The role of data centres in facilitating access to data

Aleksandra Bujnowska (aleksandra.bujnowska@ec.europa.eu)[[1]](#footnote-1), Wilhelmus Kloek (wilhelmus.kloek@ec.europa.eu)1

**Keywords:** microdata, data centres, data access,

# Introduction

The abstract aims to provide an overview of different types of data centres and modes of access provided by them. By data centres we mean here the real or virtual place facilitating access to the data. The objective of the data centres is to connect users with the data. We are mostly interested in data centres providing access to data received from multiple sources. More and more data centres constitute a single entry point to data collected by other parties.

# Data centres by functions provided

Taking into account the functions and the data available at the data centres, the following types of data centres can be identified:

* Data discovery centres
* Data exploration centres (including distributed data exploration centres)
* Data computation centres

**Data discovery centres** offer detailed catalogues of the existing data of interest to their users. The users may search the catalogue but can not access the data. The catalogues indicate the reference points where the data can be obtained. Data archives very often play the role of data discovery centres.

**Data exploration centres** provide data catalogues and access to the data. They usually store the data and administer them. These are usually own data, collected by the data centre or data confined to them by other organisations (data providers). In the latter case the data exploration centre must possess adequate rights to grant access to the data. These rights are laid down in special agreements with data providers.

Typically, national statistical institutes (and Eurostat) offer services of data exploration centres. They have their own data, collected by them and they have a legal mandate to receive the data from other parties (usually public bodies).

**Picture 1 Data exploration centre**



A special case is the **distributed data exploration centre**. It offers access to data that remain physically at the premises of the data providers. There is no transfer of data to the data centre.

**Picture 2 Distributed data exploration centre**



**Data computation centres** provide access to statistics computed "on the fly" while the data itself remain at the data providers premises. There is no transfer and no access to the data. Users receive aggregated results. Data computation centres become more and more popular as they fill the gap between users looking for data and massive amount of data collected for various purposes but normally not available for access.

**Picture 3 Data computation centre**



# Data centres by organisation type

Taking into account their organisation, the following types of data centres can be identified:

* Own data centres
* Independent data centres

The organizations collecting the data usually have their **own data centres**. Most of the statistical offices created data centres to grant access to their own data. The data centre is then a unit identified in the organigram.

**Independent data centres** specialize in facilitating access to the data. This is their core business. They excel in data description, curation and provision. In order to work efficiently, the independent data centres must be considered a trusted party. They usually collaborate with several data providers. Independent data centres more often than own centres provide access to linked data.

# Conclusions

Increase of the amount of data available for analysis requires development of data centres that play a major role in facilitating access to (their own or third party) data. The data centres aim at assembling as many data sources as possible and at offering various access modes.

Statistical offices may have their own data centres but may also confine their data to independent data centres that grant access on their behalf. Table 1 provides examples of various types of data centres.

Table 1. Data centres by data types and organisation

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Data discovery | Data exploration | Distributed data exploration | Data computation |
| Own data centres | National Statistical Institutes, Eurostat | NordMAN | Census HUB |
| Independent data centres | INEXDA, CESSDA | CASD.eu, UK Data Service, IPUMS | IDAN | Private companies offering secure computation of other parties' data |

1. Eurostat, Statistical confidentiality and microdata access team [↑](#footnote-ref-1)