Gestione informatica dei dati

Statistical Data and Metadata Excange (SDMX)

Prof. Roberto Foderà

Cattedra di Gestione informatica dei dati
Dipartimento di Giurisprudenza - Palermo

LUMSA
UNIVERSITÀ

A.A. 2021-2022

Statistical Data and Metadata Excange

- 7 organizzazioni internazionali (BIS, ECB, Eurostat, IMF, OECD, UN, World Bank) hanno unito le loro forze per sviluppare lo standard

- Mar 2007 le organizzazioni sponsor hanno firmato un Memorandum of Understanding

- Feb 2008 the UN Statistical Commission ha riconosciuto “SDMX as the preferred standard for exchange and sharing of data and metadata in the global statistical community”

Statistical Data and Metadata Excange

- Gli Istituti Nazionali di Statistica hanno chiesto più volte alle organizzazioni internazionali soluzioni comuni

- Le Organizzazioni internazionali possono espandere la collaborazione tra di essi, e con le loro agenzie nazionali

- Nuove tecnologie IT sembrano offrire chiari vantaggi

- Progressi possono essere fatti all'interno di un coerente framework e su base incrementale

Statistical Data and Metadata Excange

Cos'è lo Statistical Data and Metadata Excange (SDMX)

Serie di norme tecniche e linee guida orientati ai contenuti, comprensiva di una architettura IT e di strumenti informatici, da utilizzare per lo scambio efficiente e la condivisione di dati e metadati statistici

Migliora la qualità e l'efficienza nello scambio, condivisione e diffusione dei dati e metadati attraverso

- l'armonizzazione e coerenza dei dati
- la conservazione del significato
- l'uso di un formato open (XML) piuttosto che un formato proprietario

Può ridurre il carico di lavoro per le organizzazioni fornitrici di dati

Statistical Data and Metadata Excange

I suoi elementi possono essere utilizzati dalle Organizzazioni Statistiche come elementi chiave all'interno di un sistema informativo dalla raccolta dei dati fino alla loro diffusione

Potenzialmente può ridurre il costo di sviluppo di sistemi informativi statistici

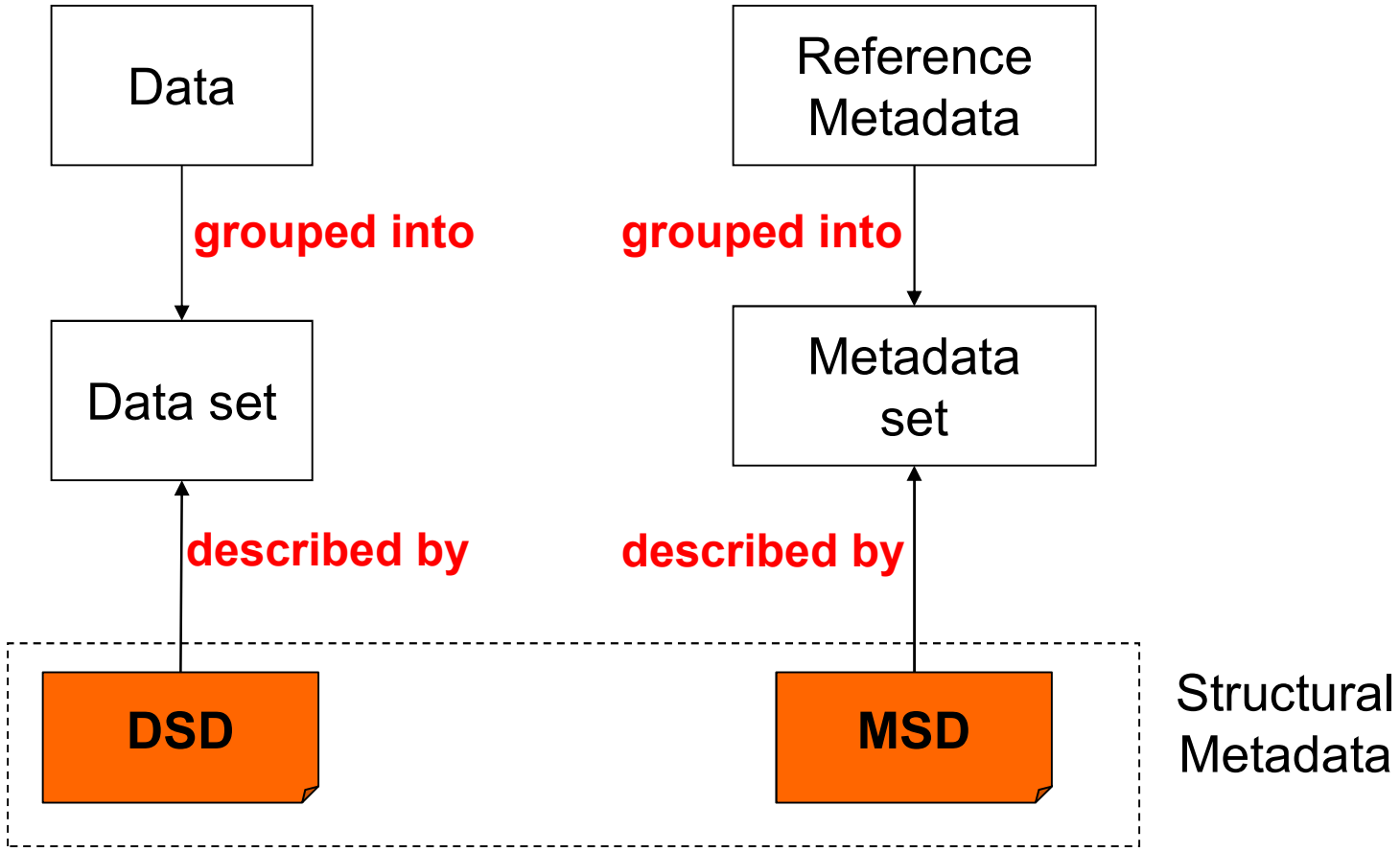
Elimina la duplicazione degli sforzi nello sviluppo e manutenzione di standard per l'elaborazione delle informazioni statistiche

E' riconosciuto dall'ISO

E' pienamente compatibile con standard usati correntemente o nel passato

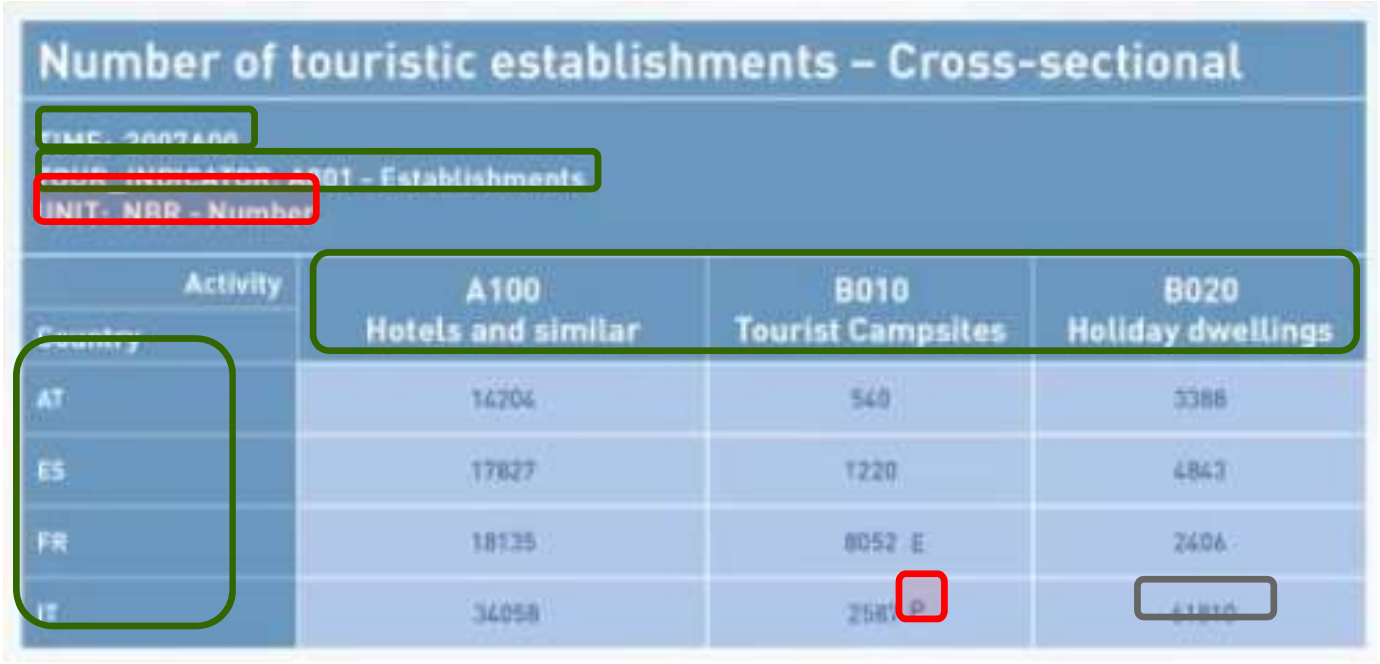
Statistical Data and Metadata Excange

Gli “oggetti” di SDMX



Statistical Data and Metadata Excange

Gli “oggetti” di SDMX



Role	Dimension or attribute name	Identifier	Attachement level	Code list
Dimension	Frequency	FREQ		CL_FREQ
Dimension	Country	COUNTRY		CL_AREA
Dimension	Tourism Indicator	TOUR_INDICATOR		CL_TOUR_IND
Dimension	Tourism activity	TOURISM_ACTIVITY		CL_TOUR_ACT
Dimension	Period	TIME		
Attribute	Unit	UNIT	Dataset	CL_UNIT
Attribute	Observation status	OBS_STATUS	Observation	CL_OBS_STATUS
Measure	Observation value	OBS_VALUE		

Concetto

Si compone di un identificatore, di un nome e di una descrizione

CONCEPT : Tourism Activity	
ID	TOURISM_ACTIVITY
Name	(English) Tourism activity
	(French) Activité touristique
Description	(English) Different types of tourism activity
	(French) Types d'activité Touristiques

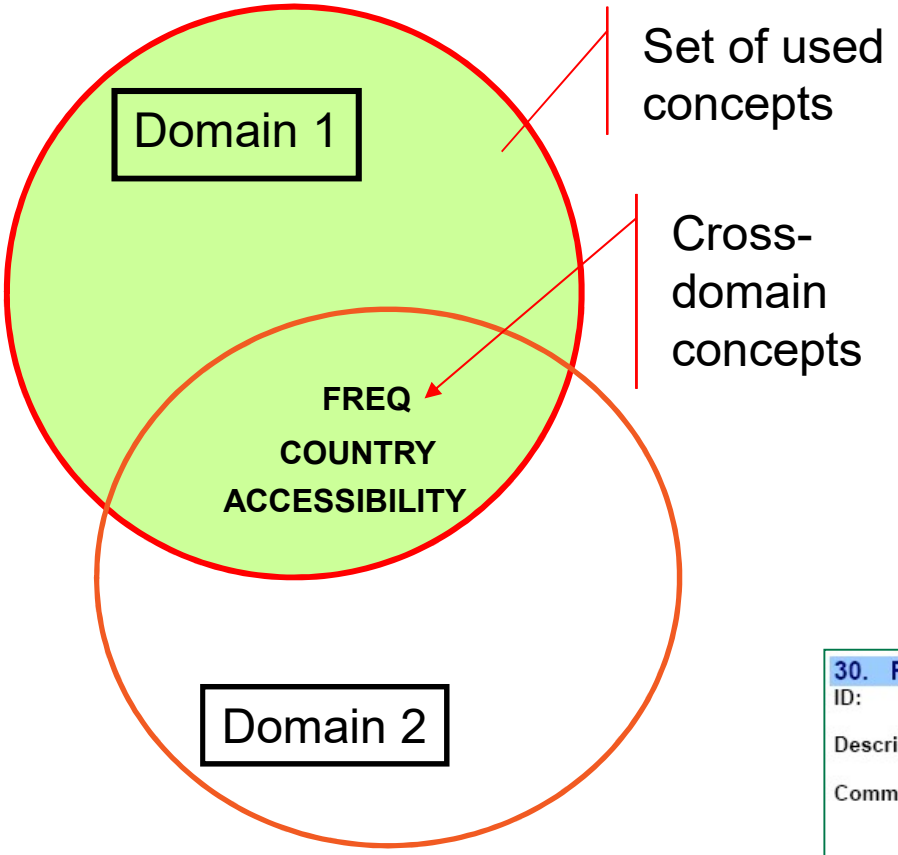
Code list

Si compone di un codice e di una descrizione

ID: CL_TOURISM_ACTIVITY	
Code value	Description
A100	Hotels and similar
B010	Tourist campsites
B020	Holiday dwellings

Statistical Data and Metadata Excange

Cross-domain Concepts and Code Lists



CL_FREQ

Name: code list for Frequency (FREQ)

Description: it provides a list of values indicating the "frequency" of the data (e.g. monthly) and, thus, indirectly, also implying the type of "time reference" that could be used for identifying the data with respect time.

Recommended code value	Recommended code description	Annotation
A	Annual	It is typically used for annual data. This can also serve cases of multi-annual data (data that appear once every two, three or, possibly, five years). Descriptive information on the multiannual characteristics (e.g. frequency of the series in practice and other methodological information can be provided at the dataflow level, as long as these characteristics are applicable for the entire dataflow).
S	Half-yearly, semester	
Q	Quarterly	
M	Monthly	
W	Weekly	
D	Daily	
B	Daily - business week	Similar to "daily", however there are no observations for Saturday and Sunday (so, neither "missing values" nor "numeric values" should be provided for Saturday and Sunday). This treatment ("business") is one way to deal with such cases, but it is not the only option. Such a time series could alternatively be considered daily ("D"), thus, with missing values in the weekend.
N	Minutely	While N denotes "minutely", usually, there may be no observations every minute (for several series the frequency is usually "irregular" within a day/days). And though observations may be sparse (not collected every minute), missing values do not need to be given for the minutes when no observations exist: in any case the time stamp determines when an observation is observed.

30. Frequency

ID:	FREQ
Description:	The time interval at which observations occur over a given time period.
Comment:	If a time series has a constant time interval between its observations, this interval determines the frequency of the time series (e.g. monthly, quarterly, yearly). "Frequency" – also called "periodicity" – may refer to several stages in the production cycle, e.g. data collection, data compilation or data dissemination. (e.g., a time series could be available at annual frequency but the underlying data are compiled monthly).
Presentation:	CL_FREQ_SDMX Free text

Statistical Data and Metadata Excange

Metadati concettuali:
descrivono i concetti usati e la loro attuazione pratica

Capacity and occupancy of tourist accomodation establishments

Reference Metadata in Euro SDMX Metadata Structure (ESMS)

Compiling agency: Eurostat, the statistical office of the European Union

1. Contact	Top
1.1. Contact organisation	Eurostat, the statistical office of the European Union
1.2. Contact organisation unit	G3: Short-term Business Statistics and Tourism
1.5. Contact mail address	2920 Luxembourg LUXEMBOURG

2. Metadata update	Top
2.1. Metadata last certified	18/07/2014
2.2. Metadata last posted	15/12/2014
2.3. Metadata last update	15/12/2014

Statistical Data and Metadata Excange

Metadati metodologici:
descrivono i metodi usati per la generazione dei dati

3. Statistical presentation

Top

3.1. Data description

Accommodation statistics is a key part of the system of tourism statistics in the EU and has a long history of data collection. Annex I of the Regulation (EU) 692/2011 of the European Parliament and of the Council deals with accommodation statistics and includes 4 sections focusing on accommodation statistics of which sections 1 and 2 include the requirements concerning rented accommodation (capacity and occupancy respectively).

Data are collected by the competent national authorities of the Member States and are compiled according to a harmonised methodology established by EU regulations before transmission to Eurostat. Most of the time, data are collected via sample or census surveys. However, in a few cases the data is compiled from a demand-side perspective (i.e. via visitor surveys or border surveys). Surveys on the occupancy of accommodation establishments are generally conducted on a monthly basis.

The concepts and definitions used in the collection of data shall conform to the specifications described in the Methodological manual for tourism statistics.

Accommodation statistics comprise the following information:

Monthly data on tourism industries (NACE 55.1, 55.2 and 55.3)

Monthly occupancy of tourist accommodation establishments: arrivals and nights spent by residents and non-residents

Net occupancy rate of bed-places and bedrooms in hotels and similar accommodation

Annual data on tourism industries (NACE 55.1, 55.2 and 55.3)

Statistical Data and Metadata Excange

Metadati della qualità:
descrivono i differenti aspetti della qualità dei dati statistici

12. Quality management

Top

12.1. Quality assurance

Tourism statistics are compiled by the competent national statistics authorities. Data are collected and compiled in line with the Council Directive 95/57/EC on tourism statistics, and Regulation (EU) 692/2011 of the European Parliament and of the Council and with the Code of Practice applicable to all processes for collecting and compiling European statistics.

After reception of the data, thorough quality control and validation checks are performed by Eurostat before releasing the data.

12.2. Quality management - assessment

All aspects of the basic principles of quality of European statistics, listed in Regulation 223/2009 on European statistics are evaluated by Eurostat.

- Timeliness and completeness: compliance with transmission deadlines and compliance with the data requirements is evaluated on a regular basis. The timeliness could still be improved by a number of Member States while the completeness is generally very good.
- Punctuality: the tool EDIT/EBB is available for the validation of micro-data. This reduces the number of exchanges between Eurostat and Member States and consequently improve the punctuality of data disseminated on Eurostat public database.
- Accessibility and clarity: with the implementation of regulation 692/2011 new breakdown are available and disseminated on Eurostat web site(since reference period 2012). Break in series, when observed, are flagged.
- Accuracy: evolution are systematically verified. When huge variations are observed, the reporting country is asked to confirm them.
- Coherence and comparability: annual data are compared with monthly data. The comparison is generally good.

Statistical Data and Metadata Excange

Metadati

Eurostat ha predisposto 21 sezioni nella descrizione del metadato per i concetti trasversali (Cross-domain) maggiormente utilizzati.

1. Contact	8. Release policy	15. Timeliness and punctuality
2. Metadata update	9. Frequency of dissemination	16. Comparability
3. Statistical presentation	10. Dissemination format	17. Coherence
4. Unit of measure	11. Accessibility of documentation	18. Cost and burden
5. Reference period	12. Quality management	19. Data revision
6. Institutional mandate	13. Relevance	20. Statistical processing
7. Confidentiality	14. Accuracy and reliability	21 Comment

Statistical Data and Metadata Excange

DSD

DIMENSIONS

Position in Key	CONCEPT					REPRESENTATION				Dimension Type	XS LEVEL
	ID	Name	CONCEPT SCHEME			CODELIST			TEXT FORMAT		
			ID	VER	AGENCY	ID	VER	AGENCY			
1	FREQ	Frequency	TOUR_CONCEPTS	1.0	ESTAT	CL_FREQ	1.0	ESTAT		Frequency	Section
2	COUNTRY	Tourism Country	TOUR_CONCEPTS	1.0	ESTAT	CL_COUNTRY	1.0	ESTAT		Measure	Observation
3	INDIC_TO	Tourism Indicator	TOUR_CONCEPTS	1.0	ESTAT	CL_TOUR_INDICAT	1.0	ESTAT			Section
4	ACTIVITY_TO	Tourism Activity	TOUR_CONCEPTS	1.0	ESTAT	CL_TOUR_ACTIVITY	1.0	ESTAT			Section
TIME	TIME_PERIOD	Time period	TOUR_CONCEPTS	1.0	ESTAT					TIME	Section

MEASURES

TYPE	CONCEPT					REPRESENTATION				MEASURE DIMENSION	CODE
	ID	Name	CONCEPT SCHEME			CODELIST			TEXT FORMAT		
			ID	VER	AGENCY	ID	VER	AGENCY			
Primary	OBS_VALUE	Observation value	TOUR_CONCEPTS	1.0	ESTAT				Decimal	N/A	N/A
Cross-Sectional	ES	Spain	TOUR_CONCEPTS	1.0	ESTAT	CL_COUNTRY	1.0	ESTAT		COUNTRY	ES
Cross-Sectional	FR	France	TOUR_CONCEPTS	1.0	ESTAT	CL_COUNTRY	1.0	ESTAT		COUNTRY	FR
Cross-Sectional	AT	Austria	TOUR_CONCEPTS	1.0	ESTAT	CL_COUNTRY	1.0	ESTAT		COUNTRY	AT
Cross-Sectional	IT	Italy	TOUR_CONCEPTS	1.0	ESTAT	CL_COUNTRY	1.0	ESTAT		COUNTRY	IT

ATTRIBUTES

ATTACH-MENT LEVEL	CONCEPT					REPRESENTATION				ATTRIBUTE TYPE	ASSIGN-MENT STATUS	XS LEVEL
	ID	Name	CONCEPT SCHEME			CODELIST			TEXT FORMAT			
			ID	VER	AGENCY	ID	VER	AGENCY				
Observation	OBS_STATUS	Status of the observation	TOUR_CONCEPTS	1.0	ESTAT	CL_OBS_STATUS	1.0	ESTAT			Conditional	Observation
Series	UNIT	Unit	TOUR_CONCEPTS	1.0	ESTAT	CL_UNIT	1.0	ESTAT			Mandatory	Section
Series	TIME_FORMAT	Time format	TOUR_CONCEPTS	1.0	ESTAT	CL_TIME_FORMAT	1.0	ESTAT		TimeFormat	Mandatory	Section

Statistical Data and Metadata Excange

DSD

```
- <message:KeyFamilies>
  - <KeyFamily id="DCCV_OCCUPATIT1" agencyID="ISTAT">
      <Name xml:lang="en">Employment (thousands)</Name>
      <Name xml:lang="it">Occupati (migliaia)</Name>
    - <Components>
        <Dimension conceptRef="ITTER107" codelist="CL_DCCV_OCCUPATIT1_ITTER107"/>
        <Dimension conceptRef="SEXISTAT1" codelist="CL_DCCV_OCCUPATIT1_SEXISTAT1"/>
        <Dimension conceptRef="ETA1" codelist="CL_DCCV_OCCUPATIT1_ETA1"/>
        <Dimension conceptRef="TITOLO_STUDIO" codelist="CL_DCCV_OCCUPATIT1_TITOLO_STUDIO"/>
        <Dimension conceptRef="CITTADINANZA" codelist="CL_DCCV_OCCUPATIT1_CITTADINANZA"/>
        <Dimension conceptRef="ATECO2002" codelist="CL_DCCV_OCCUPATIT1_ATECO2002"/>
        <Dimension conceptRef="ATECO_2007" codelist="CL_DCCV_OCCUPATIT1_ATECO_2007"/>
        <Dimension conceptRef="POSIZPROF" codelist="CL_DCCV_OCCUPATIT1_POSIZPROF"/>
        <Dimension conceptRef="PROFILO_PROF" codelist="CL_DCCV_OCCUPATIT1_PROFILO_PROF"/>
        <Dimension conceptRef="PROFESSIONE" codelist="CL_DCCV_OCCUPATIT1_PROFESSIONE"/>
        <Dimension conceptRef="PROFESSIONE1" codelist="CL_DCCV_OCCUPATIT1_PROFESSIONE1"/>
        <Dimension conceptRef="REGIME_ORARIO" codelist="CL_DCCV_OCCUPATIT1_REGIME_ORARIO"/>
        <Dimension conceptRef="CARATT_OCC" codelist="CL_DCCV_OCCUPATIT1_CARATT_OCC"/>
        <TimeDimension conceptRef="TIME" codelist="CL_DCCV_OCCUPATIT1_TIME"/>
      - <PrimaryMeasure conceptRef="OBS_VALUE">
          <TextFormat textType="Double"/>
        </PrimaryMeasure>
        <Attribute conceptRef="OBS_STATUS" codelist="CL_DCCV_OCCUPATIT1_OBS_STATUS" attachmentLevel="Observation" assignmentStatus="Conditional"/>
        <Attribute conceptRef="TIME_FORMAT" codelist="CL_DCCV_OCCUPATIT1_TIME_FORMAT" attachmentLevel="Series" assignmentStatus="Conditional" isTimeFormat="true"/>
      </Components>
  </KeyFamily>
```

Statistical Data and Metadata Excange

MSD

Organisation Role

The function or activities of an organisation, in statistical processes such as collection, processing and dissemination

Source
Statistical Data and Meta
UNSD - Metadata Commo

Context

Hyperlink
http://www.sdmx.org/

Related terms
Data collection
Data Consumer
Dissemination
Organisation

Category

Generic term for items at any level within a classification, typically tabulation categories, sections, subsections, divisions, subdivisions, groups, subgroups, classes and subclasses (UN Glossary Classification Terms):

Source
United Nations Glossary of Classification Terms: prepared by the Expert Group on International Economic and Social Classi

Context
Classification categories a
provide both a unique iden
They contain elements whi
activities, products, types of

Hyperlink
http://unstats.un.org/unsd/cl

Related terms
Category Scheme
Structure

Attachment level

A property of attributes in Gesmes/TS.

Source
GESMES/TS User Guide, Release 3.00, February, 2003; unpublished on paper

Context
For each attribute specified in a key family, it is defined whether this attribute takes:
- an independent value for each observation in the data set
- an independent value for each time series in the data set
- an independent value for each sibling group in the data set
- a single value for the entire data set.

Hyperlink
http://www.sdmx.org/Data/GesmesTS_rel3.pdf

Related terms
Attribute [Gesmes terminology]
GESMES/TS

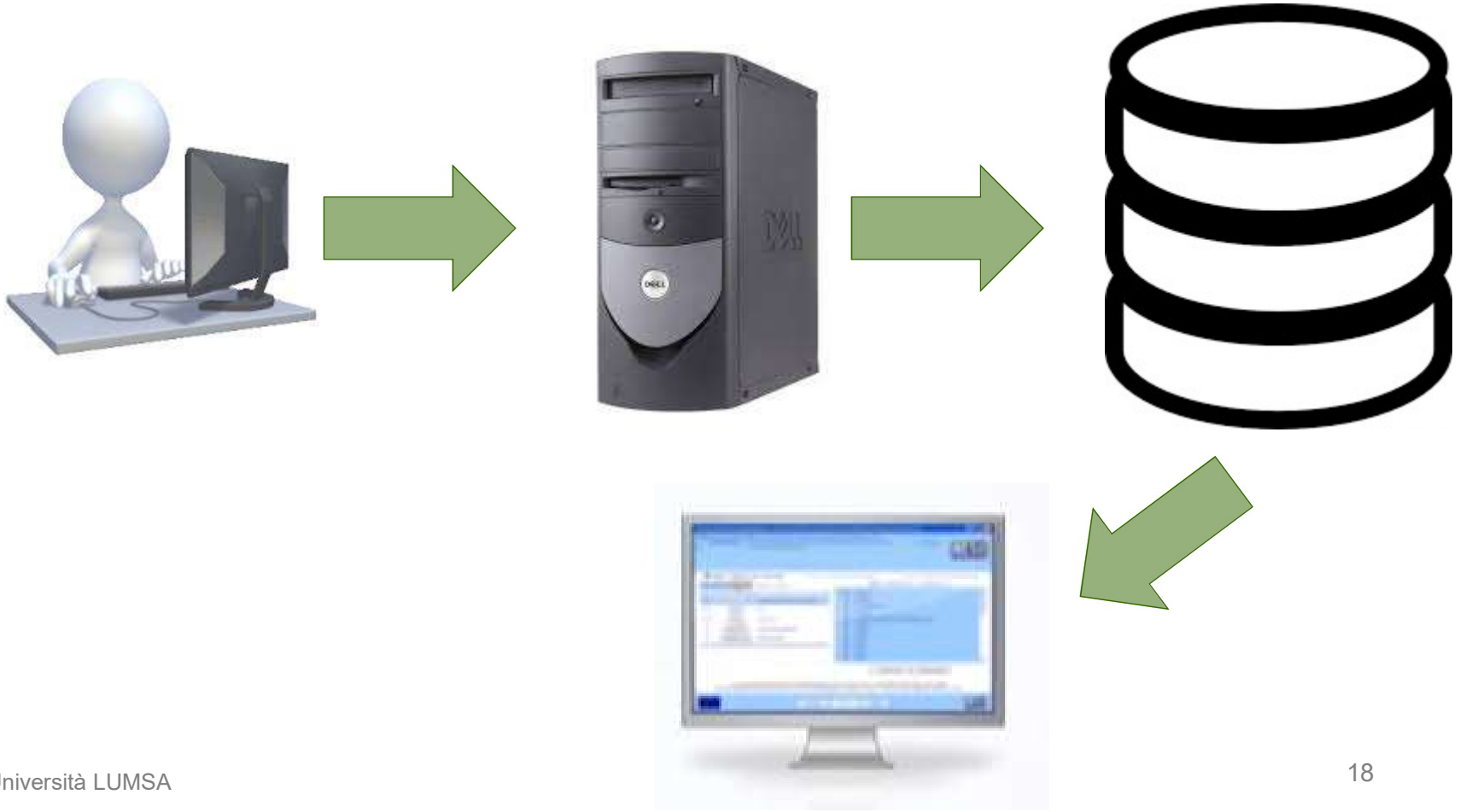
Statistical Data and Metadata Excange

Statistical Subject-Matter Domains

Domain 1: Demographic and social statistics		Domain 2: Economic statistics		Domain 3: Environment and multi-domain statistics	
1.1	Population and migration	2.1	Macroeconomic statistics	3.1	Environment
1.2	Labour	2.2	Economic accounts	3.2	Regional and small area statistics
1.3	Education	2.3	Business statistics	3.3	Multi-domain statistics and indicators
1.4	Health	2.4	Sectoral statistics	3.3.1	Living conditions, poverty and cross-cutting social issues
1.5	Income and consumption	2.4.1	Agriculture, forestry, fisheries	3.3.2	Gender and special population groups
1.6	Social protection	2.4.2	Energy	3.3.3	Information society
1.7	Human settlements and housing	2.4.3	Mining, manufacturing, construction	3.3.4	Globalisation
1.8	Justice and crime	2.4.4	Transport	3.3.5	Indicators related to the Millennium Development Goals
1.9	Culture	2.4.5	Tourism	3.3.6	Sustainable development
1.10	Political and other community activities	2.4.6	Banking, insurance, financial statistics	3.3.7	Entrepreneurship
1.11	Time use	2.5	Government finance, fiscal and public sector statistics	3.4	Yearbooks and similar compendia
		2.6	International trade and balance of payments		
		2.7	Prices		
		2.8	Labour cost		
		2.9	Science, technology and innovation		

Statistical Data and Metadata Excange

Push mode



Push mode

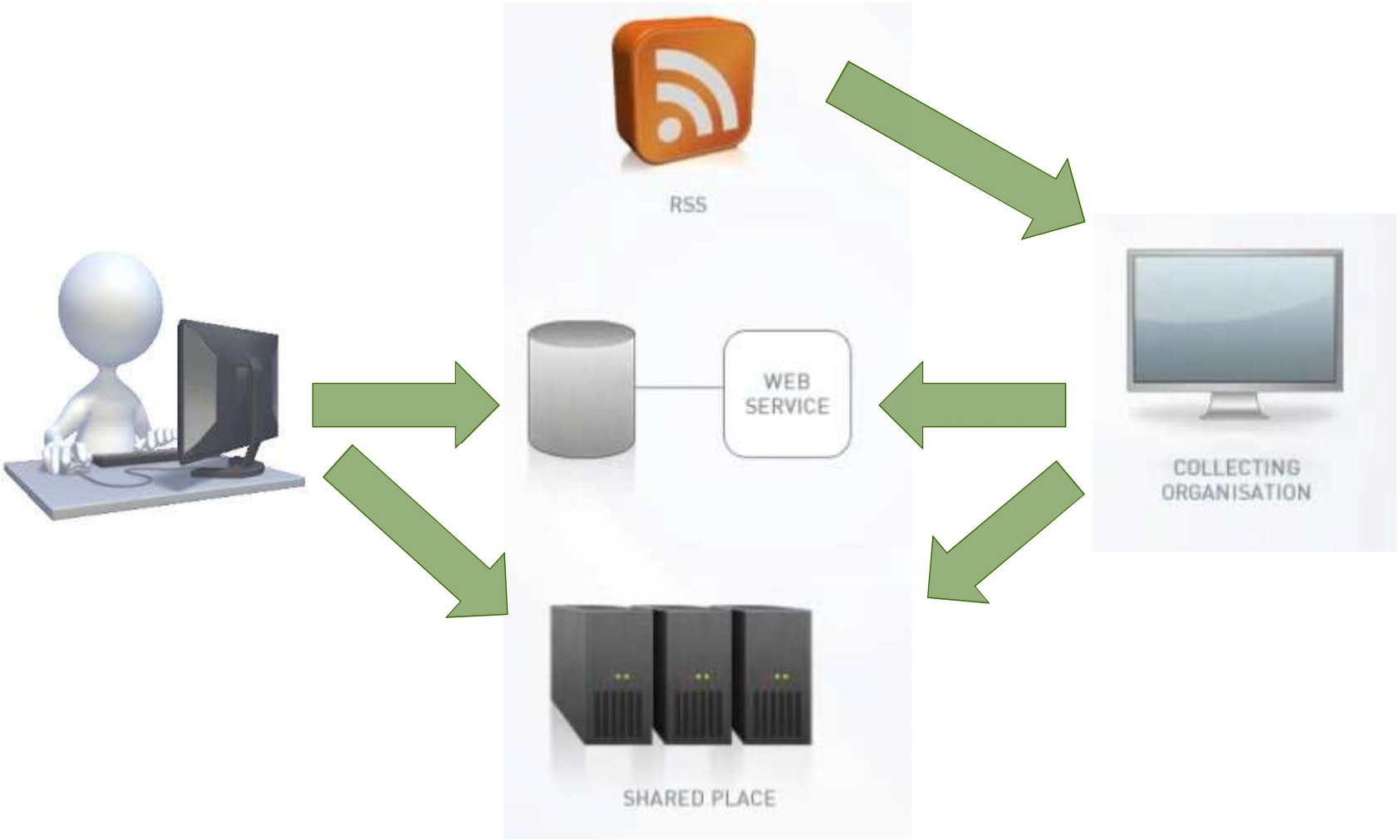
La modalità push prevede che i dati vengono trasmessi da un'organizzazione all'altra. Questa modalità comporta che il fornitore di dati interviene per inviare i dati all'organizzazione che raccoglie i dati. Ciò può avvenire utilizzando diversi mezzi, come l'e-mail o il trasferimento di file.

Il fornitore di dati inserisce i dati in un repository o su un server o, anche, li trasmette tramite email.

Queste sono state le modalità tradizionali di raccolta dei dati, eseguite da organizzazioni internazionali per molti anni.

Statistical Data and Metadata Excange

Pull mode



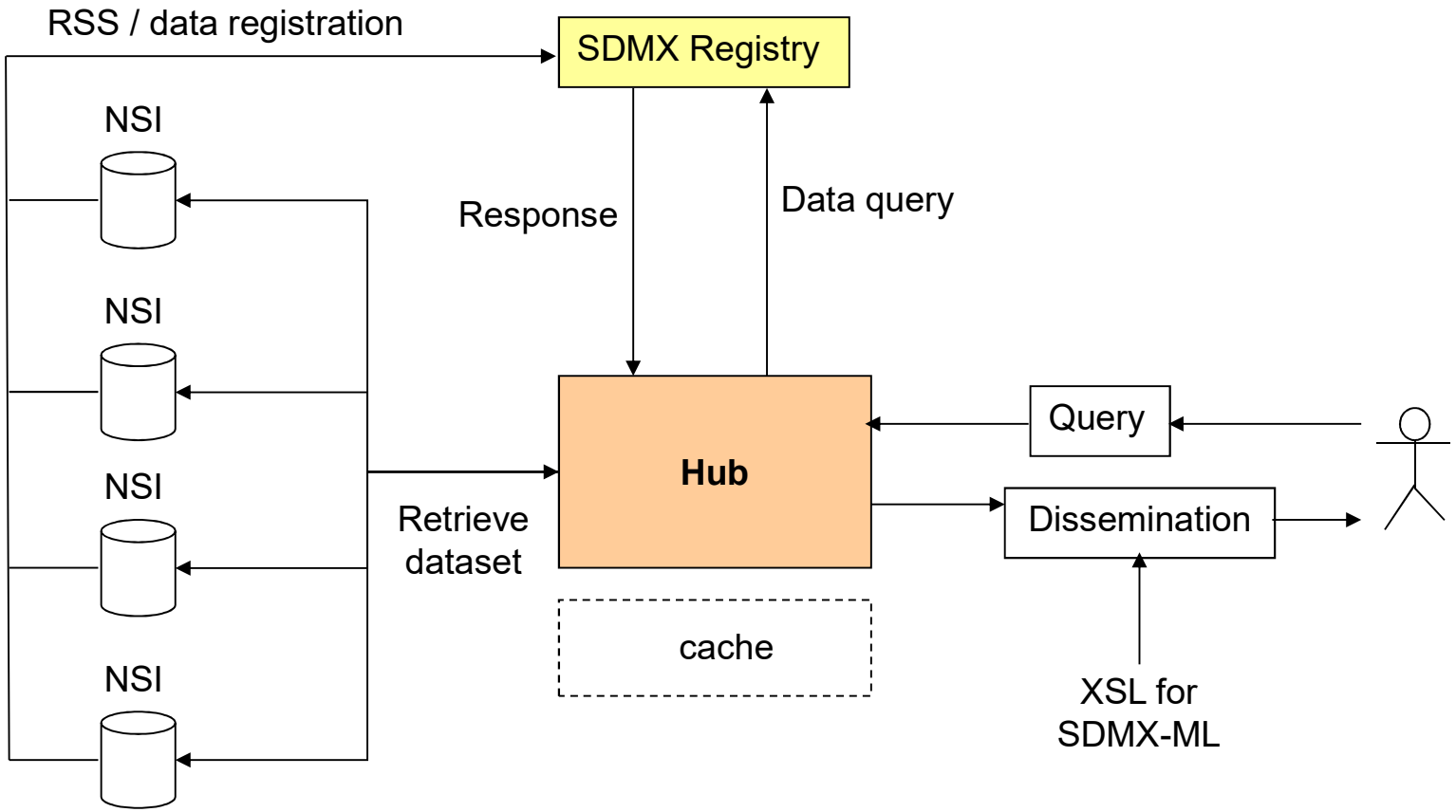
Statistical Data and Metadata Excange

Pull mode

Attraverso il protocollo SDMX e le sue implementazioni pratiche è possibile andare a prelevare i dati che un certo “nodo” del sistema-comunità espone su Web non appena siano resi disponibili. Questa modalità, che appunto viene chiamata pull, è indicata, a ragione, come un significativo passo in avanti nel favorire l’accesso e la condivisione di dati rispetto alla tecnologia push. I dati non vengono più spinti dal nodo che li detiene verso il nodo centrale, né parcheggiati in un repository comune per essere successivamente estratti, ma vengono tirati via dagli altri nodi interessati, avvertiti da un apposito messaggio RSS.

Altro vantaggio di tale approccio è proprio il passaggio dalla logica del deposito a quella del registro. Non più «pozzi» di dati che vengono facilmente alimentati ma a fatica armonizzati e integrati, ma ambienti leggeri di mappatura che lasciano i dati dove stanno naturalmente e si occupano solo di interpretarne formati, struttura e significato. Attraverso registry (e non più repository).

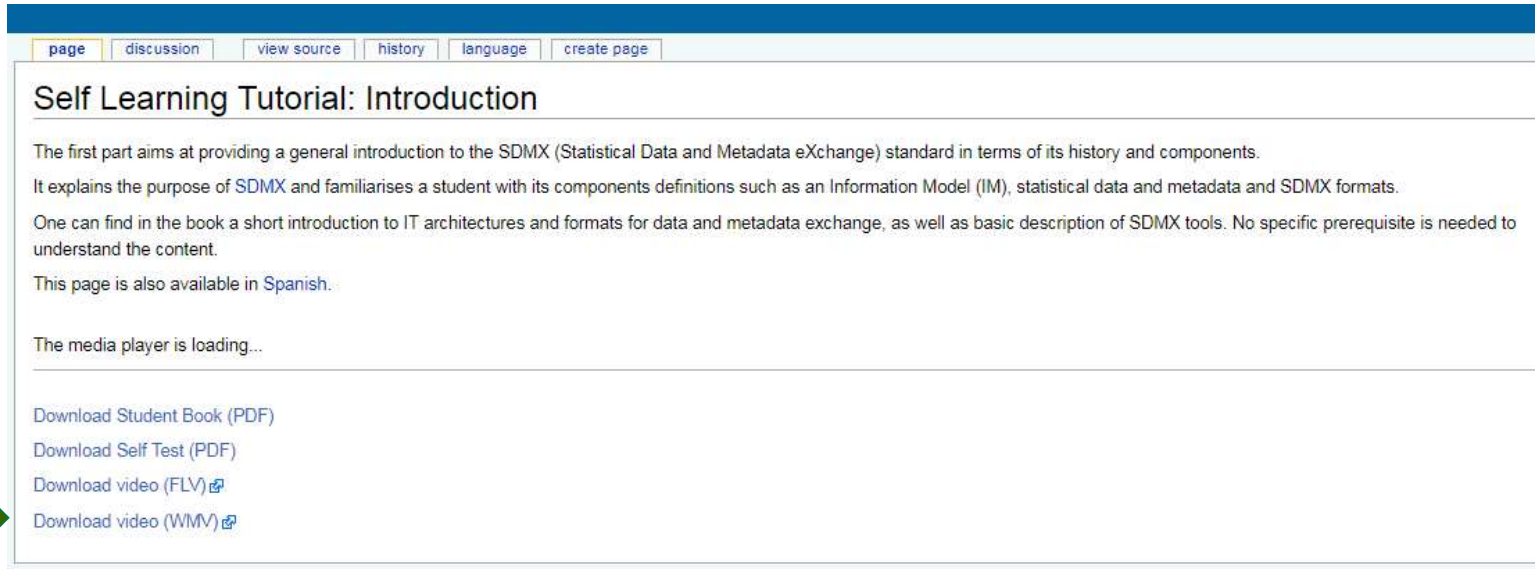
SDMX data hub architecture



SDMX data hub architecture

Per avere maggiori informazioni su SDMX si consiglia di usare lo Student Book che potete scaricare in PDF al seguente link:

https://webgate.ec.europa.eu/fpfis/mwikis/sdmx/index.php/Self_Learning_Tutorial:_Introduction



Allo stesso link troverete un interessante video tutorial che consiglio di visionare.