Statistical Journal of the IAOS 29 (2013) 301–313
DOI 10.3233/SJI-130784
IOS Press

How to fulfil user needs – from industrial production of statistics to production of knowledge[1]

Lars Thygesen* and Mogens Grosen Nielsen
Statistics Denmark, Copenhagen, Denmark

Abstract. Statistical organisations disseminate statistics to an extent never seen before. We have developed tools that handle increasing amounts of data. However, in the production of statistics we tend to look inward; our primary focus is on whether macro figures are reasonable and published in time. We don't take the user's situation sufficiently into account. According to an in-depth user-consultation, we have found that it is an urgent task to uncover needs and to give end-users better assistance when they use statistics or wish to find relevant statistics. Is it just a question of improving our services towards users or are we facing more fundamental problems due to political, social and technological changes for users and producers of statistics? Our hypothesis is that the growing political, social and technological complexity should be reflected in the way we work and in the choices of statistics we produce in order to fulfil needs in a knowledge-based society.
The paper first presents user-consultations conducted at Statistics Denmark. Hereafter follow four themes that the authors find relevant for the fulfilment of user needs in a society where knowledge plays an important role. The four themes are: 1) dynamic use of registers, 2) from silos to dynamic processes, 3) quality and user needs, and 4) metadata and user needs. Finally the insight from user-consultations and related reflections under the four themes are put together. Business Process Management is suggested as an "umbrella-perspective" uniting the various disciplines that need to be put into play in order to reach a situation, where the production of statistics can be organised in way that moves toward a fulfilment of needs identified in the user-consultations.
The insights have been put into practice in a pilot project at Statistics Denmark. A description of the project and lessons learned is provided.
The paper concludes that user-consultations and subsequent reflections call for continuously investigating user needs, enhancement of the use of registers, better organisation of processes and more focus on quality and metadata.

Keywords: Statistics, dissemination, user needs

1. Introduction

The main focus of this paper is on how we can ensure that processes, statistical quality and metadata can be defined and implemented in such a way that users of statistics get profound help in their day-to-day work. User-consultations at Statistics Denmark show that there is a strong need to focus more on users and delivering services that are less industrial, more flexible and adapted to individual needs.

The paper will, with this starting point, look into different themes that relate to fulfilling user needs.

As the first theme (Chapter 3) we look into the use of registers as a way to address problems raised in the user-consultations. Regarding the organisation of data, the paper suggests using administrative data as "generic fulfilment of user needs", since user needs – to a certain extent – can be fulfilled dynamically by combining data from different sources as needs arise.

The second theme (Chapter 4) addresses the problem of changing processes from silos to a more in-

[1] A very preliminary version with some of the ideas of this paper was presented at the Q2012 conference www.q2012.gr [16].

*Corresponding author: Lars Thygesen, Statistics Denmark, Sejrøgade 11, 2100 Copenhagen Ø, Denmark. E-mail: LTH@dst.dk.

tegrated approach in order to better fulfil user-needs. In the international statistical community many discussions on initiatives for changing processes revolve around the problem of organisation and the traditional industrial way of thinking in silos (often also labelled "stove-pipes"). Paul Harmon describes it this way: "most companies had focused on dividing processes into specific activities that were assigned to specific departments. Each department developed its own standards and procedures to manage the activities delegated to it. Along the way, in many cases, departments became focused on doing their own activities in their own way, without much regard for the overall process" [5]. We will use this understanding of silos in the paper. As part of the second theme on processes we introduce the business process management perspective as a way forward to address the increasing complexity and the problems on silo-organisation pointed out in the user-consultations. The discussion on theme two will also include standards and methodology.

The third theme is about quality (Chapter 5). The user-consultations show that we must be more careful on how to understand and publish information about quality. As part of moving toward dynamic processes, the users must have a much more prominent place in defining and capturing information about quality.

The fourth theme is about metadata (Chapter 6). The experience of the authors shows that you have to be very careful when using this concept and when deciding what should be included and what not included when discussing metadata in relation to users. In the user-consultation we focused on a very specific set of metadata. The user-consultations showed that the challenge is how to communicate metadata and subsequently how to integrate metadata into the production of statistics in the right places.

In Chapter 7 we suggest a way to unite the input from user-consultations and insights discussed in the four themes. The main idea is to have Business Process Management as the umbrella. Underneath this umbrella, we must integrate various disciplines: use of the international process standard Generic Statistical Business Process Model (GSBPM), project management model, systems development model, and metadata/quality-model. The GSBPM-model gives the overall process-structure and guides us in understanding how and when to produce the expected product for the user. The relevant disciplines must be given appropriate roles. The pilot-project focuses on how to implement the EU quality standards as part of a more general metadata framework. The project includes establishing a common understanding of, and guidelines for, documentation that is integrated with the GSBPM model. The metadata-part includes description of content (e.g., methodology and quality declarations), variables, classifications and concepts. We strive to build a system based on international standards as much as possible. In addition, we include documentation of processes, user-manuals and IT systems. The pilot-project will be evaluated at the end of the year. The results will be implemented in 2014 at all surveys in Statistics Denmark. The paper will give some preliminary lessons learned.

2. User consultations

2.1. Metadata should guide and give value to users

Statistics Denmark is presently working to integrate metadata systems. The aim is making metadata available to end users *with content, and in a form that will facilitate end users in their business processes*, in situations where they might use or consider using official statistics. To this end, we have undertaken rather deep consultations with key user segments about their business processes.

The user-consolations were conducted using focus-group methodology [19]. We used this methodology in order to enhance the discussions among users and to reduce as much as possible the interference of pre-understanding from Statistics Denmark in the discussions.

Methodological decisions: First of all we decided on a top level structure for the focus-groups reflecting the purpose of the meetings. We used the following overall agenda:

1. Describe the way you use statistics today and put words on your needs ("Demand")
2. Presentation of a preliminary model and solution developed at Statistics Denmark ("Supply")
3. Discuss the balance between needs and the suggested solution ("Match")

The background for Part 1 was based on a hunch that we need more concrete information about the use of statistics. How well do we know the preferences of our users? Statisticians and IT people may have very good ideas about what will benefit users, and there is a temptation to "just do it". Our idea was that it would be worthwhile to try to better understand how users wish to use statistics, and what role metadata could and should play in this process. This means trying to see

metadata not just as documentation that should be created because we want to have good documentation, e.g. for archiving; but seeing metadata as part of a process that gives value to our users of today. Some documentation might prove to be of lesser interest to these users than we imagined. We want to give priority to developments that users would seem to really use in their business processes.

The second part of the user-consultation Statistics Denmark introduced a prototype of an integrated metadata-system. Our hypothesis was that we could improve the situation for users by making our existing metadata-elements (quality, concepts, classifications and variables) inter-operate so that one could go from, for example, a particular concept to all of the quality declarations for statistics where this concept is used. Users should be able to navigate across the whole space of metadata. We believed that this would be valuable even if it might not be possible to improve the basic metadata.

The third part of the user-consultations consisted in discussion on the match between needs and the models presented by Statistics Denmark. The focus here also included considerations on how well the organisational setup is supporting this match.

Besides methodological decisions on over-all structure of the content of meeting we also decided on the number of groups and number of participants in each group. We wanted to have different kinds of use of statistics reflected in the discussions in the groups. We decided on the following groups:

1. Intensive users, mostly government.
2. Municipal and regional users.
3. Education and the media.

The focus groups each consisted of around 10–14 handpicked users. We wanted to make each group representative and wanted also to have enough heterogeneity in the group in order to get as much discussion in each group as possible.

A methodological decision conducting the meetings: The meetings were chaired by an external consultant in order to reduce the interference from Statistics Denmark. A number of observers from Statistics Denmark were present but not allowed to speak (which was quite difficult) unless asked. It gave a very lively and frank discussion, and many aspects of Statistics Denmark's services were constructively criticized. Observations were not limited to documentation but also included the statistics produced, revision policy, etc.

A final methodological note is about the limitations in the user-consultations: Conducting the focus-group meetings supplements the user needs investigations already carried out regularly at Statistics Denmark. The focus group methodology gives access to deeper understanding of user needs due to discussions among users. Of course there are limitations. Some user might hold back information due to the setup where participants cannot be anonymous, e.g. "how does the group react to my proposal" (group thinking). In addition we are aware that the results are to a certain degree observer dependent. We have tried to reduce this problem by using an external consultant.

Our main approach to reduce the effect of these limitations is to follow up on focus-groups. We will do this continuously as part of the strategy for handling user needs.

2.2. *The results of the consultations*

The consultations in general confirmed that we needed a much better understanding of our users – at least those users with complex uses of statistics. Many of the tasks that users face today are more complex compared to traditional uses of statistics. In order to give the user appropriate help, we need to understand the business processes of the user. An example of this is a journalist who looks for statistics to illustrate the differences between economies of southern and northern European countries. The journalist would often prefer to have only brief and easily understood data and metadata that can be directly included in the daily newspaper. Other examples are "documentation projects", where a ministry wants, for example, to have indicators that shed light on the development of living conditions for elderly people, or disabled people. Here it is crucial to understand not only the data but also the business processes in the municipalities where data are collected. It is also necessary to understand the business processes in ministries and municipalities, where the data are used in political decision making. Other highlights from the consultations are: a) the need for more global statistics; b) Statistics Denmark should be less sectorial; c) metadata should give guidance to both less skilled and expert users.

The list below shows some key conclusions:

– The metadata prototype and the method with four interconnected components (quality declarations, concepts, variables and classifications) won strong support in all three groups. It was found that such a development of metadata would provide a good and logical approach to documentation.

- There is great need for metadata among intensive users, slightly less among municipalities, and the media say they have almost no need – their deadline is *now*, they have no time to read anything.
- There are essentially two ways that users search for statistics: 1) Ad hoc, broad search, and 2) Deep search in a fixed subject area that the user knows well. By broad search, the documentation was important, but in all groups, there is a certain tendency to call a Statistics Denmark expert and ask.
- Documentation of data breaks and changes of definitions is insufficient. Each break should be mentioned and well explained in relation to the figures.
- Statistics Denmark was encouraged to produce long time series which are corrected for breaks, otherwise each user must do it himself, which is not necessarily better, and in any event costs resources and gives different results.
- Revisions and revision practices should be better documented.
- Statistics Denmark must explain uncertainty and possible error sources, explaining what data can and cannot be used for.
- Users are often unaware of the contents of quality declarations. They should be supplemented with 'pop up' messages in the Danish statistical database Statbank.dk, especially where there is a problematic figure.
- The variables documentation system was rated as very relevant; the so-called "high-quality documentation" (giving more in-depth evaluations and being prepared with external reviewers) was especially praised. Users wish to be able to distinguish between variables used in register data (micro data) and aggregated data (StatBank). One should have a filter in the search, so the user could ask only to see the desired type of variables.
- All documentation now on paper should be made digitally accessible. There is some important documentation that exists only in books and Statistical Reports and is therefore hard to find.

Some messages about statistical production:

- Statistics Denmark is generally too sectorial. It is hard to combine statistics produced by different units in Statistics Denmark. Adjacent statistics don't relate or refer to each other in publications. Users often need to contact various divisions if they are interested in cross-sectoral statistics.
- Many users need to compare Statistics Denmark's numbers with international figures, typically from Eurostat or the OECD, but this is not easy. Where are the corresponding figures? Eurostat especially has much less documentation than Statistics Denmark. Users often try to find an indicator in the OECD's statistical database (OECD.Stat) where the number for Denmark is similar to the Statistics Denmark number. It would be very helpful if Statistics Denmark could link to relevant international sites where comparable figures can be found.
- Several users expressed the desire for more development of the statistics produced in relation to tasks that users have (relevance). The statistics must keep abreast of developments in society.
- Users want to be informed whether statistics are preliminary or final numbers.
- Users would like to participate actively in development groups around new statistical domains.

Some messages regarding dissemination:

- Users would like to have one point of entry to all documentation; from the statistics there should be access or links to the right spot within the integrated documentation model.
- Be honest in the announcement if Statistics Denmark estimates are of poor quality.
- Definitions, comments to tables and figures should appear when the mouse is moved over the cells (in StatBank); likewise, warnings should appear if there are breaks or concerns about data quality.

3. Register based statistics and the generic fulfilment of user needs

3.1. From one survey per user need to fulfilling many user needs per survey

Traditionally, official statistics have been produced using surveys and censuses aiming at fulfilling well-defined user needs. This way of production can be well calibrated with needs that are known beforehand. However, outputs may also be adapted – to a limited extent – to emerging needs that were not taken into account when planning the survey.

We have witnessed over the last 20 years an explosion and diversification of user needs, calling for ad hoc production of statistics when new needs arise. This need for more ad-hoc production was expressed in the

user-consultations on needs for fast and cross-sectional statistics.

In most cases, fulfilling such needs cannot wait for the National Statistical Institute (NSI) to carry out a new survey, and the cost and response burden of such a survey would be a serious obstacle.

In Denmark and several other countries this has led to the development of a system that allows for linking and reusing statistical data across statistical domains. This information system requires the use and storage of identification numbers for statistical units (persons, enterprises, dwellings and real estate). It is based on Svein Nordbotten's notion of an *Archive Statistical System* [1,2].

In this system, we try to take a more integrated view of the need for information as an input to knowledge-processes. We build an integrated model of "reality," rather than seeing statistics as a number of isolated surveys, or islands of information. The model builds on the most important classes of entities (or objects) that our users wish to analyse. These are *persons (and families/households), business units, and dwellings/real estate.* In most cases, users are interested only in subgroups of one or more of these three classes, e.g. unemployed persons in the municipality of Copenhagen as of 1 July 2012.

The objects are interlinked by relations, the most important being *Persons living in dwellings (Habitation)* and *Persons working in workplaces (Employment).* The model holds information about states and processes affecting the objects, e.g. unemployed persons in the municipality of Copenhagen as of 1 July 2012 who were hospitalized during the preceding year. A simple information model is shown in Fig. 1.

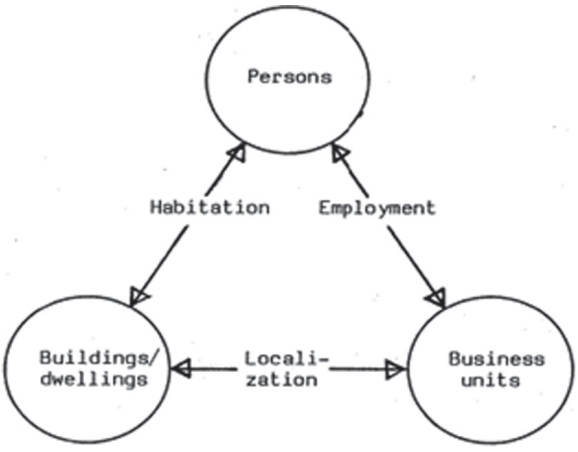

Fig. 1. A simple model [4].

Microdata must be efficiently organised in an archive that at the same time protects the information from any attempts to invade privacy and makes it possible to use it whenever needs arise. The data should thus be linked only when necessary, using identifiers for linkage.

In order to fill in most of the model, we make intensive secondary use of administrative sources stemming from all sectors of (public) administration. The cornerstones are the three *basic registers* that keep track of the populations of the most important units in our statistics: persons, business units and dwellings. A large number of other administrative registers provide additional data, allowing us to make estimates of the processes or events going on in different parts of the model (e.g. births, deaths, incomes, education), as well as the states that objects may be in (e.g. alive, dead).

In estimating the variables, registers are combined with sample surveys as sources of the system. They are linked with the other data as needed in the estimation process, and they are stored in the archive so they can be used alongside register data for other end uses.

In summary, the archive statistical system reuses and combines data from multiple sources, serving needs that had not necessarily been foreseen at the time of collection. In addition, the system offers significant advantages that are highly appreciated by end users:

- Providing for consistency across traditional statistical domains, since the same estimates of variables are used in several domains; and
- Allowing longitudinal studies where individual objects (e.g., persons) can be followed over a number of years in order to draw inferences.

A draw-back of the system is that concepts in the administrative registers don't always correspond very well with those needed in statistics. For this reason, surveys are also used to complement the registers, and they are integrated into the system.

3.2. Metadata should guide the users

When building such an integrated system based on multiple sources and with innumerable possible uses, allowing users to understand statistics becomes a huge challenge. Potential users should be able to understand whether some parts of the system can be combined in order to create useful knowledge relevant to their decisions, and with sufficient accuracy, timeliness, and so forth. The metadata we make available to potential users must enable them to make such evaluations. This requires a generic metadata system and metadata-processes that allow for many different views

of the metadata, so that the different user communities can be accommodated. As can be seen in the user-consultations it is not possible for an NSI to fully know or understand the business processes of all these diverse user communities. Therefore we must consult carefully with users in order to find out what works and what does not work. These consultations must guide us on how both to implement dissemination metadata and quality of data and how to design our internal processes.

Before going into the aspects of quality and metadata with regard to production and dissemination, we will unfold some more basic considerations concerning the role statistics and the organisation of the production of statistics.

4. Towards design of processes based on fulfilling user needs

4.1. From stable to dynamic processes

The user-consultations and many other surveys show that there is a need for more knowledge-intensive products and services than those traditionally delivered by our statistical institutions [11]. NSIs still need to produce a lot of routine statistics, but the main challenge comes from fulfilling needs where you need more knowledge about the users' more dynamic business processes. In earlier mainly industry-based societies, both the users' and our own processes were stable; the products produced were relatively stable, and they did not need to a have high knowledge content. The products were produced for a mass-market without customers interfering in the production process. This is no longer the case for many of our users.

4.2. Main challenge: complexity at global and organisational level

However, we still act as if we were living in the world of yesterday, as if we were culturally living in a world of national states with impenetrable physical borders. In our statistics we tend to miss cross-border phenomena in e.g. finance and environment. Many organisations still organise and "talk about themselves" as if the foundation were still physical industrial production, the production of knowledge being seen as a residual phenomenon mainly as a consequence of the way we organise. The main challenge in many organisations is how to handle social complexity related to the production of knowledge rather than how to convert raw materials into tangible products sold on the market [6,7,12]. To a certain extent, nationally grown statistical organisations are still influenced by this 'old days' thinking dating back to the time of their original formation.

We will include a short historical note on statistical organisation in order to better understand our present situation. This will help us identify the things we must overcome in order to move to a more dynamic way of organising processes.

The first census in Denmark was carried out in 1769 to serve user demands much different from those of today. The results were not published, as it would reveal the country's position of strength in case the country were to go to war. Later the tasks of statistics production were closely linked to the needs of the political and economic administration of society. The decision on which statistics to produce was often made by the Ministry of Finance.

According to a new Law on Statistics Denmark in 1966, statistics must support democratic processes and be impartial, so that they cannot come under suspicion of being coloured by political considerations. Statistics should be available to all and must therefore be made public immediately as they are available. With this approach, the statistical work is essentially different from that of the departments that work under direct ministerial control, and whose agenda is determined by the concerns of the Government [9]. In that respect the foundation of statistics has moved from "state secret"-decisions to decisions taken by a politically independent board. But do we need better or additional measures in order to *fulfil diverse* user needs?

Increasing complexity is the main challenge for statistical organisations today. This has at least two main consequences: First, what categories (i.e. what statistics) should we use to describe the knowledge-based society? Second, how should we be organised to fulfil user needs in a society whose production is primarily dependent on the handling of knowledge?

4.3. From "silos" to focus on business processes and users

As mentioned in the introduction many discussions in the international statistical community on initiatives for changing processes revolve around the problem of organisation and using the traditional industrial mindset of thinking in silos (often also labelled "stovepipes").

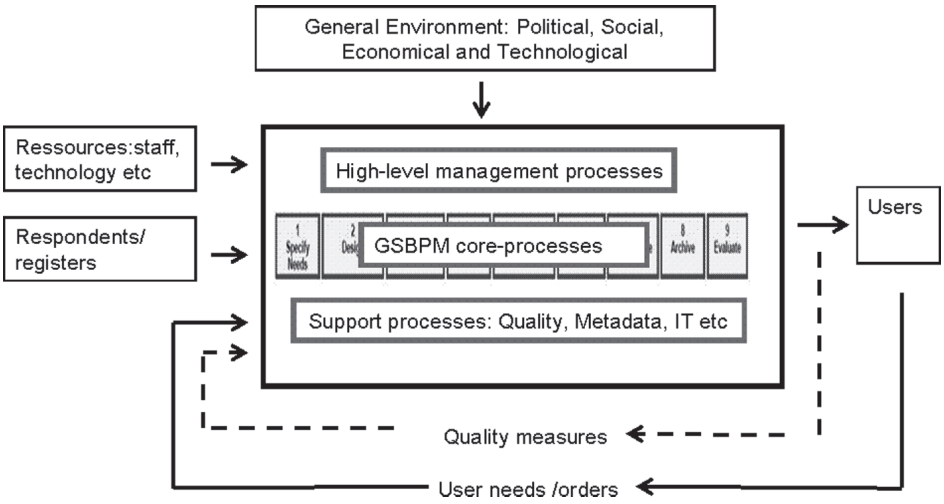

Fig. 2. Process-centric organisation and environment.

In reaction to this we see a movement toward dynamic "social-system thinking". Regarding processes, the change has gone from a focus on optimizing stable processes to focusing on what users need at the end of the production line, and adjusting the processes accordingly. Assembly lines at the Ford Model T factory are an example of the industrial way of thinking, while dynamic production lines in the Toyota factories are an example of the latter type of business processes. The main focus in the Ford model is on how to split the business processes into functional units. The main focus in Toyota is on how to build dynamic business processes as value-chains based on knowledge of user needs and feedback from users.

4.4. Process-centric organisations

We must establish processes that provide the right knowledge for the user. But more precisely, how should we build processes and interact with users? In order to move in this direction we must have a more detailed picture of statistical organisations. If we do not want silos depicted in a traditional organisational diagram, what should we have instead? The high-level model of the process-centric organisation described below is a starting point.

The construction of the model is inspired by the idea of value-chains introduced by Porter in the 1980's [14].

This way of thinking implies that you organise all processes from start to finish in such a way that each process adds value to users. This is reflected in distinctions between core-processes, high-level management-processes and support-processes. Quality, metadata, IT, etc. are seen as supporting processes. The main focus must be on core-processes that deliver value to users. Management and support processes must be designed to assist the core-processes in giving value for users.

There are two important on-going discussions related to the practical implementation of process-centric organisations. The first discussion is about which organisational structures to use. The second concerns cooperation with other organisations.

Regarding the introduction of new organisational structures, it is important not to create new silos. This may be the case in some organisations that build sub-organisations responsible for data-collection, processing, analysis, dissemination etc. Some would say that if we do not specialise like that we will just keep on doing what we do today. But how can we avoid the traps of a) introducing new silos and b) risk of just continuing to do things as we do today using the end-to-end-approach described above? There is no simple answer. A type of matrix-organisation must be introduced where a team/function is responsible for all elements in a value-chain for a survey (core processes). This team/function would draw on support processes and high-level management processes. This way of organisation ensures that the staff responsible for end-to-end-processes in a survey can evaluate and decide to improve, change or remove processes depending on whether the processes contribute value to the users.

The second important challenge regards cooperation between organisations. One case of cooperation is when sub-organisations must contribute products or services to one common organisation. An example of

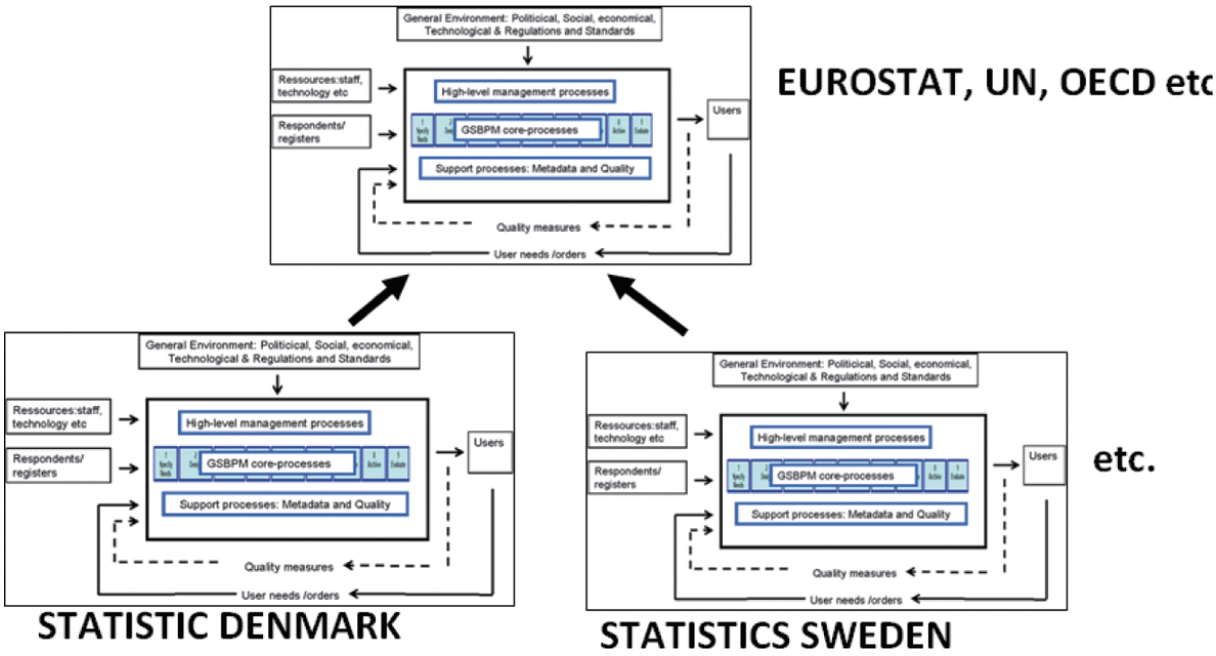

Fig. 3. Organisations with coordinated value chains using common standards. (Colours are visible in the online version of the article; http://dx.doi.org/10.3233/SJI-130784)

this is when the NSIs in EU must deliver data to Eurostat in order to create European statistics. Using the mind-set described above, NSIs and Eurostat can be seen as using the same model. NSIs and Eurostat have the same types of environments (users, staff, general environments, etc.). They also have core-, support- and management-processes. However, processes and environments differ. The political environment for the NSI includes the national parliament whereas Eurostat is more connected to the European Parliament. There is a different focus regarding inputs and users as well. Seen from Statistics Denmark, Eurostat will be regarded as an external user.

With regard to cooperation, it is important to understand that each organisation is unique with historically based management practices, social aspects and professional skills. As a consequence of this, organisations with their own built-in complexity can only respond to certain impulses from the outside world. It is not fruitful just to try to impose models from outside [10].

Given these conditions, how do we move towards building organisations that cooperate in order to provide global information as requested from users? The main tools are an appropriate set-up of inter-organisational communication and the use of common standards. This cooperation and these standards must ensure fulfilment of common goals and at the same time allow organisations to make decisions that fit their historically-based practices and "local" environment.

The process on of adopting the Generic Statistical Business Process Model (GSBPM) in many statistical organisations is a tangible example of how statistical organisations build complexity that fits their existing practices. The GSBPM-model is basically a generic model showing business processes to be included in the production of statistics.

The idea of having an abstract generic model instead of a very detailed model is the key to understanding the success of GSBPM. Its generic nature means it can be implemented in many ways. Each organisation can implement the model taking historical and local factors into account. This is OK, since it helps us move towards a common language and common knowledge without forcing organisations to implement too drastic changes.

Given that it is not possible to impose standards and detailed implementation from outside, how can EU, UN and other international organisations help NSIs to move in the same directions but with different speeds and different levels of ambition? A way forward would be to support the implementation of generic standards that respect the variety of organisations. Each organisation has specific value-chains. The challenge is to introduce and coordinate these value-chains in an appropriate way in order to fulfil user-needs. See Fig. 3.

Another case of cooperation is with users outside the statistical community. As mentioned earlier, it is important to understand the business processes of the user. As an example, Statistics Denmark must understand the processes of school-planning municipalities in order to fulfil needs for statistics for these kinds of processes. The model depicted in Fig. 2 above applies partially to these situations as well. As producers of statistics we must understand that our users in their organisations also must be aware of political, social and technological environments. Our users also face challenges on fulfilling the needs of their users, feedback from their users and challenges regarding other types of input: staff, finance, information and non-information input. The task for NSIs is to understand the users' organisations sufficiently in order to deliver the right statistics. This is by no means a trivial task.

4.5. Business process management

How can we move towards a process-centric organisation with a focus on fulfilling user needs as described above? Fortunately, we can get ideas and examples from organisations outside the statistical community. The ideas above are tested in many Business Process Management (BPM) projects. BPM is a process management method where business processes are managed, designed and continuously improved in order to increase business effectiveness and value creation. Using BPM, it is possible to create greater integration between business and IT, quality, user needs etc. Carlsberg, Arla and the Ministry of Defense in Denmark have all successfully introduced and are using BPM [18]. They are all facing the challenges described above – involvement of users, cooperation between organisations, etc. Internationally, Boeing is an example of an organisation using BPM in all processes. The implementation of BPM in Boeing is used as a case in many books on BPM [5].

An important element of business process management relates to change processes. All sources stress the importance of the high-level management focus. In addition, BPM often includes the idea of describing "as-is" and "to-be"-states. The short version of the idea behind the as-is and to-be can be seen in the Fig. 4.

One starts by examining and describing the existing "as-is" processes (management, core and support processes) and subsequently describing the desired processes based on the goals one wants to achieve. The main target is fulfilling user needs. Based on the main-target, sub-targets may cover a wide range of areas: improved documentation and knowledge sharing, implementing a quality framework, better use and management of metadata, better use of IT including stadardisation and automation, improved organisation, improved methods, etc. The next step is to decide and prioritise the changes. This is followed by the implementation. Not all changes can be carried out at once. It is therefore important to reiterate the first three steps several times. As an example of a change-process, Social Statistics at Statistics Denmark has used the GSBPM –model (introduced in Section 4.4) to document existing processes with the objective of improving flexibility and knowledge management. This example shows that one can start with small "as-is" projects to improve knowledge about surveys. The next steps will be decided based on decisions by management in Social Statistics.

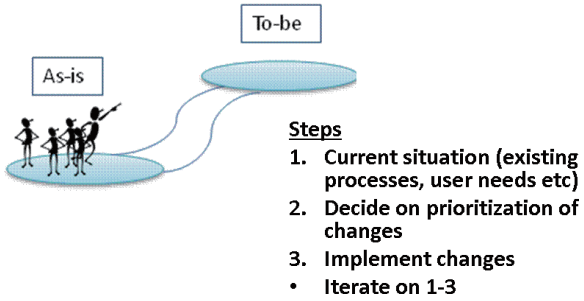

Fig. 4. "as-is" and "to-be". (Colours are visible in the online version of the article; http://dx.doi.org/10.3233/SJI-130784)

4.6. Introduction and implementation of GSBPM in the statistical community

There have been many international discussions of the application of the Generic Statistical Business Process Model (GSBPM). Should we "industrialise" statistics, as proposed by the High-Level Group for Strategic Developments in Business Architecture in Statistics [17]? Should we focus on standardisation of work-processes? How should we organise? There are many on-going initiatives, some with good results and some with less good results. The GSBPM has given us a good model of the processes, but it did not provide us with the tools and ideas for management and implementation of changes.

As described above we must be ready to organise in a way that ensures that all internal processes add value for the users. Users should not be offered "any colour of car, as long as it is black", as suggested by Henry Ford. Statistical organisations must be prepared to react to pressures and demands from outside, includ-

ing demands from users who want information personalised to their needs or taste.

We can do this by creating a flexible, process-centric organisation using the GSBPM-model and Business Process Management as the methodology. Using these tools we can move towards a situation where users are more involved in defining needs and outputs and where the organisation focuses on the horizontal contributions of each process. Only if each process gives value to users, can it be justified. This also includes management and support processes such as IT and Human Resources processes.

5. The quality concept and user needs

The user-consultations show that we must be more careful on how to understand and publish information about quality.

5.1. History

It is possible to see parallels between the movement from an industrial society to a knowledge-based society on one side, and the movement of the quality concept from having mainly a product focus to having mainly a user focus. Traditionally, "the focus of quality control is inspection and correction. From a batch of production output, a sample is selected and each item is inspected for defects. The number of defective items is measured and, if that number exceeds a certain predetermined maximum, the whole batch is rejected, meaning that it is scrapped or sent back to the production line to be reworked" [8].

In quality work of the 1940s and 1950s, more emphasis was placed on preventing defects rather than on correcting them. These new approaches, trying to prevent quality problems, was referred to as upstream quality control, and the broader range of quality measures were referred to as *quality assurance*. The main notion of quality assurance was extended to *total quality management (TQM)*. TQM principles are typically expressed along the following lines:

a) Customer focus,
b) Leadership and constancy of purpose,
c) Involvement of people,
d) Process approach,
e) Systems approach to management,
f) Continual improvement,
g) Factual approach to decision making, and
g) Mutually beneficial supplier relationships.

Much work in Eurostat and statistical organisations has followed TQM or similar lines. Eurostat defines quality in this way: "The most general and succinct definition of product quality is fitness for use" [8]. Afterwards they distinguish between output and process quality. The definitions of quality of output and process quality are reflected in the European Statistics Code of Practice (CoP) and Quality Assurance Framework (QAF), which set the standard for developing, producing and disseminating European statistics. QAF includes the traditional quality elements: Relevance, Accuracy and Reliability, Timeliness and Punctuality, Accessibility and Clarity, Coherence and Comparability. In addition to this, QAF also includes indicators related to business processes.

5.2. How to integrate quality elements into the production of statistics?

Output quality is achieved through *process quality*. Eurostat emphasises two broad aspects: *"Effectiveness:* which leads to the outputs of good quality; and *Efficiency*: which leads to their production at minimum cost to the NSI and to the respondents" [8]. But how to define and implement quality more precisely? Eurostat only gives some general guidelines referring to CoP principles and QAF.

Many NSIs define and publish quality metadata as an end product; for instance, a quality report is produced and used at the end of production. However, quality should be implemented in a way that ensures "fitness for use". This is not the case when quality is defined at the end of production. It is suggested that Relevance, Accuracy and Reliability, Timeliness and Punctuality, Accessibility and Clarity, Coherence and Comparability should be drafted together with the user as an integral part of determining user needs, which must take place in the very beginning of the production life cycle. The reason for this is that quality and methodology information is indispensable in relation to discussion with both external and internal users.

We can differentiate the users on a continuum of behaviours of beginner, experienced and expert user. Processes and applications should be designed accordingly. For the beginner, methodology and quality descriptions should be relatively easy to understand, omitting details that may be too complicated to explain. For the experienced and expert users, the descriptions and applications can be more complex. Regarding the processes, they should be designed in a way to support the needs and levels of users.

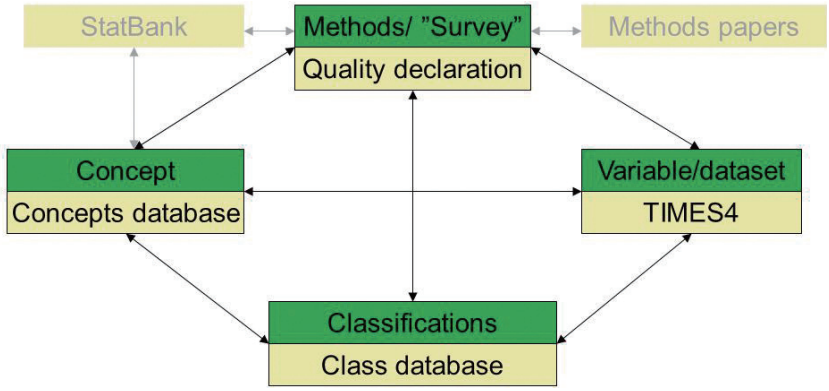

Fig. 5. High level metadata-model. (Colours are visible in the online version of the article; http://dx.doi.org/10.3233/SJI-130784)

6. The metadata concept and user needs

The third theme is about metadata. This concept is normally a concept to avoid, unless you are very specific on what is included. In the user-consultation we focused on very specific sets of metadata. The user-consultations showed that the challenge is how to communicate metadata and how to integrate metadata into the production of statistics.

6.1. History

The term "Metadata" was first introduced by Bo Sundgren in 1972 [13] and later elaborated in detail [15]. He used the following distinctions:

> "(a) the real-world phenomena that we are interested in: the object system
> (b) information about the object system
> (c) data representing information about the object system
> ... Accordingly, the data base should contain quality information and other information about the information contents of the data base. We shall refer to such information as 'information on information"' [15].

Later, ISO 11179 gave a systematic view of metadata. This standard has been adhered to by many metadata applications for statistics, for example in Sweden, Portugal and Canada. These systems all have different ways of interpreting ISO 11179. It has been popular to implement four subsystems. We suggest using the following high-level metadata model:

> The model includes descriptions of content (methodology, quality declarations, etc.), variables, classifications and concepts; the model also includes links to the dissemination database (Statbank) and independent methodology papers. The model has been implemented as separate systems in Portugal and other countries. Lately the XML-based standards *Data Documentation Initiative* (DDI) and *Statistical Data and Metadata eXchange* (SDMX) have been implemented in many places. The advantage of using DDI and SDMX is that many elements can be related to processes in the GSBPM model, so we can describe how to produce metadata as part of the whole statistical production process. Many tools are available for the SDMX and DDI standards, so it is not necessary to build a metadata system from scratch.

6.2. How to integrate metadata elements into the business process model

In the chapter on quality it was suggested that some quality-elements are defined during the needs definition process. Other types of metadata – concepts, classifications and variables – should be defined in a similar way. We must match the user's context. Various users (including internal users at the NSIs) have different needs. Some users need very general and summarised metadata. Others need content-rich or technical-rich metadata. This should be taken into account as part of the processes in the GSBPM. Each of the three types of users (beginner, experienced and expert) should be supported. Accordingly applications and supporting documents should be targeted at each user-group. Besides targeted documentation, each user group should be supported by social media (blog, Facebook, etc.) allowing them to give feedback.

7. Putting the pieces together – An interdisciplinary approach

The four themes have all given some ideas and suggestions related to the problem identified in the user-consultations. We have used this insight in a pilot-project. Before going into this we will "put the pieces together"

Figure 2 in Section 4.4 depicts the overall perspective including environments. When designing the processes the main idea is to have Business Process Management as the umbrella. Underneath this umbrella, we must integrate various disciplines: alignment of business processes with GSBPM, project management model, systems development model, metadata/quality-model. The GSBPM-model gives the overall process-structure and guides us in understanding how and when to produce the expected product for the user. Regarding the metadata and quality-processes, we have placed these as supporting processes (see Fig. 2). These processes must ensure that the metadata elements are integrated into the model by having guidelines on where to include and process quality (related to users), and the subsequent definition of concepts, classifications and variables. Part of this work must be carried out together with users. For example "User needs" in the Quality Assurance Framework must be discussed with users at the start of a survey while User satisfaction must be completed at the end of a survey.

Statistics Demark is presently conducting a pilot-project as part of implementing metadata and quality frameworks. We use the ideas from this paper. The purpose is to establish a common understanding of, and guidelines for, documentation that is integrated in the GSBPM model. The documentation includes description of content (e.g., methodology and quality declarations), variables, classifications and concepts. In addition, we include documentation of processes, user-manuals and IT. As part of the metadata pilot project, and as part of the implementation of a Register Strategy in Social Statistics, the use of registers as generic sources is integrated. This implies that the possibilities for combining sources are discussed with users. Subsequently, the registers are used in the production of the desired result for the user.

We are using Colectica [3] as a metadata tool, DDI-lifecycle and SDMX reference metadata as the standards for storing metadata. Regarding project management, systems development and processes, we use tailor-made templates that as far as possible follow established international standards (e.g. Project Initiation Document, Business Process Model Notation, standard for specifying use-cases).

Status for the project so far:

- Draft guidelines describing workflow based on GSBPM have been prepared and discussed.
- Project organisation and means for communication have been established.
- A continuation of the user-consultations has been conducted.
- Basic standard DDI-software (Colectica) has been installed. Extensions of Colectica have been made, complying with DDI and SDMX standards from Eurostat.

The pilot-project will be evaluated at the end of the year. The guidelines and accompanying software will be implemented in 2014 at all surveys at Statistics Denmark.

Lesson learned so far: due to history many NSIs are heterogeneous environments that tend to organise production in isolated silos. Therefore you must be very careful when introducing changes. The overall solution to this problem is to work stepwise. Each step must be small enough, so you communicate and "sell" the project to management and other staff. The design of each step must, on the other hand, be designed in such a way that you can later reach long-term goals. Think global, but act local, must be your guide.

The following three main considerations were included in designing the project.

1) The development and approval of strategy should take place with top-management. Subsequently it is important to prepare and discuss proposals for directors on implementing the strategy.
2) Remember to devote time for development and training in standards and tools including establishing support-processes for metadata and quality. This includes among other things translation, review and training regarding GSBPM.
3) Ensure communication and change management (Disseminate knowledge on project plans, GSBPM etc.).

In general it is important to have all three considerations above in mind continuously and not only at the beginning of the project. We found it particularly hard to introduce new standards, mainly because they have to live together with old standards and old ways of doing things. It is difficult to introduce procedures on work-processes. It is important to have top management involved not only at a strategic level but also in day-to-day processes in order to make all staff un-

derstand the implications of the project. Work should be implemented incrementally and we should continuously show and discuss benefits.

8. Conclusion

The user-consultations and subsequent reflections point toward a need for improved understanding of user needs, use of registers, organisation of processes and enhanced role of quality and metadata. This is particularly important in a society, where knowledge is playing an important role, and where it is not enough to publish routine statistics, e.g. counting how many people are living in a given geographic area.

The new way of organising statistics should be followed up in in more detailed user-consultations. The reflections introduced in the themes should be further discussed. This is to a certain degree already taking place in EU and UN statistical working groups, but it would be a good idea that the academic world at universities include considerations on the role of statistics, how to organise etc. in their research.

Statistics should contribute to solution of both local and global issues and problems. This can only take place if we understand how the statistics and their metadata are going to be utilised. The authors hope that this paper will contribute to the discussion issues related to this.

References

[1] S. Nordbotten, Elektronmaskinerne og Statistikkens Utforming i Årene Framover (*Computers and the Future Form of Statistics*). The Statistical Conferences of the Nordic Countries in Helsinki 1960. Statistical Reports of the Nordic Countries, Vol. 7, pp. 135–141. Helsingfors 1961.

[2] S. Nordbotten, *A Statistical File System*, *Statistisk Tidsskrift* **2** (1966), Stockholm.

[3] Colectica, - a tool for statistical metadata, developed by Algenta. Link: www.algenta.com.

[4] L. Thygesen, Methodological Problems Connected with a Socio-Demographic Statistical System Based on Administrative Records. Bulletin of the International Statistical Institute, Volume L Book 1, Madrid, 1983.

[5] P. Harmon, *Business Process Change – A Guide for Business Process Managers and BPM and Six Sigma Professionals*, Massachusetts, USA, 2007.

[6] P. Senge, *The Fifth Discipline: The Art and Practice of the Learning Organization*, New York, USA, 1990.

[7] L. Qvortrup, *Det lærende samfund – hyperkompleksitet og viden*, Gyldendal, København, 2001.

[8] Eurostat, *ESS Handbook for Quality Reports 2009 Edition,* Office for Official Publications of the European Communities, Luxembourg, 2009.

[9] P. Jensen, *Dansk Statistik 1950-2000 Bind 1.* Danmarks Statistik, København, 2000.

[10] R. Espejo, *Self-construction of desirable social systems* in Kybernetes, Vol. 29 no. 7/8, MCB University Press, 2000.

[11] M. Grosen Nielsen and L. Thygesen, *How do end users of statistics want metadata?* Paper presented at Workshop on Statistical Metadata in Geneve: Implementing the GSBPM and Combining Metadata Standards, 05–07 October 2011.

[12] N. Luhmann, *Europæisk rationalitet,* in Autopoisis II, Politisk Revy, København, 1992.

[13] B. Sundgren and J. Greenberg, *Metadata correspondence with Jane Greenberg about the first use of the concept metadata*, 1972.

[14] M. Porter, Competitive Advantage: Creating and Sustaning Superior Performance, The Free Press, 1985.

[15] B. Sundgren, *An Infological Approach to Data Bases*. Stockholm University and Statistics Sweden, Urval No 7, 1973.

[16] M. Grosen Nielsen and L. Thygesen, *How to fulfil user needs – metadata, administrative data and processes.* Paper presented at European Conference on Quality in Official Statistics Q2012, Athens, Greece, 29 May–1 June 2012.

[17] The High-Level Group for the Modernisation of Statistical Production and Services (HLG) was set up by the Bureau of the Conference of European Statisticians in 2010 to oversee and coordinate international work relating to the development of enterprise architectures within statistical organisation. Link to webpage: http://www1.unece.org/stat/platform/display/hlgbas/High-Level+Group+for+the+Modernisation+of+Statistical+Production+and+Services. Link to vision: http://www1.unece.org/stat/platform/display/hlgbas/Strategic+vision+of+the+HLG.

[18] Deloitte Business Consulting, Trine Bramsen, Christian Gabe and Michael Theill, *Business Process Management, Fokus på styring af processer*, 2009.

[19] B. Halkier, *Fokusgrupper*, Roskilde Universitetsforlag, Roskilde, Denmark, 2007.