Building a new topic taxonomy, which has user needs at its core

Mette-Line Jakobsen, Statistics Denmark, mjk@dst.dk

Abstract

Statistics Denmark uses a three-level topic taxonomy in our dissemination of statistical products – e.g. in our database Statbank Denmark, at our website dst.dk, in newsletters and on other communication platforms. The taxonomy is a key tool for our users to find relevant content at Statistics Denmark.

However, inquiries from users and a number of user surveys have shown that a large number of users had difficulties navigating the taxonomy. To address this issue, the Communication division in Statistics Denmark recently carried out a project developing a new and more intuitive taxonomy. The new taxonomy was launched mid-September 2021.

This paper addresses and evaluates the methods used in the development of the new taxonomy. The methods include an expert review, mapping the practices at other statistical institutes, workshops with the participation of colleagues from the statistics-producing divisions in Statistics Denmark and, last but not at all least, several user tests.

The most fundamental change carried out by the implementation of the new taxonomy is a shift from a more traditional approach reflecting the organisation of the statistical production to an approach mainly based on the users' perspective. The new taxonomy has been rolled out on all our dissemination platforms, as it is a vital principle that the taxonomy is identical and familiar to our users across all platforms.

Based on these principles, the paper takes the discussion further and considers the possibility of creating a uniform topic taxonomy that can be implemented across all national statistical institutes, and elaborates whether this is a suitable option or not.

Keywords: topic taxonomy, user perspective

1. Introduction

Since the publication of the first Statistical Yearbook in Denmark in 1896, we have organised statistical domains into groups with related content to help users find the statistics they are looking for. This early form of a topic taxonomy was quite simple and without subgroups.

In line with the development of the statistical production in both quantity and complexity, the grouping of the statistics for dissemination purposes has changed. The changes were linguistic, substantive and structural bringing subgroups into use, and always for the purpose of making it as easy for the users as possible.

With the invention of the internet and the distribution of free-of-charge databases, access to a large amount of statistical data became easier for everyone – and more complicated. It became more complicated because the very quantity that became available made it more difficult to find what you were looking for. Fortunately, the ever-improving search functions on websites and in databases can help you a long way, but they too have their limits. Today the topic taxonomy is still a key tool for our users to find relevant content with Statistics Denmark as it also helps users to gain an overall view of a statistical domain.

All national statistical institutions (NSIs) struggle to structure their large content of statistics in the best possible way for their users and most NSIs use some form of topic taxonomy in their dissemination. However, a topic taxonomy is not without drawbacks and the design of a taxonomy can cause different advantages and disadvantages.

This paper describes and evaluates the methods and considerations we have applied in the development of a new and more intuitive topic taxonomy at Statistics Denmark. Furthermore, the possibilities of a uniform taxonomy that can be implemented across all national statistical institutes are discussed.

2. Background

The latest topic taxonomy used in Statistics Denmark was established in 2011. At that time, attemps were also made to integrate the user perspective into the taxonomy. Nevertheless, the taxonomy very much came to reflect the organisation of the statistical production and for many users, it never really seemed intuitive.
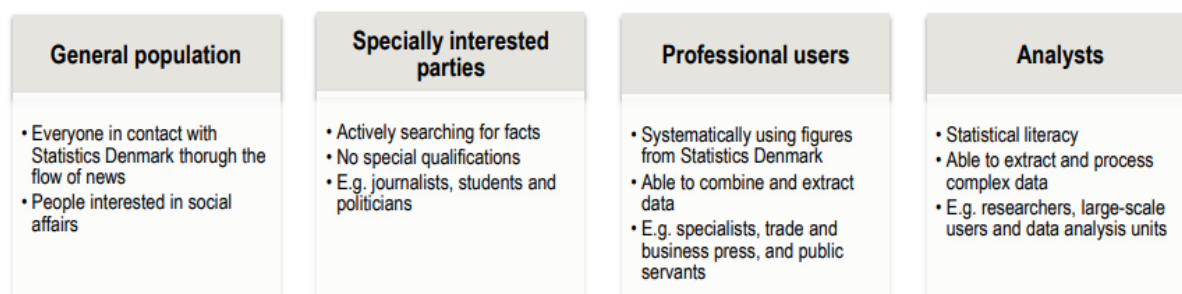
Over the years, several user surveys have shown that a large number of users had difficulties navigating the topic taxonomy. This was also reflected in direct inquiries from users who had difficulties finding the relevant content. Our users found the existing taxonomy very complex and difficult to use.

In 2017, Statistics Denmark published Strategy 2022, which included a number of formulated initiatives to strengthen Statistics Denmark's digital publishing and at the same time support an even stronger user perspective in our dissemination. One of these initiatives was the development of a new and more intuitive topic taxonomy and this project has been carried out by the Communication division in Statistics Denmark. The new taxonomy was launched mid-September 2021.

User profiles

Statistics Denmark provides data and knowledge to support decisions, debate and research at all levels of society and we strive to reach all types of users throughout the population. Our users have different needs and are not equally qualified to locate and interpret the statistics. Our communication must be adapted to these different needs, and to this end, we have defined four user profiles to use as a basis for communication, which we match with the users' different backgrounds. See figure 1 for a characteristic of the four user profiles.

Fig. 1. Four defined user profiles

General population	**Specially interested parties**	**Professional users**	**Analysts**
• Everyone in contact with Statistics Denmark thorugh the flow of news • People interested in social affairs	• Actively searching for facts • No special qualifications • E.g. journalists, students and politicians	• Systematically using figures from Statistics Denmark • Able to combine and extract data • E.g. specialists, trade and business press, and public servants	• Statistical literacy • Able to extract and process complex data • E.g. researchers, large-scale users and data analysis units

Communication and dissemination strategy 2018-2022 (2018), p. 10

3. Methods

Pilot study

The project of designing a new topic taxonomy was initiated with a pilot study, the primary purpose of which was to identify inadequacies in the existing topic taxonomy. The pilot included

- Internal mapping of experiences of colleagues in Information Service and in the statistics-producing divisions, results from previous user surveys and also user data from the search history of our website and database.

- External mapping of the use of taxonomies at several other statistical institutions, a card sorting test with a number of users and in addition, an expert review on the existing taxonomy with a focus on structure and terminology.

In connection with the mapping of practice at other statistical institutions, we were in contact with colleagues from abroad learning from their considerations and experiences in developing new topic taxonomies.

Based on the results from the pilot study, four basic principles for the new topic taxonomy were formulated:

1. Balance in width and depth
2. Same topic taxonomy on all dissemination platforms
3. Logical grouping of topics – following the users' logic and scheme of things
4. Intuitively understandable naming of topics

Fundamental in the further work with the project was also a user-oriented development process, where users from Statistics Denmark's four defined user profiles were actively involved. In addition, we wanted to involve our colleagues in the statistics-producing divisions, as we found it important that the new topic taxonomy was anchored throughout the institution to the extent possible.

Preparing and working out a new taxonomy

The project has included a high degree of user involvement to meet the need for a more user-oriented topic taxonomy. To a large extent, it has been an iterative process and has consisted of online card sorting tests, workshops with colleagues representing different statistical divisions and several user tests.

The results from the online card-sorting tests validated the results from the card-sorting test in the pilot study. The respondents predominantly distributed the content within five to nine overall topics. As to naming, there was particular consensus on the terms 'Culture', 'Economy', 'Education', 'Environment' and 'Transport'. In addition, the respondents suggested a grouping of 'Home environment', 'Personal', 'Family' or

'Domestic sphere' as well as a grouping of business-related issues. These results are also in line with the results of the pilot study.

The workshops with colleagues had several purposes. We wanted to gain insight into the needs and wishes of the statistical divisions and also increase our understanding of the content and coverage of the statistics – e.g. whether a particular set of statistics would fit into the proposed topic. Furthermore, we wanted to collect and discuss ideas with our statistics-producing colleagues and thus anchor the development of the new topic taxonomy in the institution. Therefore, colleagues responsible for statistics were continuously involved in the process. Based on the results of all card-sorting tests, all sub-topics were grouped into a smaller number of overall topics. These groupings – incl. the naming of general topics – have been the starting point for fourteen workshops with relevant colleagues, which have resulted in many constructive ideas and suggestions. The workshops were followed by dialogue meetings and other contact.

An important part of the process was the testing of proposals and drafts with external users. During the main project, three user tests of the type 'think aloud' were held on prototypes of drafts for a new topic taxonomy. Each of the tests included six respondents from three of Statistics Denmark's four described user profiles, namely: General population, Specially interested parties and Professional users. We wanted to investigate whether users could intuitively navigate to specific sub-topics on the lowest level using the topic taxonomy. The tests also included an exploratory approach to expected content within the presented topics on the top level.

The tests were carried out over a period of several months and brought about many adjustments. Potential challenges were tested, and – more or less clearly – confirmed or denied. At the same time, the tests revealed other challenges in, for example, groupings or naming. The tests provided important insights into the participants' considerations in their interaction with the presented topic taxonomy.

Both user tests and dialogue with colleagues from statistics-producing units taught us that the more experienced user also experiences difficulty navigating outside his or her own statistical domain. The expert user who operates confidently within his or her own area of expertise may be as much on thin ice as the novice user in other areas. And thus also needs guidance.

4. Results

The various studies and tests during the project period showed a clear preference for fewer topics on the top level than the 12 topics we had then at Statistics Denmark.

In literature[1], the term 'Scent of information' is used to describe how users evaluate the options they have when looking for information on a website. When users are presented with a list of options (e.g. a topic taxonomy), users choose the option that gives them the clearest indication (or strongest scent) that it will bring them closer to the information they need. It is all about making it as easy as possible for users to make choices on a website. Reducing the number of top-level topics makes it easier for users to make the first choice - the choice of the relevant, overall topic entry.

In our mapping of the practices among other statistical institutions, we saw the same trend, namely a transition from many topics to fewer topics at the top level. However, most NSIs' taxonomies with few top-level topics are only used on their websites. The majority of the NSIs we have explored did not have the same taxonomy on their website and in their statbank. Since 2011, we have an identical topic taxonomy on Statistics Denmark's website and in Statbank Denmark to ensure consistent and recognisable access for users, and it has been a fundamental premise of the project to maintain this one-to-one relation.

The previous taxonomy at Statistics Denmark had 12 topics on the top level and was three levels deep. The 12 top-level topics are shown in figure 2.

Fig. 2. 12 topics at the top level in the previous topic taxonomy

---

[1] Steve Krug, "Don't make me think, revisited" (2014), p. 43

In the new taxonomy, we have kept the three levels. Partly because we want users to avoid too many clicks, but also because technical limitations would make it difficult to expand the depth. It would not have been impossible to see beyond these technical limitations, but it would have meant a significantly larger project – and would have required additional resources. We therefore chose to delimit the project within the existing technical framework. The new taxonomy has nine top-level topics – please see figure 3.

Fig. 3. Nine topics at the top level in the new topic taxonomy

Reducing the number of top-level topics increases the number of sub-level topics. The challenge is to balance the number of topics at each sub-level. The nine top topics do not contain an equal number of sub-topics – or the same amount of tables in the statbank. Some topics contain several sub-topics, while other topics contain less.

During the project period, there was some concern in the organisation that the top topic 'Business' would contain too many sub-topics. 'Business' contains 12 sub-topics, and is thus the top topic in the taxonomy with the most sub-topics. Eight of these sub-topics are limited to business sectors according to (parts of) the NACE classification[2], while

---

[2] NACE is the EU statistical classification of economic activities

the last four are cross-sectional topics (structural business statistics) and business tendency surveys. The intention is to help the user choose the right top entry, which will make the choice at the next level easier – even though the number of topics at the next level may be relatively many. This presupposes that the topics at the top level are, as far as possible, mutually exclusive.

The user tests showed that users in many situations group the contents of topics differently than we used to do. Across user profiles, there was a high degree of consensus on the grouping of topics. At Statistics Denmark, we used to follow a tradition where a statistical product constituted a topic – e.g. the formalised survey on cultural habits was disseminated under the topic 'Cultural habits survey'. However, the user who may be looking for statistics on the use of public libraries does not realize that there might be relevant data both under the topic 'Public libraries' and under 'Cultural habits survey'. In the new taxonomy, different statistical products on a topic are no longer separate topics but grouped under the generic topic. This means that topics no longer reflect the statistical production, but instead follow the way users view a topic.

We are aware that it is not without problems to group different statistical products on the same topic together, as there may be significant differences in methods, coverage and/or statistical data periods – or other determinants. In such cases, it is extremely important that differences are clarified for the users. At Statistics Denmark, we try to do that using headings in Statbank Denmark and clear descriptions on the affiliated topic pages and in the documentation of the statistics.

This change from an approach reflecting the organisation of the statistical production to an approach mainly based on the users' perspective may also influences the way our statistics-producing colleagues perceive "their" statistics. The statistical product may now be part of a topic that may contain multiple aspects (i.e. multiple statistical products).

The user tests also showed that users might have difficulty perceiving the content of a topic as the naming is not always intuitively understandable or logical. On the recommendation of the expert review, we have paid special attention to naming. As a starting point, the topics at the upper levels are general and characterised by everyday language, while the topics at the lowest level may be more specific, and statistical

terms can be used to some degree. In addition, we have tried to adapt the naming to a more contemporary language.

Due to the multiple ways humans may categorise subjects, it is not possible to create a taxonomy that satisfies everyone. Our aim was to create a taxonomy that is easy to navigate for the majority of users. Creating a topic taxonomy that helps the less experienced users will expectedly not make it more difficult for the more experienced user. If the less experienced users can find the information they are looking for, so can the experienced one.

How the new taxonomy has been received

Feedback from colleagues has generally been positive and several have expressed that they find the new taxonomy more logical and less confusing.

Regarding the external users, we expected a number of inquiries to Statistics Denmark's Information Services, where users could not find their "old" statistics or tables – e.g. a price index, which in many cases is only visited once a year. However, this has not happened. In Information Service, we talk with users daily who need help locating relevant statistics, and spontaneous feedback from these users suggests that the new taxonomy is relatively easy to see through. For further evaluation of the new taxonomy, we look forward to the next user survey of our website and statbank.

5. Learnings and discussion

What have we learnt from the process of creating the new topic taxonomy?

During the course of the project, there were periods, especially in the beginning, where we proceeded by trial and error and tried to get hold of the project's focus and scope. Now that the project is complete, we can look back and evaluate what went well and contributed positively to the result – and what would we do differently with the knowledge we have today.

The project of developing a more intuitive, topic-based navigation at dst.dk and Statbank Denmark was specifically formulated in Statistics Denmark's official Strategy 2022. The fact that the Communication division had the mandate of the

executive board of Statistics Denmark to carry out the project has been significant for its success.

The participants in the workshops were colleagues who are responsible for the production of the statistics and who also have contact with the users to some extent. In these workshops, no colleagues at the managerial level participated and it worked really well to collaborate with colleagues close to the statistics and close to the users. The relatively extensive involvement of statistics-producing colleagues took time and was thus costly – both for the colleagues and for the project group, but it was time well spent as it gave us great insight and understanding. At the same time, our colleagues gained insight into our considerations and the user perspective – and had the opportunity to contribute with ideas and suggestions.

In the pilot study, we reviewed several taxonomies at other statistical institutions. We gained valuable knowledge and our findings strengthened the results we obtained in other studies. However, I am afraid that we spent an unnecessary amount of time on this mapping. Looking back, it had not been necessary to be quite as thorough as we actually were. On a positive note, we achieved insightful dialogue with several colleagues abroad who had been or were working on similar projects.

In the pilot, we had a university-employed expert in information architecture and user behaviour review the existing taxonomy with emphasis on structure and terminology. The expert review was quite clear in its mapping of the difficulties of the existing taxonomy, but was unfortunately less clear in its recommendations. Especially in terms of structure, we had probably hoped for clearer guidelines, while the review was more distinct in its recommendations of terminology. It was reassuring though, that an external professional expert review consolidated our own considerations on the issues with the existing taxonomy.

This project included a significant amount of user involvement. Both the initial card sorting tests and later the think-aloud-tests on prototypes of the new taxonomy have had a huge impact on the outcome. The basic idea throughout the process has been, if you want to understand the needs of the users, ask the users.

The most annoying thing throughout the course of the project was the fact that the process was protracted. For various reasons – including the COVID-19 pandemic –

the project was dormant for periods. Periods of stagnation do not further the commitment or the progress of the project. At the same time, the world is changing and colleagues are changing jobs, which requires a greater effort when the project is to be restarted. In an ideal world, we would compress the process by allocating more dedicated resources to the project. But of couse, it is not that easy in the real world.

Is it possible to create a uniform statistical topic taxonomy that can be used by all NSIs?

From an immediate user perspective, it seems tempting to have all NSIs disseminate their statistical content in a similar topic taxonomy. In the pilot study, we saw several varieties of taxonomies with varying numbers of topics on the top level and differences in the depths of the taxonomies. Also different taxonomies on a website and in the affiliated statbank.

In the Nordic countries and Europe, a large part of our statistics are compiled on the basis of EU regulations and legislation, which could be an argument for the possibility of a uniform grouping of topics. However, several national statistics are also compiled – and there are different traditions in different countries. Moreover, differences in traditions mean differences in approaches and perspectives, just as there may be different traditions for what topics are perceived as related. In addition, web pages and statbanks are structured differently and are based on different technical solutions, which can have an impact on what is possible with the individual NSI.

In the pilot study, we investigated whether we could find a taxonomy from another NSI, which we could, so to speak, "adopt" and apply with adjustments at Statistics Denmark, but no one adequately met our needs. At Statistics Denmark, we have moved from topic taxonomies of earlier times that were primarily based on the organisation of the statistical production to a much greater focus on the users' perspective. We know that other NSIs are undergoing similar changes in approach, but it is development processes that do not proceed identically and to the same extent in all institutions.

With a large degree of user involvement, we have developed a more intuitive topic taxonomy that also complies with our requirement for the one-to-one relation between the dissemination at dst.dk, Statbank Denmark and other dissemination platforms. The new taxonomy has similarities and overlaps with the topic taxonomies of other NSIs, but there are also unique differences.

6. References

Krug, Steve, "Don't make me think, revisited" (2014)

Statistics Denmark, "Strategy 2022" (2017): www.dst.dk/pubomtale/47196

Statistics Denmark, "Communication and dissemination strategy 2018-2022" (2018):

www.dst.dk/pubomtale/31494

June 27, 2022