2008/6

Documents

Statistics Norway

Statistisk sentralbyrå

Anne Gro Hustoft and Jenny Linnerud

Managing metadata in Statistics Norway

Department of Management Support/Statistical Methods and Standards

Joint UNECE/Eurostat/OECD work session on statistical metadata (METIS)
(Luxembourg, 9-11 April 2008)

Managing metadata in Statistics Norway

I. INTRODUCTION

1. Statistics Norway's metadata strategy was approved in 2005. The strategy as such advocates for metadata(systems) in Statistics Norway , and several measures were recommended to support this. Two of the recommendations were to document definitions of key metadata concepts, and to connect variables, classifications and file descriptions within a metadata portal. In addition the need for clear roles and responsibilities was stressed.

2. Development of the metadata portal began late 2005 and user testing was completed in 2007. The portal was released on the Internet in February 2008.

3. The portal displays our key metadata concepts and the contents of our master metadata systems making them more accessible and easier to use both for researchers and external metadata experts, and for internal users. Metadata managers can use the portal to follow up the coverage and quality of the contents of the underlying metadata systems. The design is flexible so that the contents of other metadata systems, e.g. a questionnaire server, can be added as these become available.

II. Implementation of recommendations from the metadata strategy

A. Definition of key metadata concepts

1. During our work on a metadata strategy for Statistics Norway, it became clear that while our statistics are based on a good common understanding of concepts, we lacked formal definitions for the most central metadata concepts. This was a potential source of misunderstandings amongst people working in different subject matter areas or on different types of tasks. This again could lead to misunderstandings and lower efficiency and quality when exchanging data and metadata and filling up Statistics Norways metadata or metadata-driven systems such as Stabas (classifications) and Vardok (definitions of variables) or StatBank (dissemination of tables).

2. The establishment and documentation of key concepts related to metadata is an important part of the implementation of our metadata strategy. It was made in close contact with those working in statistical methods, IT, production and dissemination of statistics. The document was subject to a hearing round in all these departments. The document has also been discussed in our metadata forum, in the steering group for our metadata strategy and in our standards committee. Finally, it was approved by the director general.

3. The documentation work started by looking at international definitions and common practise in Statistics Norway and elsewhere. The SDMX-initiative has collected together many definitions used by international organisations such as Eurostat, the European central bank, BIS, OECD, IMF, UN and the World Bank. When relevant for our definitions the source is mostly listed as SDMX even if this was not the primary source. In the few cases where we did not find a definition in SDMX, or the one we found did not seem relevant for our purposes, we made our own definitions (e.g. statistical metadata, indicator, register). The first version of this document was produced in Norwegian for use within Statistics Norway but it has later been translated to English in a reduced version more appropriate for an international audience.

4. Concepts listed in the document are grouped under statistics, quality, metadata and registers and thereafter concepts related to the units that statistics builds upon and their properties.

5. A useful, ongoing work is the discussion and recommendation on how these concepts are to be used in the practical, everyday work of filling up our metadata systems. The key concepts are usually a part of the agenda when metadata issues are discussed with the subject matter divisions. It takes time to implement a "new" vocabulary (traditionally other terms have been used for some of these concepts), but the document on our key concepts, which is officially approved, is very useful in this respect.

The figure below shows the concepts in the order in which they are presented in our key concepts document.

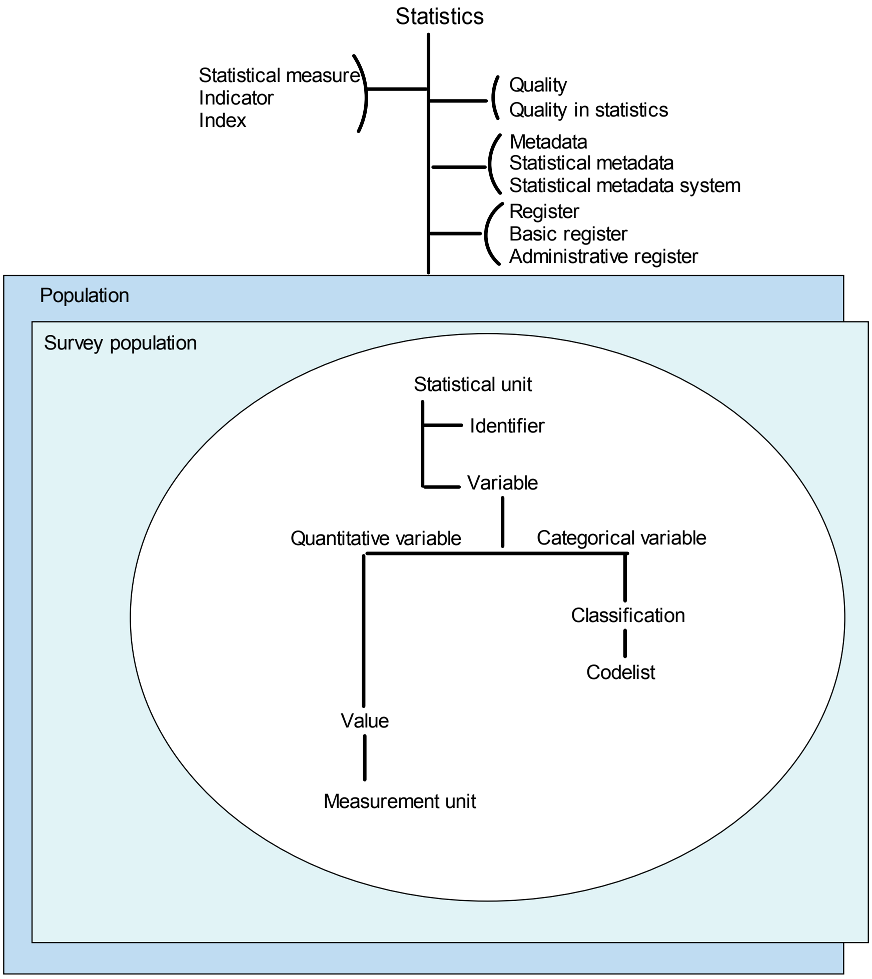


Figure 1. Defined concepts

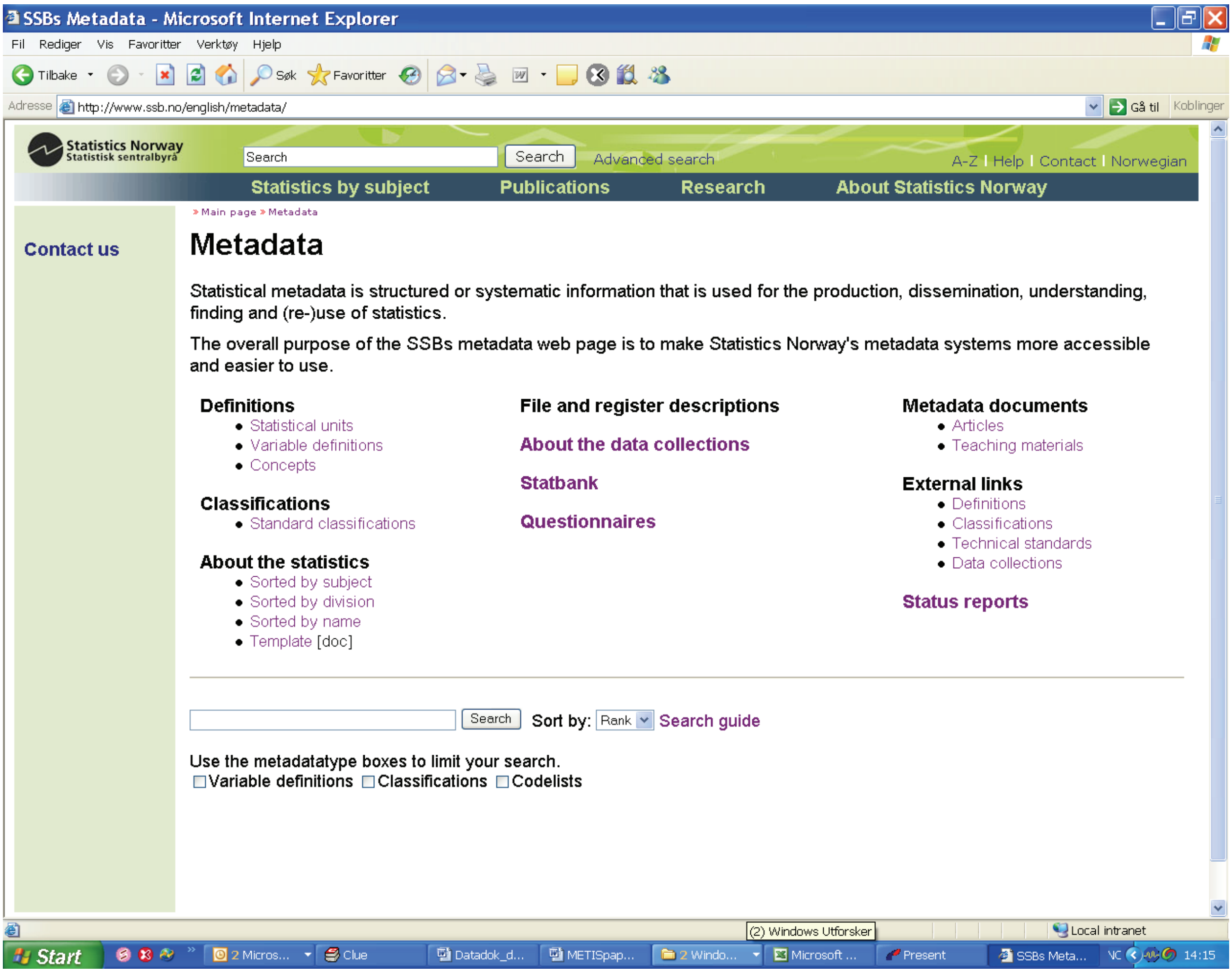
B. Metadata portal

1. The overall purpose of the metadata portal is to make Statistics Norway's metadata systems more accessible and easier to use. Both internal and external users get easier access to the metadata by displaying the contents of these systems in a common web page. Our work within this area has been inspired by the corresponding web pages of Statistics Canada (www.statcan.ca/english/concepts) and Statistics New Zealand (www.stats.govt.nz/statistical-methods/).

2. The main purpose of the metadata portal is to give access to information stored in the metadata systems and delivered by web services, but the page also contains links to other relevant metadata. At present the portal gives access to classifications, variable definitions, codelists, file descriptions, register descriptions and file variables collected from our different metadata systems. The file descriptions, register descriptions and file variables are only shown in the internal version, among other things because they contain sensitive information.

3. In addition to the internal version there is an Internet version both in Norwegian and English. At the national level this can make a contribution to semantic interoperability. At the international level we hope this will make a contribution to fruitful discussions on standardisation and harmonisation of metadata.

4. The metadata portal also contains metadata that are not yet stored in metadata systems (e.g. definitions of statistical units), links to other relevant Statistics Norway web pages (e.g. About the statistics and Statbank) and external links to relevant international metadata web pages. The home page of the Internet version of the metadata portal in English is shown below:



Metadata

Statistical metadata is structured or systematic information that is used for the production, dissemination, understanding, finding and (re-)use of statistics.

The overall purpose of the SSBs metadata web page is to make Statistics Norway's metadata systems more accessible and easier to use.

Definitions
- Statistical units
- Variable definitions
- Concepts

Classifications
- Standard classifications

About the statistics
- Sorted by subject
- Sorted by division
- Sorted by name
- Template [doc]

File and register descriptions

About the data collections

Statbank

Questionnaires

Metadata documents
- Articles
- Teaching materials

External links
- Definitions
- Classifications
- Technical standards
- Data collections

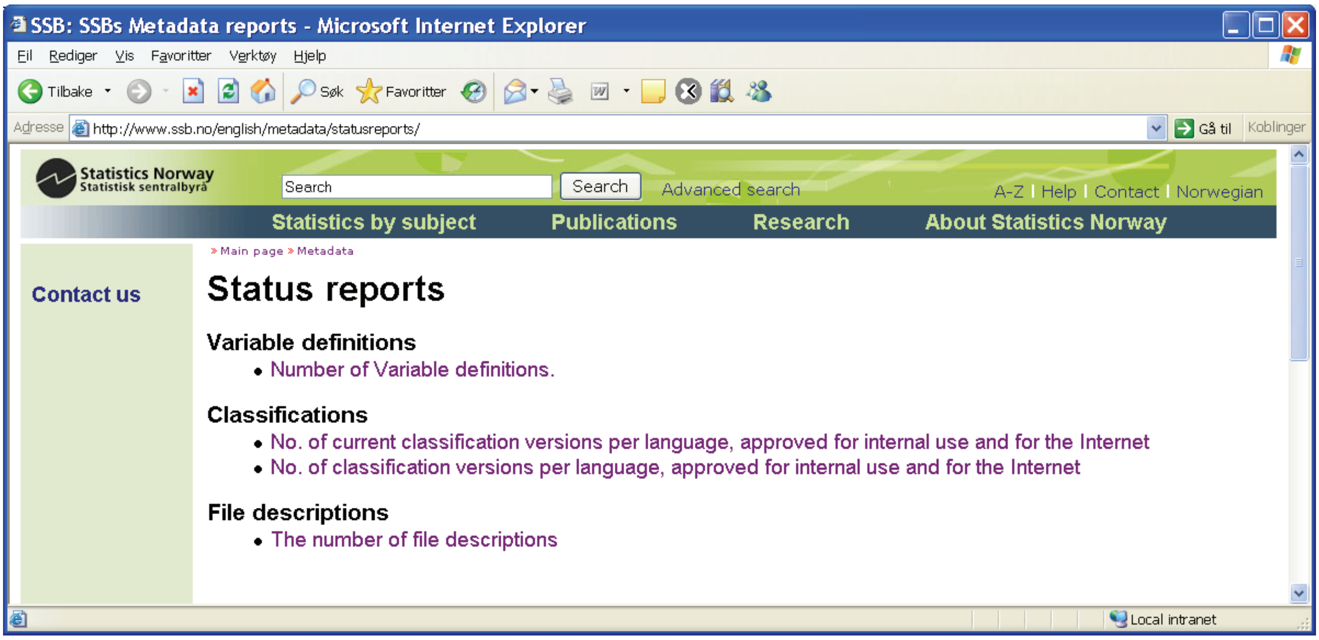
Status reports

Search Sort by: Rank Search guide

Use the metadatatype boxes to limit your search.
□ Variable definitions □ Classifications □ Codelists

Fig. 2 Metadata portal – home page

6. The metadata portal makes it possible for managers to check the progress of the metadata work. The status reports in fig. 4 are available on the Internet, while the internal version has several more status reports to satisfy the needs of internal users (e.g. Number of variable definitions by language, Number of variable definitions by subject area, Number of variable definitions linked to file descriptions, Number of variable definitions linked to tables in Statbank and Classification versions approved for internal use, but not for external use – at present a total of 25 status reports).



Status reports

Variable definitions

- Number of Variable definitions.

Classifications

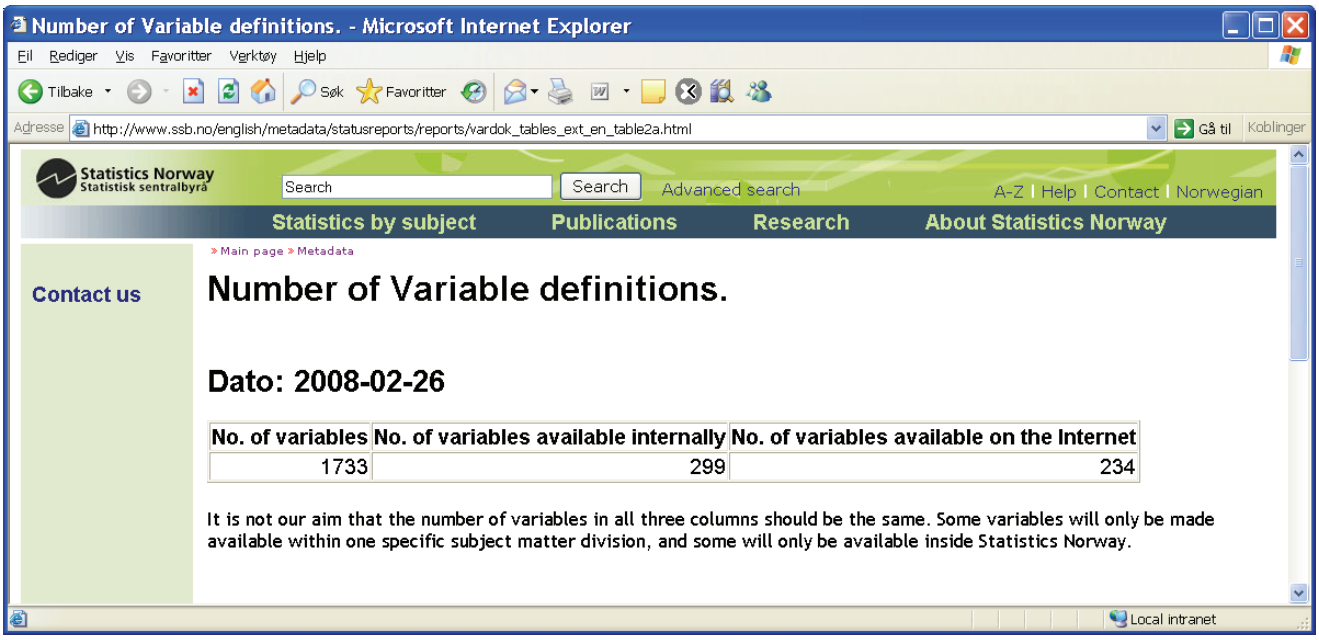
- No. of current classification versions per language, approved for internal use and for the Internet
- No. of classification versions per language, approved for internal use and for the Internet

File descriptions

- The number of file descriptions

Fig. 4 Available status reports

7. If we choose the status report for variables, we get the following information:


Number of Variable definitions. - Microsoft Internet Explorer
Fil Rediger Vis Favoritter Verktøy Hjelp
Tilbake
Søk
Favoritter
Adresse http://www.ssb.no/english/metadata/statusreports/reports/vardok_tables_ext_en_table2a.html
Gå til
Koblinger
Statistics Norway
Statistisk sentralbyrå
Search
Search
Advanced search
A-Z | Help | Contact | Norwegian
Statistics by subject
Publications
Research
About Statistics Norway
» Main page » Metadata
Contact us
Local intranet

Number of Variable definitions.

Dato: 2008-02-26

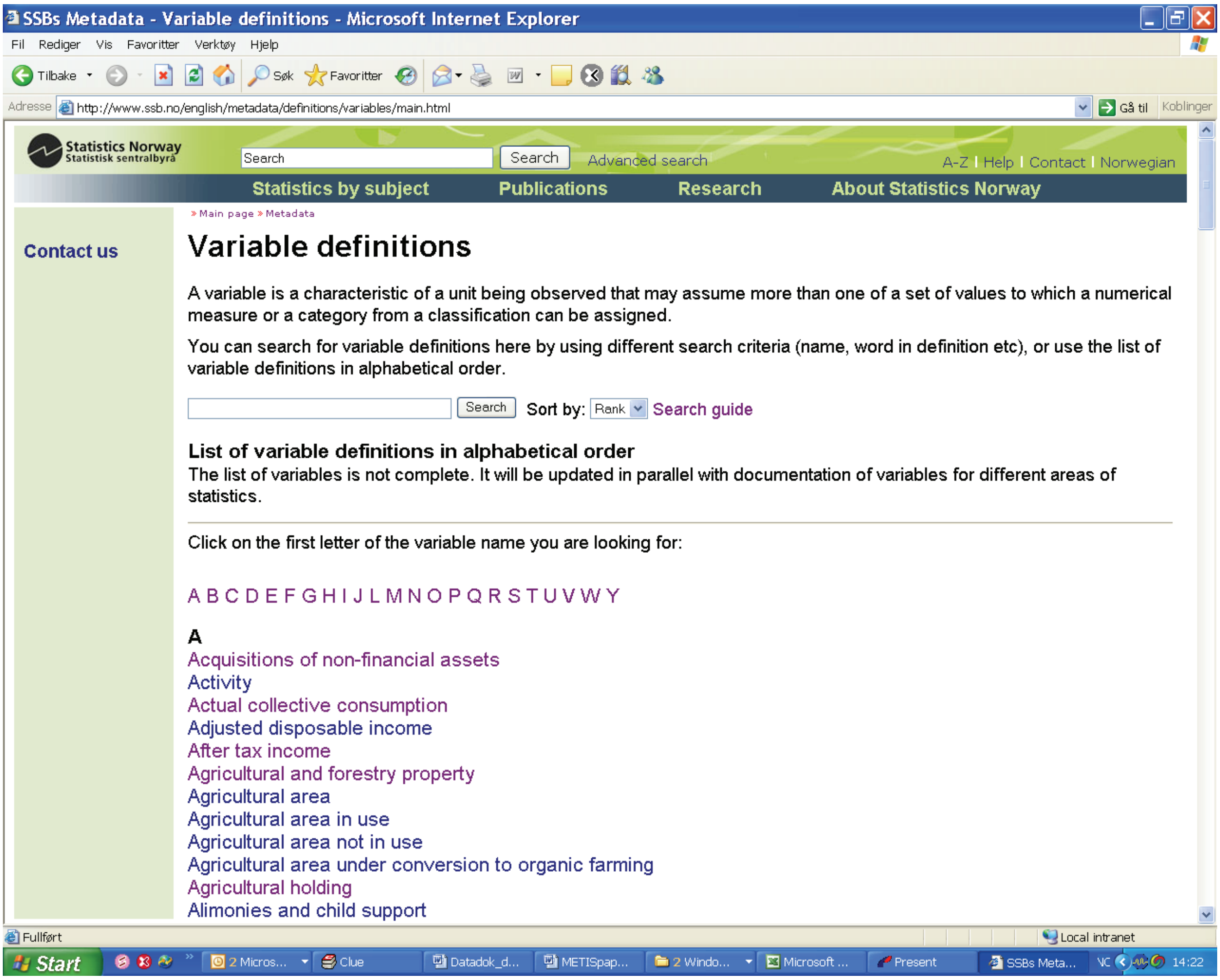
No. of variables	No. of variables available internally	No. of variables available on the Internet
1733	299	234

It is not our aim that the number of variables in all three columns should be the same. Some variables will only be made available within one specific subject matter division, and some will only be available inside Statistics Norway.

Fig. 5 Status for Variable definitions

8. The second and third column give the number of variables available in English (total number of variables in Norwegian is 1733) so there is still some translation work to be done. In the internal portal version the status for the different metadata systems is given per subject matter division which makes it a more relevant tool for the managers.

9. If the user is searching for a variable definition, they can click on the Variable definitions link, and they are then linked to the window below where they can search for a variable definition by using different search criteria (name, word in definition etc.), or use the list of variable definitions in alphabetical order..



Variable definitions

A variable is a characteristic of a unit being observed that may assume more than one of a set of values to which a numerical measure or a category from a classification can be assigned.

You can search for variable definitions here by using different search criteria (name, word in definition etc), or use the list of variable definitions in alphabetical order.

Search Sort by: Rank Search guide

List of variable definitions in alphabetical order

The list of variables is not complete. It will be updated in parallel with documentation of variables for different areas of statistics.

Click on the first letter of the variable name you are looking for:

A B C D E F G H I J L M N O P Q R S T U V W Y

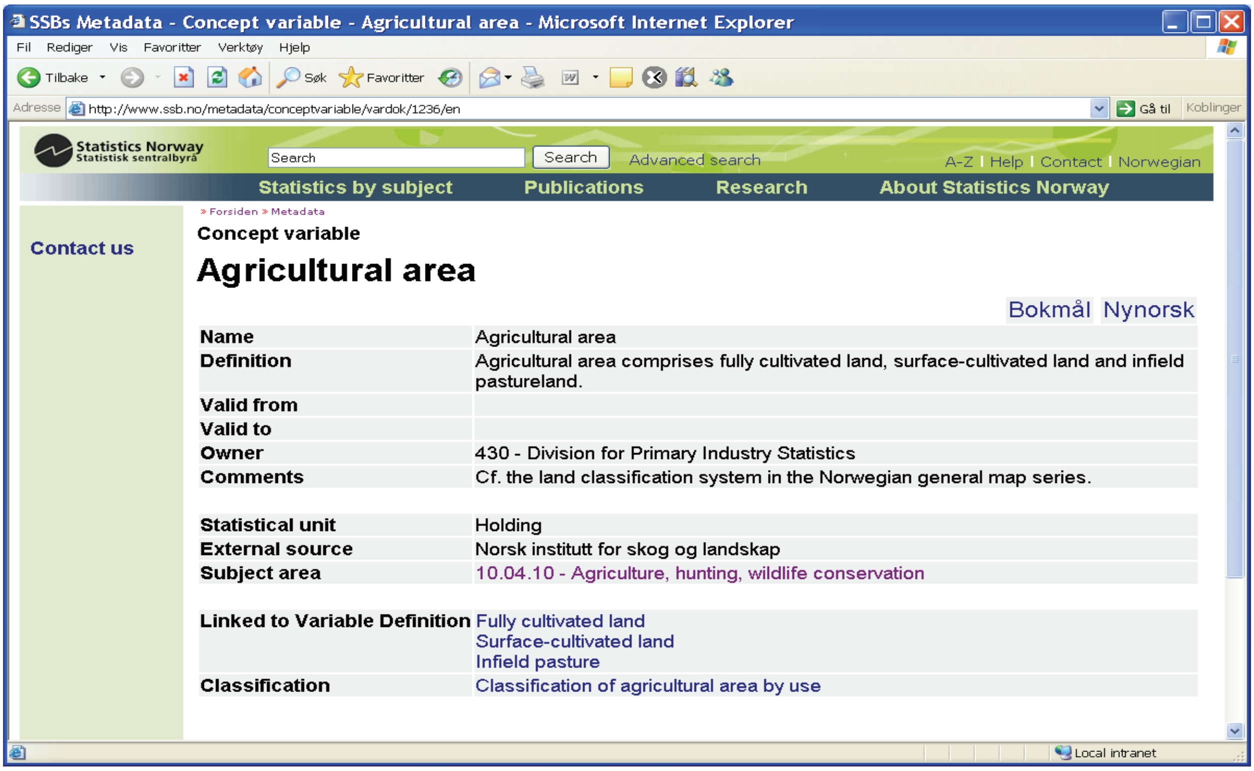
A

Acquisitions of non-financial assets
Activity
Actual collective consumption
Adjusted disposable income
After tax income
Agricultural and forestry property
Agricultural area
Agricultural area in use
Agricultural area not in use
Agricultural area under conversion to organic farming
Agricultural holding
Alimonies and child support

Fig. 6 Variable definitions

10. Some variables have the same name, but are defined differently (different laws/regulations require the same variable name but define the variables differently according to different subject areas). This is visible in the alphabetical list where you can see the number of definitions belonging to the same variable name in parenthesis behind the variable if this is greater than one. Sometimes, however, these duplicates do not arise from real differences, but from errors, and we have been able to discover some errors by checking this list.

11. If we choose agricultural area as our variable, we get the information in fig.7. The external source is the institution that is the origin of the definition (if the variable has not been defined by Statistics Norway). Often a definition will mention other variables (fully cultivated land, surface cultivated land and infield pasture land) and the definitions of these variables are linked in the “Linked to Variable Definition”-field.



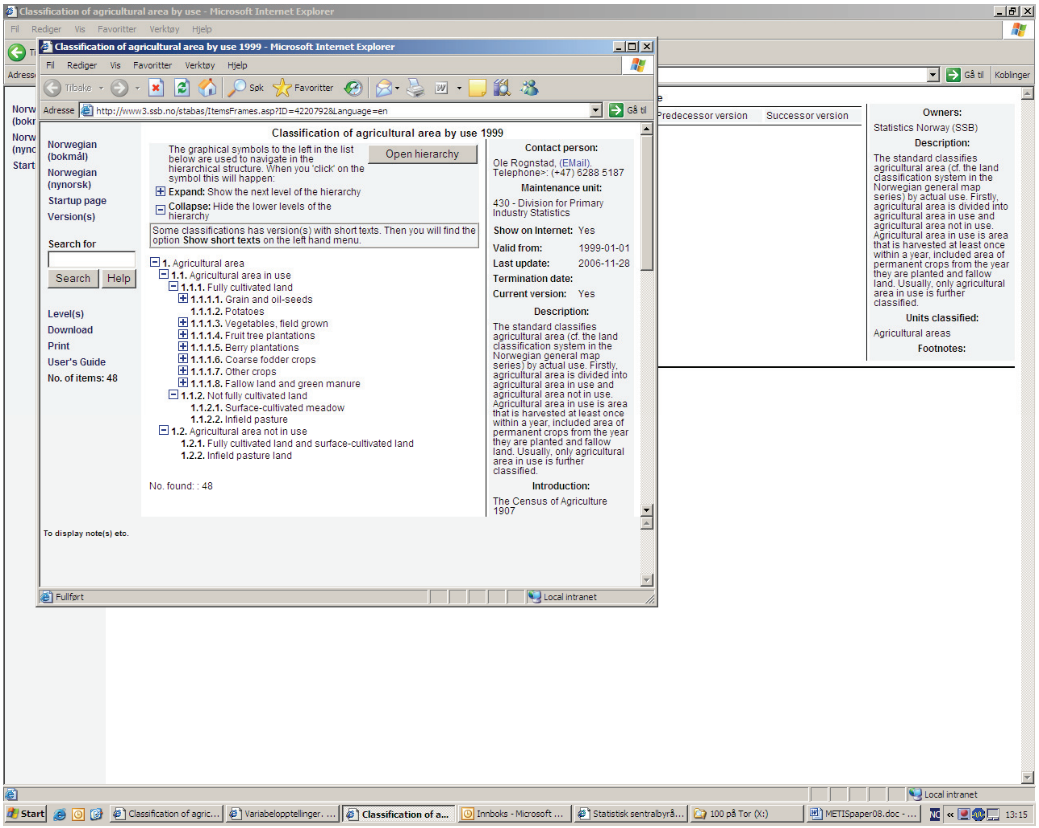
Concept variable

Agricultural area

Name	Agricultural area
Definition	Agricultural area comprises fully cultivated land, surface-cultivated land and infield pastureland.
Valid from	
Valid to	
Owner	430 - Division for Primary Industry Statistics
Comments	Cf. the land classification system in the Norwegian general map series.
Statistical unit	Holding
External source	Norsk institutt for skog og landskap
Subject area	10.04.10 - Agriculture, hunting, wildlife conservation
Linked to Variable Definition	Fully cultivated land Surface-cultivated land Infield pasture
Classification	Classification of agricultural area by use

Fig. 7 Documentation of agricultural area

12. If the variable chosen is a categorical variable, it will be linked to the relevant classification or classification version in the Classification database, and by clicking on the link in the Classification field, the user will get access to this (see fig. 8). By using the two clickable links at the top right in fig 7 the user can also get access to the variable documentation in both versions of Norwegian (Bokmål and Nynorsk).



Classification of agricultural area by use - Microsoft Internet Explorer
Classification of agricultural area by use 1999 - Microsoft Internet Explorer
http://www3.ssb.no/stabas/ItemsFrames.asp?ID=4220792&Language=en
Norwegian (bokmål)
Norwegian (nynorsk)
Startup page
Version(s)
Search for
Search
Help
Level(s)
Download
Print
User's Guide
Open hierarchy
Predecessor version
Successor version
Fullført
Local intranet
Start

No. of items: 48

Classification of agricultural area by use 1999

The graphical symbols to the left in the list below are used to navigate in the hierarchical structure. When you 'click' on the symbol this will happen:

⊞ **Expand:** Show the next level of the hierarchy

⊟ **Collapse:** Hide the lower levels of the hierarchy

Some classifications has version(s) with short texts. Then you will find the option **Show short texts** on the left hand menu.

```
⊟ 1. Agricultural area
  ⊟ 1.1. Agricultural area in use
    ⊟ 1.1.1. Fully cultivated land
      ⊞ 1.1.1.1. Grain and oil-seeds
        1.1.1.2. Potatoes
      ⊞ 1.1.1.3. Vegetables, field grown
      ⊞ 1.1.1.4. Fruit tree plantations
      ⊞ 1.1.1.5. Berry plantations
      ⊞ 1.1.1.6. Coarse fodder crops
      ⊞ 1.1.1.7. Other crops
      ⊞ 1.1.1.8. Fallow land and green manure
    ⊟ 1.1.2. Not fully cultivated land
        1.1.2.1. Surface-cultivated meadow
        1.1.2.2. Infield pasture
  ⊟ 1.2. Agricultural area not in use
      1.2.1. Fully cultivated land and surface-cultivated land
      1.2.2. Infield pasture land
```

No. found: : 48

To display note(s) etc.

Contact person:

Ole Rognstad, (EMail).
Telephone>: (+47) 6288 5187

Maintenance unit:

430 - Division for Primary Industry Statistics

Show on Internet: Yes

Valid from: 1999-01-01

Last update: 2006-11-28

Termination date:

Current version: Yes

Description:

The standard classifies agricultural area (cf. the land classification system in the Norwegian general map series) by actual use. Firstly, agricultural area is divided into agricultural area in use and agricultural area not in use. Agricultural area in use is area that is harvested at least once within a year, included area of permanent crops from the year they are planted and fallow land. Usually, only agricultural area in use is further classified.

Introduction:

The Census of Agriculture 1907

Owners:

Statistics Norway (SSB)

Description:

The standard classifies agricultural area (cf. the land classification system in the Norwegian general map series) by actual use. Firstly, agricultural area is divided into agricultural area in use and agricultural area not in use. Agricultural area in use is area that is harvested at least once within a year, included area of permanent crops from the year they are planted and fallow land. Usually, only agricultural area in use is further classified.

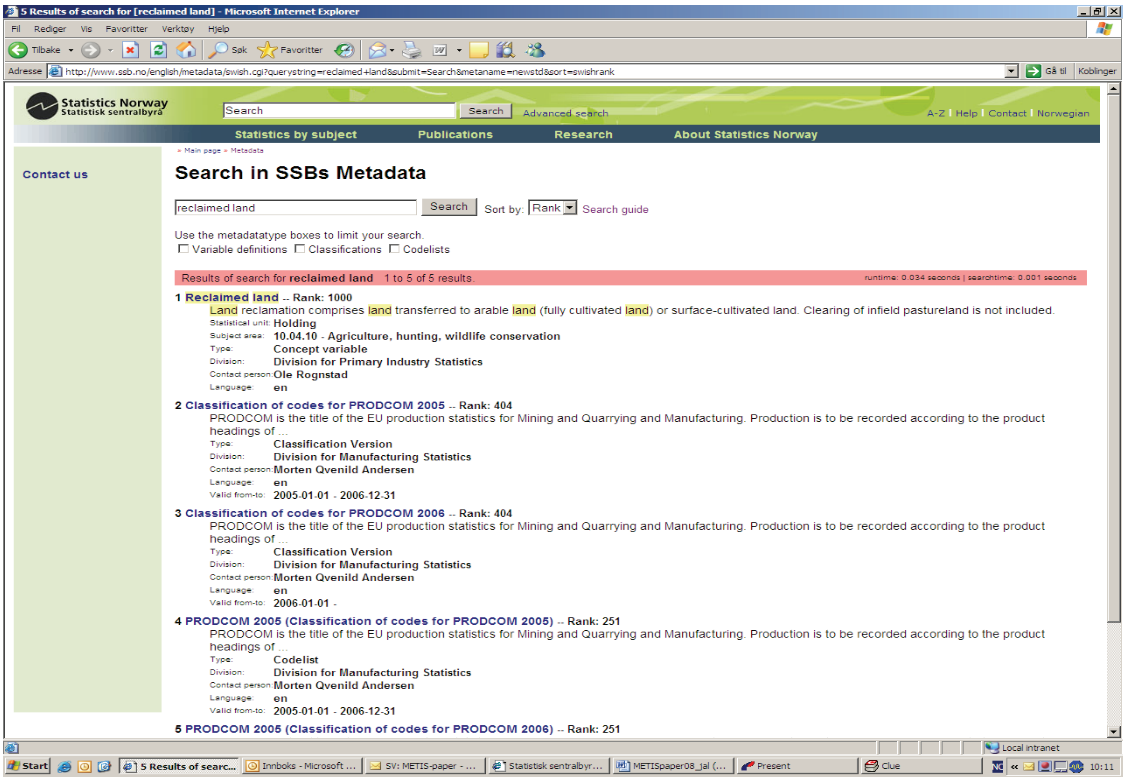
Units classified:

Agricultural areas

Footnotes:

Fig. 8 The classification linked to Agricultural area

We can also use the home page for searching across the different metadata types. If we search for e.g. "reclaimed land" without ticking off any of the metadata type boxes, we will get hits among all metadata types where "reclaimed land" is found.



Search in SSBs Metadata

reclaimed land [Search] Sort by: Rank Search guide

Use the metadatatype boxes to limit your search.
☐ Variable definitions ☐ Classifications ☐ Codelists

Results of search for **reclaimed land** 1 to 5 of 5 results. runtime: 0.034 seconds | searchtime: 0.001 seconds

1 Reclaimed land -- Rank: 1000
Land reclamation comprises land transferred to arable land (fully cultivated land) or surface-cultivated land. Clearing of infield pastureland is not included.
Statistical unit: Holding
Subject area: 10.04.10 - Agriculture, hunting, wildlife conservation
Type: Concept variable
Division: Division for Primary Industry Statistics
Contact person: Ole Rognstad
Language: en

2 Classification of codes for PRODCOM 2005 -- Rank: 404
PRODCOM is the title of the EU production statistics for Mining and Quarrying and Manufacturing. Production is to be recorded according to the product headings of ...
Type: Classification Version
Division: Division for Manufacturing Statistics
Contact person: Morten Qvenild Andersen
Language: en
Valid from-to: 2005-01-01 - 2006-12-31

3 Classification of codes for PRODCOM 2006 -- Rank: 404
PRODCOM is the title of the EU production statistics for Mining and Quarrying and Manufacturing. Production is to be recorded according to the product headings of ...
Type: Classification Version
Division: Division for Manufacturing Statistics
Contact person: Morten Qvenild Andersen
Language: en
Valid from-to: 2006-01-01 -

4 PRODCOM 2005 (Classification of codes for PRODCOM 2005) -- Rank: 251
PRODCOM is the title of the EU production statistics for Mining and Quarrying and Manufacturing. Production is to be recorded according to the product headings of ...
Type: Codelist
Division: Division for Manufacturing Statistics
Contact person: Morten Qvenild Andersen
Language: en
Valid from-to: 2005-01-01 - 2006-12-31

5 PRODCOM 2005 (Classification of codes for PRODCOM 2006) -- Rank: 251

Fig. 9 Search result for "reclaimed land"

C Lessons learnt

1 One of the biggest challenges in management of metadata is allocating the necessary resources. Releasing good quality statistics within the planned time schedule is the primary task for the subject matter divisions and documentation will often have a lower priority. It is therefore crucial that the management stresses the importance of documentation and increases the status for this kind of work.

2. It is important that we can refer to formal documents like the metadata- and IT-strategy (which have been approved by the board of directors) in our metadata work. In the same way it is useful that the list of key metadata terms promoted for use within the statistical office has been officially approved.

3. Use step-wise development of metadata systems with active user involvement and regular delivery of functionality.

4. Ensure continuous follow-up of progress and quality with direct feedback to users and regular reports to middle and top management.

5. Harmonising variables between subject matter divisions is also a considerable challenge and an important tool to improve the quality of metadata. Several subject matter divisions may use the same variable names, but define them differently. In some cases this is necessary because of laws and regulations, but this is not always the case. We have meetings where contact persons from divisions using variables with similar names come together and discuss the definitions, e.g. a division could change the wording of their definition to such an extent that other divisions might use it as well, which would allow us to reduce the number of definitions to one instead of e.g. three. This is a time consuming task which we have started, but which will require a lot more of resources, both to monitor where harmonisation is needed and to do the job.

6. The possibility to release metadata on the Internet makes it easier to motivate subject matter divisions to document metadata and improve metadata quality.

7. We think that to really make metadata work a natural part of everyday life in the subject matter divisions, we have to include the metadata systems in the production cycle. Then we can establish routines where the handling of metadata is included in all relevant production steps. So far the metadata work in Statistics Norway has been focused on implementing metadata systems and filling them with relevant documentation. This year we will have focus on investigating the role of metadata(systems) in the production cycle.

D. Concluding remarks

1. We find that the metadata portal is a valuable tool for managing the large amounts of metadata stored in metadata systems and elsewhere, and a tool (especially for managers) to help monitor the development in our metadata work.

2. Version 1 of the metadata portal was released both on the Intranet and the Internet in February 2008. The portal will be further developed this year by extending functionality and including other metadata when relevant. We will also focus on documenting more metadata in the existing metadata systems that are being accessed through the portal. Last, but not least, we hope to get feedback from internal and external users that will help us improve the portal further.