

UNITED NATIONS  
ECONOMIC COMMISSION FOR EUROPE  
CONFERENCE OF EUROPEAN STATISTICIANS

EUROPEAN COMMISSION  
STATISTICAL OFFICE OF THE  
EUROPEAN UNION (EUROSTAT)

ORGANISATION FOR ECONOMIC CO-OPERATION  
AND DEVELOPMENT (OECD)  
STATISTICS DIRECTORATE

**Work Session on Statistical Metadata (METIS)**  
(Geneva, Switzerland, 10-12 March 2010)

## THE ADOPTION OF THE METIS GSBPM IN STATISTICS DENMARK

Helle Stender and Jesper Moltrup-Nielsen, Statistics Denmark

### ABSTRACT

*Abstract* **Statistics Denmark (SD) started in 2009 to work extensively with the METIS Generic Statistical Business Process Model (GSBPM). The project is driven SDs new multi-annual corporate strategy (“Strategy 2015”), which states the strategic objective of gradually achieving a considerably higher degree of standardisation and harmonisation of processes and IT systems within our organisation in the coming years – and to base it on the GSBPM. The project has a dual purpose: The first is to test the extent to which the GSBPM is a useful tool in analysing and documenting the work processes in different areas of statistical production and how it – in daily practice – can be used in a fruitful manner in projects that cut across SDs organisation. Especially, the phases 4-7 in the GSBPM have been in focus. The lessons learned so far are promising. The second purpose of the project is to analyse the current coverage of metadata systems in our portfolio, based on the framework given by the GSBPM.**

### I. STRUCTURE OF THE PAPER

1. The following paper has 4 main parts. The first part gives an overview of the background for working in depth with METIS GSBPM in SD, and the context in which this work has been conducted. Afterwards there are two parts that considers working with METIS GSBPM from two different points of view, that has been in focus. These are firstly from the point of view of working processes, and secondly from the aspect of documentation and metadata. Each of the two parts is concluded by a section considering future work within the area. Finally there is a part that concludes the paper by considering what the lessons learned by working with METIS GSBPM has been so far in SD.

### II. BACKGROUND AND CONTEXT

*Strategy 2015* 2. In 2009 Statistics Denmark (SD) has been developing a new multi-annual corporate (“Strategy 2015”). An important part of this strategy focus on gradually achieving a considerably higher degree of standardisation and harmonisation of business processes and IT systems within the organisation.

3. In the strategy it says:

*“Increasing standardisation is the most important initiative. This implies that homogeneous tasks and job functions are performed in accordance with best practice applying common methods, tools and systems.*

*Standardisation gives rise to a higher quality via a higher level of documentation for work processes, statistics and systems. Moreover, this will in the long term result in a reduction of the consumption of resources for producing statistics and for operating, maintaining and developing IT systems. A prerequisite of standardising work processes is that they are identified and subsequently optimised and standardised, wherever possible and expedient. The standardised work processes are supported by common standard systems, tools and methods.*

*This will imply that SD becomes less vulnerable to staff replacements and opens up possibilities of a more flexible use of employees across the institution. Targeted efforts and investments are required to achieve these advantages.”*

*Working group*

4. To accommodate the objective listed above, a project was started in the spring of 2009, with the forming of a working group on standardisation which spans across all departments in our organisation and which refers directly to the Board of Directors.

*Working with METIS  
GSBPM*

5. It was quickly decided to start using GSBPM model as a frame for the work in this group. The purpose for this was twofold, as it should give us a head start by adopting a model that has been created on basis of input by the best statistical organisations in the world, this was also a perfect opportunity to test the extent to which the model that should now be considered an international standard, could be adopted to describe working processes in SD now and in the future.

*Dual focus*

6. The working group has had a dual focus in working towards standardisation. The first is to test the extent to which the GSBPM is a useful tool in analysing and documenting the work processes in different areas of statistical production and how it – in daily practice – can be used in a fruitful manner in projects that cut across SDs organisation. Especially, the phases 4-7 in the GSBPM have been in focus. The lessons learned so far are promising. The second purpose of the project is to analyse the current coverage of metadata systems in our portfolio, based on the framework given by the GSBPM.

### **III. ACTIVITIES IN ANALYSING WORK PROCESSES**

7. This part will describe the work done in order to identify and document work processes in the light of METIS GSBPM. First there will be a description of the method and the selection of cases.

#### ***Cases and workshops***

*Selection of cases*

8. In the spring of 2009 a list of statistics were selected to be the first cases subjected to an in-depth analysis. The first case studies were started in the department of business statistics, where 7 cases were chosen. Later on the concept continued in the 3 other departments of our organisation, who each contributed with 2 cases, giving the project 13 case subjects.

9. The goal behind the selection of these cases was to cover as much as possible the variety of tasks performed in our organisation, while still limiting the number of cases to a manageable task with resources available.

10. Also the cases were selected in a fashion that the amount of work to be done, by each of the divisions of statistical production, was limited. Furthermore, we tried to select cases where the analysis could be a helpful stepping stone in pursuing other strategic goals.

*Workshop concept* 11. All associated employees in the selected statistical fields, were then invited to participate in 2-3 workshops, facilitated by two members of the working group on standardisation. Each workshop would typically last for 3 hours. In that time span every (manual) working process that would be done by the statistical team for each iteration of the statistical production cycle, was identified and scrutinized.

*Documentation* 12. The outcome of the workshop has been documented centrally, and subsequently reviewed by the statistical team. SD has tested different software tools designed for business process modelling that quickly and easily can make usable diagrams that follow the international standard for Business Process Modelling Notation (BPMN). An example of the level of documentation can be seen in Appendix A.

*Additional cases* 13. Additionally, we tried to test the model on a different kind of generic processes that were also characterized by the fact that they cut across divisions and departments.

14. These were, firstly, developmental projects regarding digitalization, and secondly processes in phase 4 in METIS GSBPM, that were generic across statistical domains. The latter were made because SD has split out data collection in a separate division, which conducts the data collection for an increasing number of statistical domains.

15. Especially, in the area of digitalisation of questionnaires, there are several repeated processes that cut across divisions. These – and their many interfaces – were also analysed and diagrammed with the tools adopted by this project.

*Coverage* 16. As the work with business processes was meant to cover as much of SDs production as possible cases were selected according to a number of central background variables such as: Frequency, Statistics using/not using standardized IT-systems, Complexity, Type of statistical unit, Method for error detection, Coverage, Confidentiality scheme, Cost, Stability, Maturity and Type (Primary statistics vs. Derived statistics). In order to cover the different types of cases, it was attempted to cover at least two cases within each of these categories<sup>1</sup>.

### ***Results of working process analysis***

*Phases 4-7* 17. The facilitators of the workshops decided to have their main focus on testing the extent to which the phases 4-7 of the METIS GSBPM could be adopted to the production in SD. Along side this test the outcome of the workshops also had the ability to solve “problems” or challenges for the statistical teams involved.

18. The main reason for selecting phases 4-7 was, that many statistics that is well established, and has been carried out for many years, will be very stable, and their

---

<sup>1</sup> See Appendix B for a list of cases including background variables.

generic outlook, that could have a potential for becoming more efficient, lies within these phases that are carried out for each iteration of the statistical cycle.

*Immediate impact* 19. The process itself therefore had an immediate impact in improving the production in SD. One outcome of the workshops was a list of observations and possible improvement actions for the statistical team. Within half a year – after getting back to the teams that had been going through the process – it showed that a number of the teams had already implemented some of the suggestions from their list. Also some teams had acknowledged that an improvement in an area would be positive, and that they would implement it when the step was reached the next time in their production cycle.

*Subcategories* 20. The list of observations was divided into 3 subcategories, according to whether the observation had the potential of:

- improving standardisation,
- improving efficiency, or
- improving the quality.

Some of the observations could have more than one potential. Also, a distinction was made between improvements for the particular statistic in question, or if it were suggestions that had a broader perspective for SD as a whole.

21. The observations are also varied, and some will require deeper analysis. There are huge difference in scope and whether the change would be costly for SD. Some will be small “quick hits”, where others might need planning and allocation of new resources in order to be implemented. This also means that some improvements can be made by the team in question, where others need decisions at a higher level in the organisation. Finally, some of the areas would have to go through a deeper analysis to determine if there would be a positive business case.

*Improvement in communication* 22. The work that was done in analysing and documenting processes in the area of digitalisation of questionnaires also had some immediate effects. It became clear that there was a need for better coordination and communication across teams in various divisions that each contributed to the development of new digital questionnaires. There was then established a new coordinator role that was institutionalised in one of the divisions in question. The person in this new role would then have the responsibility of making sure that teams working on a project together would all have all necessary information at all times.

23. Also, this new coordinator would improve the communication towards the “customers” in the statistical divisions that would end up with the outcome of each project. In this way the statisticians would only have to ask one person regardless of which information they needed at any step of the project. Previously, it had been difficult for the statisticians to know where they should go for the information they needed because there were so many project managers involved in different aspects of the project.

### ***Future work***

*Paper on usefulness* 24. After finalizing the case studies of work processes, the working group will submit a paper to our Board of Directors that considers the usefulness of METIS GSBPM in SD.

*Suggestions for improvements in SD* 25. It is foreseen that more statistics will have their work processes analysed and documented according to the tools and experiences adopted in the first phase of the project. This will lead towards identifying suggestions for areas where there will be a positive business case in adopting/developing standardised methods and systems.

*Implementations of improvements* 26. The final phase of the project will be implementing improvements based on the suggestions made. These improvements however, might be made both in the short term as well as in the longer term. Smaller improvements might be implemented directly, while it is foreseen that there will be drafted a plan for bigger improvements, that will span across the period of our Corporate strategy (2010-2015).

#### IV. ACTIVITIES REGARDING DOCUMENTATION AND METADATA

27. SD uses a range of applications containing documentation and metadata. The current situation was examined by a Working Group on metadata and documentation applications in 2008. The working group mapped the applications used to the key processes in the GSBPM model. The purpose of the work was to determine whether the applications overlapped, and whether there were areas uncovered by the applications.

##### The applications

28. The applications mentioned can be divided into three groups:

- applications containing only metadata
- the data collection applications
- applications used to publish data.

29. These systems are relevant when metadata and the GSBPM model are discussed.

##### *Applications used for metadata only*

30. These systems are characterized by not containing actual statistical data. The main examples of this type of systems are:

- |                                   |  |
|-----------------------------------|--|
| <i>Declarations</i>               | • (Product) Evidence of statistical level. Information on the quality of the statistics. The information is intended for external users, and is available on SDs website.  |
| <i>Times</i>                      | • (Data) Documentation at the micro level. Information about tables, variables and values. The information is found primarily at unit level. This does not take into account the calculated variables, aggregation, index etc. The information is intended for internal and external users, and is available on SDs website. |
| <i>Conceptual Database</i>        | • (Concepts) Textual descriptions of the meaning of statistical terms used. Not in a formalized format. The information is intended for external users with little knowledge of statistical terms, and is available on SDs website.  |
| <i>Classifications</i>            | • (Value) Hierarchical values at the micro level such as NACE. The information is intended for internal and external users, and is available on SDs website.   |
| <i>High quality documentation</i> | • (Data) Documentation at micro level. Contains detailed information selected variables. It contains information on population, break, etc. The information is intended for researchers and is available on SDs website.   |

### ***Applications used to collect data***

31. These systems are used to collect the input data. Applications contain data, meta-data about input data and metadata on the delivery and processing of input data. The most important of these systems are:

- XIS2* • XIS2 is used to collect input data from different sources as the Internet portal called Virk.dk, from e-questionnaires, system-to-system applications and OCR. The application contains different sorts of metadata (about data, transformation, respondents and storage of data)
- CEMOS* • CEMOS is used to collect data from administrative registers. The application contains different metadata (data transformation, storage, and delivery frequency)
- IBS* • IBS is a filing system that is used to gather information about the received data from respondents.

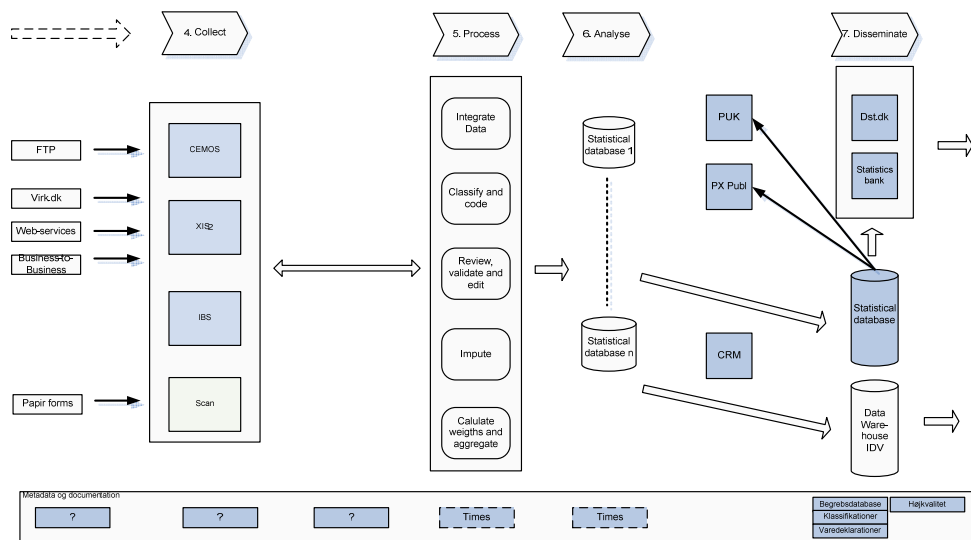
### ***Applications used for dissemination***

32. These systems are primarily used in the dissemination phase. The systems are used in different contexts and therefore have very different content and very different metadata. The most important of these systems are:

- PUK* • PUK is a calendar for planning and management of statistics from SD. The documentation application contains links to the statistics.
- PX-Publ* • PX-Publ is used for extraction from the statistical database, for tabulation and preparation of publications. It contains metadata about the extracted data from the Statistical Database.
- CRM* • CRM is used to manage customer relationships. The system contains data about customers, sales opportunities, quotations, orders, products and invoices. The system is used primarily in connection with customized statistics.
- The Statistical Database* • The Statistical Database contains all the data published by SD. The system contains metadata and data at the aggregated level.
- The Statistics Bank* • The Statistics Bank is part of the Statistical Database. The system is accessible to the public, and contains data and metadata on the part of the statistical database that is published.

### **An overview of applications and key processes**

33. The chart below shows the results from the working group's studies. It gives an overview of the main processes in the GSBPM model and the applications containing metadata in SD:



34. The Chart shows how data is collected from different sources and placed in the data collection systems (CEMOS, IBS and XIS2). The working group found that this process is standardized to some extent, but needs further standardization, as it uses a lot of manpower to ensure that data is handled properly.

35. From the data collection systems the data is processed and stored in different statistical databases. These processes needs standardization, as there currently are few guidelines to how they should be handled.

36. From these databases data is analysed, and transferred to the statistical database and disseminated in publications and the Statistics Bank. The data analysis needs standardization, but the transfer process is standardized, and well documented.

37. The bar at the bottom of the chart shows the applications only used for metadata. As the charts also shows, the dissemination phase is very well covered by metadata applications with different scopes (by the declarations, the classifications, the conceptual database, and by Times). These systems are not connected. The working group discovered that the applications are not overlapping either. The scopes are very different.

38. The data processing and the data analysis phases are more or less covered by the Times system. The Times system is widely used in the social statistics department, somewhat in the business statistics department, and hardly at all in the economics statistics department. This is among other reasons due to the fact that the Times system is most fit for register data, and therefore is unsuitable for e.g. calculated indexes.

**Future work**

39. One of the conclusions from the working group’s studies is that even though the applications are not overlapping, there is almost no connection between them.

40. If we look at the data collection system, they all contain metadata on the delivered data and the delivery itself. These metadata could be used in a common application, to document both the data delivery and the data delivered.

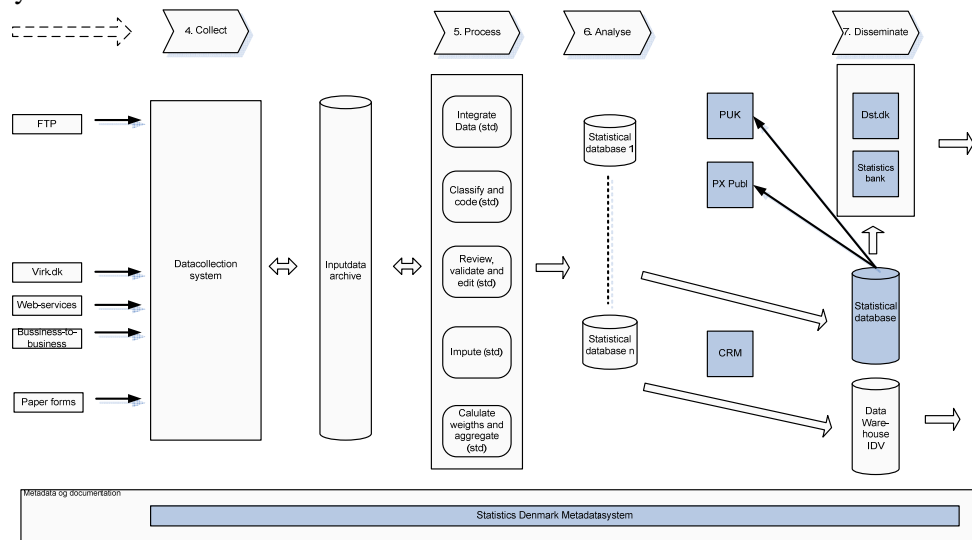
41. The data processing and the data analysis phases demand more advanced facilities, such as the opportunity to indicate how derived variables are calculated from

other variables, how imputation is carried out, how an index is calculated and development over time.

42. The statistical registers can be covered by the existing Times application. The dissemination phase is covered by several applications, which should be collected into one, or at least made “looking like one system” to the users.

43. As mentioned earlier, the existing Times application is not used by the economic statistics department. An important task is to consider how this can be achieved. This means that a common metadata system must contain the functionality demanded by macro economic statistics.

44. One common metadata system, covering all (or nearly all) of the phases in the GSBPM model is the ultimate (but long term) goal, so in the ideal world the metadata system would look like the chart below.



## V. LESSONS LEARNED ABOUT USING THE GSBPM MODEL

### *Usefulness of GSBPM*

45. As the tasks of the working group on standardization have not yet been finalized, and their report not has been approved by the Board of Directors, the final conclusions on the usefulness of GSBPM in SD is not yet drawn.

46. However there are some key lessons that have been drawn that would be important to bring forward as a conclusion to this paper.

### *Planning*

47. The basis for planning a strategy for further development of the overall metadata coverage of SD is better using GSBPM. This is because a common framework leads to common understanding across various areas of production.

### *Better overview*

48. The GSBPM model has given us a clearer view of the different need for metadata and documentation, in the different phases of the statistical production.

### *Standardization*

49. One of the major challenges that SD is facing is the standardization of methods. The GSBPM model helps planning and organizing this work as well.

50. However, while this statement holds true, it has become clear, that there is a lot of variation in different areas of the statistical production cycle. Based on the analysis and documentation of work processes some areas showed to be standardized to some

extent, whilst others were in lack of a common framework regarding methods for production.

*Dissemination* 51. As an example, most aspects regarding dissemination of statistical outputs, has been standardized for many years, and is institutionalized in a division for this purpose. Thus SD can be said to be fully functional in this area, and the methods surrounding dissemination is well documented.

*Data Collection* 52. For some areas in SD, working with statistics based on questionnaires, there has in the past years been developed a standardized framework for data collection. While not all statistics lives up to the concepts of the common framework as of yet, this can still be regarded as standardized to a large extent. This covers as well a common database for collected data, a common framework for registration of data suppliers, and a functionally institutionalized division for data collection.

*Archiving* 53. In some areas of SD, in the department of Social Statistics, there has also been developed a common framework for data storage. This is developed in a manner that the same data is archived in many forms, as data is processed through the cycle of statistical production.

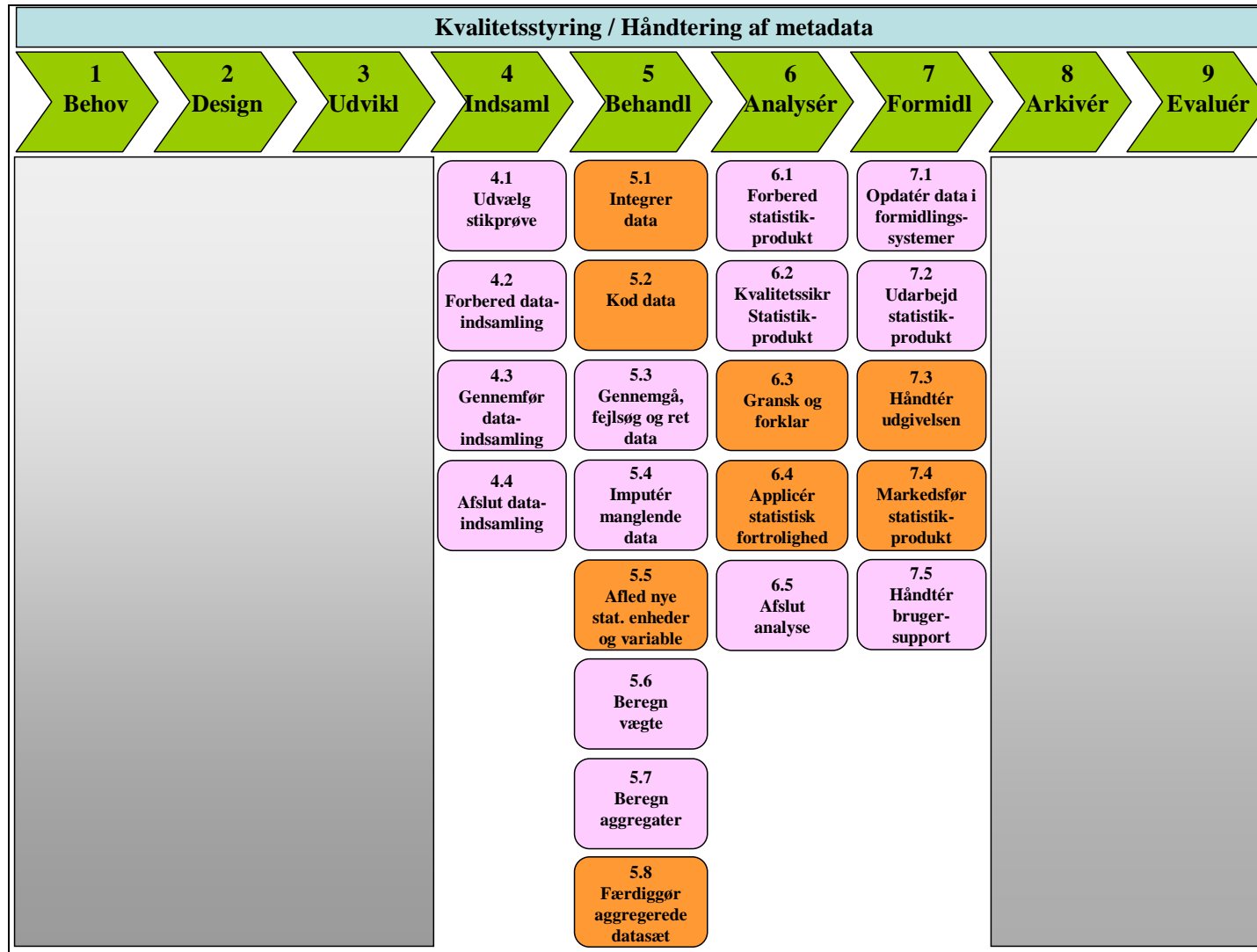
*Processing data* 54. On the other hand it has become clear, that there are still major challenges facing SD regarding steps where data is processed. Historically there has been a wide range of freedom to the statisticians in each statistical field to develop their own methods and applications for production. This has led to a broad field of methodological approaches, which in turn creates challenges when it is sought to develop common tools for these areas of the production cycle, in SD as a whole.

*Strong and weak areas* 55. It is foreseen as of this moment that on the basis of METIS GSBPM SD will have a solid foundation for identifying areas of interest for improvement initiatives.

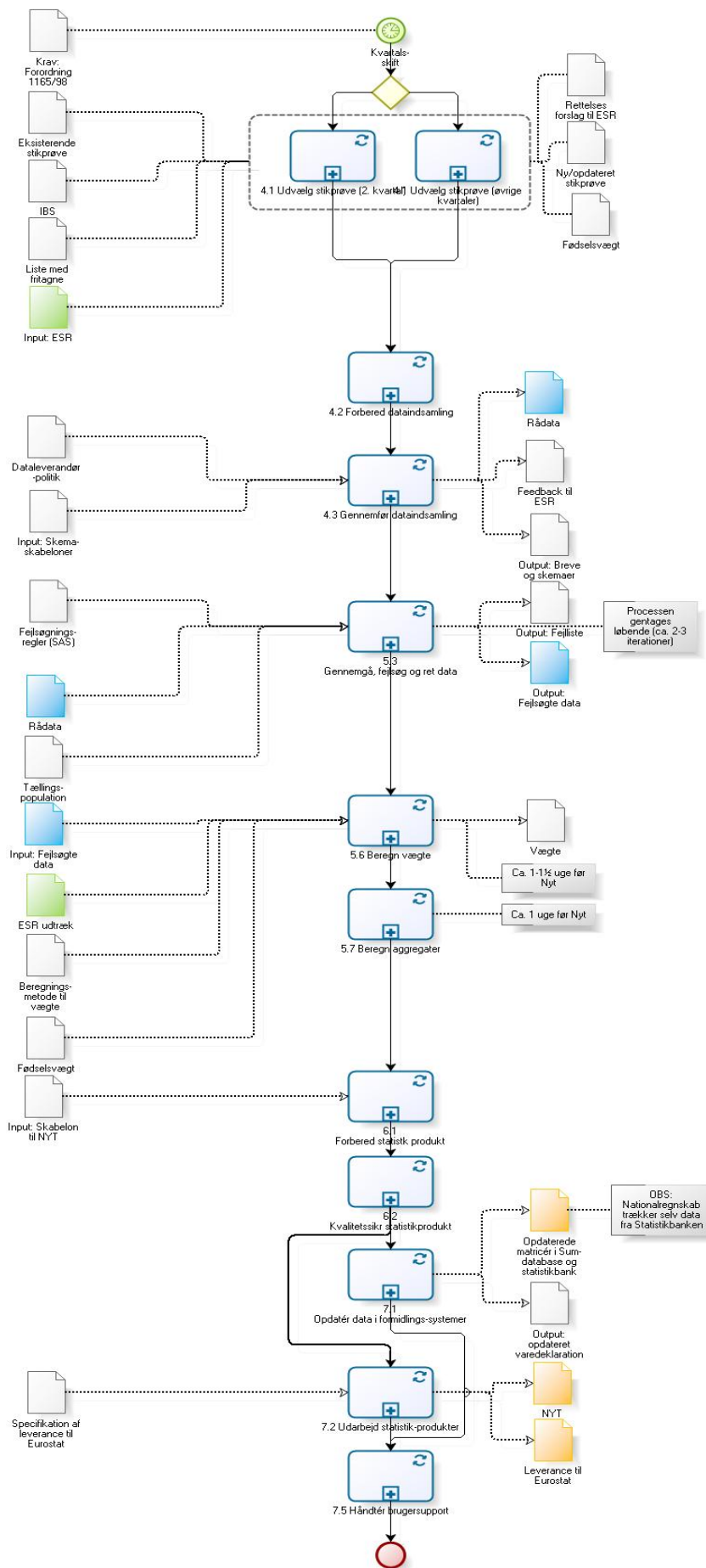
56. The GSBPM thus also has provided a better overview of the strong and the weak areas of our metadata applications. The collection phase is not very well covered in comparison to the dissemination phase.

## Appendix A

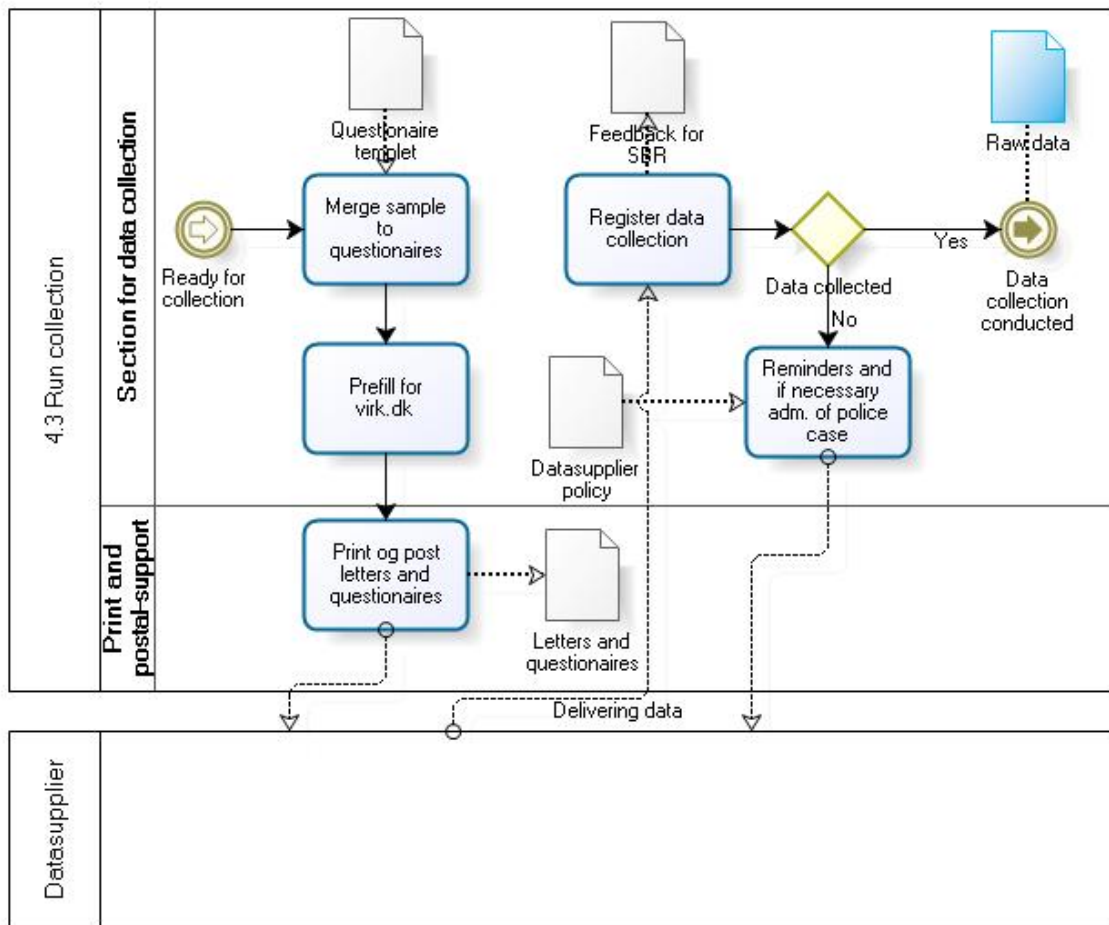
### METIS phases for: Employment in construction industries (marked with pink)



# Process control-flow



### Step 4.3 Run collection



*Condition at start* In the starting condition we are ready to collect data.

*Condition at the end* At the end the collection has been conducted.

*Input* There is a questionnaire template as input for the process.

The data supplier policy is input, as for guidelines for reminders etc. in the collection phase.

*Output* Letters and questionnaires are output of this process.

The raw data for the statistic is the ultimate output of this process.

*Note that...* The sample is stratified. This means that there are 6 different letters and questionnaires when merging with the letters and sample.

## Appendix B: Cases

Statistic	Frequency	Standardised system (if any)	Complexity	Type of Statistical Unit	Method for error detection	Coverage	Confidentiality scheme	Cost	Stability	Maturity	"Type"
Agricultural and horticultural survey	Annual	Genstat	Complex	Not SBR	Micro	Population	Positive	High	High	Well established	Primary statistics
Gen. economic statistic	Annual	"Stand-alone"	Complex	SBR	Micro	Sample	Positive	High	High	Well established	Primary statistics
Carregister	Annual	"Stand-alone"	Complex	Not SBR	Macro	Population	Positive	Low	High	Well established	Derived statistics
Employ. in construction industries	Quarterly	IBS	Simple	SBR	Micro	Sample	Positive	Low	High	Well established	Primary statistics
Industrial commodity statistic	Quarterly	IBS	Complex	SBR	Micro	Cut-off	Negative	High	High	Well established	Primary statistics
Retail Trade Index	Monthly	"Stand-alone"	Simple	SBR	Micro	Sample + cut-off	Positive	Low	High	Well established	Primary statistics
Use of ICT in enterprises	Annual	IBS	Complex	SBR	Macro	Sample	Positive	Low	Low	Well established	Primary statistics
Foreign trade in services	Quarterly	"Stand-alone"	Complex	Not SBR	Macro	Sample+ cut-off	Negative	High	High	Well established	Primary statistics
Consumer price index	Monthly	"Stand-alone"	Complex	Not SBR	Macro	Sample	Positive	High	Low	Well established	Primary statistics
Key figures in housing (std. product in Customer Centre)	IDV/Ad-hoc	"Stand-alone"	Complex	Not SBR	None	Population	Positive	Low	High	Well established	Derived statistics
Student register (register updates)	Annual	Moduldata	Complex	Person	Macro	Population	Positive	Low	High	Well established	Primary statistics
Population register	Annual	Moduldata	Simple	Person	None	Population	Positive	Low	High	Well established	Primary statistics