



High Value Datasets under the Open Data Directive

**EuroGeographics General Assembly webinar
8 October 2020**

Jiri PILAR, DG CNECT, European Commission
jiri.pilar@ec.europa.eu

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).

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

Progress so far and next steps

- The Open Data Directive to be transposed in Member States by 17 July 2021
- Work towards the definition of the List of HVDs (2019-21)
 - Open Data Committee (comitology: for the Implementing Regulation, 2nd meeting in June 2020)
 - Public consultation – until 31 May 2020
 - Inception impact assessment – was open for comments in August 2020
 - Impact assessment – supported by the ongoing study till October 2020
 - Dedicated actions (workshops, focus groups)
 - Regulatory Scrutiny Board by end 2020
 - Proposal of the list of HVDs to be submitted to the Open Data Committee in Q1 2021

Recommendations by the "Impact Assessment study on the list of high-value datasets to be made available by the Member States under the PSI Directive"

Deloitte.



theLisborcouncil
think tank for the 21st century



Geospatial

List of HVDs and their value

Administrative Units	<p>Units of administration, dividing areas where Member States have and/or exercise jurisdictional rights, for local, regional and national governance, separated by administrative boundaries.</p> <p>Land Administrative Units and Maritime Units are the basic units. Land Administrative Units are covering mostly land surface, while Maritime Units are covering territorial waters.</p>	<p>Mapping or use as statistical units, manage emergency rescue, waste management plans, protect water ecosystems, find responsible party for policy implementation and administration, forest management, subsidies for farmers, forecast agricultural production, spatial planning, monitoring of regional and urban policy implementation using territorial typologies based on administrative units, maritime spatial planning, integrated coastal management</p>
Place Names	<p>Geographical names or place names (or toponyms) are the proper nouns applied to topographical features and settled (and used) places and spaces on the earth's surface. Toponyms represent an important reference system used by individuals and societies throughout the world.</p>	<p>Emergency response Economic, social and environmental analysis Cultural identity and heritage Mapping and navigation Providing a link / index function to other spatial and aspatial data</p>
Addresses	<p>Location of properties based on address identifiers, usually by road name, house number, postal code.</p> <p>The basic unit of addressing is a building; a permanent construction, intended or used for the shelter of people, having at least one entrance from publicly-accessible space.</p>	<p>Geocoding of statistical surveys, manage emergency rescue, locate where people are, accessibility studies, manage incidents; locate economic activities in ecosystem accounting</p>
Buildings	<p>Geographical location of buildings.</p> <p>Constructions above and/or underground, intended or used for the shelter of humans, animals, things, the production of economic goods or the delivery of services that refer to any structure permanently constructed or erected on its site [from INSPIRE Data Specifications on Buildings].</p>	<p>Buildings are 3D topographic objects and, as such, may influence the propagation of physical phenomena. These data are required for serving citizens (e.g. school, hospital), assessments for air and noise pollution or risk assessments to various kinds of risks (earthquake, fire, flood etc.), monitoring of land consumption, population concentration and access to services.</p>
Cadastral Parcels	<p>Single areas of Earth surface (land and/or water), under homogeneous real property rights and unique ownership, real property rights and ownership being defined by national law.</p>	<p>Protect state lands, reduce land disputes, facilitate land reform, agriculture, land management, taxation, disaster management, real Estate Market, Taxation, LPIS (Agriculture), Land consolidation, Infrastructure Management, Spatial Planning, Protection of Soil and Water, Statistics</p>

Geospatial

Lower intensity intervention – Recommended measures

	Description	Administrative units	Place Names	Addresses	Buildings	Cadastral parcels
Openness	License and terms of use	CC-BY 4.0				
	Format	GeoPackage; GeoJSON; INSPIRE requirements.	GeoPackage; CSV; GeoJSON; INSPIRE requirements.	GeoPackage; CSV; GeoJSON; INSPIRE requirements.	GeoPackage; GeoJSON; INSPIRE requirements.	GeoPackage; GeoJSON; INSPIRE requirements.
	Machine-readability	Mandatory				
	Availability of API, bulk download	Bulk download; INSPIRE distribution services; RestAPI (e.g. OGC API, ArcGIS RestAPI, Carto API).				Read-only mode (WMS service defined by INSPIRE).
Documentation	Metadata (dataset content description)	INSPIRE				
	Documentation (incl. structure and semantics)	INSPIRE; GeoDCAT-AP.				
	Data linking					
	Shared vocabularies/taxonomies	INSPIRE				
Completeness	Traceability	National Geodata Catalog and/or open data catalog				
	Update frequency and timeliness	Annual update	When necessary	When necessary	When necessary	When necessary
	Granularity	From municipalities to countries; sea-frontiers.	National coverage.	Partial National coverage (e.g. most populated cities).	Partial National Coverage (e.g. most populated cities); Level of scale 1:5000.	National coverage; Level of scale 1:5000.
	Key attributes	National identification code; identification code of the upper administrative level; official name; country code; name in multiple languages (only for countries with more than one official language).	Name; name in multiple languages (only for countries with more than one official language); category; latitude and longitude (INSPIRE)	Latitude and longitude; house number; suffix of the number; name of the street; name of the municipality; national identification code of the municipality; last update.	Footprint of the building; entrances; floors; type of use.	Geometry of cadastral parcels; type of particle; particle code; references to the administrative area to which the particle belongs.

Geospatial

Lower intensity intervention – Recommended measures

- Concerning **licences and terms of use**, the recommended licence for the lower intensity intervention is the Creative Commons Attribution 4.0 International - CC-BY 4.0. This recommended option will have no impact since this type of licences is widely used across Member States (MS). In the same time, this type of licence is also preferred by re-users.
- APIs and bulk download**: Regarding the cadastral parcels - the recommendation is to ensure at least access through WMS service as defined by INSPIRE, in read-only mode, due to the difficulties in the release of this dataset. The WMS remains a data product which allows the creation of new georeferenced data and the view of the geo-data as a whole. The WMS protocol provides feature information (as XML) by identifying a point on a map. For the other datasets, the download through API, bulk download, OGC services listed by INSPIRE should be guaranteed.
- When it comes to **formats**, an option for the publication of datasets is to follow INSPIRE requirements. According to the stakeholders' suggestions, the recommendation is to use Geopackage and GeoJSON, which are two relatively recent open and low-cost formats. GeoJSON is commonly used, the re-users have shown a strong preference for new open standards like Geopackage. Using these formats increases the machine-readability of the data. The proposal includes the use of dedicated services for automatic conversion is added.
- Granularity**: concerning Addresses and Buildings datasets the recommendation is to guarantee at least a partial coverage (e.g. most populated cities). Based on our interviews, the findings already show a diversified situation across the MS and main issues are related to the data ownership at local level, the costs of management and the frequency of update that strongly affect the lack of a full national coverage. However, the full national coverage is recommended for the rest of the datasets. The level of scale to be guaranteed for the Buildings and Cadastral parcels is 1:5000 or beyond. The increase of detail in the level of scale corresponds to a raise in the costs of the dataset production.
- The **metadata** should match the INSPIRE requirements as the transformation to geoDCAT-AP can be done automatically from this basis. It is important that the data are also displayed in geoDCAT-AP in order to integrate them automatically into national open data portals.
- When it comes to **key attributes**, the values identified correspond to the common characteristics available across the majority of the Member States, for each type of dataset in scope. This is due to the implementation of the INSPIRE directive. Therefore, the impact on the organisations will remain rather low and limited. The interviewed stakeholders highlighted the relevance of having names in multiple languages as a key attribute, because it significantly improves the reuse. Regarding the Buildings dataset, entrances and floors are recommended in the list of the key attributes, as based on these elements it's possible to calculate the height of the building. The type of use could be generic and very basic. The recommended attributes for the Cadastral Parcels are the basic ones needed to release cadastral data with respect to the GDPR and to guarantee a good level of reusability.

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. 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.

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)

Earth observation and environment

Higher intensity intervention Recommended measures

Dimensions		E-reporting priority data and Environmental Monitoring Facilities	INSPIRE Themes
<i>Openness-data specification</i>	<i>License (terms of use)</i>	Creative Commons 0 or BY (or equivalent open license) No terms of use	Creative Commons 0 or BY (or equivalent open license) No terms of use
	<i>Format</i>	Cf. existing INSPIRE specifications	Cf. existing INSPIRE specifications
	<i>Machine-readability</i>	Obligatory	Obligatory
	<i>Availability of API, bulk download</i>	Both API and bulk download, as prescribed in e-reporting priority list	INSPIRE download services (e.g. WFS or OGC:API) and bulk download
<i>Documentation</i>	<i>Metadata (dataset content description)</i>	Complete (INSPIRE)	INSPIRE, or DCAT-AP2
	<i>Documentation (incl. structure and semantics)</i>	Complete and web-available (INSPIRE)	INSPIRE
<i>Completeness</i>	<i>Update frequency and timeliness</i>	As collected, for EMF highest collected temporal resolution	INSPIRE
	<i>Granularity</i>	Highest collected temporal and geographic resolution	INSPIRE
	<i>Key attributes</i>	All attributes mentioned in e-reporting priority data list and EMF locations, parameters measured and complete observations data	All features within INSPIRE theme's scope

No specific recommendations for data linking, shared vocabularies, taxonomies, or traceability, other than existing INSPIRE specifications

Mobility

Lower intensity intervention – Datasets in scope (in addition to inland waterways)



INSPIRE features under the Transport Networks data theme, including:

Common	Road	Rail	Water			Air	Cableways
<ul style="list-style-type: none">• Access Restriction• Condition Of Facility• Maintenance Authority• Marker Post• Owner Authority• Restriction For Vehicles• Traffic Flow Direction• Transport Area• Transport Link• Transport Link Sequence• Transport Link Set• Transport Network• Transport Node• Transport Object• Transport Point• Transport Property• Vertical Position	<ul style="list-style-type: none">• E-Road• Form Of Way• Functional Road Class• Number Of Lanes• Road• Road Area• Road Link• Road Link Sequence• Road Name• Road Node• Road Service Area• Road Service Type• Road Surface• Road Width• Roundabout• Speed Limit• Vehicle Traffic Area	<ul style="list-style-type: none">• Design Speed• Nominal Track Gauge• Number Of Tracks• Railway Area• Railway Electrification• Railway Line• Railway Link• Railway Link Sequence• Railway Node• Railway Station Area• Railway Station Code• Railway Station Node• Railway Type• Railway Use• Railway Yard Area• Railway Yard Node	<ul style="list-style-type: none">• Beacon• Buoy• CEMT Class• Condition Of Water Facility• Fairway Area• Ferry Crossing• Ferry Use• Inland Waterway• Marine Waterway• Port Area• Port Node• Restriction For Water Vehicles• Traffic Separation Scheme• Traffic Separation Scheme Area	<ul style="list-style-type: none">• Traffic Separation Scheme Crossing• Traffic Separation Scheme Lane• Traffic Separation Scheme• Traffic Separation Separator• Water Link Sequence• Water Node• Water Traffic Flow Direction• Waterway• Waterway Link• Waterway Node	<ul style="list-style-type: none">• Aerodrome Area• Aerodrome Category• Aerodrome Node• Aerodrome Type• Air Link• Air Link Sequence• Air Node• Air Route• Air Route Link• Airspace Area• Apron Area• Condition Of Air Facility• Designated Point• Element Length• Element Width	<ul style="list-style-type: none">• Field Elevation• Instrument Approach Procedure• Lower Altitude Limit• Navaid• Procedure Link• Runway Area• Runway Centreline Point• Standard Instrument Arrival• Standard Instrument Departure• Surface Composition• Taxiway Area• Touch Down Lift Off Area• Upper Altitude Limit• Use Restriction	<ul style="list-style-type: none">• Cableway Link• Cableway Link Sequence• Cableway Link Set• Cableway Node





Mobility

Lower intensity intervention – Recommended modalities (*INSPIRE transport network data*)

Dimensions		Road transport	Rail transport	Water transport	Air transport	Cableways
Openness-data specification	License (terms of use)	CC-BY 4.0 No terms of use				
	Format	GML, GeoPackage, GeoJSON				
	Machine-readability	Mandatory				
	Availability of API, bulk download	Webservice, (OGC) API and bulk download				
Documentation	Metadata (dataset content description)	Complete (*.csv document available)				
	Data linking	Links to national INSPIRE Geoportals and datasets				
	Documentation (incl. structure and semantics)	Complete and available				
	Shared vocabularies	INSPIRE data specifications are recommended but not mandatory.				
	Taxonomies	N/A				
Completeness	Traceability	N/A				
	Update frequency and timeliness	When necessary				
	Granularity	From local to national, including links with cross-border networks, if and where available				
	Key attributes	Any national identification code; latitude and longitude				

Multi-Criteria Analysis (MCA)

Results and Conclusions

- Applying equal weights to all five criteria within the algorithm used, **PO1 (lower intensity intervention)** is identified as preferred option in the thematic areas of **Company & Company Ownership**, **Geospatial data and Mobility**. In the thematic areas **Meteorological Data**, **Earth Observation & Environment** and **Statistics Policy Option 2 (higher intensity intervention)** is the preferred option.

	Preferred PO (Equal Weight 0.20)
Company & Company Ownership	PO1
Geospatial Data	PO1
Meteorological Data	PO2
Earth Observation & Environment	PO2
Statistics	PO2
Mobility	PO1

Supporting activities

- **Open Data digital infrastructure: European Data Portal and EU Open Data Portal**
- **Connected Europe Facility (CEF):** actions funded from the 2019-2020 Work Programme will improve the availability for re-use of existing open data falling into the high-value datasets categories, paving the way to more targeted actions funded under the Digital Europe Programme as of 2021, focusing on the actual datasets
- **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.