

Select Language ▼

Powered by  Google Translate

Data Integration / Data Integration Home / Data Integration Inventory

# Data Integration Survey

Created by Jenine Borowik, last modified on 01 Nov, 2017

## 1 Your Information

**Organisation****Country (if applicable)****Number of Employees**

- 1-99
- 100 - 499
- 500 - 1499
- 1500 or above

**Use of the information provided**

- The information provided can be shared publicly as part of the guide
- Don't share this information publicly, but only among colleagues in official statistics
- Only use this information in aggregates or as anonymous information

Unless otherwise indicated, information may be shared publicly as part of the guide.

**Contact Person****Email address**

## 2 Strategies

**If your organisation has a definition of data integration, please provide it here and/or provide a link****Do you have, or are you developing, organisation wide, national or international strategies for data integration?**

- Organisation wide strategy in place
- Organisation wide strategy under development
- National/whole of government strategy in place
- National/whole of government strategy under development
- Multi-country/international strategy in place
- Multi-country/international strategy under development
- No

**If so, please describe and/or provide links****Some official statistics organisations have a leading role in the development of national/whole of government/international data integration practices. Can you identify the main advantages of official statistics organisations taking this role?****Can you identify challenges of official statistical organisations taking this role?**

**Do you have specific units/functions in your organisation responsible for data integration strategies and/or operations?**

- Data Integration strategy unit
- Data Integration operations unit
- Data integration governance committee
- Management/support of whole of government data integration activity
- Management/support of international data integration activity
- Other
- No

**If other, please describe and/or provide links**

**Why were the units/functions created?**

### 3 Barriers

**Please rate the following issues as barriers to successfully using data integration for the ongoing production of statistics**

**Public acceptance and trust issues**

- Significant barrier
- Moderate barrier
- Slight barrier
- Not a barrier
- No opinion

**Lack of supporting legislation or legislation that blocks data integration**

- Significant barrier
- Moderate barrier
- Slight barrier
- Not a barrier
- No opinion

**Access to new data sources**

- Significant barrier
- Moderate barrier
- Slight barrier
- Not a barrier
- No opinion

**Maintaining access to data sources (e.g. when data provider changes availability or format)**

- Significant barrier
- Moderate barrier
- Slight barrier
- Not a barrier
- No opinion

**Quality issues**

- Significant barrier
- Moderate barrier
- Slight barrier
- Not a barrier
- No opinion

**ICT issues**

- Significant barrier
- Moderate barrier
- Slight barrier
- Not a barrier
- No opinion

**Lack of methodologies**

- Significant barrier
- Moderate barrier

- Lack of meta information describing definitions, statistical units, etc used in data sources**
- Slight barrier
  - Not a barrier
  - No opinion
- Differing definitions**
- Significant barrier
  - Moderate barrier
  - Slight barrier
  - Not a barrier
  - No opinion

- Budget/resources**
- Significant barrier
  - Moderate barrier
  - Slight barrier
  - Not a barrier
  - No opinion

- Skills**
- Significant barrier
  - Moderate barrier
  - Slight barrier
  - Not a barrier
  - No opinion

**Please describe any other barriers not mentioned above and/or provide links**

**4 Legislative Environment**

- Is your organisation allowed by law (e.g. stated in the Statistical Act or similar) to use administrative data for statistical purposes?**
- Yes, in all cases
  - Yes, in most cases
  - Yes, in some cases
  - No
  - Don't know

Reference "Using Administrative and Secondary Sources for Official Statistics: A Handbook of Principles and Practices"  
<https://unstats.un.org/unsd/EconStatKB/KnowledgebaseArticle10349.aspx>

- Does a legal basis to access data exist?**
- Yes, in all cases
  - Yes, in most cases
  - Yes, in some cases
  - No
  - Don't know

- Is data free to access for statistical purposes?**
- Yes, in all cases
  - Yes, in most cases
  - Yes, in some cases
  - No
  - Don't know

- Is your organisation prohibited from access or linking of data due to privacy issues (e.g. stated in a Personal Data Protection Act or similar)?**
- Yes, in all cases
  - Yes, in most cases
  - Yes, in some cases
  - No
  - Don't know

**Please describe any legislative limitations or barriers not**

mentioned above

Please list any legislative supports not mentioned above

### 5 Identifiers

Is there a unified personal identity system in the country?

- Yes
- Yes, partly (e.g. not comprehensive)
- No
- Not relevant

Is there a unified business identity system in the country?

- Yes
- Yes, partly (e.g. not comprehensive)
- No
- Not relevant

Is there a unified farmers identity system in the country?

- Yes
- Yes, partly (e.g. not comprehensive)
- No
- Not relevant

Is there a unified address system in the country?

- Yes
- Yes, partly (e.g. not comprehensive)
- No
- Not relevant

### 6 Institutional Environment

What pre-integration practices does your organisation use or plan to use with data providers or other partners?

- Cooperation agreements for transferring the data
- Collaboration in the preparation of legal documents establishing and/or maintaining use of data
- Collaboration for determining coverage, concepts and/or definitions in the data
- Collaboration in the preparation of administrative or statistical classifications used for data
- Long-term partnerships (formal or informal) which consist of two or more institutions using the same data
- Other
- None

If other, please describe and/or provide links

### 7 Statistical Domains and Data Integration

Please advise any data integration activity in the statistical domains listed below (if known)

Agriculture, forestry, fisheries (2.4.1)

Banking, insurance, financial statistics (2.4.6)

Business statistics (2.3)

Culture (1.9)

<b>Economic accounts (2.2)</b>	<input type="text" value="▼"/>
<b>Education (1.3)</b>	<input type="text" value="▼"/>
<b>Energy (2.4.2)</b>	<input type="text" value="▼"/>
<b>Entrepreneurship (3.3.7)</b>	<input type="text" value="▼"/>
<b>Environment (3.1)</b>	<input type="text" value="▼"/>
<b>Gender and special population groups (3.3.2)</b>	<input type="text" value="▼"/>
<b>Globalisation (3.3.4)</b>	<input type="text" value="▼"/>
<b>Government finance, fiscal and public-sector statistics (2.5)</b>	<input type="text" value="▼"/>
<b>Health (1.4)</b>	<input type="text" value="▼"/>
<b>Human settlements and housing (1.7)</b>	<input type="text" value="▼"/>
<b>Income and consumption (1.5)</b>	<input type="text" value="▼"/>
<b>Indicators related to the Millennium or Sustainable Development Goals (3.3.5)</b>	<input type="text" value="▼"/>
<b>Information society (3.3.3)</b>	<input type="text" value="▼"/>
<b>International trade and balance of payments (2.6)</b>	<input type="text" value="▼"/>
<b>Justice and crime (1.8)</b>	<input type="text" value="▼"/>
<b>Labour (1.2)</b>	<input type="text" value="▼"/>
<b>Labour cost (2.8)</b>	<input type="text" value="▼"/>
<b>Living conditions, poverty and cross cutting issues (3.3.1)</b>	<input type="text" value="▼"/>
<b>Macroeconomic statistics (2.1)</b>	<input type="text" value="▼"/>
<b>Mining, manufacturing, construction (2.4.3)</b>	<input type="text" value="▼"/>
<b>Political and other community activities (1.10)</b>	<input type="text" value="▼"/>
<b>Population and migration (1.1)</b>	<input type="text" value="▼"/>
<b>Prices (2.7)</b>	<input type="text" value="▼"/>
<b>Regional and small area statistics (3.2)</b>	<input type="text" value="▼"/>
<b>Science, technology and innovation (2.9)</b>	<input type="text" value="▼"/>
<b>Social protection (1.6)</b>	<input type="text" value="▼"/>
<b>Sustainable development (3.3.6)</b>	<input type="text" value="▼"/>

Time use (1.11)

Tourism (2.4.5)

Transport (2.4.4)

Yearbooks and similar compendia (3.4)

Other

If other, please describe

Please indicate the most important or prominent ways you use integration of additional information or alternate data sources in your organisation

- As a source for sample frames
- To supplement surveys (e.g. for a part of population, for a set of variables)
- To supplement traditional censuses (e.g. register based population census, agricultural census)
- For replacing sample surveys
- For replacing traditional censuses (e.g. register based population censuses, agricultural census)
- For maintaining registers
- For data validation
- For data editing and/or imputation
- For estimation (e.g. small area estimation)
- To provide geospatial products
- For customised data services
- To create statistical products in partnership with other organisations
- To meet the requirements of measuring the Sustainable Development Goals (SDGs)
- Other

If other, please describe and/or provide links

Please indicate the types of data being integrated with other data for the production of statistics in your organisation

- Survey data
- Census data
- Automatic identification systems
- Banking/stock records (2220)
- Blogs and comments (1200)
- Car/vehicle location (3122)
- Commercial transactions (2210)
- Credit cards (2240)
- Data from public administration (2100)
- E-commerce (2230) E-Mail (1900)
- Home automation (3111)
- Internet searches (1600)
- Logs (3110)
- Medical records (2110)
- Mobile data content: text messages (1700)
- Mobile phone location (3121)
- Mobile phone: call/text times and positions (312.)
- Personal documents (1300)
- Pictures: Instagram, Flickr, Picasa etc. (1400)
- Satellite images (3123)
- Scientific sensors (3114)
- Security/surveillance videos/images (3115)
- Smart Energy Meters (311?)/Smart gas meters
- Social Networks: Facebook, Twitter, Tumblr etc. (1100)
- Traffic sensors/webcam (3113)
- User-generated maps (1800)
- Videos: Youtube etc. (1500)

- Weather/pollution sensors (3112)
- Web logs (3220)

Source Classification of types of big data

(<https://statswiki.unece.org/display/bigdata/Classification+of+Types+of+Big+Data>) plus additional items

If other, please describe and/or provide links

## 8 Information Technology

Which tools (applications, software, etc.) do you use for linking and/or matching data?

- R
- SAS
- SPSS
- SQL
- Oracle
- G-Link
- FEBRL
- RELAIS
- CDA
- ATINE
- LinkageWiz
- LinkSolv
- Link Plus
- The Link King
- ChoiceMaker 2
- Other commercial tools
- Other in-house developed tools
- Other outsourced services
- Other open source tools

Please provide additional details and/or links about the tools you use

## 9 Methods

What methods do you use?

- Deterministic record linkage (links based on individual identifiers that match among the available data sets)
- Probabilistic record linkage (linking two pieces of information together using multiple, possibly non-unique, keys)
- Other

Please provide additional details and/or links about the methods you use

## 10 Quality

How do you measure the quality of statistical information produced as a result of data integration activities?

- As integral part of quality reporting
- On an ad-hoc basis
- Do not measure the quality of the integrated dataset

Do you use a quality framework for data integration activities?

- Yes
- No quality framework is used but there is case by case consideration of quality issues
- No

If you use a quality framework, please describe and/or provide links

## 11 Skills

Does your organisation have the required skilled resources, or access to resources, to undertake data integration activities?

- Yes  
 Partly  
 No

Has your organisation developed or sourced specific training or other forms of skills development for data integration activities in the last 3 to 5 years?

- Yes  
 Partly  
 No

Would you be interested in obtaining or providing data integration training with other official statistics organisations?

- Yes - obtaining training  
 Yes - providing training  
 No

Please indicate the areas where you think training is most required

- Strategies for success  
 Developing effective partnerships  
 Governance  
 Legislation issues  
 Pre-integration practices  
 Tools / Information Technology  
 Methods  
 Quality frameworks  
 Other

Please provide any relevant information and/or links

## 12 Integrating Geospatial and Statistical Data

Does your organisation use geospatial data in the production of statistics?

- Yes  
 No  
 Don't know

If yes, what type of geospatial data is used?

- Points  
 Polygons  
 Lines  
 Don't know

Does the geospatial data meet statistical needs in terms of

- Resolution  
 Scale  
 Quality  
 Accuracy  
 Update processes  
 Don't know

Please indicate any registers you use which have geospatial attributes / data items

- Person  
 Address  
 Building  
 Dwelling  
 Business  
 Cadastral parcels  
 Statistical units



**To what level are address database(s) geocoded?**

- Other
- Don't know
- Nomenclature of Territorial Units for Statistics (NUTS)
- National administrative division
- Statistical units
- Enumeration area / Meshblocks
- Coordinates
- Other
- Don't know

**If other, please describe**

**Is it possible to pair statistical data with external spatial data by identifiers (e.g. personal identifiers, addresses, real estate codes, building ids, names, etc)?**

- Yes
- No
- Don't know

**If yes, who conducts the pairing?**

- Your organisation
- Another organisation with responsibility for official statistics
- A commercial partner
- The data provider

**What are the main barriers to geocoding of geospatial or administrative data**

- No demand for spatial statistics
- No legal support for spatial statistics
- Lack of knowledge
- Geospatial data is available but is too expensive
- Lack of data resources
- Lack of other resources
- Legal or bureaucratic restrictions on availability of geospatial data (e.g. public institutions don't cooperate well)
- Legal or bureaucratic restrictions on availability of administrative data (e.g. public institutions don't cooperate well)
- Administrative data stored in a way that makes geocoding impossible (lack of identifiers to connect to geographical locations)
- Statistical information is not collected in a way that makes geocoding possible or meaningful
- No uniform reference system between different administrative data sources
- Other

**If other, please describe**

**What is the lowest possible geographical level to which you can geocode statistical data?**

- Small geographical areas such as enumeration districts, blocks or small administrative units
- Single points (coordinates) such as address locations, buildings or locations of real estates (cadastral parcels)
- Combination of both (different data in different parts of the country)
- Don't have the ability to geocode information
- Other

**What are the main threats to your organisation's current geocoding practices?**

- No threats
- No big problems but there is room for improvement
- Weak internal support - the benefits are contested
- Scarce resources
- Restricted access to geospatial information needed for geocoding
- Inconsistencies in geospatial information needed for geocoding
- Poor cooperation and coordination between organisations responsible for different geospatial information and administrative data
- Inconsistencies in statistical information or administrative data resulting in bad match with geospatial data
- Restricted access to administrative data from other organisations
- Other

**If other, please describe and/or provide links**

### 13 Other information

**Thank you for completing the survey. Please provide any other information, links, comments or suggestions here**



Save

Print

 Form: **DataIntegrationInventory** ▾

*Powered by ConfiForms add-on for Confluence (free license, not for commercial use)*