Changes in the model for meta database 1.06-> 2

Updated with decisions made at the data model meeting 1-2 July 2003 and some additional updates after the meeting.

**General**
In order for a change in the data model to be fully successful within the database area, many of the changes must also be carried out in *PC-Axis format*, so that the metadata in the SQL database can be transferred to the PC Axis family and affect handling there.

All tables and columns described below are included in the common model. If a column is described as voluntary, this means it does not have to be filled in with any values and can be NULL.

**Subject area - main table**

1. **New structure**
   *Decision:* A solution that allows an optional number of levels above the main table. Each organization can freely decide the number of levels. At Statistics Sweden there will be 5 levels - subject area, statistics area, product, sub-product and main table. The solution allows for a variable number of levels within an organization, but it is easier to handle with a fixed number.

   It was decided that we will apply this in the following way: a) A menu will be able to contain objects from different levels. Louise will study if this is possible in PC-Axis. If it is difficult to implement there, this option will have to be postponed. (At Statistics Sweden we will probably not allow objects on different levels in one and the same menu.) b) A type of object, e.g. product, always has the same level number, even if an overlying level is missing. No "dummy" levels are set up.

   Other effects of the new structure are that the presentation text on all levels above the main table is in 100 characters, that sorting is possible on all levels, and that a description text can be added on all levels. Linking can also be done to several different alternatives on the same level, e.g. a main table can be shown under several products.

   *Changes in the data model:*
   1) The tables Subject, Subject_Eng, Product and Product_Eng will be omitted. The table ProductTable will also be omitted. At Statistics Sweden, Subject and Product will be renamed with the prefix Mi_ and will be used by those programs that handle micro material.

   2) The table MenuSelection will be added.
Menu varchar(20) obligatory at level 1 = START
Selection varchar(20) obligatory
PresText varchar(100) obligatory if LevelNo not equal to Menu levels in MetaAdm, otherwise NULL
PresTextS varchar(20)
Description varchar(200)
LevelNo char(1) obligatory 1-
SortCode varchar(20)
Presentation char(1) obligatory
A = active, visible and can be selected
P = Passive, cannot be selected
N = no, not shown in the menu
To be filled in the same way for all levels.
InternalId varchar(20)
UserId
Date

In Statistics Sweden's case, InternalId is an obligatory identity that can be linked to the Product Database (PDB).

3) MenuSelection_Eng is added. Same appearance as MenuSelection, but the columns LevelNo and InternalId are not needed.

4) Since the table Product is omitted, a new table is needed to keep a check in which database the data tables of the product will be placed. At Statistics Sweden the table will contain all products including those that not yet have any tables in the database.
The new table DataStorage contains:
ProductId varchar(20) obligatory
Server varchar(8) obligatory
Database varchar(30) obligatory
UserId
Date:

As a result of this, Server and Database are removed from MainTable.

5) The columns SubjectCode char(2) and ProductId varchar(20) are added after SpecialCharExists (see point 11) in MainTable. The column SubjectCode is obligatory and ProductId is voluntary (obligatory at Statistics Sweden).
AreaCode is used for example when a PC-Axis file is created and therefore must always exist.
6) AreaCode and ProductId are removed from Contents, but may remain at Statistics Sweden. If the column ProductId in Contents remains, it is changed
at Statistics Sweden to varchar(20) so that ProductId will have the same format at all places in the data model.

7) The number of levels is entered in MetaAdm, which gets a new structure. See point 29.

Example:

<table>
<thead>
<tr>
<th>Property</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Menu levels</td>
<td>5</td>
</tr>
</tbody>
</table>

8) Levels are entered in the ColumnCode and ColumnCode_Eng.

Example:

<table>
<thead>
<tr>
<th>Meta Table</th>
<th>ColumnCode_Eng</th>
</tr>
</thead>
<tbody>
<tr>
<td>MenuSelection</td>
<td>LevelNo</td>
</tr>
<tr>
<td>Code</td>
<td>3</td>
</tr>
<tr>
<td>PresText</td>
<td>Statistical product</td>
</tr>
</tbody>
</table>

**PC-Axis format affected?** Yes, the subject area text is changed from 80 to 100 characters.

2. **Numbered main tables**

**Decision:** A short unique identity for main tables (numbers and/or letters), which can be used upon inquiries from end users and the like.

A table identity is already in use at Statistics Denmark and Statistics Norway. The decision involves a coordination of our table identities in the data model.

**Changes in the data model:** A new obligatory column TableId varchar(20) is added after PresTextS in MainTable. (The column PresCode varchar(8) remains in Contents.)

At Statistics Sweden, TableId will be created automatically when a new main table is set up according to the formula one letter + 3 digits. PresCode will contain TableId + one letter to differentiate between the various columns.

**PC-Axis format affected?** No

3. **Link between tables**

**Decision:** In case of a break in the time series, which requires a new table to be set up, information referring to one or more other tables can be added in a footnote.

**Changes in the data model:** No Information is added in a footnote, which is shown when the main table is selected, as a normal text (reference).

**PC-Axis format affected?** No

4. **The short presentation text of MainTable is extended**

**Decision:** The text is used in the table heading if a table has several contents
columns. Today the text is varchar(80) but is extended to varchar(150). However, long texts should be avoided as much as possible.

Changes in the data model: Pres TextS in the MainTable and MainTable_Eng is extended to varchar(150).

PC-Axis format affected? No

Footnotes and links to further information

5. Connection of footnotes and information texts to new levels

Decision: New connections are added for all levels in subject areas, including main table.

Changes in the data model:
1) The connection table FootnoteSubject will be omitted.

2) The connection table FootnoteMenuSel is added and will include footnotes for all levels above the main table. See the data model.

3) The connection table FootnoteMainTable is added. See the data model.

4) The connection table FootnoteSubTable is added. See the data model.

5) Each new connection will have its own footnote type (FootnoteType in the table Footnote).
   Level 1 (=subject area) will keep type 1.
   Level 2 (= statistics area at Statistics Sweden) will have A, Level 3 (= product at Statistics Sweden) will have B, Level 4 (= sub-product at Statistics Sweden) will have C and so forth for each new level.
   MainTable will have 7 and SubTable 8

6) Previously there has been a column Footnote in the table Subject which was used by the retrieval programs to easily see if there were footnotes or not. This column is not considered necessary. In the other tables, Footnote remains as previously.

PC-Axis format affected? Yes, but only with regards to contents since the new footnotes will be placed together with the other footnote texts in the existing key words. However, it is not clear if the footnotes for all levels above the main table will be included. Today, the subject area text is not included.

6. Information on when a footnote will be shown

Decision: Information on when a footnote will be shown - upon selection, presentation or both - will be added.
Changes in the data model: The obligatory column ShowFootnote char(1) is added after FootnoteType in the table Footnote.
S = shown upon selection
P = shown upon presentation (for all presentation forms, screen, files, maps etc.)
B = shown upon both selection and presentation

PC-Axis format affected? No. Footnotes with ShowFootnote P and B will be included when a PC-Axis file is created from the SQL database and put into the existing key words.

7. Links to the product-specific document, other pages on the website or other websites from a footnote

Decision: Possibility to have links in the footnote text to more product-specific documents, publications etc. with more detailed explanations than those given in the information texts. More documents can be connected to the same footnote.
Links can also be made to other pages on Statistics Sweden's website such as the Statistical Reports, product pages or to other statistical authorities.

Changes in the data model: One or more links can be put into the footnote text. The link is written in HTML format.
Example: <a href=http://www.scb.se>See Statistics Sweden's website for more information!</a>
(PLEASE NOTE: no quotation marks (”) since they cause problems in PC-Axis.)

PC-Axis format affected? Yes, but only with regards to content of existing key words. Link information is reformed when a PC-Axis file is created from the SQL database, and will only be available as information and not as a link.

8. Links to the product-specific document, other pages on the website or to other websites directly from a menu alternative

Decision: Possibility to have links directly without footnote. The link can be connected with a row in the new table MenuSelection. This means that links can be connected with for example a subject area or a main table. To make it possible to connect links with other objects in the data model in the future, the connection is placed in a special table (compare with footnote). This also allows the same link to be connected with several main tables for example, thus facilitating maintenance of the links.
PresCategory can for example allow links to be shown for internal users only.

Changes in the data model:
1) A new table Link is added with the following contents:
LinkId        int obligatory
Link           varchar(250) obligatory
LinkText       varchar(250) obligatory

<a href=http://www.scb.se>See Statistics Sweden's website for more information!</a>
(PLEASE NOTE: no quotation marks (”) since they cause problems in PC-Axis.)
9. **Editing characters in the footnote text**

*Decision:* Possibility to insert editing characters in the footnote text for row change, bold style and italics.

*Changes in the data model:* Notation for HTML tags is used. No attributes can be given. Only the following is allowed:

- `<b>` **Bold** *Bold text*
- `<i>` *Italic* *Italic*
- `<BR>` *New row*

*PC-Axis format affected?* Yes, but only because HTML tags can be found in the footnote text.

---

**Processing information**

10. **Aggregation possible**

*Decision:* To prevent the user from aggregating things that cannot be aggregated, information on whether the contents variable can be aggregated or not is added. Internet interface and PC-Axis family can then prevent aggregation and will not show the possibility to group. However, this does not prevent the user to transfer material to for example Excel and aggregate it there.

Another advantage is that double value sets are not needed. To prevent groups from being shown for statistical material that cannot be grouped, today we
must set up a value set with groups and one without groups.

PLEASE NOTE that the adopted solution applies to all spread variables for a contents column.

Changes in the data model: The obligatory column AggregPossible char(1) is added after PresCellsZero in Contents

Y = Yes
N = No

PC-Axis format affected? Yes. A key word is added in the PC-Axis format that applies to the whole table, i.e. it is not possible to have different ones for the different contents. PLEASE NOTE! This change will not be introduced in PC-Axis format until October 2003.

11. NULL, dot, missing

Decision: A voluntary option to add a new column to the data tables for special characters will be introduced. Since special characters are found in a separate column in the data table, the figures can still be in the ordinary columns but are not shown. It is also possible to state whether a data cell marked with special characters can be summed or not. This means that even if the separate data cell is not shown, it can still be included in a sum.

This solution makes it possible to have an optional number of special characters and also have different special characters for different installations of the data model.

Changes in the data model:

1) A new obligatory column is added in the MainTable SpecCharExists char(1) Added after PresCategory.
   Y = Yes
   N = No

2) A new table SpecialCharacter is added with the following contents:
   CharacterType  varchar(2) obligatory primary key for example 2
   PresCharacter  varchar(20) obligatory for example ..
   AggregPossible char(1) obligatory Y = Yes, N = No
   DataCellFilled char(1) obligatory
   Y = The data cell must be filled
   N = The data cell may not be filled, must be NULL
   V = optional, i.e. the data cell may be filled or be NULL
   PresText  varchar(200) optional explanation to what is stated in PresCharacter
   UserId
   Date
If a number of cells will be added and some cell has AggregPossible=N, the cells are not summed in the interface. By setting AggregPossible=Y, the figures can then be included in a sum even if they are not shown one by one.

The corresponding table for English texts is SpecialCharacter_ Eng. Same columns except AggregPossible.

3) If MainTable.SpecCharExists=Y there is an extra contents column for all contents columns in the data table. These special columns have the same names as the contents column they belong to, with the suffix x, for example Born and Born_x. The format is char(2), i.e. the same as SpecialCharacter.CharacterType. To facilitate downloading of data tables, we at Statistics Sweden will put all special columns last and in the same order as the contents columns. However, this has no significance for the retrieval programs.

Placement of special columns is optional, i.e. they do not have to be put immediately after each contents column. All contents columns in a main table do not have to have a special column.

Special columns shall only contain characters that are found in SpecialCharacter.CharacterType or null.

4) SpecialCharacterSum is entered in MetaAdm, which gets a new structure. See point 29.
Example:
Property  SpecialCharacterSum
Value  99
SpecialCharacterSum will also be put into the table SpecialCharacter and SpecialCharacter_Eng so it can give a value that will be presented in the table.

PresCharacter for SpecialCharacterSum shall be used if some figure that is included in the sum has AggregPossible=N and figures with several different CharacterTypes included in the sum. If some figures included in the sum have AggregPossible=N and only one CharacterType is included, PresCharacter is used.

PC-Axis format affected? Yes, if more than 4 different character types are allowed. For the time being, we suggest to use 4 types.

12. Current map
Decision: Information on the current map does not have to be different for MainTable that uses the same ValueSet. It can be moved from SubTableVariable to ValueSet. The information will then be easier to fill in. To show the current map after grouping, the table Grouping is supplemented with map information.
Changes in the data model: GeoArea is moved from SubTableVariable to ValueSet(before KDBid) and Grouping (before KDBid). Column is optional. The column is redone to int and changes its name to GeoAreaNo. The number will then be again found in the new table TextCatalog with TextType = GeoArea. See point 25. VariableType shall include G if GeoAreaNo is filled in ValueSet.

Handling in the retrieval interface: When selecting several databases, the keyword MAP is not created.

PC-Axis format affected? No

Value sets, groups, values, time periods

13. Select code/text for presentation of values in a value set

Decision: To increase flexibility, information on presentation of values is added in the value set since we found a problem in certain cases, for example in value pool Region where show code+text was selected. Sometimes it is not relevant to show the code, e.g. when you have made "own regions" such as SE01excl0030. Since the new flexibility made it more difficult to get the producers to follow a standard for a value pool and also more difficult to join different tables in for example PC-Axis, this use should be restrictive.

Changes in the data model: A new column ValuePres char(1) is added after ValuePool in ValueSet. The column is obligatory and can contain C for code form, T for text form, B for both, and V if ValuePres in ValuePool shall apply.

Handling in the retrieval interface: When several value sets are selected for the same variable (several sub tables) ValuePres is taken from ValuePool.

PC-Axis format affected? No

14. English description for ValueSet

Decision: When supplementing the English tables in version 1.06, we forgot to include Description of ValueSet. It is used in PC-Axis at the classification window to present value sets and search for values.

Changes in the data model: The table ValueSet_Eng is added and will include the column Description. SortCodeExists refers to both SortCode in VSValue and VSValue_Eng. That is, if SortCodeExists is Yes, sort code shall be filled in both VSValue and in VSValue_Eng.

PC-Axis format affected? No
15. **Longer presentation texts for values**

*Decision:* The long presentation text for value is changed from 100 to 250 to, among other things, handle long English texts.

*Changes in the data model:* The column ValueTextL in Value and Value_Eng is changed to (250)varchar.

*PC-Axis format affected?* No. Is already extended to 256.

16. **Sorting groups in a grouping**

*Decision:* Option to sort groups in a grouping in a logical order in the same way as values in a value set.

*Changes in the data model:* A new column SortCode varchar(20) is added after ValuePool in VSGroup. The column will be optional. All groups in a grouping for a value set will either have a sorting code or else none will have a sorting code.

2) A new table VSGroup_Eng is also inserted with the same appearance as VSGroup.

*Handling in the retrieval interface:* If SortCode is in VSGroup, the retrieval interface will use it. Otherwise, the sorting code from the value pool will be used as before.

*PC-Axis format affected?* No

17. **Sorting of groupings for a variable**

*Decision:* Option to sort groupings for a value pool in logical order. At Statistics Sweden, this applies to different groupings for countries.

*Changes in the data model:* The tables Grouping and Grouping_Eng are supplemented with ValuePool (primary key) first in the table and SortCode varchar(20) before UserId. SortCode is optional.

The solution means that the sorting is done for all groupings that belong to a certain value pool and that the same sorting is for all main tables that use this value pool. Among other things, a link to ValueSet has been discussed, but was ruled out since there would be problems with PC-Axis. There you can select for a whole main table, and groupings with different value sets with different sorting orders would complicate the situation.

*Handling in the retrieval interface:* If SortCode exists in Grouping, the retrieval interface will use it. Otherwise sorting of the presentation text will be as before.

*PC-Axis format affected?* No
18. Presentation of groups in a grouping

*Decision:* A solution that can control how the groups are shown, and in certain cases, that gives the grouping itself a clearer text. See the following example.

*Changes in the data model:* The obligatory column GroupPres char(1) is added after Description in the table Grouping.

A = aggregated value is shown
I = integral values are shown
B = both aggregated value and integral values are shown

*Handling in the retrieval interface:*

**Alternative 1.** Information found in the GroupPres in Grouping is used to steer presentation of the groups. In the interface it is possible to present several alternative groupings with clear texts that state whether an aggregated value and/or integral values will be shown.

Example:
1) Grouping=Continents1, PresText=Continents, GroupPres=A
   For example, if the user selects Africa, only Africa will be shown.

2) Grouping=Continents2, PresText=Continents (separate countries), GroupPres=I
   For example, if the user selects Africa, the countries in Africa will be shown.

3) Grouping=Continents3, PresText=Continents (continents with integral countries), GroupPres=B
   For example, if the user selects Africa, both Africa and the integral countries will be shown.

**Alternative 2.** The presentation is handled completely by the interface in combination with the user who can decide himself how the presentation will appear. That is, the interface gives the user the opportunity to select the grouping Continents and then the opportunity to decide whether or not aggregated values and/or integral values should appear. The aggregating alternative can be taken away if the contents are not summable.

*PC-Axis format affected?* No.

**Person and organization**

19. Several contact persons/persons responsible for updating.

*Decision:* Possibility to an unlimited number of contact persons and persons responsible for updating as well as a couple of new roles. The persons (groups) are linked to the MainTable instead of to Contents like is done today. One person can have several roles for a main table.
Changes in the data model:
1) The columns Contact1, Contact2, UpdatePerson1 and UpdatePerson2 are removed from the Contents.

2) A new table called MainTablePerson is added between MainTable and Person.

<table>
<thead>
<tr>
<th>Field</th>
<th>Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>MainTable</td>
<td>varchar(20)</td>
<td>obligatory</td>
</tr>
<tr>
<td>PersonCode</td>
<td>varchar(20)</td>
<td>obligatory</td>
</tr>
<tr>
<td>Role</td>
<td>char(1)</td>
<td>obligatory</td>
</tr>
<tr>
<td>M</td>
<td></td>
<td>main contact person</td>
</tr>
<tr>
<td>C</td>
<td></td>
<td>contact person</td>
</tr>
<tr>
<td>U</td>
<td></td>
<td>updating person</td>
</tr>
<tr>
<td>I</td>
<td></td>
<td>person responsible for international reporting</td>
</tr>
</tbody>
</table>

UserId
Date:

PC-Axis format affected? No

20. Changes in Person

Decision: The column Name is divided into two columns - forename and surname - to allow for better sorting possibilities. The English table for Person is not needed and thus can be omitted. A new column for telephone prefix is added instead.

Changes in the data model:
1) The column PersonName is removed from Person. (50)Is replaced by Forename varchar(50) and Surname varchar(50). Forename is optional and Surname Obligatory. Groups are only put into Surname, but persons shall always have both Forename and Surname.

2) Person_Eng is omitted.

3) A new column PhonePrefix varchar(4) is added to Person before PhoneNo. For example, could include +46. To be used when presenting phone numbers in the English version of the interface. Column is obligatory.

PC-Axis format affected? No. Forename and surname are combined before they are put into the PC-Axis file. In the English version, PhonePrefix is combined with PhoneNo.

21. X400 is removed from Person

Decision: Is removed as soon as possible when Statistics Sweden can coordinate it with Metadok and the local use of the table of the producers within Statistics Sweden.

Changes in the data model: The column X400 is removed from the table.
Person.

PC-Axis format affected? No

22. **New information in Organization**

*Decision:* At Statistics Sweden, even in this table, we need to have a link to the Product Database. It would also be advisable to give links by producers and statistical authorities if these are other than Statistics Sweden.

*Changes in the data model:* Two optional columns will be added. The column `WebAddress` varchar(100) is added after `Unit` in the table `Organization`. The column `InternalId` varchar(20) is added after `WebAddress`.

PC-Axis format affected? No

Other

23. **Change of format for UserId**

*Decision:* To have longer User ID codes, the column UserId is extended.

*Changes in the data model:* The column UserId is changed to varchar(20) in all tables.

PC-Axis format affected? No

New points from the meeting 1-2 July.

24. **Better marking of cell footnotes**

*Decision:* Supplementary information is added to the data model to mark which footnotes, linked to the contents, are cell footnotes. A cell footnote is defined as a footnote that exists for at least two of the contents column's variables, e.g. region and time or region and sex.

*Changes in the data model:* The column `Cellnote` char(1) is entered in `FootnoteContentsValue` and `FootnoteContentsTime` before `UserId`. Column is obligatory and can have the values Y and N.

PC-Axis format affected? No

25. **Text catalog**

*Decision:* To make it easier for those who will be adding metadata and at the same time obtain more consistent texts, certain texts such as name are placed on different maps in a table. The different alternatives can then be shown in the input program for metadata in a list box where the person in charge of input can choose an alternative and perhaps add new alternatives.
Changes in the data model:
Two new tables TextCatalog and TextCatalog_Eng are set up with the same structure.

<table>
<thead>
<tr>
<th>Property</th>
<th>Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>TextCatalogNo</td>
<td>Int</td>
<td>obligatory</td>
</tr>
<tr>
<td>TextType</td>
<td>varchar(30)</td>
<td>obligatory established text</td>
</tr>
<tr>
<td>PresText</td>
<td>varchar(100)</td>
<td>obligatory optional text</td>
</tr>
<tr>
<td>Description</td>
<td>varchar(200)</td>
<td>optional</td>
</tr>
<tr>
<td>UserId</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Date</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

PC-Axis format affected? No

26. Presentation texts for time periods
Decision: To obtain presentation texts for time periods, a value pool can be set up in ValuePool. The name of the value pool will then be found in MetaAdm.

Changes in the data model:
The following is needed to show the text of the point in time:
1) The name of the value pool is entered in MetaAdm, which gets a new structure. See point 24.

Example:
<table>
<thead>
<tr>
<th>Property</th>
<th>Value</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>VPTime</td>
<td>VPTime</td>
<td>Optional name of value pool for time</td>
</tr>
</tbody>
</table>

2) ValuePres in ValuePool is T, which means that the text for time period will be shown.

3) All time periods will be available in the value pool in Value, and they must also have a text in ValueTextL.

4) Points 2 and 3 are repeated for English texts in ValuePool_Eng and Value_Eng.

PC-Axis format affected? No

Handling in the retrieval interface: If VPTime is in MetaAdm, and the value pool has ValuePres=T and the time period has a text in ValueTextL in Value, the time period’s text is shown. Otherwise the time period text from ContentsTime is shown.

27. Time scale instead of time in table heading
Decision: A possibility to show the time scale in the heading instead of the word "time" is added.

Changes in the data model: A new column TimeScalePres char(1) is added
after PresText in TimeScale. Column is obligatory and can have the values Y and N.

**PC-Axis format affected?** No

**Handling in the retrieval interface:** If TimeScalePres is Y, TimeScale.PresText is shown in the heading.

### 28. Variable name when selecting several contents

**Decision:** In the text catalog, every installation shall add the variable name of choice. There is also a possibility to add a variable name for each main table.

**Changes in the data model:**

1) In the text catalog, a row with variable text, is added, e.g.:

   - TextCatalogNo 67
   - TextType ContentsVariable
   - PresText type (shall begin with a small letter)
   - Description ..........

   The same procedure is done in TextCatalog_Eng.

   A new column ContentsVariable varchar(2) is added in the MainTable before TableId. The column is optional and can contain a text or null. The same applies for MainTable_Eng.

   **PC-Axis format affected?** No

   **Handling in the retrieval interface:** If MainTable.ContentsVariable contains a text, it is used. Otherwise, PresText in TextCatalog is used.

### 29. Change of the table MetaAdm

**Decision:** Up to now, MetaAdm has only been used to keep track of the most recently used footnote number. Table will be changed so that it can also be used with other system variables.

**Changes in the data model:** The table will have the following appearance:

- **Property** varchar(30) obligatory
- **Value** varchar(20) obligatory
- **UserId**
- **Date:**

   The following property can be found:

   - LastFootnoteNo optional
   - MenuLevels obligatory
   - SpecCharSum obligatory
   - VPTime optional
PC-Axis format affected? No

30. StoreColumnNo, StoreFormat and StoreNoChar in Contents will be obligatory.

31. Column in ColumnCode and ColumnCode_Eng is changed to varchar(30).