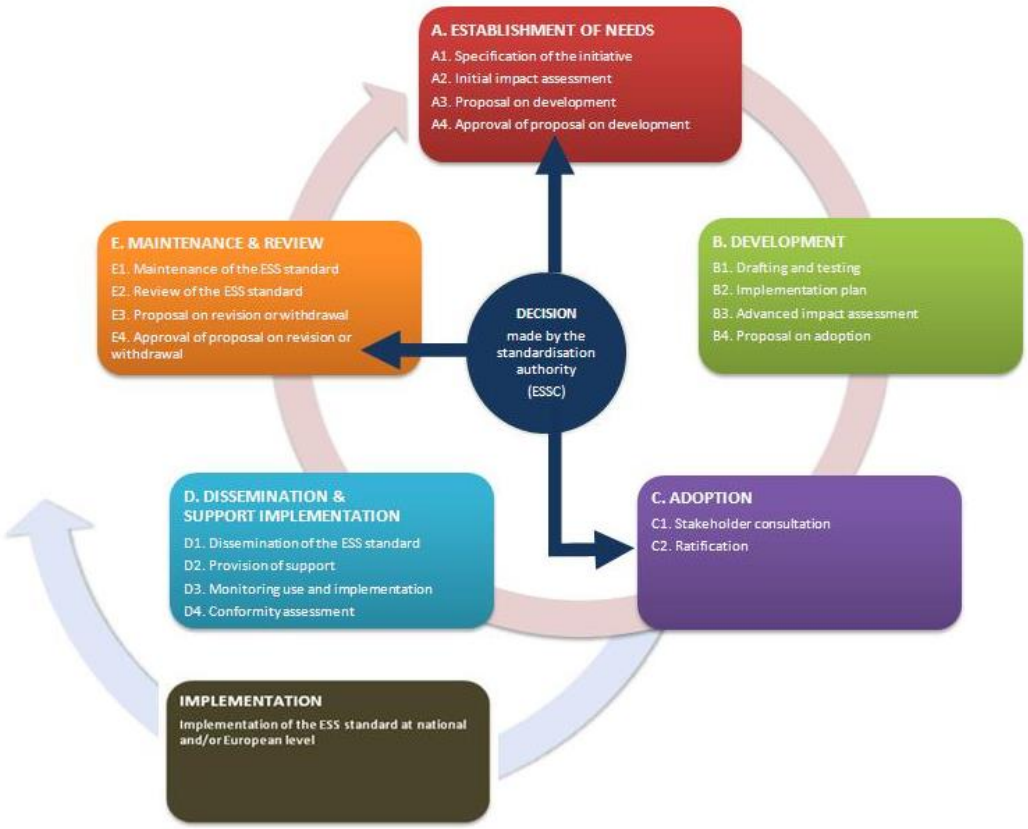
Standardisation in the European Statistical System

ESS standards for data and metadata reporting

European Commission

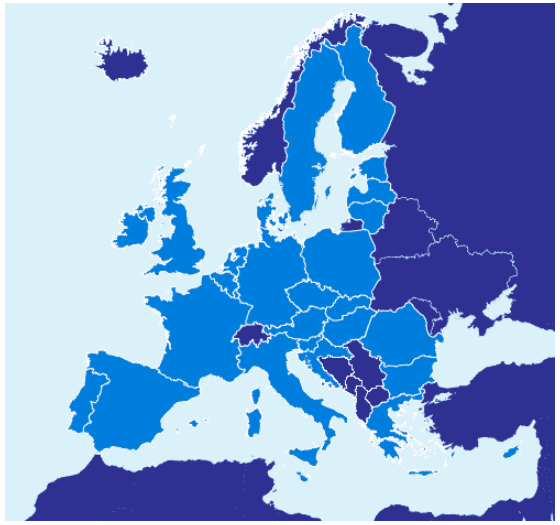
Standardisation in the European Statistical System



- The European Statistical System (ESS) has a long-standing history of developing and adopting standards for the production of Official Statistics.
- Process and principles of ESS standardisation adopted by the 26th Meeting of the ESSC on 25 September 2015.
- Catalogue of ESS standards is available online. It currently contains 12 standards. Half of these standards refer to data and metadata exchange

European Commission

ESS Standards for Data and Metadata Exchange



Data

sdmx

eurostat

SIMS

Metadata and quality reports

No data without metadata!

Legal act (223/2009): Member States shall provide the Commission (Eurostat) with reports on the quality of data transmitted, including any concerns they have regarding the accuracy of the data.

European Commission

ESS Standards for Data Exchange - SDMX

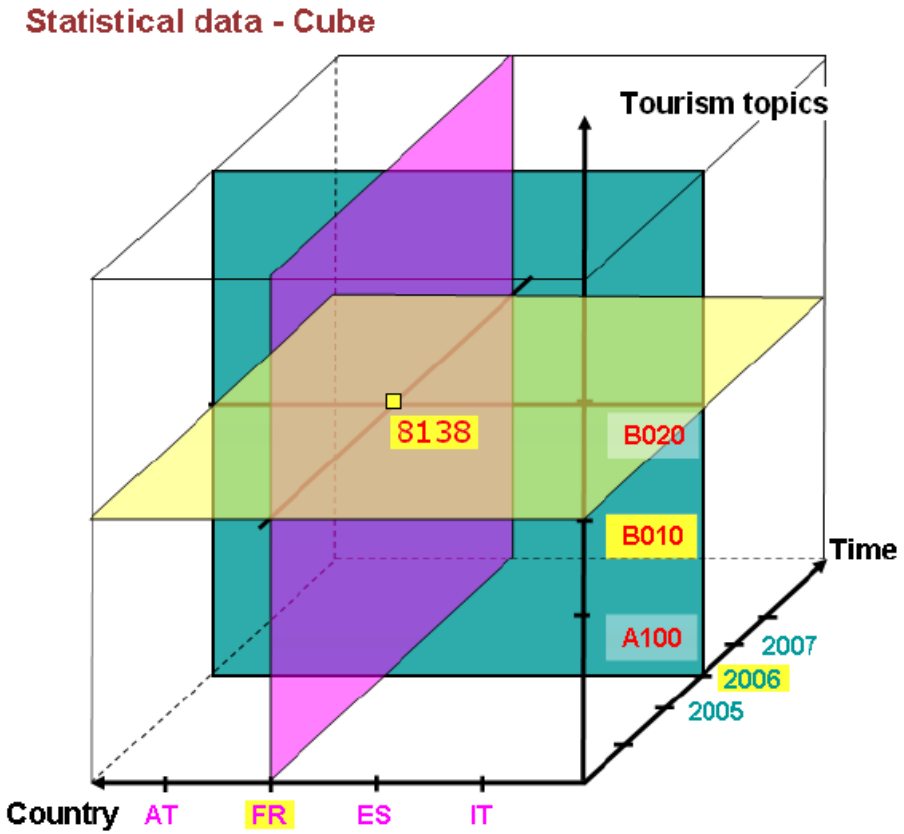
The ***S****tatistical* ***D****ata and* ***M****etadata e****X****change* is an international initiative aimed at developing and employing more efficient processes for the exchange and sharing of statistical data and metadata among international organisations and member countries.

It consists of technical and statistical standards, guidelines, an IT service infrastructure and IT tools.

sdmx
Statistical Data and Metadata eXchange

European Commission

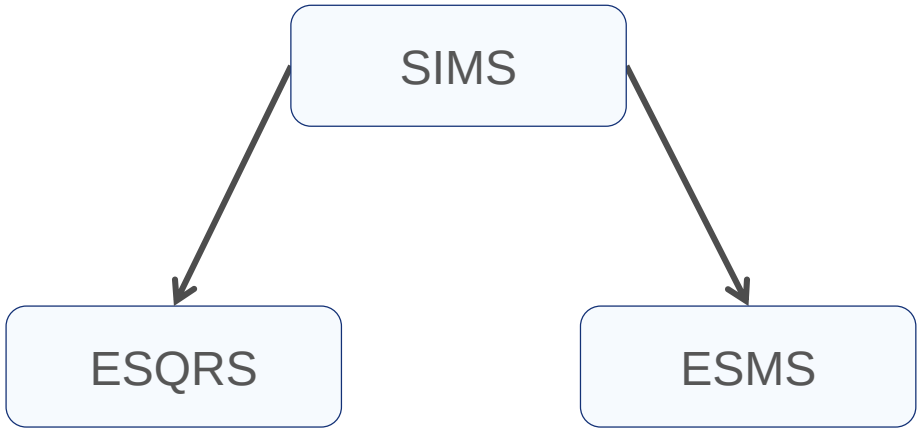
ESS Standards for Data Exchange - SDMX



At its core, SDMX provides a standard language to describe the structure of statistical data and metadata, specifying the codes and concepts to be used.

This standardised description is called Data Structure Definition (DSD) for data and Metadata Structure Definition (MSD) for metadata.

ESS Standards for Metadata Exchange - SIMS



SIMS: set of concepts relevant for metadata/ quality reporting

ESMS: Subset for user-oriented report

ESQRS: Subset for producer-oriented reports

Objective: Once for all purposes reporting using SIMS

Technical infrastructure powered by sdmx Statistical Data and Metadata eXchange

European Commission

SIMS 2.0

Item No	Concept name
S.1	**Contact**
S.1.1	Contact organisation
S.1.2	Contact organisation unit
S.1.3	Contact name
S.1.4	Contact person function
S.1.5	Contact mail address
S.1.6	Contact email address
S.1.7	Contact phone number
S.1.8	Contact fax number
S.2	**Metadata update**
S.2.1	Metadata last certified
S.2.2	Metadata last posted
S.2.3	Metadata last update
S.3	**Statistical presentation**
S.3.1	Data description
S.3.2	Classification system
S.3.3	Sector coverage
S.3.4	Statistical concepts and definitions
S.3.5	Statistical unit
S.3.6	Statistical population
S.3.7	Reference area
S.3.8	Time coverage
S.3.9	Base period
S.4	**Unit of measure**
S.5	**Reference period**
S.6	**Institutional mandate**
S.6.1	Legal acts and other agreements
S.6.2	Data sharing
S.7	**Confidentiality**
S.7.1	Confidentiality - policy
S.7.2	Confidentiality - data treatment
S.8	**Release policy**
S.8.1	Release calendar
S.8.2	Release calendar access
S.8.3	User access
S.9	**Frequency of dissemination**
S.10	**Accessibility and clarity**
S.10.1	News release
S.10.2	Publications
S.10.3	On-line database

Item No	Concept name
S.10.3.1	AC1. Data tables - consultations
S.10.4	Micro-data access
S.10.5	Other
S.10.5.1	AC 2. Metadata - consultations
S.10.6	Documentation on methodology
S.10.6.1	AC 3. Metadata completeness - rate
S.10.7	Quality documentation
S.11	**Quality management**
S.11.1	Quality assurance
S.11.2	Quality assessment
S.12	**Relevance**
S.12.1	User needs
S.12.2	User satisfaction
S.12.3	Completeness and R1. Data completeness - rate for U
S.12.3.1	R1. Data completeness - rate for P
S.13	**Accuracy and reliability**
S.13.1	Overall accuracy
S.13.2	Sampling error and A1. Sampling errors - indicators for U
S.13.2.1	A1. Sampling errors - indicators for P
S.13.3	Non-sampling error and A4. Unit non-response - rate for U and A5. Item non-response - rate for U
S.13.3.1	Coverage error
S.13.3.1.1	A2. Over-coverage - rate
S.13.3.1.2	A3. Common units - proportion
S.13.3.2	Measurement error
S.13.3.3	Non response error
S.13.3.3.1	A4. Unit non-response - rate for P
S.13.3.3.2	A5. Item non-response - rate for P
S.13.3.4	Processing error
S.13.3.5	Model assumption error
S.14	**Timeliness and punctuality**
S.14.1	Timeliness and TP2. Time lag - final results for U
S.14.1.1	TP1. Time lag - first results for P
S.14.1.2	TP2. Time lag - final results for P
S.14.2	Punctuality and TP3. Punctuality - delivery and publication for U
S.14.2.1	TP3. Punctuality - delivery and publication for P
S.15	**Coherence and comparability**
S.15.1	Comparability - geographical
S.15.1.1	CC1. Asymmetry for mirror flows statistics - coefficient
S.15.2	Comparability - over time and CC2. Length of comparable time series for U
S.15.2.1	CC2. Length of comparable time series for P

Item No	Concept name
S.15.3	Coherence- cross domain
S.15.3.1	Coherence - sub annual and annual statistics
S.15.3.2	Coherence- National Accounts
S.15.4	Coherence - internal
S.16	**Cost and burden**
S.17	**Data revision**
S.17.1	Data revision - policy
S.17.2	Data revision - practice and A6. Data revision - average size for U
S.17.2.1	A6. Data revision - average size for P
S.18	**Statistical processing**
S.18.1	Source data
S.18.2	Frequency of data collection
S.18.3	Data collection
S.18.4	Data validation
S.18.5	Data compilation
S.18.5.1	A7. Imputation - rate
S.18.6	Adjustment
S.18.6.1	Seasonal adjustment
S.19	**Comment**

Legend

	Common concepts in SIMS, ESMS and ESQRS
	Common concepts in SIMS and ESMS
	Common concepts in SIMS and ESQRS

ESS Data and Metadata Standards – Implementation status

Implementation of SDMX for data exchange: 61.2% of data transmissions to Eurostat are structured ad validated according to SDMX Data Structure Definitions.

Implementation of SIMS for metadata exchange: for 90% of datasets covered by EU legal acts (i.e. are mandatory according to EU law) metadata and quality reports are also collected according to SIMS.

Lessons Learned in the implementation of standards

Meet your users where they are

Eurostat exchanges data with over 400 national and international organisations. Not all organisations have the same degree of technical competence or interest for the implementation of standards.

Rather than attempting to impose the same standardisation level to everyone, standardisation can be structured as a tiered, level-based approach. Standardisation should not be viewed as a series of obstacles, but as a series to steps in a ladder that give growing benefits.

Lessons Learned in the implementation of standards

Legislation is a double-edged sword

Imposing the use of certain standards via legislation has benefits in terms of guaranteeing compliance, but also important drawbacks in terms of buy-in by stakeholders and flexibility. A combination of a high-level legal framework complemented by consensus-based technical-level implementation agreements has worked well in the ESS.

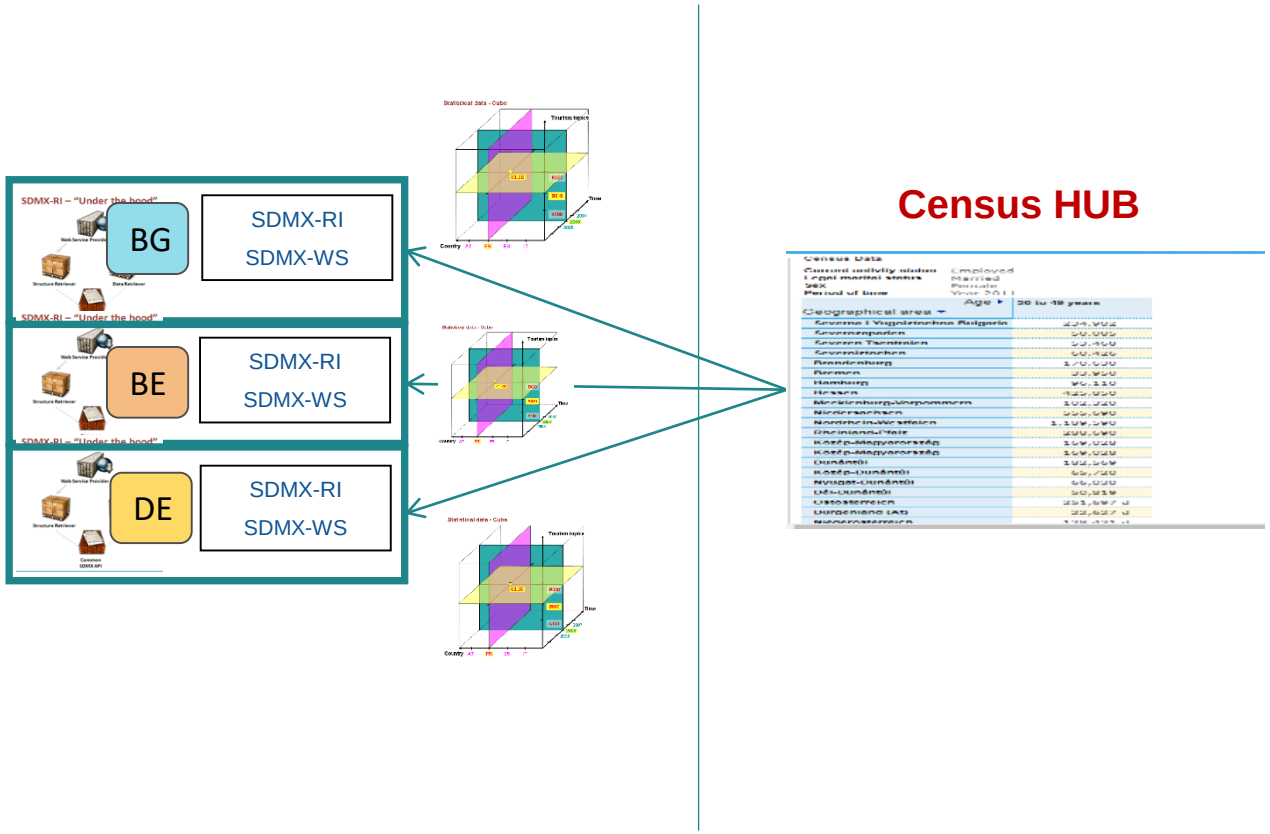
Example for metadata and quality reporting:

- The need to provide quality reports according to standards established via Regulation 223/2009
- The establishment of the SIMS standard taken as a decision by the ESSC (not a legal act).

ESS experiences with the exchange of geospatial data and metadata: the example of the 2021 Census

European Commission

2021 Census: the role of SDMX



Census data disseminated via the Census Hub

- National Statistical Institutes structure their data according to SDMX Data Structure Definitions and expose their data via an SDMX-compliant web service
- Users query the different national web services via a central application hosted by Eurostat – the Census Hub

Census 2021 – Grid-level population data

- Census 2021 foresees the collection of data on the following 13 variables for all 1km^2 grid cells in the EU and EFTA

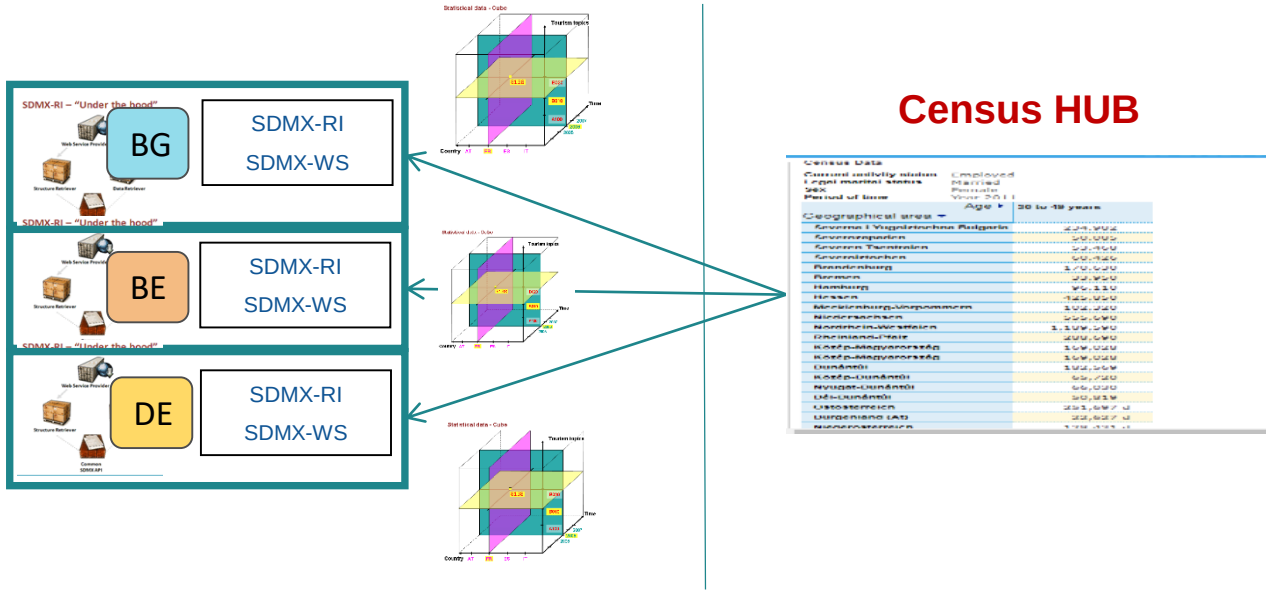
T	Total population
M	Male
F	Female
Y_LT15	Under 15 years
Y15-64	15 to 64
Y_GE65	65 years and over
EMP	Employed persons
NAT	Place of birth in reporting country
EU_OTH	Place of birth in other EU Member State
OTH	Place of birth elsewhere
SAME_1y	Usual residence unchanged
CHG_IN	Usual residence: move within reporting country
CHG_OUT	Usual residence: move from outside reporting country

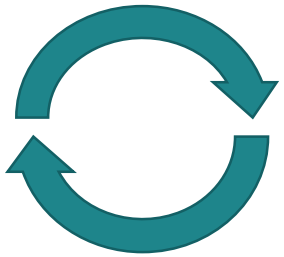
The INSPIRE Directive: requirements for the dissemination of geospatial data

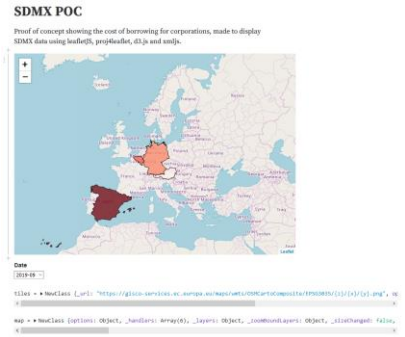
- The INSPIRE Directive aims to create a European Union spatial data infrastructure for the purposes of EU environmental policies and policies or activities which may have an impact on the environment. This European Spatial Data Infrastructure will enable the sharing of environmental spatial information among public sector organisations, facilitate public access to spatial information across Europe and assist in policy-making across boundaries.
- INSPIRE is based on the infrastructures for spatial information established and operated by the Member States of the European Union. The Directive addresses 34 spatial data themes needed for environmental applications.

Overall approach

- Data collected via from Member States via the Census Hub using the SDMX standard.
- Eurostat converts the data to INSPIRE specifications and publishes the datasets on the INSPIRE Geoportal



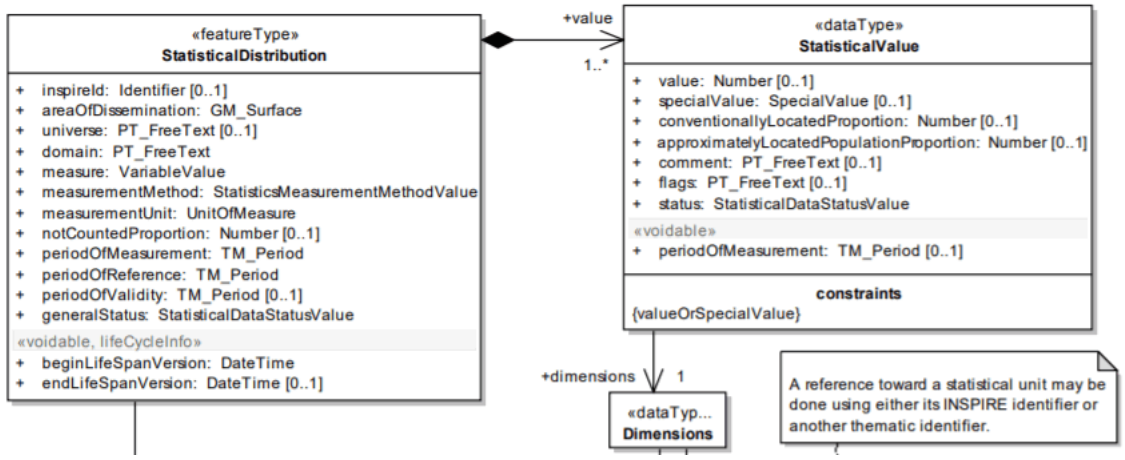




European Commission

Combination of INSPIRE and SDMX - Data

INSPIRE Directive outlines specifications for the dissemination of geospatial data, including certain fields / attributes that should be included



These requirements were included in an SDMX-compliant DSD, to ensure a framework for expressing the data that is both SDMX and INSPIRE-compliant

Combination of INSPIRE and SIMS - Metadata

ESS requirements

Metadata to be supplied according to ESS metadata standards (SIMS)

Census requirements

Census regulation outlines certain metadata elements which must be collected

INSPIRE requirements

The INSPIRE directive specifies that certain metadata should accompany grid-level data

One integrated metadata structure combining all requirements

Timeline

- First geo-located data and metadata on the 2021 Census (on total population only) to be provided by the end of 2022.
- SDMX can work in tandem with standard for geospatial data, but the currently used version of the standard (SDMX 2.1) does not have specific elements to mark / deal with geospatial information. The new version of the standard (SDMX 3.0) introduces such elements.

How to contact us:

estat-data-metadata-services@ec.europa.eu

Data and Metadata exchange portal on CROS

Training and courses

IT Tools

User guides

Explanatory recordings

European Commission

Thank you

European Commission