



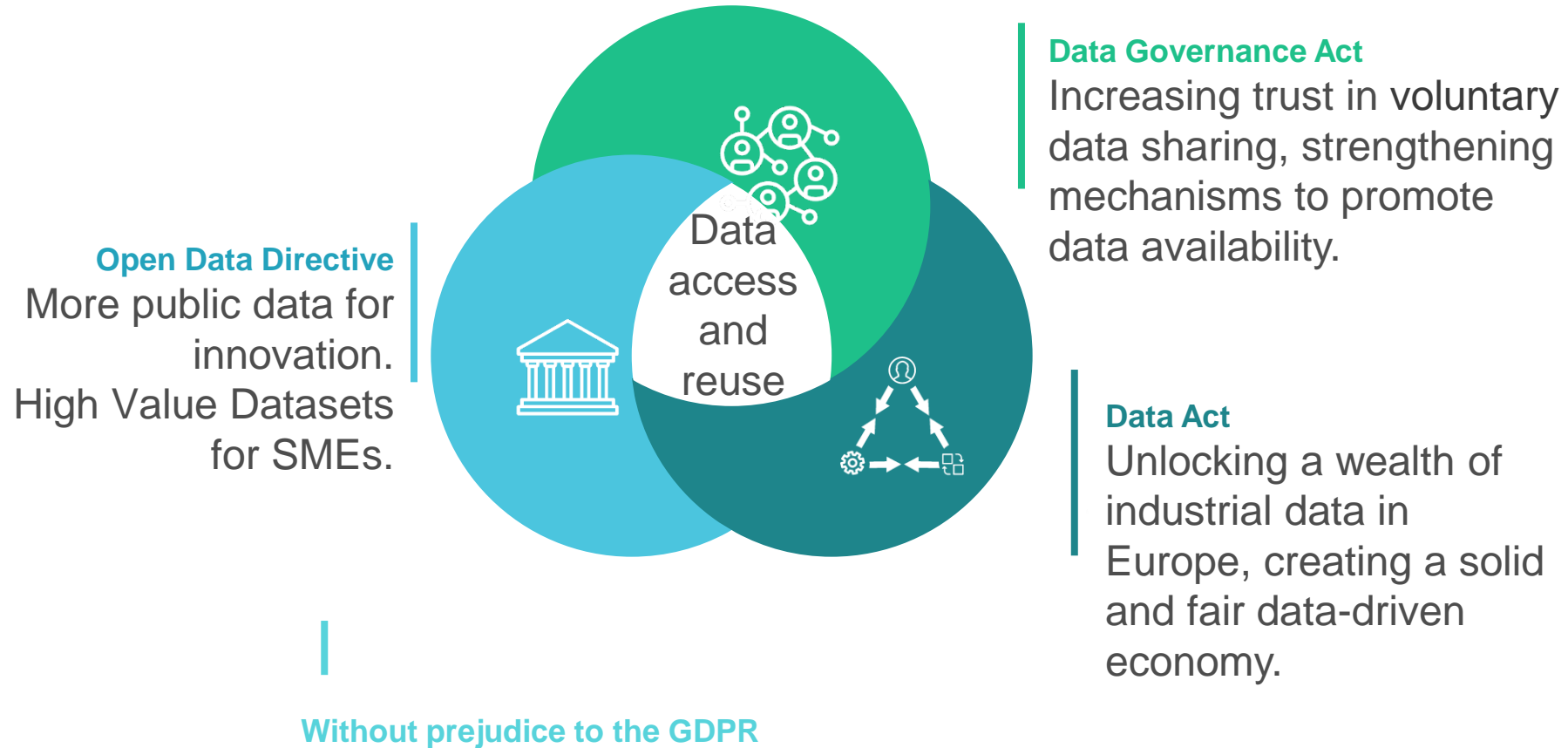
High-value datasets: open data and their linkages

Jiri PILAR, European Commission

DG for Communications Networks, Content and Technology, unit G1

Policy goal - Creating a single European market for data

Legal instruments from the perspective of the type of data



The beginning: public sector data open for re-use

Open Data Directive

Open Data Directive

of 2019

(its predecessor
in place since 2003)

- **The Directive**: all publicly accessible data from the public sector should be easier and cheaper to re-use for innovation.
- Access to information rules are mostly set by the MS.
- Application of the Directive must fully respect the data protection rules, confidentiality, security, IPRs etc.

Public sector bodies' obligations:

- be transparent on conditions for re-use;
- avoid any form of discrimination between re-users, including a re-use by the public sector body itself;
- deal with applications for re-use within a time limit;
- not enter into exclusive arrangements other than in exceptional circumstances;
- limit charges (Free re-use by default; covering marginal cost of reproduction is possible; narrowly defined exceptions that allow full recovery of costs of data).

Way forward: public sector data to be made really open

Open Data Directive => high-value datasets

Open Data Directive

of 2019
(its predecessor
in place since 2003)

- **The Directive**: all publicly accessible data from the public sector should be easier and cheaper to re-use for innovation.
- Access to information rules are mostly set by the MS.
- Application of the Directive must fully respect the data protection rules, confidentiality, security, IPRs etc.
 - **Next level:** Commission Implementing Regulation starting to apply from 9 June 2024.
 - List of public sector datasets with a high reuse potential for the society and economy that **have to be made available**:
 - for free, in machine-readable formats, via APIs and (where relevant) as bulk downloads.

Geospatial

*Earth observation
and environment*

Meteorological

Statistics

*Companies and
company
ownership*

Mobility

Main Act

- Subject matter
- Definitions
- Arrangements for publication (e.g. details on APIs)
- Arrangements for reuse (e.g. charging exception, time series, licences, personal data issues)
- Reporting by Member States: **from February 2025** every 2 years
- Applicability: 16 months after entry into force, i.e. **9 June 2024**

Annex

- Datasets or their description and arrangements for the publication and re-use for individual categories

Statistics – Annex to the Implementing Regulation 1/2

Datasets

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 (see Tables 1 and 2 below for variables in scope)

Harmonised Indices of consumer prices

National accounts – GDP main aggregates (see Tables 6-7 below for variables in scope)

National accounts – key indicators on corporations (see Table 8 below for variables in scope)

National accounts – key indicators on households (see Table 9 below for variables in scope)

Government expenditure and revenue (see Table 10 below for variables in scope)

Consolidated government gross debt (see Tables 11 and 12 below for variables in scope)

Environmental accounts and statistics

Population, Fertility, Mortality

Legal acts laying down the key variables of the datasets in scope and their breakdowns

Regulation (EU) 2019/2152 of the European Parliament and of the Council [\(44\)](#)

Table 26 of the Annex to Commission Implementing Regulation (EU) 2020/1197 [\(45\)](#)

Table 5 of Annex I to Implementing Regulation (EU) 2020/1197

Table 7 of Annex I of Implementing Regulation (EU) 2020/1197

Regulation (EU) 2019/2152

Section 2 of Annex I to Regulation (EU) No 692/2011 of the European Parliament and of the Council [\(46\)](#)

Article 1 of Commission Delegated Regulation (EU) 2019/1681 [\(47\)](#)

Article 3 of Regulation (EU) 2016/792 of the European Parliament and of the Council [\(48\)](#)

Regulation (EU) No 549/2013 of the European Parliament and of the Council [\(49\)](#), in particular Annex B

Regulation (EU) No 549/2013, in particular Annex B

Regulation (EU) No 549/2013, in particular Annex B

Regulation (EU) No 549/2013, in particular Annex B

Chapter I of Council Regulation (EC) No 479/2009 [\(50\)](#)

Regulation (EU) No 549/2013

Annex I to Regulation (EU) No 691/2011 of the European Parliament and of the Council [\(51\)](#)

Annex I and II to Regulation (EC) No 2150/2002 of the European Parliament and of the Council [\(52\)](#)

Regulation (EU) No 1260/2013 of the European Parliament and of the Council [\(53\)](#)

Statistics – Annex to the Implementing Regulation 2/2

Datasets

Current healthcare expenditure

Poverty (see Table 13 below for variables in scope)

Inequality (see Table 14 below for variables in scope)

Employment (see Tables 15-16 below for variables in scope)

Unemployment (see Table 17 below for variables in scope)

Potential labour force (see Table 18 below for variables in scope)

Legal acts laying down the key variables of the datasets in scope and their breakdowns

Annex II to Regulation (EC) No 1338/2008

Annex II to Commission Regulation (EU) 2015/359 [\(61\)](#)

Regulation (EU) 2019/1700 of the European Parliament and of the Council [\(62\)](#)

Commission Implementing Regulation (EU) 2019/2180 [\(63\)](#)

Commission Implementing Regulation (EU) 2019/2181 [\(64\)](#)

Commission Implementing Regulation (EU) 2019/2242 [\(65\)](#)

Regulation (EU) 2019/1700

Implementing Regulation (EU) 2019/2180

Implementing Regulation (EU) 2019/2181

Implementing Regulation (EU) 2019/2242

Regulation (EU) 2019/1700

Commission Implementing Regulation (EU) 2019/2240 [\(66\)](#)

Regulation (EU) 2019/1700

Implementing Regulation (EU) 2019/2240

Regulation (EU) 2019/1700

Implementing Regulation (EU) 2019/2240

Arrangements for the publication and re-use (example of statistical data)

- The statistics thematic category includes statistical datasets **except microdata** related to reporting obligations as defined in the legal acts listed in the table.
- **Time series** shall begin no later than at the date of application of the respective legal act listed in the table below.
 - a) The datasets shall be made available for re-use:
 - as **frequently** as required by the corresponding legislation (e.g. monthly, quarterly, yearly);
 - under the conditions of the **Creative Commons BY 4.0 licence** or any equivalent or less restrictive open licence;
 - in CSV, XML (SDMX), JSON or another publicly documented, Union or internationally recognised **open, machine-readable format**;
 - through **APIs and bulk download**;
 - b) The **metadata** describing the data set shall be available as a well-developed structured file containing at least a description of the statistical data, the statistical concepts, methodologies and information on data quality.
 - c) The datasets shall be described in a complete and publicly available **online documentation** describing at least the data structure and semantics.
 - d) Datasets shall use Union or internationally recognised and publicly **documented controlled vocabularies and taxonomies**, where available.

DCAT-AP for HVDs: metadata in key role

- DCAT-AP – metadata standard
- Useful for: (i) findability, (ii) reporting: from February 2025.
- <https://semiceu.github.io/DCAT-AP/releases/2.2.0-hvd/> - Guidelines on how to use DCAT-AP for a dataset that is subject to the requirements imposed by the implementing regulation.
- Ongoing: guidance to Member States on how to automate their reporting thanks to metadata

DCAT-AP High Value Datasets

14 December 2023

▼ More details about this document

Latest published version:

<https://semiceu.github.io/DCAT-AP/releases/2.2.0-hvd>

Latest editor's draft:

<https://semiceu.github.io/DCAT-AP/releases/2.2.0-hvd>



Reporting: builds on existing (improved) common practices

Use existing metadata **standards** and metadata **sharing networks** of data portals

DCAT-AP High Value Datasets
14 December 2023

More details about this document

Latest published version:
<https://semiceu.github.io/>

Latest editor's draft:
<https://semiceu.github.io/>

History:
[Commit history](#)

Editors:
Makx Dekkers
Pavina Fragkou (SEMICEU)
Natasia Sofou
Bert Van Nuffelen (TenForce)

Author:
Pavina Fragkou (SEMICEU)

Feedback:
[GitHub SEMICEU/DCAT-AP](#)

Copyright © 2022

Field	Value
Title	Technical Guidance for the implementation of INSPIRE dataset and service metadata based on ISO/TS 19139:2007
Creator	Temporary MIWP 2021-2024 sub-group 2.3.1
Date of publication	2024-01-31
Subject	Technical Guidance for INSPIRE metadata for datasets and services
Status	Version 2.2.0 This document has been endorsed by the INSPIRE maintenance and implementation group (MIG).
Publisher	INSPIRE Maintenance and Implementation Group (MIG)
Type	Text
Description	Implementation specification for defining metadata for INSPIRE datasets and services in ISO/TS 19139 based XML format in compliance with the INSPIRE Implementing Rules for metadata.
Format	AsciiDoc
Licence	Creative Commons Attribution (cc-by) 4.0
Identifier	http://inspire.ec.europa.eu/ids/document/00/metadata-iso19139
Changelog	https://github.com/INSPIRE-MIF/technical-guidelines/releases/tag/v2024.1
Language	EN

European Union

Log in English

EU Vocabularies

Home EU law European data EU tenders EU research results EU Whoiswho EU publications

Home EU Vocabularies Controlled vocabularies Dataset

Controlled vocabularies Models Business collections Online tools Releases Help

Permanent link Add to my lists Create alert

Asset

High-value dataset categories

Dataset URI: <http://publications.europa.eu/resource/dataset/high-value-dataset-category>

Dataset type: Code list

Author: Directorate-General for Communications Networks, Content and Technology

Publisher: Publications Office of the European Union

Description: List of thematic categories of high-value datasets, as set out in Annex I of Directive (EU) 2019/1024 of the European Parliament and of the Council of 20 June 2019 on open data and the re-use of public sector information. The concepts listed in this concept scheme are representations of the data categories as set out in Commission Implementing Regulation (EU) 2023/138 http://data.europa.eu/eli/reg_impl/2023/138/oj. It is not recommended to use these data categories outside the context of legislation on the re-use of public sector information.

High-value dataset categories is maintained by Directorate-General for Communications Networks, Content and Technology and published by the Publications Office of the European Union on the EU Vocabularies.

Version: 1.0 LATEST

Downloads Documentation Alignments Release notes Diff files Advanced view

INSPIRE Geoportal

Home Priority Datasets High-Value Datasets Thematic Data Find out more

European Commission

European data

data.europa.eu The official portal for European data

Home Data Academy Community Publications Documentation

Enabling the INSPIRE Geoportal allows

- monitoring
- discovering
- accessing

The metadata used

Feedback register

Explore regional similarities with EU Twinings

Compare different regions and get to know your European neighbours better

Find out more

Application procedures datasets and eReporting

Search Datasets

Metadata standards such as [DCAT-AP HVD](#),
INSPIRE HVD guidelines

Shared code lists such as [HVD categories](#), [licences](#), ...

Harvesting networks and data portals such as data.europa.eu and [INSPIRE data portal](#).

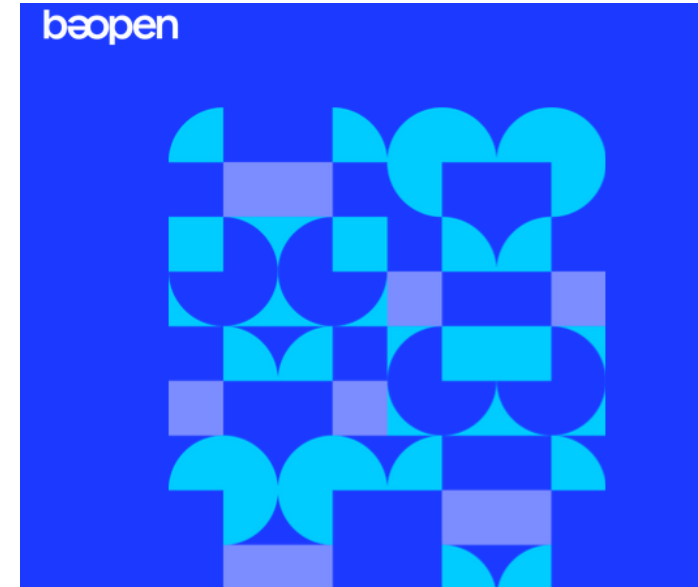
Ongoing activities which are being improved to match the HVD requirements

Extension of thematic categories of HVDs

- Final study published in 2023
- Potentially for the next Commission.
- Candidates:
 - Climate loss;
 - Energy;
 - Financial;
 - Government and public administration;
 - Health;
 - Justice and Legal;
 - Language.



Projects under Digital Europe Programme



MareGraph
Towards an Interoperable
Marine Knowledge Graph

Data Governance Act

Regulation (EU) 2022/868 of the European Parliament and of the Council of 30 May 2022, [applicable since 24 September 2023](#))

Leveraging the potential of non-public data for the economy and society

Pillar 1:

Re-use of **closed data** held by **public sector bodies**

- **European single access point to public sector's non-open data**



[European register for protected data held by public sector \(ERPD\)](#)

Pillar 2:

Framework for new **data intermediaries** in the Single Market: Data broker

- [EU register of recognised data intermediation services](#)

Pillar 3:

Corporate and Individual **data altruism** for purposes of general interest

- [EU register of recognised data altruism organisations](#)

Pillar 4:

Co-ordination and interoperability: **European Data Innovation Board**

- [First meeting in December 2023](#)
- [Subgroup of the DGA for competent authorities \(2 meetings\)](#)

Horizontal framework for the creation of a single market for data

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:

<http://data.europa.eu/eli/dir/2019/1024/oj>

http://data.europa.eu/eli/reg_impl/2023/138/oj

<https://digital-strategy.ec.europa.eu/en/policies/open-data>

<https://digital-strategy.ec.europa.eu/en/factpages/open-data-and-high-value-datasets-step-step-access-guide>