involvement in a study commissioned by the European Commission in 2020. The study aimed to identify data themes for the extensions of public sector HVDs based on a **multi-criteria analysis that considered aspects such as demand, supply, impact potential, feasibility and alignment with EU priorities**. The study resulted in a **final report published in 2023** that proposed **7 additional data themes for HVDs: health, energy, justice & legal, government and public administration, financial, language, climate loss**.

.

High Value Open Data & Ireland – The Future

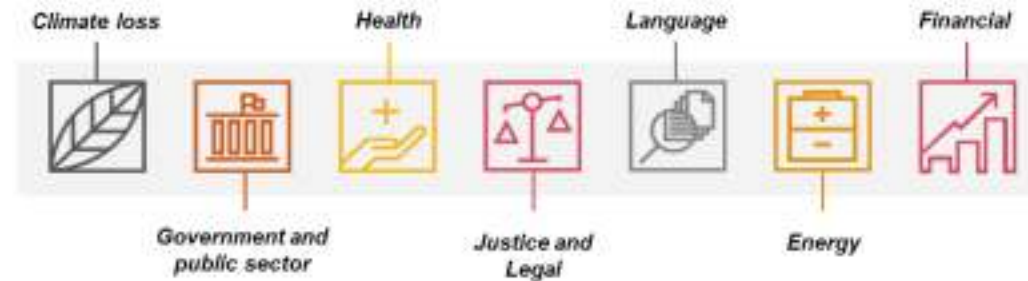
Identification of data themes for the extensions of public sector High-Value Datasets | Final Study

Preliminary list of data themes identified at the beginning of the study

#	Theme		
1	Energy	10	Information and Communication Technology
2	Health	11	International cooperation
3	Public procurement	12	Election results
4	Legal information	13	Government Accountability and Democracy
5	Language corpora	14	Social Protection and Inclusion
6	Agriculture, food, fisheries	15	Education and Training
7	Climate loss	16	Working conditions
8	Tourism	17	(Social) Equality
9	Taxation	18	Justice system information and activity

Source: PwC EU services

High Value Open Data & Ireland – The Future



High Value Open Data & Ireland – The Future

List of new potential HVD themes and sub-themes

Theme	Sub-themes
Financial data	<ul style="list-style-type: none">● Indexation and inflation data● Taxes and economic activities● Asset and interest declarations data● Government budget and spending information
Government and public administration	<ul style="list-style-type: none">● Procurement● Election data● Lobbying and interest representation data● Other (national legislations, register of registers, etc.)
Energy	<ul style="list-style-type: none">● Real time energy information● Production and consumption of households and industrial buildings with their own energy production● Share of electricity production from renewable energies
Health	<ul style="list-style-type: none">● Healthcare system statistics and capacity● Medical devices and medicines● Data on infectious diseases● Data on treatments
Climate loss	<ul style="list-style-type: none">● Economic losses and damages● Environmental vulnerability data● Biodiversity data● Other - Waste management data

High Value Open Data & Ireland – The Future

Language data	<ul style="list-style-type: none">• Official documents from public administrations• Paper and newspaper articles, reports, magazines, newsletters, etc• Language corpus in multiple languages and key domains (i.e. those relevant for other HVD themes)• Language corpus in non-verbal languages (signs language, braille)
Justice and Legal	<ul style="list-style-type: none">• Judgments• Enforcement of sentences• Draft legislation and parliamentary data• Legal system and regulations

Source: PwC EU services

High Value Open Data & Ireland – The Future

Theme	Theme Potential dataset categories and supporting datasets
Financial data	<ul style="list-style-type: none"> • Indexation and inflation data <ul style="list-style-type: none"> ○ House price and rent indexes (regional level) ○ Data on properties (e.g. house prices per region/district/city; loans) • Taxes and economic activities <ul style="list-style-type: none"> ○ Tax data on natural persons ○ Enterprises economic activity • Asset and interest declarations data <ul style="list-style-type: none"> ○ Financial disclosures with structured data on income, spending, assets, and debts. ○ Declared financial and non-financial interests of public officials • Government budget and spending information <ul style="list-style-type: none"> ○ Public administration budget ○ Public expenditure
Government and public administration	<ul style="list-style-type: none"> • Procurement <ul style="list-style-type: none"> ○ Tenders per government office ○ Awards per government office ○ Contract and implementation data • Election data <ul style="list-style-type: none"> ○ Local/regional elections results over the last ten years ○ National election result data ○ Plebiscites and referendums ○ Elected officials' declared conflicts of interest • Lobbying <ul style="list-style-type: none"> ○ Activities of lobbyists: clients and public officials ○ List of accredited lobbyists to institutions • Other <ul style="list-style-type: none"> ○ List of government entities and their subsidiaries ○ National legislation ○ Register of registers

High Value Open Data & Ireland – The Future

Theme	Theme Potential dataset categories and supporting datasets
Energy	<ul style="list-style-type: none"> • Real time energy information • Production and consumption of energy generated own production by households and industrial buildings • Share of electricity production from renewable energies
Health	<ul style="list-style-type: none"> • Healthcare system statistics and capacity <ul style="list-style-type: none"> ○ Number of healthcare facilities ○ Number of beds ○ Standard medical classifications used for healthcare administration in each country • Medical devices and medicines <ul style="list-style-type: none"> ○ Market approval of medical device and medicines, ○ Medical device monitoring and pharmacovigilance (PMS) ○ List of manufacturers and importers (or Authorised representatives) • Data on infectious diseases <ul style="list-style-type: none"> ○ Vaccinations ○ Tests & positivity ○ Confirmed cases and deaths ○ Reproduction rate ○ Hospital & ICU ○ Policy responses • Data on treatments <ul style="list-style-type: none"> ○ Treatment process (diagnoses, types of treatments, duration, discharge or death, certificates of death) ○ Data on admissions, ICU admissions and re-admissions ○ Data on rehabilitation
Language data	<ul style="list-style-type: none"> • Official documents from public administrations • Paper and newspaper articles, reports, magazines, newsletters, etc • Language corpus in multiple languages and key domains (i.e. those relevant for other HVD themes) • Language corpus in non-verbal languages (signs language, braille)

High Value Open Data & Ireland – The Future

Theme	Theme Potential dataset categories and supporting datasets
Justice and Legal	<ul style="list-style-type: none">• Judgments<ul style="list-style-type: none">○ Judgement text and title○ Proceedings duration (overall, per stage, etc.)○ Plaintiffs and defendants○ Criminal data (e.g. rate per type of crime, recidivism including human rights and environmental violations)○ Judgments concerning EU-wide rules○ Cross-border procedural justice data• Enforcement of sentences<ul style="list-style-type: none">○ Data on sentences and penalties (by type: civil, commercial, criminal) and anonymised data connected to it.○ Enforcement processes and detentions (duration, location, amendments to sentences, preventive detentions)○ Cross-border law enforcement• Draft legislation and parliamentary data<ul style="list-style-type: none">○ Data on bills (status, author, votes, transcripts of debates, records of committee proceedings)○ Information generated through the legislative process○ Text of introduced legislation and amendments○ Parliamentary agenda and schedule○ Historical records of legislation• Legal system and regulations<ul style="list-style-type: none">○ National laws and statutes○ Regional laws○ Jurisdictions○ Courts and judge

High Value Open Data & Ireland – The Future

Theme	Theme Potential dataset categories and supporting datasets
Climate loss	<ul style="list-style-type: none">• Economic losses and damages<ul style="list-style-type: none">○ Financial accounts related to natural disasters○ Statistics on the civil assurance amounts and calculation/estimation mechanism○ Budget required per mitigation / adaptation / resolution activity• Vulnerability data<ul style="list-style-type: none">○ Urban water (quality, access, and scarcity)○ Land use monitoring○ Data on extreme weather events○ Population access to early warning systems for disease vectors and natural calamities○ Infrastructures vulnerability (territories exposed to potential floods)○ Data on scope of coastal protection• Biodiversity data<ul style="list-style-type: none">○ Endangered species and ecosystems○ Agricultural practices○ Crop varieties• Other - Waste management data

High Value Open Data & Ireland – The Future

- PO1 - to be applied to those data themes presenting the higher maturity level for which, the attainment of the stronger data openness requisites is considered more feasible in the analysed 9-year target timeframe (2024 - 2032). The data themes potentially covered by this PO would include:
 - Government and public administration data;
 - Financial data;
 - Language data;
 - Climate loss data.
- PO2 - preferred policy options for less mature data themes in which a more gradual approach could be more feasible and proportionate, considering that no similar requirements were previously imposed for such data themes. The data themes potentially covered by this PO would include:
 - Justice and Legal;
 - Health;
 - Energy.

High Value Open Data & Ireland – The Future

Country	Health	Justice & Legal	Financial data	Energy	Gov. and public admin.	Climate loss	Language data
EEA average	55.2	52.9	63.4	57.3	66.3	54.3	69.4
Austria	16.7	0.0	68.8	50.0	68.8	18.8	48.4
Belgium	62.5	37.5	65.6	56.3	56.3	33.8	63.2
Bulgaria	50.0	56.3	63.5	43.8	62.5	58.8	67.0
Croatia	56.3	62.5	48.4	31.3	62.5	56.3	64.1
Cyprus	50.0	50.0	59.4	56.3	43.8	53.8	63.4
Czechia	56.3	62.5	53.1	50.0	62.5	50.0	66.9
Denmark	58.3	56.3	62.5	62.5	62.5	52.5	70.3
Estonia	60.4	81.3	74.5	68.8	81.3	73.8	84.5
Finland	47.9	50.0	40.6	50.0	50.0	31.3	56.2
France	70.8	68.8	67.7	68.8	75.0	66.3	80.8
Germany	33.3	0.0	37.5	43.8	50.0	38.8	45.1
Greece	66.7	93.8	59.9	62.5	75.0	42.5	77.9
Hungary	72.9	68.8	78.1	68.8	75.0	72.5	83.9
Iceland	62.5	0.0	70.3	68.8	75.0	73.8	69.6
Ireland	50.0	62.5	29.7	62.5	62.5	46.3	63.4

Maturity scores of selected data themes across Europe

Conclusions

The policy and strategic context for open data nationally and internationally is therefore complex and dynamic, with multiple actors, initiatives and challenges involved. **The EU's Open Data Directive, HVDs and Implementing Regulation represent a significant opportunity for Ireland to enhance its open data ecosystem and to reap the benefits of data-driven innovation and public value creation.** However, they also pose some **challenges** for Ireland in terms of **ensuring compliance, quality, interoperability and re-use of HVDs across different sectors and levels of government.**



Questions