Exploring the potential of high-value datasets: connecting with real use cases

WEBINAR

14 June 2024
10.00 — 11.30 CEST
Rules of the game

The webinar will be recorded

For questions, please use the ClickMeeting chat.

Please reserve 3 min after the webinar to help us improve by filling in our feedback form.
Introduction

Martin Page PhD
ODM Study
Lead, data.europa.eu

Martina Ramponi
Research and Policy,
data.europa.eu

Lenka Kováčová
Open Data Specialist, Digital
and Information Agency

Antti Helen
Ministerial Adviser for
EU Legal Affairs,
Ministry of Finance of
Finland

Jiří Pilař
Data Policy &
Innovation, European
Commission

Georg Hittmair
Head of Legal and Public
Affairs, Compass-Gruppe

Claudius Determann
Legal and Public Affairs,
Compass-Gruppe
<table>
<thead>
<tr>
<th>Time</th>
<th>Event</th>
<th>Presenter(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>10.00 – 10.10</td>
<td>Opening and introduction</td>
<td>Martin Page</td>
</tr>
<tr>
<td>10.10 – 10.20</td>
<td>Main takeaways from the research report</td>
<td>Martina Ramponi</td>
</tr>
<tr>
<td>10.20 – 10.30</td>
<td>High-value datasets' state of play from Czech Republic</td>
<td>Lenka Kováčová</td>
</tr>
<tr>
<td>10.30 – 10.40</td>
<td>High-value datasets' state of play from Finland</td>
<td>Antti Helin</td>
</tr>
<tr>
<td>10.40 – 10.50</td>
<td>Q&amp;A session</td>
<td></td>
</tr>
<tr>
<td>10:50 – 10:55</td>
<td>Introduction by DG CNECT</td>
<td>Jirka Pilar</td>
</tr>
<tr>
<td>10:55 – 11:15</td>
<td>Use case and experiences on evaluating the accessibility of company data</td>
<td>Georg Hittmair &amp; Claudius Determann</td>
</tr>
<tr>
<td>11:15 – 11:30</td>
<td>Q&amp;A session</td>
<td></td>
</tr>
<tr>
<td>11:30</td>
<td>Closing remarks</td>
<td>Martin Page</td>
</tr>
</tbody>
</table>
Agenda

Introduction to the report  3
Methodology  5
Key highlights for best practices  8
Conclusion  10
Introduction to the Report
The HVD implementing regulation
The high-value datasets implementing regulation entered into force on 9 June 2024

• What are high-value datasets (HVDs)?
  • Datasets that are particularly valuable thanks to their social and economic benefits.
  • List of six categories, provided in the Annex I: Geospatial data, Earth observation and environment, Meteorological data, Statistics, Companies and company ownership, Mobility.

• The HVD Regulation and the new requirements introduced:
  • Recital 2: free of charge;
  • Article 3: APIs (Application Programming Interfaces) + bulk downloads;
  • Article 4: machine readable format + open data licenses.
The aim of the Report

• Aim of the HVD Best Practices Report:

1) Learn about challenges.

2) Identify best practices and actionable recommendations to support the Member States in the implementation of HVD requirements.
Methodology

*Interviews with Member States*
How did we gather information about best practices?

• **Qualitative interviews** with Member States representatives that are involved with the implementation of HVD requirements

  • The Member States are: **Czechia, Denmark, Estonia, Italy, The Netherlands, Romania, and Finland**.

  • Interviews were based on a **questionnaire** which followed the legal analysis and used the ODM questionnaire as a starting point.

  • We interviewed first the less mature countries and, based on their answers, we tailored the questions for the more mature countries to identify **best practices**.

*The order of the countries follows the EU protocol*
Findings

*Key insights of the identified best practices*
Key highlights for best practices

• Policy and legal framework:
  • Practice prioritisation and focus on achievable goals (Czechia, Denmark).

• Technical aspects, metadata quality and new requirements:
  • Adapt the business model gradually to provide free data (Denmark).
Key highlights for best practices

• Governance and processes:

• Promote sustainability through individual agency responsibility for their respective datasets (Czechia, Denmark).

• Utilise open data to address your challenges and find solutions (Estonia).
Conclusion

Remaining Challenges &
Suggested Roadmap
Remaining challenges

• Balancing data protection standards:
  • Estonia, Italy and the Netherlands

• Licence ambiguity:
  • Estonia, Denmark
Suggested roadmap

- The report includes a suggested roadmap to guide the Member States in the compliance with HVD implementing regulation.
OVERVIEW_

1. WHAT_SHOULD_HAVE_BEEN_DONE_?  
2. WHAT_WE_DID_?  
3. WHAT_CHALLENGES_HAVE_WE_ENCOUNTERED_?
**MAIN_TASKS**

**WHY**
- MOTIVATION
- WORKING GROUP
- MINIMAL FINANCIAL SUPPORT

**WHAT**
- IMPLEMENTING REGULATION
- LIST OF DATASETS
- INSPIRE IN/CONSISTENCY

**WHO**
- RESPONSIBLE PUBLIC BODIES
- LEGAL ANALYSIS
- DIFFERENT DATA HOLDERS

**HOW**
- QUALITATIVE REQUIREMENTS
- WEBINAR AND STANDARDS
- LACK OF TECHNICAL CAPACITIES
WHY
MOTIVATION
WORKING GROUP
MINIMAL FINANCIAL SUPPORT

WHAT
IMPLEMENTING REGULATION
LIST OF DATASETS
INSPIRE IN/CONSISTENCY

WHO
RESPONSIBLE PUBLIC BODIES
LEGAL ANALYSIS
DIFFERENT DATA HOLDERS

HOW
QUALITATIVE REQUIREMENTS
WEBINAR AND STANDARDS
LACK OF TECHNICAL CAPACITIES
CHALLENGES:

**WHY**
- MOTIVATION
- WORKING GROUP
- MINIMAL FINANCIAL SUPPORT

**WHAT**
- IMPLEMENTING REGULATION
- LIST OF DATASETS
- INSPIRE IN/CONSISTENCY

**WHO**
- RESPONSIBLE PUBLIC BODIES
- LEGAL ANALYSIS
- DIFFERENT DATA HOLDERS

**HOW**
- QUALITATIVE REQUIREMENTS
- WEBINAR AND STANDARDS
- LACK OF TECHNICAL CAPACITIES

DIGITÁLNÍ
A INFORMAČNÍ
AGENTURA_
THANK YOU

LENKA.KOVACOVA@DIA.GOV.CZ
OTEVRENADATA@DIA.GOV.CZ
Implementing the HVD Regulation in Finland: Observations

ANNTI HELIN, MINISTERIAL ADVISER
PUBLIC ICT DEPARTMENT
14 JUNE 2024
Background

• The Open Data Program (2013-2015)
• Utilizing and Opening Data Program (2020-2023)
• Act on Openness of Government Activities (1999)
• Act on Information Management in Public Administration (2019)
Act on Information Management in Public Administration

The objectives of the Act are:

1. to ensure harmonised and high-quality management and data secure processing of datasets of authorities to implement the principle of openness;

2. to enable secure and efficient exploitation of the datasets of authorities so that an authority may attend to its tasks and provide its services to public administration clients successfully and in a qualitative manner in compliance with good governance;

3. to promote the interoperability of information systems and information pools.

The Act also mandated the creation of the Information Management Map of Public Administration
Tracking down the HVDs and their holders

• Locating the datasets and identifying the critical stakeholders using
  • the information management map
  • materials and reports from the 2021 implementation of the Open Data Directive
  • working groups and forums set up on the basis of the Act on Information Management
Getting to work: the ’HVD community’

• After identifying the critical stakeholders, we established three groups:
  • Working group for legal issues (ministries)
  • Cross-sectoral coordination group for practical and technical issues (ministries and public sector bodies)
  • Network of interested parties and stakeholders for feedback and discussion (all of the above + researchers, others in the open data community)
• Also involved: the Geospatial Data Advisory Board and its secretariat
At the end of the day, the easiest and fastest solution for tracking HVD status was...
Outcomes and challenges

• Most HVD datasets, once tracked down and analyzed, were already nearly or completely compliant with the HVD Regulation.
• No major legal issues or questions were identified.
• The biggest challenges were related to understanding the interplay between INSPIRE and HVD/ODD.
  • The INSPIRE community was used to stricter and more technical specifications than what are in the HVDR
• Land use plan datasets, collected and maintained by Finland’s 300+ municipalities, remain a challenge.
• Lack of technical know-how and expertise at the policymaker/ministry level remains another challenge.
Q&A

Antti Helen
Ministerial Adviser for EU Legal Affairs, Ministry of Finance of Finland

Martin Page PhD
ODM Study Lead, data.europa.eu

Martina Ramponi
Research and Policy, data.europa.eu

Lenka Kováčová
Open Data Specialist, Digital and Information Agency
High-value datasets: application and reporting by Member States

Data.europa.eu Academy webinar, 14 June 2024

Jiri PILAR, Michal KUBAN, European Commission
DG for Communications Networks, Content and Technology, unit G1
The beginning: public sector data open for re-use.

Open Data Directive

- **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.
The beginning: public sector data open for re-use.
Open Data Directive => high-value datasets.

- **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 started 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.
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://www.gtsc.wiev.de
- Latest editor's draft: https://www.gtsc.wiev.de
- History: https://www.gtsc.wiev.de

**Technical Guidance for the implementation of INSPIRE dataset and service metadata based on ISO/TS 19139:2007**

**EU Vocabularies**

- Technical Guidance for the implementation of INSPIRE metadata for datasets and services based on ISO/TS 19139:2007
- Technical Guidance for the implementation of INSPIRE metadata for datasets and services based on ISO/TS 19139:2007

**Harvesting networks and data portals such as data.europa.eu and INSPIRE data portal.**

Metadata standards such as **DCAT-AP HVD**, **INSPIRE HVD guidelines**

Shared code lists such as **HVD categories**, **licences**, …

Ongoing activities which are being improved to match the HVD requirements
Member States to report on High Value Datasets

- Under Article 5(3) of the Implementing Regulation each Member State must „provide the Commission with a report on the measures they have carried out to implement this Implementing Regulation“.

- The Commission made available tools to make the reporting simple, thanks to concrete SPARQL queries.

- By 9 February 2025 and then every 2 years.

Article 5 (3): The report shall contain the following information:

(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 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 carried out in accordance with Article 35 of Regulation (EU) 2016/679;

(f) the number of public sector bodies exempted in accordance with Article 14(5) of Directive (EU) 2019/1024.
Beyond reporting: everyone will benefit from easier findability

- Harmonised metadata from national portals harvested by data.europa.eu

- Data users will be able to search for HVDs either through filters available at data.europa.eu or through SPARQL queries

- The process and tools (SPARQL queries) are described at https://dataeuropa.gitlab.io/data-provider-manual/hvd/what_hvd/; ready to be tried and tested by anyone.

NB: As countries are still in the process of denoting high-value datasets on their national portals and making them accessible for harvesting, properly denoted high-value datasets are only gradually becoming available on data.europa.eu.
High-value datasets (HVDs) as a foundation for business models
Nothing really new at Compass for 150 years
Compass collects data from various sources...

The most important elements are data from the Austrian company register.
... combines interesting information,

150 years ago customs tariffs and the costs for telegrams were essential for doing business
…transfers the data to modern storage media,
...and sells it to our customers!

Some customers buy our data since more than 150 years.
The business model is longstanding - storage media vary
Our service includes financial statements and bylaws

Compass collects all Austrian data licensed by the Ministry of Justice in accordance with the PSI Directive.
Over the years the number of sources has grown…

More than 40 data origins request complex matching mechanisms to present consistent datasets

…as has the number and the expertise of our IT specialists.
High-value datasets?
They offer(ed) the opportunity to scale up on a European level

- 2019: Assumption that the data category „Companies and company ownership“ will contain all the disclosed company data according to European Law.
- Those data should be released as PSI according to the PSI directive anyway.

→ The Annex to the Implementing regulation (EU) 2023/138 was a therefore a big disappointment, as only 8 key attributes were included to be released as HVDs.
Using HVDs for a business model

**Precondition:** A high level of consistency between Member States with regard to

- Data formats (API)
- Update cycle
- Included data categories
Using HVDs for a business model

**Precondition:** A high level of consistency between Member States with regard to

- Data formats (API)
- Update cycle
- Included data categories

**Problem:** Unharmonized APIs from all Member States require investments and manpower for maintenance:

- Although Metadata are standardised, the mandatory APIs are not:
  
  "To provide information about the use of the API platform, the publisher provides OpenAPI technical documentation and an SLA to document the quality of service."

- organizing access to 27 different registers
- merging 27 different sources
- maintaining 27 different update cycles and temporal coverage
With BRIS a European solution already exists

Member States deliver key attributes of company data on a daily basis in an EU-wide harmonized format!

Data is accessible and therefore should be reusable!
Advantages of using BRIS data

- High degree of topicality
- Uniform consolidation across Europe
- Harmonized data formats
- Ready to use

BRIS data is not available at the moment.
Reuse of bylaws and financial statements

According to recent Q & As for the Member States, the **obligation to provide HVDs is weakened**:

*The HVD Implementing Regulation does not oblige Member State to create datasets (or key attributes) that do not exist in that Member State.*

It is up to the Member States to decide how much personal data they provide for reuse:

*Member States should make use of appropriate methods and techniques (such as generalisation, aggregation, suppression, anonymisation, differential privacy or randomisation), thus making as much data as possible available for re-use.*
Reuse of business register information apart HVDs

- Potential reusers have to search for the data and very often the obstacles described by “access info europe“ are confirmed by the relevant public sector bodies:
  - No response to requests
  - Free access is possible but only for record-by-record searches
  - Record-by-record access is possible but must be paid for
  - You have to know what you’re looking for
  - Bulk access must be paid for
  - Personal privacy is an obstacle... unless you can pay for the data
  - There are limits on reusing the data
HVDs could be found easily, if available

So far, little data on companies can be found
Barriers when reusing HVDs on companies and company ownership

- No substantive coverage of HVDs on the central platform
- No usage of harmonized solutions
- Legal and administrative barriers
Solutions for reusing HVDs on companies and company ownership?
Solutions for reusing HVDs on companies and company ownership?

- Europeanwide data on a daily basis
Solutions for reusing HVDs on companies and company ownership?

- Europeanwide data on a daily basis
- Easy implementation: Members States should provide common interfaces if available
Solutions for reusing HVDs on companies and company ownership?

- European-wide data on a daily basis
- Easy implementation: Members States should provide common interfaces if available
- EU-wide scope without legal exceptions. Data protection impact assessments are necessary on a European basis that apply to all member states.
Thank you!

Compass-Verlag GmbH
Schönbrunner Straße 231, 1120 Wien

www.compass.at
Stay up-to-date on our 2024 activities!
Webinar

The European Drug Report: using an open data approach to improve data visualisation

21 June 2024
10.00 – 11.30 CET

data.europa academy

Webinar

Stories from the Use Case Observatory - Volume 2

28 June 2024
10.00 – 11.30 CET

data.europa academy
Please provide your feedback!
Thank you