# Guidelines for High Quality Documentation (HQD)

Please be aware this is not an official document from Statistics Denmark, but a document translated by the RTAA as part of the Twinning project on Statistics in Israel (IL/12 CRIS 2015/370-467)

### **Preface**

This guide provides a description of the structure, content and workflow for high quality documentation of variables for researchers in Statistics Denmark.

The documentation takes place using the software in Times4 and will therefore refer to the template used here. Examples of "good" documentation are included in the end of the manual.

In addition to the purely technical descriptions, the instructions also contain procedure descriptions for the high quality external and internal quality assurance work.

Statistics Denmark March 2010

### Content

Preface	2
Quality assurance of High quality documentation	4
External assessment	
Internal assessment	
Initial meeting	5
Examination of reviews	5
Preparation of documents in connection with the quality assurance process Documents with external reviewers comments	6
The content of the documentation	7
Example of a good documentation	10
Annex 1 - Checklist for documentation content in high quality documentation	14
Annex 2 – HQD actors	15
Annex 3 – Evaluation sheet	16
Annex 4 – Certification	17

#### Introduction

High quality variables are created as a result of researchers' requests for detailed documentation of the most widely used variables. The project was launched in January 2006, and the first high quality variables were presented on Statistics Denmark's website in April 2006. By 01 January 2010, approximately 500 high quality variables were documented at:

https://www.dst.dk/da/TilSalg/Forskningsservice/Dokumentation

The work on high quality documentation is coordinated by Research Services, which is also part of the external steering committee set up for the High Quality Documentation Project.

### Quality assurance of High quality documentation

Quality assurance of high quality documented variables The quality assurance of the High Quality Documentation (HQD) consists of an external and internal quality assurance including an external and internal assessment. All high quality variables are going through both ratings. The external evaluation takes place before the internal assessment.

The external and internal quality assurance process is described below.

### **External assessment**

External assessment

The external quality assurance consists of evaluating all variables within a single area by two external reviewers. These reviewers are persons who have special insight into the area concerned. The reviewers review the documentation and make suggestions for corrections that they list in a template to be filled in for each variable. The template contains 8 questions that the external evaluators should relate to.

Annexes and general comments

In cases where there are annexes, in the form of documents produced, attached to a variable, the external evaluators will also comment on these. In addition, external reviewers may collect general remarks to a registry area in a single document, which is sent to Research Services together with external assessments.

### Internal assessment

Internal assessment

The internal quality assurance implies that the High Quality Documentation produced by expert from the subject unit is reviewed by an internal professional reviewer and a linguistic reviewer.

The internal quality assurance process is initiated when the two external reviews are available in Research Services. The internal quality assurance process includes the High Quality Documentation administrator, an internal reviewer and a linguistic reviewer.

The internal reviewer are usually from same unit as the one whom has produced the High Quality Documentation or be a former employee who is now employed in another unit. The internal reviewer must be able to assess the academic content of the documentation produced. The internal reviewer must review the drafted high quality documentation carefully as well as ensure that the external reviewer's comments have been incorporated when appropriate in order to ensure a high standard of documentation.

It is the office manager's task to find the internal reviewer; however, the Research Services will assist if needed. If it is not possible to find an internal reviewer, with the required professional skills, the Head of Department must be responsible for the professional assessment of the drafted high quality documentation.

The linguistic reviewer is from the dissemination unit and is involved in order to ensure a readable and comprehensible representation of the high quality documentation so that it can easily be read by people without particular insight into the high quality area concerned.

### **Initial meeting**

Initial meeting

When external assessments are available, a meeting will be held at the Research Services initiative, where the HQD responsible and their Head of the Department, the internal and linguistic reviewers, and the responsible in Research Service Unit will be present. The invitation to the meeting will include external assessments.

Individual process

At the meeting, the content of the external review will be discussed shortly and the further process of internal quality assurance between the three parties involved will be agreed upon. The exact further process is agreed upon for each area and adapted to the individual needs for the area in question. It is up to the individual statistical unit whether the internal or linguistic reviewer is done first or at the same time. However, the assessment will only take place after the HQD responsible has implemented the external assessors' comments in a Word document that corresponds to the documentation (see later).

Results from the initial meeting

The meeting between the parties involved must result in an agreement on how the internal quality assurance process must be conducted and a schedule for when it is completed.

### **Examination of reviews**

Examination of external reviews

When reviewing the external assessments, the HQD responsible must clearly state whether the comments of the external reviewers are incorporated into the documentation or not. This means that the HQD responsible must review the evaluators comments by 1) correcting the documentation so that the comments of the external reviewers are taken into account (the documentation must clearly indicate where and how the external reviewers comments has been met) or 2) provide a reason for why the HQD responsible may not find it relevant to take into consideration.

Once the HQD responsible has incorporated the corrections in the word document corresponding to the documentation, the internal and linguistic assessor must review this.

Research Service recommends that the internal evaluator's comment on the documents before the linguistic evaluation is performed. Any corrections incorporated into the documentation, as a result of the comments of the external and internal assessors, will thus also have a linguistic assessment. This way it will be the final documentation that is assessed linguistic corrected.

Internal reviewer's tasks

When the internal reviewers has reviewed the documentation for a high quality variable as well as the implementation of the external reviewers comments, the internal reviewer ensure that the comments of the external reviewers have been taken into account, or is accompanied with sufficiently substantiated statement that this is not relevant, cf. above.

Once the linguistic reviewer has reviewed the documentation, the HQD administrator should review this again. The HQD responsible and the internal reviwer must

ensure that the linguistic editing does not change unnecessarily with the professional content of the documentation.

Preparation of recommendation

The Head of Department of the relevant statistical offices has the responsibility for internal quality assurance. Once the quality assurance process has been completed, the HQD administrator will inform this to the Head of Department that has to certify that the documentation has the required high and uniform quality, including ensuring that the reviewers comments have been taken into account in a qualified manner or, alternatively, that a well-founded reasoning has been given for why a specific comments has not been taken into account.

For the purpose of the attestation for completion, the Research Services prepares a template in which the must be signed by The Head of Department (Appendix 4)

Completion of process

Finally the external steering committee must approve the documentation of the individual subject areas before it can be considered as fully completed.

Paid by the Research and Innovation Agency The resources used for quality assurance are paid by the Research and Innovation Agency. The employees involved will thus be refunded for the hours spent on high quality documentation as soon as the registry areas are approved by the external steering committee.

### Preparation of documents in connection with the quality assurance process

#### **Documents with external reviewers comments**

For each high quality variable, as previously written, a completed template is received from two external reviewers, explaining their comments on the completed documentation. The HQD administrator can either find the comments relevant to incorporate or not relevant to incorporate.

For comments that the HQD responible does not find relevant to take into account, the HKD responible must provide a sufficiently thorough explanation for this in response to the reviewer's comments. The proofreading feature in Word must be used in order to ensure that all changes can be seen.

The comments that the HQD responsible considers relevant does not need to be commented since, if no comments are added then it is assumed that they have been incorporated.

If there is a document with general remarks produced, the HQD responsible person must comment on these comments as well, i.e. it must be clarified whether the comments have been taken into account and, if not, justified.

### Word-document with final documentation

The corrections made in the documentation of the individual variables must be made in a word document prepared by the Research Service for each individual HQD responsible. The Word document will contain all high quality variables within the specific area in alphabetical order. The Word document is printed based on the Software Times4 and will contain the general and detailed description, long name, and population definition.

Proofreading function

When corrections are added to the document, the proofreading function must be enabled so that it will be possible to see which corrections have been made in relation to the original documentation.

In the word document, each variable will be followed by a checklist for documentation content in high quality documentation. When the HQD responsible considers that the documentation is finally completed, the checklist must be completed as part of the preparation of the documentation.

The final document is handed to the Head of Department as an attachment in order to be certified before being sent to the external steering committee.

#### The content of the documentation

The documentation of High Quality variables takes place using the Software Times 4. Below a screen dump for a high quality variable can be found.



The upper six fields

In the first six fields the following information need to be typed in variable name, Type, length, name of dataset, creation date (done automatically), initials of the responsible person (done automatically), date for last update (done automatically).

Descriptions There are two different fields of descriptions that always have to be filled in: (i) the General Official Description, (ii) Detailed Official Description.

## General Official Description

Here's a brief definition of the variable. The text should provide an immediate picture of what the variable contains without further technical details.

### Detailed Official Description

This description must contain a much more in-depth historical description of the variable.

The content of the variable must be described in details, i.e. it must be explained how the variable is defined and how it is designed. Any calculations must be presented and explained.

If there are special conditions or conditions that apply to the variable, which must be taken into account when using the variable, this should also be mentioned.

It must be specified when the records are missing - and possibly explain the difference between 0 and missing record.

Please also include reference to other variables if relevant.

For the variables, where it is regarded as relevant, the Research Services compiles a table and graph that are attached as attachments to each variable. If there are sequences in those that look "weird" then this should be commented and explained.

### Data breaks between variables

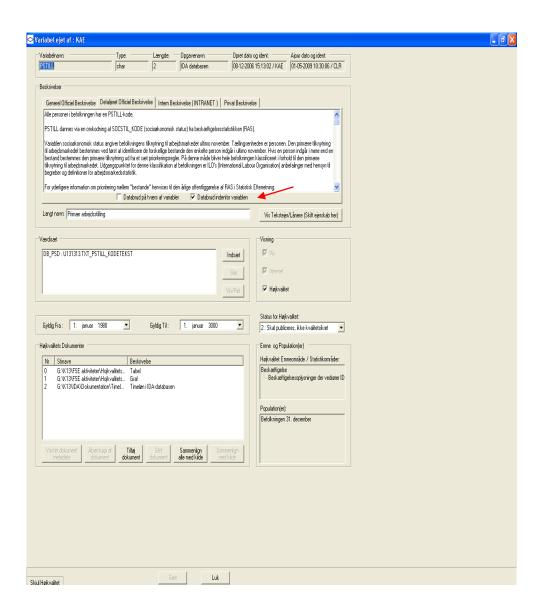
Data breaks across variables occurs when variables are replaced by another variable e.g. due to major changes in definitions or legislation over time. In such cases it has to be marked (see next page). In the detailed description, reference must always be made to the new variable that is replacing a historical variable and vice versa. For example for employment Status, referenced must be made between BESKST (Employment Status using definitions applicable until 2002) and BESKST02 (Employment Status using the 2002 definitions)

### Data breaks within a variable

It is very important that you carefully explain how the definition of the variable may have changed over time. For example, there may be changes in the legislation on which the variable is based, changes in calculations, changes in associated value set (codes text) etc.

Breaks within the variable must be described carefully both "before" and "after" the break.

The detailed description must be written in a language so the text is also understandable to a person who has no prior knowledge about that area. Abbreviations must be explained so that everyone can understand what it is referring to.



Long name In the field "Long Name", the variable label is specified (long explanatory name).

Codes The next box of fields contains the codes and associated text - also called code-text table.

Valid from Please specify the data the variable is available from.

*Valid to* Please specify the data the variable is available to. If the variable is still valid please indicate that the variable is valid until 1. January 3000

Status for HQD This field has three values;

1 = Created, not published

2 = Published, not quality assured

3 = Published is quality assured

Additional documentation

In this box you can attach attachment relevant documents containing detailed information about conditions for understanding the variable. These may be of the type; pdf, jpg, gif, doc, htm, htlm, png and xls

Area Please indicate publication area (Level 1 and level 2)

Population Please define and desribe population(s)

### Example of a good documentation

Short Name HFFSP		HFFSP
Long Name Classification code for highest completed education		Classification code for highest completed education

Creation date	15 January 2015
Last updated	12 May 2017
Owner of the	SMS
variable	Educational Statistics

Data Type	Numeric
Precision	Eight
Length	22

First Year	1981
Last Year	Still valid

Data breach in variable	YES
Data breach across variables	NO

Internal quality	YES
assurance	
<b>External Quality</b>	YES
assurance	

## General description

The classification code for highest completed education has been developed to support comparability of education statistics - both at national and international level. Thus, the code serves as a useful classification system across statistics-producing authorities in the education sector in Denmark, while ensuring a clear connection to the international classification system ISCED. In order to ensure flexibility, the classification code is created in a four level hierarchical structure.

The classification code for highest completed education is 8-digit long. The first two digits describe the level (main group) of the education. The next two digits further provide a professional grouping (intermediate group), which is further detailed in the following two digits (subgroup). Finally, the last two digits indicate the specific training or qualification (element level).

An example of an education could be a carpenter with the 8-digit the following code: 35533520.

Main group: 35 Vocational education

Intermediate group: 3553 Construction

**Subgroup:** 355335 Carpenter and similar

**Elements level:** 35533520 Carpenter

The first two digits of the classification code are the most commonly used groupings for educational statistics published by Statistics Denmark.

The 8-digit code indicates a particular education and therefore corresponds to the 4-digit AUDD code (see AUDD documentation), and the 8-digit HFFSP code thus gives the highest completed education for

each individual of the Danish population. In some cases, however, several UDD / AUDD codes are assembled in the same 8-digit prefixed code. This will be the case if there are more than one AUDD codes for the same education.

From the two first digits of the classification codes the following 12 main groups can be created:

- 10 Primary school
- 15 Preparatory programs
- 20 General upper secondary educations
- 25 Vocational upper secondary educations
- 30 Vocational basic training
- 35 Professional internships
- 39 Further training of specialists / skilled workers
- 40 Short higher education
- 50 Middle-level higher education
- 60 Bachelor
- 65 Longer higher education
- 70 Researcher Programs

Educations in subdivisions 15, 30 and 39 are not used in the creation of the most completed education.

### Detailed description

For research purpose, a dataset is established once a year providing the highest completed education for all individuals in the entire Danish population. The population consists of all individual having a permanent address on 01 January each year, and their education status is recorded on 01 October the year before, which is the latest available educational information.

In connection with the establishment of register for highest attained education (UDDA), only completed programs are included, since the completed education must be compulsory, ie. Competence variable COMP = 1 (Competency Education), 3 (Business Qualification) or 7 (Higher Education). The exception from this is education / grade at primary school level, where completed grade levels are all included regardless of the value of KOMP. However, in order to have a primary education as the highest completed education, it is required that the 6th grade is completed (equivalent to the International Standard Classification of Education ISCED Level 1).

When the register for highest attained education is created the records for people with unknown education is reduced by allocating the previous grade as the highest education for people who has a discontinued general education in Statistics Denmark's student register.

For individuals that, at the status time, has stopped or departed their education in group 10 (Primary school), the previous grade level by 01 June is used.

For individuals that, at the status time, has stopped or departed their secondary education (main groups 20 and 25 the highest completed education is to 9th grade.

The purpose of granting students with discontinued education a highest education is to show that these persons are not unknown in the registers.

Up to and including the school year 2005/06, information on educational level has only been collected from elementary school's 8th, 9th

and 10th grade, which means that the age group from 0 to approx. 14 years will have an insurmountably highest education. As of the school year 2006/07, information is gathered from the kindergarten class, which in 2009 means persons with completed primary education increase.

Information on the highest completed education is limited for persons under 13 years, since a maximum of completed education requires at least 6th grade.

In addition, education information for persons born before 1921 is limited, as there is no information about this group in the 1970 population and housing census (FOB70). FOB70 is the primary source of educational information about the oldest part of population.

#### Description of the allocating process:

The concept of highest completed education is, as the most important criterion, determined in relation to the main groups in the Danish education classification and follows the order described in the general description. In this regard, the following rule applies in the selection of highest completed education:

- 1) Selected by highest group, with groups 20 and 35 being considered equal.
- 2) Selected within the same main group after highest PRIA (= number of cumulative training months).
- Recent education is chosen within the same main group and same PRIA.

The above rule applies with the exception of situations where an education qualification meets another educational qualification from a higher priority source (see HF\_KILDE). Here, the higher priority source will always exceed the level of education.

### Changes over time:

Codes for education and final education (UDD and AUDD) are stable over time; however, the associated classification codes may change over time. An education can be misplaced, and the code will be changed when the error is recognized or an education may have evolved over time so that it will at some point be moved to another level. Therefore, it is recommended use the education code in historical material and subsequently apply the latest classification to all historical material to avoid inconsistencies associated with time series

### **Example of changes over time:**

Nurse

Year 1990: Classification code = 40903010 AUD=5166 Year 2006 Classification code =50903010 AUD=5166

The nursing education has shifted in level from being a short secondary education to being a middle secondary education. Therefore there is change over time in the classification code, while the AUDD codes remained the same.

Also note that education that develops and moves to a new level will normally maintain its education code if there is no urgent reason to differentiate between the original and the new edition of the education. Thus, one does not want to look for nurses in different locations

	in the classification, even though the nurse has developed changed the level. Nurses with the shorter education are lifted u the new, higher level.				
	A reason for distinguishing between the original and the new version of the education will be that the content of the education is changed so much that it is actually a new education.				
	Converting AUDD codes to classification codes:  The AUDD codes can be converted to 8-digit classification codes using a conversion key. The conversion key contains the 4-digit code pleted education code and the corresponding 8-digit classification code.				com-
	New classification codes:  New courses are continuously being created, so classification codes				
	are constantly being updated.				
	In some cases, an education is also moved from a classification to				
	another (se example above). This can happen if an education has been				
	misplaced, or if the education may switch level. If it happens, a new				
	classification code will be created and the original classification code will be removed				
	will be removed				
	Sources for creation of classification codes:				
	Under the variable HF_KILDE, the sources appear, and the year from				
	which they have been used.				
Annexes	The education manual				
Population	Danish population				
Population de-	The population consist of all individual having a permanent address on				
scription	01 January each	n year			
Codes			T		
	Code	Text	Valid from	Valid until	
	10021010	Elementary	01-01-1981	Still valid	
		school – 1 grade			
	10022010	Elementary	01-01-1981	Still valid	
		school – 2 grade			
	•	•	•	•	
	•	•	•	•	

# Annex 1 - Checklist for documentation content in high quality documentation

The list below is part of the quality assurance process and must be completed as part of the documents to be delivered to the Head of the Department.

The list is used as a check in order to ensure that all points are met for "good" documentation.

All fields are required to be filled out - even if the point is irrelevant to that variable. X for completed documentation and - for irrelevant.

Documentation content	Please indicate (+/-)
Long name (label)	
General description (short definition)	
Detailed Description - Including; - content (definition) - construction (possible calculations are described) - is the meaning of missing indicated - references to other variables - Description of special conditions for that variable - data break across variables - Data break within the variable described	
Values (code text data set / table)	
Population description approved	
Validity periods approved	
Text to Statistics Area Approved	

### Annex 2 - HQD actors

### **HQD-respsonsible**;

The professional employee who has prepared the HQD documentation in the statistics office.

### Internal reviewer;

An additional professional employee who is capable of assessing the quality of the documentation, thus ensuring a high standard. The internal reviewer, who may be from the same statistical office or be a former employee who is currently employed in another statistical office, must be familiar with the academic content in the field.

### Linguistic reviewer;

A staff member from the dissemination unit

### **External reviewers**;

Persons who are not employed in Statistics Denmark and who have special insight into the area concerned (experts in the field).

### **Internal steering committee;**

Office managers from the involved statistical offices and from the dissemination office, chairman of the steering committee, employees from Research Services

### External steering group;

Steering Group Chairman Hans Hummelgaard, Helena Skyt Andersen and Charlotte Leolnar Reif

### Annex 3 - Evaluation sheet

#### Variabelnavn: XXXXXX

- 1. Is the content of the variable defined sufficiently?
- 2. *If there are data break, is it explained carefully?*
- 3. Is there a clear definiation of the population demarcation?
- 4. Is the variable set of values clear to the reader?
- 5. Does graphs and tables give a good and a quick overview of changes over time?
- 6. Is the text written in an easy-to-understand way, so that users without prior knowledge of that specific variable can read the text without any problems? Otherwise, if so, please make suggestions for changes directly in the text (using track-changes)?
- 7. Are there any other matters regarding the documentation of the respective variable that you wish to comment on?

### **Annex 4 – Certification**

It is hereby certifying with my signature that the procedure have been completed for	s for quality assurance
Step of Quality Assurance Step	Please cross
External reviews examined	
Internal reviews examined	
Linguistic reviews examined	
This certificate is attached to the HQD responsible documentatings for the register area  The reviewer's comments have been taken into account in a professional well-founded reasoning has been given if commes are not regarded relevant. Attached are annexes where cappears.	a qualified manner or a nents/suggested chang
Deapertment Head	