



TWINNING CONTRACT

CRIS 2015/370-467



Support to the Israeli Central Bureau of Statistics in Improving the Quality of Official Statistics

MISSION REPORT

on

Component B **Micro Data Services to Researchers**

Activity B.7 **Building and Integrating Data Catalogue and Documentation**

Implemented by

- Charlotte Leolnar Reif, Senior Adviser, Research Services, Statistics Denmark. clr@dst.dk
- Karin Holst Duer, Chief Adviser and Project Manager for Strategic Projects in Statistics Denmark, The Management Office, Statistics Denmark. khd@dst.dk

Jerusalem

28-31 January 2018

Version: Final

Authors' names, addresses, e-mails

Ms. Charlotte Leolnar Reif

Senior Adviser

Research Services

Statistics Denmark

Sejrøgade 11

DK-2100 Copenhagen Ø

Denmark

Email: clr@dst.dk

Ms. Karin Holst Duer

Chief Adviser and Project Manager for Strategic Projects in Statistics Denmark

The Management Office,

Statistics Denmark

Sejrøgade 11

DK-2100 Copenhagen Ø

Denmark

Email: khd@dst.dk

Table of Contents

Executive Summary	5
1. General Comments.....	7
2. Assessment and Results.....	8
2.1 Overall goal (vision)	8
2.2 Project Initiating Document (PID)	8
2.3 Organizational structure	9
2.4 Policy for a data catalogue and documentation.....	10
2.5 The Catalogue	10
2.6 Documentation of datasets	12
2.7 Documentation of variables	13
2.8 Benefits for researchers and the ICBS	15
3. Conclusions and Recommendations	16

Annexes produced for the mission (external to the report):

Annex B7 - 1 Terms of Reference
Annex B7 - 2 Meeting program
Annex B7 - 3 Persons met
Annex B7 - 4 Current status for building a data library at ICBS (BC Presentation)
Annex B7 - 5 The data library at Statistics Denmark (BC Presentation)
Annex B7 - 6 Template for overview of datasets (Output from the Pilots)
Annex B7 - 7 Template for overview of variables (Output from the Pilots)
Annex B7 - 8 Template and guideline for metadata for datasets <i>in English</i> (Material prepared for the Pilots)
Annex B7 - 9 Template and guideline for metadata for datasets <i>in Hebrew</i> (Material prepared for the Pilots)
Annex B7 - 10 Template and guideline for metadata for variable <i>in English</i> (Material prepared for the Pilots)
Annex B7 - 11 Template and guideline for metadata for variables <i>in Hebrew</i> (Material prepared for the Pilots)
Annex B7 - 12 Example of metadata for datasets <i>in English</i> (Material prepared for the Pilots)
Annex B7 - 13 Example of metadata for variable <i>in English</i> (Material prepared for the Pilots)
Annex B7 - 14 Guidelines for High Quality Documentation in Denmark (Background material)

List of Abbreviations

BC	Beneficiary country
EU	European Union
ICBS	Israeli Central Bureau of Statistics
IT	Information technology
MUC	Micro-data Under Contract
MS	Member State (of the EU)
NSS	National statistical system
OECD	Organisation for Economic Cooperation and Development
PCS	Public Council for Statistics
RAS	Remote Access System
PID	Project Initiating Document
PUF	Public Use File
RDC	Research Data Centre
RSU	Research Service Unit
SIMS	Single Integrated Metadata Structure
SO	Statistics Ordinance
TA	Technical Assistance
ToR	Terms of Reference
UNECE	United Nations Economic Commission for Europe

Executive Summary

The current mission report is the outcome of the sixth activity within Component B: Micro Data Services to Researchers as part of the Twinning Project "Support to the Israeli Central Bureau of Statistics (ICBS) in Enhancing the Quality of Official Statistics". The mission was devoted to building and integrating data catalogue and documentation.

The mission was initiated with the ICBS presenting the current situation for providing micro-data files from dedicated research rooms to authorized researchers. For each research project one customized research dataset and codebook are created according to "the need to know principle" and available microdata from the Subject Units. However, currently no common institutional data catalogue is available internally or externally. In addition, no policy, format structure, content and naming convention for a data catalogue are in place at an institutional level and sometimes not even within a unit. Therefore, the researchers need to consult multiple pages at the ICBS website, which in several cases are incomplete or do not contain sufficient information, in order to enable the researchers in selecting datasets and variables for their research project. Consequently, meetings between the researchers and the unit staff are, in most cases, needed in order for the researchers to select datasets and variables suitable for their research objectives.

In addition the ICBS presented some of the on-going initiatives expected to influence the availability of metadata for researchers like the Single Integrated Metadata Structure (SIMS) as well as the process for providing micro-data in a more sufficient manner by the establishment of a dedicated Research Service Unit (RSU) on 01 October 2017.

The MS experts presented the policy and process for establishing and maintaining the data catalogues at Statistics Denmark including the production and display of associated metadata. The presentations focused on the tasks and roles of the RSU. The MS experts also presented the benefits of providing an integrated data catalogue and documentation both for researchers and other users of data e.g. ministries and not at least the internal benefits to the organization both in terms of saving time on consulting and in terms of sharing data and documentation between subjects units.

The process of establishing a complete and fully integrated data catalogue and documentation in the ICBS can be a relatively laborious process. The aim of the mission was therefore to discuss possible options for tools that can be implemented immediately without awaiting e.g. IT solutions and other various dependencies. A prerequisite for putting the proposed tools into use is that it can be used regardless of the final IT model that still has to be defined. The tools must be flexible, simple and easy to implement and use.

During the mission, two pilots were carried out in order to test the suitability of the Danish framework for a data catalog with integrated documentation of datasets and variables. The two pilots were carried out by staff members from the Demography and Census Department and from the Micro Economic Statistics Department respectively. The pilots took outset in Excel and Word templates as well as guidelines created by the Danish experts before the mission. The proposed framework was discussed and adjusted to Israeli settings.

It was stressed by the MS experts that the documentation of research data can be done to varying degrees of details provided and that the ICBS has to agree on an appropriate standard/level at the organizational level. Furthermore, it was recommended that existing or planned documentation such as e.g. SIMS, the content of codebooks and the glossary of terms should be reused as much as possible and include links to other suitable metadata. From the experts point of view the most important task ahead is to get started and standardize the process and output so that the work can be expanded gradually and combined in future.

Support to the Israeli Central Bureau of Statistics in Improving the Quality of Official Statistics

At the end of the mission the MS experts came up with a number of conclusions and recommendations. The most central of these are the following:

- *Organize the process of building an integrated data catalogue and documentation in a formal project structure*
- *Draft a proposal for a common institutional data catalogue policy to be presented to the top management in order to get their approval for the project*
- *Decide on the structure of the catalogue and keep to it*
- *Start creating a catalogue using the most demanded datasets and variables*
- *Re-use existing documentation as much as possible*
- *Involve and motivate staff members from the subject units to take part in the project*

1. General Comments

This mission report was prepared as part of the the Twinning Project "Support to the Israeli Central Bureau of Statistics (ICBS) in Enhancing the Quality of Official Statistics". It was the sith mission in component B: Micro Data services to researchers and was devoted to building and integrating a data catalogue and documentation.

The aim of the activity was to get input and discuss options for building an integrated data library, data catalogue and documentation in a professional, user-friendly and competent manner by the Israeli Central Bureau of Statistics (ICBS) with a focus on policy, structure and content of the data catalogue. The catalogue and the associated documentation should be available both internally and externally.

Processes and services in focus were:

- The Danish Register Strategy for internal users of social statistics in Statistics Denmark and how the strategy assists researchers in using microdata for research purposes
- Policy and structure of the data catalogue provided by the Research Services Unit (RSU) at Statistics Denmark
- Policy for researchers regarding using the data of the ICBS
- A possible framework for a future data catalogue in the ICBS
- A possible framework for future standardized documentation of datasets and variables
- Overall benefits and challenges of providing a data catalogue and documentation
- Critical steps and decisions to be initiated in the ICBS

The presentations assisted the ICBS and the Twinning Project experts in forming the basis for developing a long-term and a short-term plan for providing microdata and services to the researchers in a user-friendly and a cost-effective way.

The experts would like to express their thanks to all officials and individuals met for the kind support and valuable information which they received during the stay in Israel and which highly facilitated the work of the experts.

The views and observations stated in this report are those of the consultants and do not necessarily represent the views of EU, ICBS or Statistics Denmark.

2. Assessment and Results

The process of the mission was conducted in close cooperation of two MS Experts from Statistic Denmark, Staff members from the International Relations and Statistical Coordination Department, relevant Subject Units as well as Twinning Staff; cf. Terms of Reference (Annex B7 - 1).

During the mission the issues and tasks listed below were discussed and a preliminary consensus of understanding was reached.

2.1 Overall goal (vision)

As stated in the ToR:

The vision for future services for researchers provided by the ICBS is to provide microdata and services in a secure, professional, user-friendly and competent manner meeting international standards and the best practices and to increase effectiveness in order to reduce the time spent from when a request is placed and until access to data can be obtained.

The fulfillment of this ambitious vision is a process with several steps along the way and will most likely take a long time before it is fulfilled with several technological and organizational dependencies influencing the result. One of the dependencies is to develop a supporting IT-system that will substantiate e.g. the degree of user-friendliness of providing the microdata for researchers.

In brief, the goal is for the ICBS, to build a standardized and fully integrated data catalogue and documentation in a stepwise approach that should be available to both internal and external users including researchers. The current mission aims to outline some of the first steps of the process both on a conceptual and a practical level.

In the preparation of this mission it has been emphasized that role of the MS experts should be to contribute to set up specific deliveries which can be implemented in a stepwise approach with each step implemented being operational. The creation of a complete data catalogue linked to the associated documentation requiring both technological and process changes and therefore might take some time to implement and establish. However, if the structure and content of the catalogue is carefully described and harmonized across domains, many producers of statistics from Subject Units can contribute at the same time.

An important result of the mission was to achieve a common understanding and consensus about a proposal outlined by the MS experts.

2.2 Project Initiating Document (PID)

To ensure that the planned activities will be implemented and that the goals will be met it is recommended to establish a project dedicated to establishing a data catalogue with integrated documentation, adequate resources and clear leadership.

It is recommended by the MS experts that the first step should be to write a Project Initiating Document (PID) that will be presented to Top Management of the ICBS for approval. The PID should be used in order to ensure that required resources necessary for the implementation of the project are allocated as well as a common understanding and commitment for the project's existence at an institutional level,

It is recommended that the common PID template implemented in the ICBS should be used. However, on the next page some points that the MS experts find important to raise as part of the PID:

- **Project period** – Time Assessment (length of the project).
- **Project Goals and objectives** – Integration with organizational goals/ connection to organizational goals. Provide a realistic number of subject areas, datasets and variables to be included in the data catalogue. As part of setting the goals it is recommended to create a list of the most demanded datasets and variables over the last couples of years.
- **Scope** – e.g. degree of detail of the documentation.
- **Project Organization** – appointment of project owner (responsible for making the required resources available to the project), Project Leader, Steering Committee and Working Groups. For further details please see section 2.2.2.
- **Business Case** – Feasibility and cost/benefit. For further details please see below.
- **Constraints** – e.g. available resources, dependencies regarding the development of a supporting IT system.
- **Stakeholders** – internal and external users, stakeholders.
- **Interfacing systems** – current and future systems.
- **Risks** – must be described and prioritized in order to raise awareness of the highest risks
- **Milestones** – Milestone need to be SMART (Specific, Measurable, Attainable, Relevant, and Time-bound).
- **Reporting** - Specifying type and frequency of reporting frameworks.

The PID must be adjusted to the needs of researchers and other users of microdata from the ICBS – but the PID must also consider the extent to which it is realistic to implement the project during the project period. The scope of the project depends on a number of conditions, most of all availability of sufficient resources from the Subject Units who are suppliers of input to the documentation and the data catalogue. Required resources should cover the working time and qualifications needed.

An important part of the **Business Case** is to analyze the financial consequences of implementing the project. The cost of developing the documentation and the catalogue depends on the scope of the project. However, the main aim for the project is to streamline the production process in the RSU when servicing researchers. The documentation and data catalogue will contribute to a more lean and harmonized handling of researchers' demands for data. The expected savings from the streamlining must be estimated so that the net-costs can be calculated. The proposed list of the most requested datasets will help highlight the costs and benefits when preparing the business case.

2.3 Organizational structure

On basis of the PID the ICBS Top management must decide on the scope for the project and after the approval of the Business Case, the project can be initiated. This entails:

- **Steering Committee.** The Steering Committee must be officially designated. It is recommended to establish the Steering Committee with representatives of the main stakeholders including important (external) users e.g. members from the Eckstein Committee and representatives of important suppliers (internal) from Subject Units and the Research Service unit. The representatives should be selected on the basis of a stakeholder analysis.

Support to the Israeli Central Bureau of Statistics in Improving the Quality of Official Statistics

- *Working groups.* It is recommended to establish working groups on an ad hoc basis when there is demand e.g. regarding:
 - Establishing documentation guidelines
 - Compiling experiences from Subject Unit pilots in order to adjust the dataset catalogue and documentation framework.

It should be noted that developing and implementing a remote access system for researchers admittance to the data catalogues and the variable documentation is out of the scope for this mission. Nevertheless, this task is necessary for the exploitation of researchers use of the metadata and could be done in parallel to the establishment of the data catalogue.

2.4 Policy for a data catalogue and documentation

A policy is necessary in order to build a standardized framework for the data catalogue and the documentation. The target groups for the policy and the frameworks are (i) internal suppliers of information for the data catalogue and the documentation as well as (ii) the users of the data catalogues and the documentations.

As part of the current mission the participants from Subject Units, experts from Statistics Denmark and the Twinning Staff came up with the following draft for policy to be considered as a preliminary starting point:

- The data catalogue should support data from (i) the statistical program at ICBS, (ii) special deliveries of data from other ministries and agencies (iii) research datasets created for former projects (for re-use of data).
- Requirements for documentation need to be defined for each of the three data groups listed above.
- A naming convention for both datasets and variables needs to be agreed on and described. The naming must be uniform and harmonized over time. Variables with different content must be named differently e.g. age, if different status dates are applied, for example in census and in crime statistics (Age_census, Age_crime etc.).
- Subjects Units are responsible for the documentation of datasets and variables including keeping the documentation updated.
- RSU do not own the data – these are owned by subject units or external owners. The RSU is only responsible for developing and maintaining the data catalogue.
- Resources in the RSU must be allocated for maintenance of the catalogue.

2.5 The Catalogue

During the first day of the mission the MS experts made a presentation of the data catalogue produced and used on a daily basis by the researchers and the RSU at Statistics Denmark (Annex B7 – 5). The presentation included;

- The data policy for the data catalogue at Statistics Denmark, focusing on content considered to be most relevant for the BC, harmonization of dataset names and variable names over time within the datasets over time
- The purpose and the content of the data catalogue
- A demonstration of the data catalogue for the internal and external users
- The link between the data catalogue to the Statistical Documentation (SIMS) and the High Quality variable documentation
- The organizational framework for the maintenance of the data catalogue

On the second day two pilots were held with participation of the BC and the MS. The main purpose of the pilots was to try to fill in draft catalogue templates, that the MS experts had produced as part of their preparation before the mission for both the dataset and the variables in order to evaluate and adjust the templates to the Israeli needs and settings (Annex B7 – 6 and Annex B7 – 7).

At the first pilot the subject unit was represented by Ms. Hadas Yaffe from Demography and Census Department and for the second pilot the subject unit was represented by Ms. Orly Furman from Micro Economic Statistics Department. It was decided after the first workshop only to follow through with two workshops instead of three (as originally planned) in order to ensure enough time to address the many questions raised during the pilot.

Creation of the catalogue will assist the BC in getting an overview of datasets and variables available for the researchers. The form of presentation of the data and variables catalogue on the ICBS homepage is to be decided on the future.

The following activities still remain to be completed:

1. List the most demanded datasets and variables

Additional pilots are recommended to be performed in order to ensure that all possible issues and challenges have been addressed thoroughly. It is recommended to perform the additional pilots with the supplier of the most requested datasets. Thus listing the most requested datasets will become part of the preparation of building the data catalogue in the future since the recommendation is to start with the most demanded datasets and variables.

2. Perform additional pilots

It is recommended to perform additional pilots in order to ensure thorough testing of proposed content and structure of the data catalogue. The pilots should, if not already represent in the list of the most demanded datasets also to include surveys in order to ensure the most optimal coverage.

The pilots should include representatives from the RSU and the Subject Units, producing the most demanded dataset and variables originating from surveys as well as administrative files. The pilots will result in a further clarification of the structure and content for a future catalogue.

3. Draft the final content and structure of the catalogue

It is recommended that the draft include guidelines for harmonizing dataset names and variable names over time

It is recommended that the dataset catalogue will be organized in groupings corresponding to the first three levels of the new hierarchical subject structure outlined for the new ICBS webpage. The assignment to groupings of the new hierarchical subject structure will enable sorting the catalogue into subjects at different levels and thereby assist the researcher in getting a quick overview of available datasets/variables in a specific area of interest.

Harmonization of the dataset names and the variable names over time and across domains was of great interest for the BC and was debated on at both pilots. The MS experts recommend that part of establishing the catalogue will consist of harmonization of datasets and variable names over time. The harmonization should be part of the policy of the data catalogue.

It is also recommended in the future to ensure a link between the variable catalogue and the variable documentation. In this way documentation of the variable will become obtainable from the variable catalogue ensuring a direct connection between the variable catalogue and the variable documentation.

4. Write catalogue guidelines for Subject Units

It is recommended to prepare guidelines for the Subject Units to assist them in filling in the content of catalogue. In this way standardization is assured regarding the content of the catalogues which will be of great benefit for both internal and external users.

Also, when an IT-system will be developed to support the Research Services and/or a general metadata system will be implemented, it will be of great importance that the documentation is made in a homogeneous structure as this saves resources and simplifies the process.

5. The revised and approved catalogue templates are filled in by the subject units (Implementation phase)

The pilots will result in a clarification of the structure and content of the catalogue including input for guidelines for filling out the templates.

It is recommended that the Subject Units responsible for the most demanded datasets and variables will be selected as the first Subject Units to fill in content of the catalogue, however, naturally respecting resources and other obligation of the subject unit.

The completion of the catalogue must be in compliance with the convention of naming described in the data catalogue policy.

2.6 Documentation of datasets

The MS experts recommended that datasets are documented using the SIMS format since the ICBS is currently in the process of implementing SIMS in the organization. This means that approval for this aspect of the project is already granted by the top management and that the organizational structure and resources for this part of the project are in place. There will also be the clear benefit that, if planned in a clever way, the dataset documentation part of the catalogue project will not require additional work from the subject units since SIMS will then serve multiple purposes – supporting the principle of “Write once –use many times”. Naturally, in order to obtain success, close coordination and cooperation between the SIMS project and catalogue project is a prerequisite. However, since both projects are headed by Ms. Sigalit Mazeh from the International Relations and Statistical Coordination Department the coordination between the two projects is expected to be relatively straight forward.

As preparation for the mission the MS experts prepared template and guidelines for documentation of datasets. The template took outset in the SIMS front page fields with a few additional fields implemented from the catalogue provided at the RSU at Statistics Denmark (Annex B7 – 8). However, due to the tight program, the template was not tested as part of the mission. It was regarded as less important since the ICBS already has a lot of knowledge and experience in SIMS. However, at the mission, the pros and cons of requesting filling out a full SIMS (all fields) in contrast to requesting only filling in the front page fields were discussed. In addition, requirement for data that is part of the statistical program in contrast to other data was discussed. It was generally agreed to differentiate between documentation requirements for data that are part of the statistical program of the ICBS and those that are not. For data that are not part of the statistical program a description of the source and population should be aimed for as a minimum.

It is recommended that when introducing the catalogue to the subject units, emphasis should be made on the expected long term benefit gained per unit by integrating the SIMS project to the catalogue project and the benefit of reusing metadata instead of tailor made metadata. When the SIMS project is fully implemented – the subject units should be able to entirely avoid compiling documentation of research data sets.

The following activities still remain to be completed:**1. Linking the data catalogue to SIMS**

The interface between the datasets and the Single Integrated Metadata Structure (SIMS) can be ensured by hyperlinks in the catalogue of datasets to the SIMS documentation. In this way documentation of the datasets will become obtainable directly from the dataset catalogue ensuring a direct connection between the dataset and the SIMS-documentation.

2. Determine the interface to SIMS for datasets that are part of the Statistical Programme

In the dataset catalogue a hyperlink to the SIMS will be present for datasets produced as part of the official Statistical Programme. The Subject Units will fill in the hyperlink for each dataset connecting the dataset to the SIMS-documentation relating to the dataset in question. One challenge experienced at Statistics Denmark, due to a lack of coordination between the data catalogue project and the SIMS documentation project is that, although in most cases, there is not always a 1:1 relation between datasets and SIMS. In some cases one SIMS can relate to more than one dataset and in other cases one dataset can relate to more than one SIMS. This challenge must be addressed as part of the project and a clear policy must be defined. If possible a 1:1 relation is recommended by the experts.

3. Determine the documentation requirements for datasets that are NOT part of the Statistical Programme

For datasets not being produced as part of the Statistical Programme no SIMS documentation will be produced. The recommendation is that requirements for documentation of these datasets will be reduced e.g. by only requesting the general technical fields and from SIMS only “Source” and “Population”. A hyperlink to the dataset documentation is to be filled in the dataset catalogue in line with the linking to the SIMS for the dataset that are part of the statistical programme.

2.7 Documentation of variables

In Statistics Denmark variables are documented on two different levels. All variables used as basis for publications or variables exchanged internally, must be documented on a general level. In addition Statistics Denmark also provides “High Quality Documentation” for a number of highly demanded variables for which an external review process by prominent researchers has been performed. Special attention is given to changes over time in these variables e.g. changes in codes, changes in sources or if changes in legislation affect the variable, this has to be carefully described in order for the researcher to take this into account when using the data in their analysis.

During the first day of the mission the MS made a presentation of the High Quality Documentation produced and used on a daily basis by the researchers and by the RSU at Statistics Denmark. The presentation included a presentation of:

- The background and history of the high quality documentation
- The outline of high quality documentation displayed at the RSU homepage on the internet
- The outline of IT system TIMES at Statistics Denmark used by the subject units to fill in both the documentation at a general level and for the high quality documentation of variables.

As part of the two workshops held with participation from the BC and the MS, the participants from the BC tried to fill in drafts which the MS had produced for the variable documentation inspired by the high quality documentation (Annex B7 – 10). The purpose of this was to find out whether the structure was relevant for the BC and including all the information demanded by the researchers.

During the workshops the structure of the high quality variable documentation was discussed. It was generally agreed that the framework presented by the MS experts fitted the need of the BC. If the high

quality documentation template is used by the BC it will assist both the subject units in producing standardized documentation and the researchers when using the documentation. This will result in a more effective working process leaving resources for other tasks.

In the future process, it is essential that already existing documentation will be reused. This can be done by copy-paste as much as possible from existing codebooks and linking to specially prepared documentation. Also, this ensures that available metadata and documentation are centralized and available.

It is also recommended by the MS experts that in the future, the ICBS must create a link between the variable catalogue and the variable documentation. In this way documentation for the variable will become obtainable from the variable catalogue ensuring a direct connection between the variable catalogue and the variable-documentation.

Using the variable documentation will help the researchers in obtaining more knowledge of the variables both when selecting variables for their research project and when analysing the data. In addition it will also benefit the subject units since once the variable documentation has been written by the subject units, they can refer to this and in many cases save time on consulting and creating customized documentation for each individual research project. It is recommended by the MS experts that before any data are published by the subject unit it should be required to complete and update the documentation. In the long run such an institutional policy will release resources for the subject units leaving them time to fulfill other demands. Many NSI's including Statistics Denmark has written procedures / policies requiring subject matter statisticians to update and correct the documentation before new figures can be released. Such procedures are usually either defined in the dissemination or quality policies.

The following activities still remain:

1. Additional Pilots with Subject Units

The template was presented and filled in during the two workshops. The BC found that all the fields of the template to be of great relevance to the ICBS. However, it is recommended that the RSU perform additional pilots in order to test the proposed framework carefully. The MS experts suggest that the participants of the pilots should include the subject units providing the most requested dataset and variables for research projects. During the pilots, existing documentation from codebooks and glossary of terms should be tested in order to evaluate how can be reused in the new framework outlined. The pilots will result with a further clarification of the proposed structure and content of the template as well as guidelines.

2. Agree on a common standard template for variable documentation

Based on the two pilots performed, the BC needs to outline and agree on a common standard template for variable documentation to be presented to the top management. In the process of creating a new standard for variable documentation, the Israeli settings, needs and available sources should be taken into account.

3. Decide if quality assurance of variable documentation should be performed

In Statistics Denmark all high quality documentation of variables goes through a quality assurance process. The quality assurance process consists of (i) an external review performed by two leading researcher of the specific area in question (ii) an internal review by staff members from statistics Denmark as well as (iii) a linguistic review (Please find manual in Appendix B7 – 14).

The resources used for quality assurance are financed through a yearly grant by the Research and Innovation Agency.

During the workshops an ongoing debate took place regarding different ways to go through a quality assurance process as part of the variable documentation. The BC need to consider and decide if quality assurance should be part of the project or not. If quality assurance is decided the process must be described carefully e.g. by having staff from the subject units performing quality assessments to the documentation using written guidelines in this process and/or having current users of the research services in Israel involved in the quality assurance process as is the case in Denmark. The MS experts highly recommend to include quality assurance as a natural part of the documentation process in the future.

4. Draft guideline for variable documentation

A draft of the Guidelines for High Quality Documentation used today in the MS was presented to the BC. It is highly recommended that when choosing the fields and processes to be implemented by the ICBS that guidelines for the subject units will be produced.

2.8 Benefits for researchers and the ICBS

Based on Statistics Denmark's experience from the last decade, the following benefits were presented to the RSU and the ICBS:

- ✓ Documentation of variables is done only once by Subject Units – who also are responsible for keeping the documentation updated
- ✓ Harmonized documentation is user friendly, increases the accessibility and saves time for users
- ✓ Harmonized documentation makes it easier to transfer the information to any new platform
- ✓ Harmonized documentation collects all documentation about a variable in one place, possibly with links to, for example, special quality analyses
- ✓ A data catalogue gives an overview of data-sets and the potential users have the possibility to design their project on their own based on available data and need for the project
- ✓ The RSU may refer to metadata when contacted by researchers and ask them to investigate themselves how their need for data may be met
- ✓ The quality of the documentation of variables may be increased with the help from researchers if the RSU establish a process for it
- ✓ Internal users of data across Subject Units will benefit from the structured data catalogue and the documentation if it is made available for internal use also
- ✓ Greater transparency and visibility helps, for example, to increase the quality assurance with the help of external variables/data when producing statistics
- ✓ Deliveries of datasets for researchers are documented and can easily be reused or updated
- ✓ Uniform naming of datasets and variables will contribute to increasing the transparency of the catalogue.

3. Conclusions and Recommendations

It is recommended by the experts to follow the steps listed below:

1. Organize the process in a formal project structure starting with making a PID (Project Initiating Document)
2. Present the policy to the management
3. Decide on the structure of the content of the catalogue (which datasets, variables)
4. Decide on the harmonization over time and across domains on dataset names and variable names
5. Decide on which level the documentation shall be integrated into the data catalogue
6. Compile guidelines for Subject Units
7. Decide on centralization of the data catalogue in the RSU
8. Decide on maintenance and updating of the data catalogue in the RSU
9. Decide on allocation of resources in order to build and maintain the data catalogue
10. Decide on how the Subject Units are motivated to produce the documentation according to the guidelines
11. Start with the most demanded data
12. Develop an IT-system to support the data catalogue
13. Decide how to involve users (researchers) in developing the IT-system and the quality of the documentation.