**Development of a Statistical product portfolio tool**

Rasmus Anker Kristiansen, Statistics Denmark, rkr@dst.dk

**Abstract**

*A regularly asked question is; how many statistical products does Statistics Denmark produce? To answer this question, it is necessary to ask, what is a statistical product? One answer could be that a statistical product is a qualitative measurement of some phenomenon. A less philosophical answer could be that a statistical product, from Statistics Denmark that is, consists of a name, one or more tables in the StatBank, a quality report with documentation, a press release, a page on the website, a person responsible and an administrative placement in the organisation. Statistics Denmark is currently creating a complete product portfolio tool, which provides this overview of all our statistical products.*

*Statistics Denmark has in recent years been working systematically with coherent metadata, including adoption of the Single Integrated Metadata Structure (SIMS). The idea is write once, use everywhere, but this raises the question; use where? The portfolio tool links all information about a given statistical product, in terms of naming, allocation in the subject structure, across dissemination platforms and administrative placement within the organisation. The portfolio tool will retrieve real time information from existing administrative systems such as the metadata repository (Colectica), from different dissemination-, HR- and financial management systems and from the StatBank Application Programming Interface (API).*

*There are multiple benefits and possibilities of such a tool. We get a better overview of our inventory, so it is clearer to ourselves and our users what we actually produce. We can raise credibility when what we say we do in the annual work plan, aligns with what we actually do. We ensure coherent and consistent naming of the individual statistics and we can improve our management accounting procedures with a more precise activity-based costing (ABC) in the production of our statistical products.*

**Keywords:** Product overview, Product portfolio, Coherent Metadata, Business Administration

**1. Introduction**

Imagine walking in to a supermarket looking for diapers, only to find out that diapers are located in the wine section. Meanwhile, the product information sheet can be found and read in the bread section, under the title ‘baby toilets’, which – by the way – describes the content of a carton of milk. Moreover, what about the price? As a consumer, shopping for groceries could take forever. The example is obviously rather extreme, but it does serve a valid point if translated to official statistics.

*1.1. Statistical products in Statistics Denmark*

In essence, a statistic is an estimated value of some phenomenon, obtained through an aggregation of observed data. National Statistical Institutions (NSIs) like Statistics Denmark specialise in producing statistics.

Today, there are 200-and-something statistical products coming out of Statistics Denmark. The exact number is hard to come by, partly for organisational reasons and partly because of a somewhat large variety in the definition of what a statistical product actually is. Some statistical products rely on many data collection sources – like the National Accounts. In other cases, one data collection source can be converted into several types of output – like the agricultural census.

From an output perspective, a statistical product from Statistics Denmark can be identified by a recurring title, contains one or more tables in the StatBank under a given topic, is documented in a quality report, in some cases has its results published in a press release and has its own sub-page on the website collecting all this information. Internationally well-known examples of a statistical product are e.g. the Labour Force Survey (LFS), Purchasing Power Parities (PPP) or the Consumer Price Index (CPI).

Once produced, the statistics are packaged and made available to users through different dissemination channels – mainly on the official website. Statistics must be formatted and loaded into the output database together with relevant metadata. A press release should be prepared with tables, charts and explanatory text etc., and a quality report that documents the whole thing must be written/updated. The final product is then checked to ensure that relevant standards and guidelines are met.

Several processes are in this way handled in different systems, which all contain some information about the given statistical products.

*1.2. Connecting the dots*

The purpose of this project has been to identify all the statistical products that come out of Statistics Denmark, map the associations and connect the information from the different systems and repositories. To obtain any value from this work, these information is shown in a presentable way, in a dashboard-type management information system.

Statistical products exists in a hierarchy. At the top level, is the statistical product identified by a title e.g. *Consumer Price Index (CPI).* At the second level are the quality reports which document and describe the statistical output. There is typically one quality report for every statistical product. However, in some cases it makes sense to have more than one quality report per statistical products. In case of the CPI, the related Net Price Index (NPI), i.e. the CPI without indirect taxes and duties and including price subsidies, could warrant the need for the NPI to have its own quality report for users. At the third level, are the tables in the StatBank containing the actual statistical figures and results. Every table is linked to a quality report where its reference metadata is described in more detail, in accordance to the Single Integrated Metadata Structure (SIMS). Finally every statistical product ‘live’ somewhere within a subject structure. For the CPI that subject is firstly *Prices and Consumption,* secondly, Consumer Prices, and lastly *Consumer Price Index.*

**Figure 1. Statistical product hierarchy**

**Statistical Product**

**Quality Report**

**Quality Report**

**Table**

**Table**

**Table**

**Table**

Subject

First level

Second level

Third level

**2. Statistical product portfolio tool**

The statistical product portfolio tool is an application that connects metadata and administrative information about a given statistical product. To minimize the need for manual maintenance, the information is mainly gathered from existing systems, via the common metadata repository and from an API for the StatBank.

*2.1. Proof of concept*

The Statistical product portfolio tool is basically a pivot table in Excel, based on information loaded from the StatBank API and Colectica overnight. The tool itself shows these information in a dashboard view, which can then be sorted and filtered upon request.

**Figure 2. Statistical product portfolio tool**



Figure 2 shows a snapshot of the dashboard, using the CPI as an example. Each column represents an attribute for the individual statistical product. The aforementioned hierarchy in Figure 1 is also visible in this figure, starting with the first column, which covers all five rows whereas the columns on the right are separate for each row in the spreadsheet. Further development of the tool, interconnecting the statistical products to information from other systems, simply means adding additional columns.

*2.2. The technical bit*

The technical hub are the individual StatBank tables, their metadata, and their place in the subject structure. The tables containing a table-id and structural metadata for those tables, are stored in an application called SUMTOOL. In SUMTOOLS, the placement of the tables in the subject structure are also stored. Furthermore, as a rule, every table must be linked to a quality report in SUMTOOLS. Once a table is complete in SUMTOOLS, the content is loaded into the StatBank, which has an API.

The information, which we cannot get from the StatBank API, we gather from our general metadata system called Colectica, where the quality reports are stored. Similar to the tables, the quality reports also has structural metadata attached. Finally, every quality report is attached to a concept list of all our statistical products (the highest level) in Colectica as shown in Figure 1.

For this project, if we could not retrieve the metadata from Colectica, we simply added it, e.g. like we did with organisational information, information on staff and divisional ownership/placement etc. The logical data model in Figure 3 shows the connections between the repositories where metadata is stored.

**Figure 3. Logical data model**

**Statistical Product Portfolio Tool**

**SUMTOOLS**

**StatBank**

**Colectica**

**Time Management System**

**Other Common Systems**

*2.3. The work ahead*

The Statistical Product Portfolio tool is still under construction and must be further developed and automatized. The tool as-is, is so far (mostly) constructed by connecting the information that we could already gather from Colectica and through the StatBank API. The next steps are to connect the missing links in the logical data model in Figure 3.

Firstly, we need to find a way to link the statistical products to the time registration products in the time management system. A quick review of the statistical products versus the time registration products shows that approximately two thirds of the statistical products, as presented to the users, has a one-to-one relation to a product in the time registration system.

These discrepancies can be as simple as the naming in the time registration system versus the disseminated product, i.e. the disseminated *Consumer Price Index* could carry the name *Price Index for Consumers* in the time registration system*.*

The difference can on the other hand be more complex, like the in section 1.1 aforementioned difference in the way a product is perceived by its input rather than its output. The time registration products in the time management system therefore in its current state makes sense in their own right in their own way.

For the individual, these discrepancies might not be an issue, however collectively this could be a breeding ground for bigger problems. Because, as the organisation becomes more and more functional through standardisation and centralisation, more persons are involved in the production of a statistical product. Therefore, ambiguous naming is undesirable and can eventually create unnecessary misunderstandings.

Secondly, the other missing link is the connection between statistical products and common systems that are developed to be used by statistics across the whole organisation. These systems e.g. count the common systems for reporting, the common data archive and the common handling the actual statistical products to name a few.

The more we work through the silos and outsource part of the statistical production from the statistical domain to a centralised function, the need for unambiguous titles and references to the persons responsible becomes ever more important. I other words, the Consumer Price Index might as well be registered under the same name in all our common systems with a clear indication of personnel responsibility.

**3. The case for a product based approach**

The benefits of getting our inventory of statistical products in line are that it becomes clearer to ourselves, and more importantly, to our users what we actually produce. Much of this can be solved by a higher level of consistent naming and coherent metadata.

However, the most crucial currency of an NSI is trust – trust in that what we say we do aligns with what we actually deliver and that what we actually deliver can be trusted. So why not figure out how much these individual statistical products costs to produce in this process?

*3.1. Activity-based costing (ABC)*

Activity-based costing (ABC) is an accounting method that identifies activities that an organisation performs and assigns the cost to each product.[[1]](#footnote-1) In a field of limited resources, these resources (input) can be allocated more directly to the final statistical products (output) and further, to users.

If, and when, we succeed in fully bringing together the statistical products with the products in the time registration system, this tool can provide management information that enhance decision-making and eventually merit the possibility to settle the price of producing the individual statistics.

*3.2. ESS Code of Practice*

Finally, this project of course, also has to do with fulfilling the principles in the Europeans Statistics Code of Practice.[[2]](#footnote-2) More specifically Principle 10: Cost effectiveness and Principle 15: Accessibility and Clarity.

Principle 10 states that; *resources are used effectively*. Indicator 10.1 states that; *internal and independent external measures monitor the statistical authority’s use of resources.* This tool will provide this information for decision makers.

Principle 15 states that; *statistics are presented in a clear and understandable form, released in a suitable and convenient manner, available and accessible on an impartial basis with supporting metadata and guidance.* This tool enables coherence across dissemination platforms and eventually, makes the trip to the statistical supermarket less of a headache.

**4. References**

Robert S. Kaplan at: <http://kfknowledgebank.kaplan.co.uk>

The European Statistics Code of Practice at: <http://ec.europa.eu/eurostat/>

1. [Kaplan](http://kfknowledgebank.kaplan.co.uk/KFKB/Wiki%20Pages/Activity%20Based%20Costing%20%28ABC%29.aspx) [↑](#footnote-ref-1)
2. [European Statistics Code of Practice](http://ec.europa.eu/eurostat/web/quality/european-statistics-code-of-practice) [↑](#footnote-ref-2)