

## **Integrating survey design and data quality management**

Daniel Scheuregger  
European Centre for the Development of Vocational Training  
Daniel.Scheuregger@cedefop.europa.eu

### **Data quality issues in applied survey research**

Survey research aims at producing error free and comparable data. However, the reality holds restrictions that hinder the realisation of this ideal. The comparability of data can be limited because different survey projects do not follow a coherent framework and variables that measure the same construct are operationalised in different ways. Other issues may result from changing measurement quality of items in case of repeated surveys, for instance due to changing connotation of words in a language. An effective data quality monitoring helps to recognise these issues and to implement proper countermeasures in future projects.

The implementation of data quality monitoring depends on the way projects are managed and how data quality monitoring is considered in the workflow. Often, survey projects are conducted autonomously with changing staff or contractors and the documentation of projects is likewise implemented individually by storing a jumble of information in different files and formats. This approach leads to various consequences, such as:

1. Survey designs and operationalisations of different studies are incoherent and thus limit the data comparability (as aforementioned).
2. From 1 it follows that knowledge accumulation is excluded, because every survey starts from the scratch.
3. From 2 it follows that process rationalisation is limited. In-house resources are repeatedly stressed, because all work has to be redone, such as operationalisation, conceptual work etc. Thus, also cost savings are exacerbated. For instance, reusing translated items could save translation costs.
4. Tracking data quality over different survey projects or in a longitudinal perspective is practically excluded. Even if we assume that an item is reused in the same way in different projects, it is hardly realistic that after some years the required data will be reanalysed considering the necessary research in the codebooks, questionnaires, files and folders.
5. It is hardly possible to link items, data and survey documentation or to add further information such as meta- or process data from fieldwork if everything is stored in software specific formats (such as word, pdf, statistical software files or excel files).

6. Replicability of results is likewise exacerbated if scripts of statistical analyses are not documented and stored in a coherent system but scattered over subfolders of individual expert's computers.

Altogether it turns out, that a documentation of single projects might enable analysing and retracing data quality in theory. Yet, practically data quality monitoring and accumulating knowledge in the organisation becomes increasingly difficult in practice with the number of surveys.

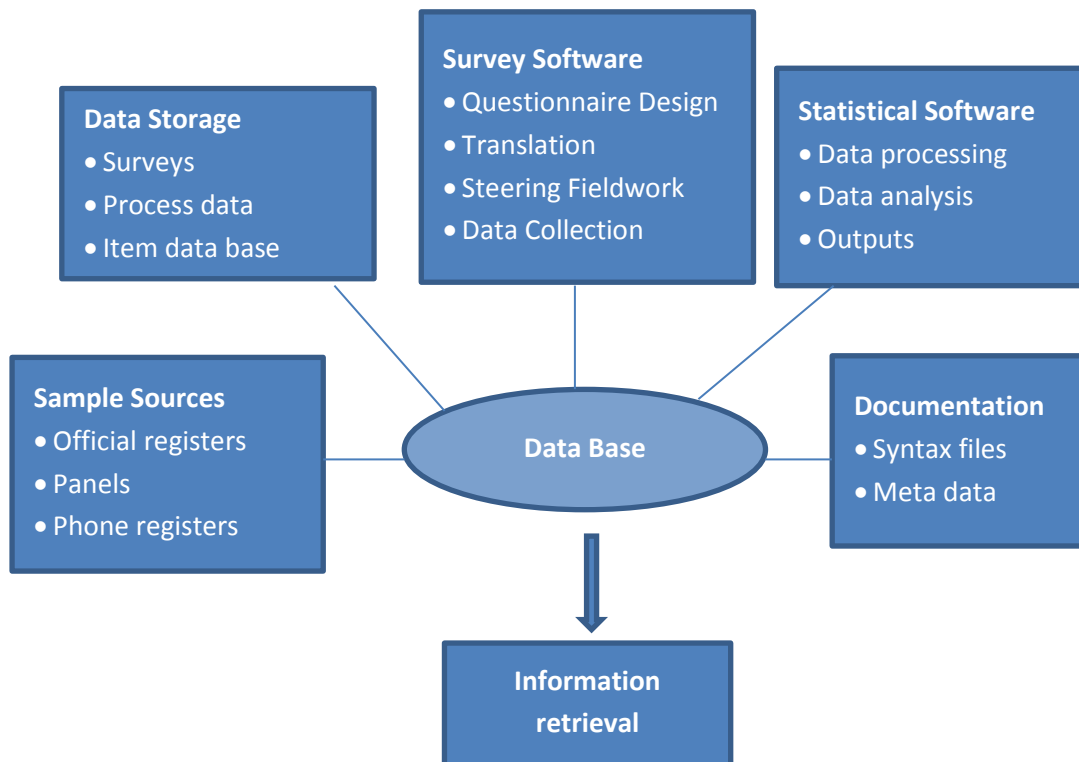
A solution to these problems could be a framework that integrates all work steps of a survey project in the current workflow.

## **The solution**

Such kind integration of workflows can be achieved with relational data base applications that allow adding, merging and storing the information in a defined data scheme. Once the information is stored, a data base only requires knowing the structure of the data schema to assemble the desired information. In case of recurring information requests this can even be automated with SQL scripts.

Likewise, a data base also allows integrating the outputs of different work steps by linking the respective data and applications. Data can be matched with meta-data, statistical software can connect to analyse survey data, surveys can be linked over time or levels and so on. These properties of data bases are particularly interesting when connecting it with survey software that allows questionnaire design (question wording, filters, loops etc.), sampling administration, fieldwork administration and reporting. Survey software offers its functionalities in user friendly frontend menus, yet, basically it is a data base in the backend. Thus, it is possible to integrate the survey software with a data base application, link it with any other information of interest and also import information from the data base into the survey software (such as stored items).

To sum up the following graph exemplifies the relations of information in the system:



### What's the benefit?

Any organisation has particular needs and therefore particular benefits of this integration approach may vary over organisations. Yet, some general gains can be identified:

**First**, the simplification of information integration basically provides the required infrastructure to setup an effective data quality monitoring.

**Second**, this procedure allows process rationalisation, saves resources and improves coherence of projects, because all information can be selected as needed for new projects repeatedly.

**Third**, such kind of system increases the independency and the control of the organisation over the workflow. In principle, the survey software allows to conduct a complete survey in most modes (CAPI, CATI, CAWI, PAPI). It is possible that a survey institute connects to the survey software, so that it can directly conduct the survey as programmed (see applied

examples in the links below<sup>12</sup>). In such a case, the data base system serves as a hub where contractors dock to only for the fieldwork. This allows establishing a standardized work flow in-house that remains identical no matter which external contractor is chosen.

The presentation will demonstrate with a practical example the implementation of this concept using *Limesurvey*<sup>3</sup> (see a demo<sup>4</sup>) as the survey software and *MariaDB*<sup>5</sup> as the relational data base application. Both applications are open source and allow the implementation for an organisation with any budget. Further, both applications provide the flexibility to tailor them to the particular needs of an organisation.

## **Questions**

What kind of problems might appear regarding the practical implementation?

---

<sup>1</sup> <http://www.kai-arzheimer.com/cati-survey-cloud>

<sup>2</sup> <http://www.kai-arzheimer.com/cati-cloud-quexs/>

<sup>3</sup> <https://www.limesurvey.org/>

<sup>4</sup> <https://demo.limesurvey.org/index.php?r=admin/authentication/sa/login>

<sup>5</sup> <https://mariadb.org/>