



Nesstar Cubes: The cube publishing process

November 2006



Contents

1	Introduction	1
2	An exploration of Cube Data	1
2.1	Dimensions and Categories	1
2.2	Hierarchical Dimensions.....	2
2.3	Additivity	3
2.3.1	Non-additive	3
2.3.2	Stock additivity.....	4
2.3.3	Flow additivity.....	4
2.3.4	Additive measures.....	4
3	Preparing Data for Input into the Publisher	4
3.1	General.....	4
3.1.1	Importing Data.....	5
3.1.2	Enter data directly in the Publisher.....	5
3.2	Hierarchical Dimensions.....	5
3.2.1	Non-additive data	5
3.2.2	Additive data.....	6
3.3	Transforming the Data.....	7
4	Using the Nesstar Publisher.....	7
4.1	Data Import.....	7
4.1.1	Import data as numerical variables.....	7
4.1.2	Import data as alphanumeric variables.....	7
4.1.3	To change the Data Type	7
4.2	Add Metadata	8
4.3	Edit Dimensions and Categories	8
4.3.1	Create categories for all dimensions	8
4.3.2	Create a hierarchy	9
4.3.3	Expand dimension	10
4.4	Measures to Dimensions.....	10
5	Using the Cube Builder	12
5.1	Cube Definition.....	12
5.1.1	Single measures.....	12
5.1.2	Multiple measures	12
5.1.3	Grouping measures	13
5.2	Default View	13
5.3	Publishing cubes	14
6	Adding map functionality	14
6.1	Creating map files	14
6.1.1	The shp and shx files.....	14
6.2	Placing the map files in an accessible area on a Server	15
6.3	Adding the map link to the data file	15
6.4	Adding more than one map for one dataset	16
7	More Complex Cubes.....	17
7.1	Multiple Tables	17
7.2	Multiple Measures	17
7.3	Incomplete Cubes	18

8	Hints and Tips	21
8.1	Nesstar Exporter Excel Macro.....	21
8.1.1	Security settings	21
8.1.2	Category values and labels in Excel.....	21
8.2	Additive Cubes	21
8.2.1	Higher-level data are calculated automatically	21
8.2.2	Missing data	21
8.3	Access to Maps when running the Nesstar Server off-line.....	22
	Appendix A.....	23
A.1	The Nesstar Exporter Excel Macro	23
A.1.1	Prepare the data for the Exporter	23
A.1.2	Using the Nesstar Exporter Excel Macro.....	24
A.1.3	Check your data	25
A.1.3	Multiple measures	25
A.2	Using Data from a Database (e.g., MS Access).....	25

1 Introduction

This guide provides information on how to publish cubes (tables) using the Nesstar Publisher and the Nesstar Cube Builder. Published cubes are available to view via Nesstar WebView.

To publish cubes you need to:

1. Create a data file for input into the Nesstar Publisher, see [Section 3](#) for further details
2. Use the **Nesstar Publisher** to:
 - a. Add metadata (such as: title, ID, Abstract)
 - b. Create hierarchical levels
 - c. Add a link to a map (optional).
3. Use the **Nesstar Cube Builder** to:
 - a. Define the time and geographical dimensions, if present,
 - b. Define the measure(s), and
 - c. Create the default view of your cube
 - d. Publish the cube onto a Nesstar server.

Sections 2 to 5 of this guide provide a quick overview on how to publish a simple cube. More advanced options are described in Section 6, and Section 7 contains details about more complex cubes.

2 An exploration of Cube Data

2.1 Dimensions and Categories

A cube is a multi-dimensional table consisting of aggregated data for a number of **dimensions** (variables) and at least one **measure** (variable). The dimensions describe the data whereas the measure represents the data itself. For example, a measure can represent average ages, the number of cars sold, or population totals. A typical representation of a cube is shown in Table 1 below.

GENDER	Males			Females		
AGE GROUP /AREA	0-14	15-64	65+	0-14	15-64	65+
Dudley	29465	98833	21416	27493	98521	29427
Sandwell	29280	88210	19007	28417	90426	27564
Walsall	26028	79887	17274	25513	81060	23737
Wolverhampton	23964	74740	17154	22326	75567	22831

Table 1. Population data by gender and age group, for the four areas of the Black Country.

In Table 1, three dimensions are present (GENDER, AREA and AGE GROUP). Two of these, GENDER and AGE GROUP, are present in the columns (*header*) of the table, and the dimension AREA is present in the rows (*stub*) of the table. The measure variable represents the actual data, i.e. the numbers in the table cells, which in this example are population figures.

Dimensions contain **categories**. For instance, the dimension GENDER has two categories: *Males* and *Females*; the dimension AGE GROUP has three categories, and the dimension AREA has four.

A **measure** variable has no categories, is continuous and needs to be classified as a numeric, scaled variable within the Nesstar Publisher.

A dimension represents the whole of a table whereas a 'category' only applies to part of it. In the example above, all data points (cells) have a GENDER but only half of the data are *Males*.

In a cube, each data point is described by all dimensions. This is clearly illustrated in Table 1 where for any given data point, there is an AREA, a GENDER and an AGE GROUP associated with it. This means that it is also possible to reconstruct the cube by moving the dimensions around, as shown in Table 2 below.

The capacity to move dimensions around is a powerful feature of cubes and it allows the user to make direct comparisons between categories of interest. For example, a user who wishes to compare the number of males in each age group will prefer a table such as Table 1, but a user who wishes to compare gender will prefer Table 2.

In the Nesstar system, the person publishing the cube decides on the default view for the table, i.e. how it initially appears to the user. The user can then rearrange the table by moving the dimensions as required.

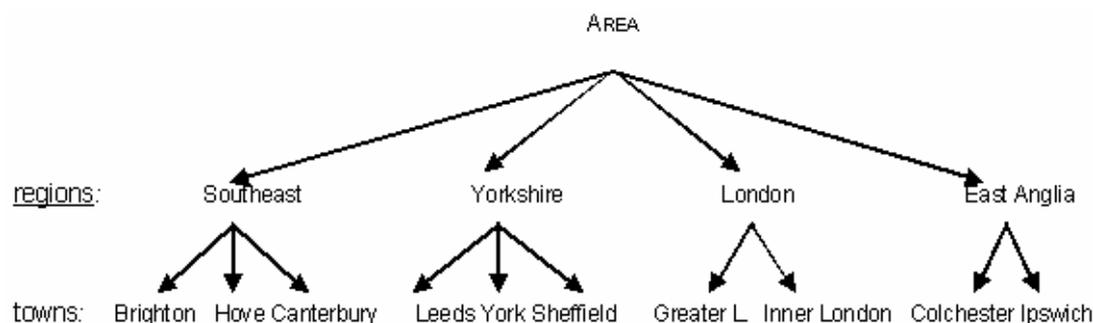
AGE GROUP	0-14		15-64		65+	
GENDER / AREA	Males	Females	Males	Females	Males	Females
Dudley	29465	27493	98833	98521	21416	29427
Sandwell	29280	28417	88210	90426	19007	27564
Walsall	26028	25513	79887	81060	17274	23737
Wolverhampton	23964	22326	74740	75567	17154	22831

Table 2. The dimensions GENDER and AGE GROUP from Table 1 have been interchanged.

2.2 Hierarchical Dimensions

It is possible, and sometimes preferable, to group certain categories together to form a hierarchy. For instance, for a geographical dimension, instead of providing a long list of towns, the towns could be grouped into regions. The dimension AREA then consists of two levels: *regions* and *towns*, as in Figure 1 below.

Figure 1. Example of a hierarchical dimension.



In Figure 1, AREA is the dimension, and *regions* and *towns* are categories. *Regions* and *towns* are not dimensions themselves, but represent different levels of the hierarchy.

Section 2.1 above stated that it is possible to move dimensions around in a table. However, this was only possible because for each AGE GROUP category, there were the same GENDER categories (i.e. Males and Females), and for each GENDER category, there were the same AGE GROUP categories.

In Figure 1, there are no identical *towns* categories for each *region* category. For example, for *Southeast* there are no *towns* categories *Leeds*, *York*, and *Sheffield* etc., and for *Leeds*, there are no *regions* categories *Southeast*, *London*, or *East Anglia*. The individual towns cannot be seen separate from their region, and therefore *towns* and *regions* cannot be separate dimensions.

In the Nesstar system, data arranged hierarchically enables the user to select a *region* level, for example, which will then open up to display the data for its underlying *towns*.

2.3 Additivity

The Nesstar Cube Builder expects a cube's '*measure*' to be defined in terms of additivity.

A **non-additive measure** means there will be no automatic aggregation of the data and no further applied operation is chosen. This is the default setting in the Cube Builder.

If a measure can be aggregated it falls into one of the following categories:

Stock additivity - which refers to a measure of activity at a given point in time, or

Flow additivity - which refers to an activity measured per unit of time.

2.3.1 Non-additive

Non-additive data cannot be aggregated at all. Examples include percentage and rates data. For example, if 65% of females smoke and 52% of males smoke, we

cannot say that 117% of the population smokes. Other examples include life expectancy ages and average house prices.

Note: 'Mean' measures can be published as additive, but only if a further measure, either Sum or Count is also available.

2.3.2 Stock additivity

Stock measures cannot be aggregated over time, whereas *Flow measures* can. A cube has **stock additivity** if the data can be aggregated along all dimensions except the time dimension.

For example: If there were 20,000 inhabitants in Newtown in 2001 and 30,000 inhabitants in 2002, it is not correct to say that the total number of inhabitants in Newtown over 2001 and 2002 was 50,000. That is, it is meaningless to aggregate 'stock' measures over time.

However, stock measures can be aggregated over other dimensions. For example, if the number of inhabitants for Newtown in 2001 was 20,000 and the number for Oldtown was 30,000, then the total number of inhabitants for the two towns in 2001 was 50,000, i.e. the 'Town' dimension has been aggregated

2.3.3 Flow additivity

Flow measures can be aggregated over all dimensions. For example, 'the number of new cars' is a flow measure. It is correct to say that the total number of new cars for 2001 and 2002 is the sum of the number of new cars for each separate year. Thus, aggregating the data over the YEAR dimension will result in the total number of new cars for the full range of years. Aggregating the data over the AREA dimension will result in the total number of new cars for the combined areas.

2.3.4 Additive measures

Stock and Flow measures require an additional selection of Applied Operations (sum, mean, count, min, max). This refers to the type of measure that is used, e.g., whether the data represent totals (sum) or mean values.

The advantage of publishing a cube as additive, where possible, is that within a dimension, higher-level categories can be created to contain the aggregated values of the associated lower-level categories.

The Nesstar Cube Builder automatically creates and calculates the higher-level category, *All*, which contains the aggregated data at the highest level of each dimension.

3 Preparing Data for Input into the Publisher

3.1 General

There are two ways to produce cubes in the Publisher:

1. Import existing data

2. Enter data in the publisher via the 'Variables' and 'Data entry' tabs.

3.1.1 Importing Data

Before cube data can be imported into the Publisher, it needs to be transformed so that the information for each data point (cell) is contained on a separate line in the input file. See Table 3 below.

Area	Gender	Age group	number
Dudley	males	0-14	29465
Dudley	males	15-64	98833
Dudley	males	65+	21416
Dudley	females	0-14	27493
Dudley	females	15-64	98521
Dudley	females	65+	29427
Sandwell	males	0-14	29280
Sandwell	males	15-64	88210

Table 3: Structure of input file for some of the data for Table 1.

Each data point (cell) in a table should be represented as a separate line in the input file. Table 3 shows the first few entries of the input file for Table 1. The first line contains the name of the dimensions (although this may be omitted and entered at a later stage when using the Publisher). The other lines describe each data point in terms of the dimensions and the measure. When there is more than one measure, extra columns need to be added, one for each measure (see [Section 7.2](#) on multiple measures for further information).

The input file can be saved as a tab-delimited (.txt) file or as a comma separated (.csv) file.

The order in which the data points are entered is not important, and neither is the order in which the categories are entered as long as the same order is followed on each line. However, the order in which the category names first appear in the input file will decide the order in which they are displayed within the Nesstar system. Thus, in Table 3 above, the dimension AREA will appear with Dudley first, followed by Sandwell, and this order cannot be altered at a later stage.

3.1.2 Enter data directly in the Publisher

Please refer to the Publisher HELP function for more information on how to enter data directly into the Publisher.

3.2 Hierarchical Dimensions

If a cube contains hierarchical dimensions, data for the higher levels should only be included if the cube is non-additive (see [Section 2.3](#) on additivity).

3.2.1 Non-additive data

For non-additive cubes, data needs to be present at all levels of a hierarchical dimension. If data at the higher levels are not present, dashes (-) will appear where the data are missing, when the cube is published.

REGION / GENDER	Male	Female
England	50%	60%
North	70%	80%
East	40%	45%
South	30%	35%
West	80%	70%

Table 4: Percentage of smokers in regions of England (fictional).

Table 4 presents a non-additive cube containing percentage data, including a hierarchical dimension REGION. England contains the regions, North, East, South and West. As the data are non-additive, (i.e., summing the region data does not give the percentage for England) the data for England needs to be present in the input file, in the same way as the data for the underlying regions.

Table 5 below, shows the format of the input file for some of the data in Table 4.

Area	Gender	% smokers
England	male	50
England	female	60
North	male	70
North	female	80
East	male	40

Table 5: Input file for a sample of Table 4 data.

In some cases it may be useful to create higher-level categories in a non-additive cube even when no data for these categories is available. For example, where there are a number of categories that can be grouped together using a higher-level category as a heading. However, when published, the cells for these higher-level categories will contain dashes (-) as the data for these cells is missing, i.e. it was not present in the original table. The (empty) higher-level categories do not need to be present in the input file as they can be created within the Nesstar Publisher.

3.2.2 Additive data

For additive cubes, data should only be present for the lowest-level categories in a hierarchical dimension. That is, the higher-level categories should be added after the data file has been imported into the Publisher. These will be virtual categories, with no actual data points associated with them. The Nesstar system will automatically calculate the data for these categories after the cube is published. If data are included for the higher-level categories, these may not be correctly re-calculated by the Nesstar system.

The exception to the above is where no lower-level data are available for a higher-level category, and no lower-level categories are present. Such a category should be included in the input file, and the Publisher, at the correct hierarchical level.

3.3 Transforming the Data

The Nesstar Exporter Excel Macro can be used to transform data in table format to the format required in the input file for the Nesstar Publisher. See [Appendix A](#). Appendix A also describes how to transform data for a database such as Access into the required format.

4 Using the Nesstar Publisher

When the data is in the appropriate format open Nesstar Publisher for the next step in the publishing process.

4.1 Data Import

4.1.1 Import data as numerical variables

If category values do not need to be alphanumeric, follow the procedures below. If you need to use alphanumeric category values in the data (e.g. area codes that are used in you map file contain alphanumeric values), follow the procedures in 4.1.2 below.

1. Import the data file [**File | Import Dataset**].
2. Ensure that the delimiter is correct (e.g., comma or tab).
3. Select the option '**Convert string variables to numeric**'.
4. Select the option '**Variable labels on first record**' if appropriate.
5. Check the preview.
6. Click '**OK**'.

4.1.2 Import data as alphanumeric variables

1. Import the dataset [**File | Import Dataset**].
2. Select the delimiter (e.g., comma or tab).
3. Select '**Variable labels on first record**' if appropriate.
4. Do not choose to 'Convert string variables to numeric'.
5. Click '**OK**'.

4.1.3 To change the Data Type

Once imported you may wish to change the data type of specific variables to become numeric, e.g. categorical variables.

1. Select the **'Variables'** tab and locate **'Variable Information'**.
2. Select the Data Type (Numeric, Fixed String, Dynamic String).
3. Select one of the conversion options and click OK.

Note: In order to create hierarchies of categories, the 'Data Type' needs to be set to 'Fixed String' and the length set by entering the appropriate value in the 'String length' box.

4.2 Add Metadata

Enter appropriate metadata into the required fields.

The **Study description | Title** field is used as the title for the cube. If no entry is provided for this field then the file name will be used as the title.

Note: the title cannot contain /, ?, *, or begin with [.

To import metadata from a different dataset, use **Documentation | Import from dataset**.

To import metadata from an xml file, use **Documentation | Import from DDI**.

Please refer to the Publisher 'Help' for further information.

4.3 Edit Dimensions and Categories

The following steps are required to publish a cube:

4.3.1 Create categories for all dimensions

1. Select the **Variables** tab.
2. Check that the dimension labels are correctly assigned.
3. Ensure that all dimension variables have categories. These can be checked in the **Variable Description** section.

Note: Categories that were numeric in the input file will not contain any category labels. For example, if a variable called YEAR, with categories 2000 and 2001 were present in the imported file, these values will not be labelled and so need to be added. Use the **Documentation | Create Categories From Statistics** option.

4. Ensure that the 'measure' variable is set to numeric (Scale) and does not have any categories.
5. Specify the number of decimal places to appear in the table for the 'measure' variable.

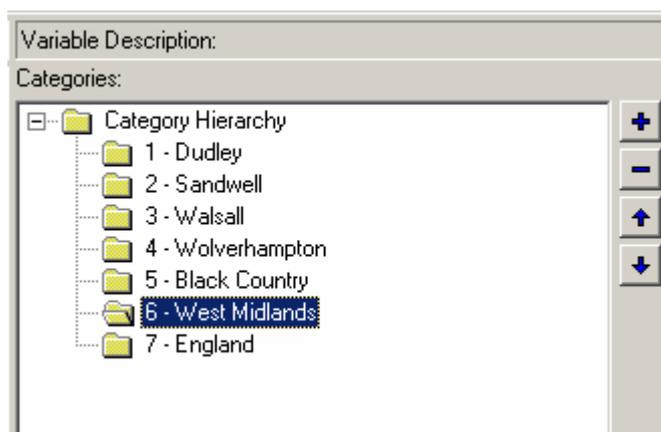


Figure 2. 'Variable Description' in the Publisher.

4.3.2 Create a hierarchy

For each hierarchical variable:

1. Under 'Variable Description' build your hierarchy. Use '+' to create new categories.
2. Drag and drop lower level categories into higher level ones to build your hierarchy. Use 'Ctrl' to select several categories at a time.
3. A 'Level Name' needs to be added for each level of the hierarchy. Select any category within a level and enter an appropriate name. A level name only needs to be specified for one category within the level.

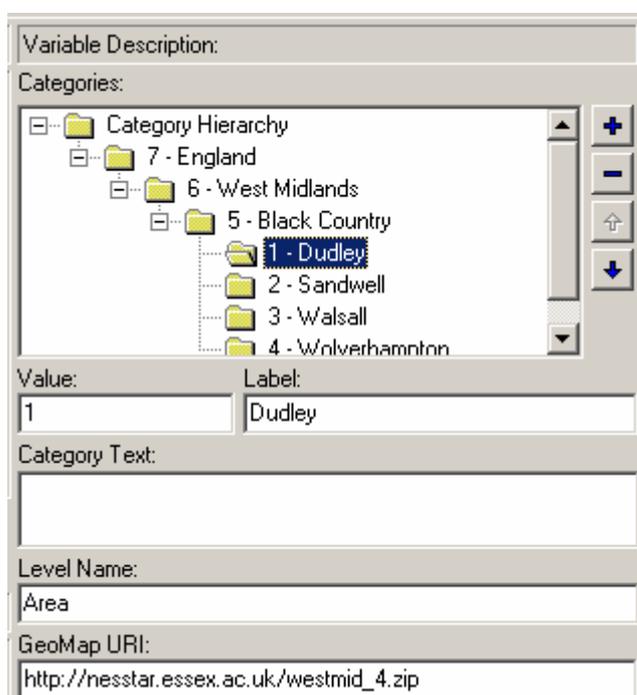


Figure 3. Publisher level name and map reference (for that level). Please refer to Section 6 for more information about maps.

Note: A 'Level Name' must be added for all dimension variables, even if they are not part of a hierarchy.

The next step is to use the Nesstar Cube Builder, see refer to [Section 5](#) below to add further information about your cube.

If you wish to add a map link to your data, please refer to [Section 6](#) before proceeding.

If you wish to add further data, see Section 4.3.3 below.

Please remember to save your file before proceeding.

4.3.3 Expand dimension

To add new data to an existing cube

1. Select 'Expand Dimension' [**Data | Cubes | Expand Dimension**].
2. Enter the category value and label.

The 'data entry' section is updated with new rows that include all the combinations of the new categories with existing categories in the other dimensions.

3. In 'Data Entry' add the data to the measure variable(s).

4.4 Measures to Dimensions

Within the 'Data' menu of Publisher there is an option under 'Cubes' to create a dimension variable using 'measure' information. This option can be used if you are preparing a cube and have a table that contains a large number of measures that you wish to treat as one dimension. If you have a number of dimensions already in your table, using the NesstarExporter Excel macro will often not work as the Excel limits on table size will soon be reached. Using the Publisher option is therefore very useful.

Example:

The table below shows a few rows from a much larger table containing information for a number of years, and for a number of areas. It also contains a number of columns of information that we want to treat as another dimension, e.g. 'Pop-Total', 'Did not work', 'Worked', and 'Worked-part year'.

Nesstar Cubes: The Cube publishing process

Year	Area	Gender	Age	Pop-Total	Did not work	Worked	Worked-part year
2000	Country	Total	15 to 24 years	3988200	1196035	2792170	1826465
2000	Country	Total	Average age	43.9	54.7	39	36.2
2000	Country	Total	Median age	42.9	61.5	39.5	35.4
2000	Country	Total	Total	23901360	7459415	16441945	6281740
2000	Country	Male	15 to 24 years	2034290	591820	1442465	943195
2000	Country	Male	Average age	43.1	54.1	39.5	36.5
2000	Country	Male	Median age	42.3	62.7	39.9	35.5
2000	Country	Male	Total	11626790	2922635	8704160	3133985
2000	Country	Female	15 to 24 years	1953910	604210	1349700	883265
2000	Country	Female	Average age	44.6	55.1	38.5	35.8
2000	Country	Female	Median age	43.4	60.7	39.1	35.3
2000	Country	Female	Total	12274570	4536785	7737790	3147755
2000	District	Total	15 to 24 years	117345	37720	79625	55975
2000	District	Total	Average age	44.8	55.6	38.8	36.2
2000	District	Total	Median age	43.8	61.5	39.4	35.9
2000	District	Total	Total	732365	259990	472375	196045
2000	District	Male	15 to 24 years	58825	18715	40110	28345
2000	District	Male	Average age	44	55.1	39.4	36.9
2000	District	Male	Median age	43.4	62.6	40	36.4
2000	District	Male	Total	350885	101950	248935	99855
2000	District	Female	15 to 24 years	58520	19005	39510	27630
2000	District	Female	Average age	45.5	55.9	38.1	35.5
2000	District	Female	Median age	44.2	60.6	38.9	35.3
2000	District	Female	Total	381480	158035	223445	96190

To prepare the above table for Nesstar Publisher:

1. From Excel save this table in tab or 'csv' format.
2. Import into Nesstar Publisher. You will then find that all column headings are now listed as variables in Publisher. The columns containing numerical information will be treated as separate measures if the cube is left as it is. To make these 'measures' into a dimension.
3. Select **Data | Cubes > Measure to Dimension**.
4. Select the Measures to be categories in the new variable/Dimension, e.g. in the above table you would select 'Pop-Total', 'Did not work', 'Worked' and 'Worked-part year'.
5. Click 'OK' and new Dimension and Measure variables will be created. These will be labelled 'Merged Dimension' and 'Merged Measure'.
6. Rename the new Dimension and Measure variables as required.
7. Continue to prepare cube as detailed within this guide.

5 Using the Cube Builder

Open the Cube Builder from the **Tools** menu in the Publisher. The Nesstar Cube Builder can be opened from the shortcut on your PC desktop, if available, and then selecting the required '.NSDstat' file.

5.1 Cube Definition

5.1.1 Single measures

1. Click on the '**Cube definition**' tab.
2. Select the **Time Dimension** if there is one (e.g. 'Year'), using the dropdown list.

Note: The time dimension is used to create time series graphs. However, when used for 'Stock' measures, the dimension variable entered here will not be aggregated.
3. Select the **Geographical Dimension**, if present.
4. Specify a 'Measure' by clicking the '+' button. A new window will open from which your measure can be chosen.
5. For '**non-additive**' cubes you do not need to do anything else and you can continue by defining a default view for your cube ([Section 5.2](#)).
6. For additive cubes: choose either '**Stock**' or '**Flow**' from the Additivity dropdown box.
7. Then select the type of measure you wish to publish: Sum, Mean, Count, Min, or Max (see [Section 7.2](#)).

5.1.2 Multiple measures

When you wish to publish several measures for the same dataset (see [Section 7.2](#)), you need to repeat steps 4-7 (above) for each measure:

1. Add the measures to the '**Measures**' box.
2. Highlight each measure in turn, and choose the additivity from the 'Additivity' box.

Note: A mixture of non-additive and additive measures can be published.

3. For additive measures, also choose the type of measure.

5.1.3 Grouping measures

Measure types that go together, e.g., Sales Min, Sales Max and Sales Sum, can be grouped together. The 'Measures' box defines the measure whereas the 'Measure definition' window, groups measure types together to be part of the same measure.

To do this follow these steps instead of the instructions in section 5.1.2 above:

1. Add the first measure type to the 'Measures' box (e.g., 'Sales Min').
2. Choose additivity (Stock or Flow), and select the same measure type (e.g. Sales Min) for the 'Documentation Variable'.
3. Now select the relevant measure types for the other variables, i.e., Sales Sum for the Sum variable, Sales Min for the Min variable, and Sales Max for the Max variable.
4. If another series of measures is included in the data set, e.g., Profit Max and Profit Sum, repeat steps 1-3.

5.2 Default View

The Default View is used to define the table that users will see when they initially open the cube. Dimensions need to be placed in columns, rows, or in layers.

The 'layers' dimension acts as a filter on the cube data. For example, if GENDER is selected to go into layers, this dimension will not be visible in the table, but depending on which subset is chosen (e.g., male or female) the table will present data for that one category only, e.g. 'Males'.

The end user can still choose 'Females' or move the GENDER dimension into the table to see data for 'Males' and 'Females' at the same time.

By using 'Default View', the data publisher is able to present the cube in the format they find most appropriate, and the flexibility of the system allows the end user to alter that view if necessary.

To set the default view of a cube:

1. Click on the '**Default view**' tab.
2. Add the dimensions that are to be displayed into the default view using the '+' button next to the **Columns** and **Rows** boxes.
3. In '**Layers**' add the dimensions that are not to be visible in the table initially.

Note: These will still be available in a dimension box above the table and can be added to the table by the user if required.

4. For each dimension in the **Column**, **Row** and **Layers** boxes, select the 'level' or 'subset' that will be displayed. To do this, click on a dimension name and in the '**Select by dimension**' box choose either '**Select by level**' (if only one level of a hierarchical dimension is to be displayed) or '**Subset**', and check the required categories.

Note: 'All' cannot be used for non-additive data, in 'Select by level'.

5. Finally, add the measure variable in the 'Measures' box. For multiple measures, add all measures into this box. The first measure will be shown in the published table whereas the other measures will be available for the user to select.

Note: If you have a large number of categories and/or dimensions, think carefully about the contents of the default view. If you try to include everything when you have a large table, the information may not be displayed within WebView. Select those categories that will provide a useful picture of the data available.

5.3 Publishing cubes

1. Click on '**Publishing**' and select the Server you wish to publish to. If the correct server is not listed, use 'Add server...' to add it. Then use '**Publish cube...**' to publish the cube.
2. Use 'Manage server' under '**Publishing | The <server name> Server**' in the Cube Builder to add catalogues or move datasets around.
3. If the cube is not to be published immediately, save the files created by the Cube Builder using **File | Create Cube File**. Two files, with the extensions '.NSDstatCube' and '.xml', will be created. These files can be manually dropped into the data folder of the Nesstar server if necessary.

For further information on publishing and managing your Nesstar Server, please see the Help function in the Nesstar Publisher and Nesstar Cube Builder.

6 Adding map functionality

The Nesstar Publisher provides support for maps through the use of 'Shape' files. See www.esri.com/software/opencv/openpdf.html for further information.

Nesstar expects a link to be added to the Publisher which points to a zip file containing the following three files:
a '.shp' file, a '.dbf' file that contains a numeric identification code for each polygon (geographic unit) which has to match the value of the corresponding geographic data contained in the cube, and an index '.shx' file.

6.1 Creating map files

6.1.1 The shp and shx files

The usability of the map will depend on the map detail. The map file will be larger the more detail it contains. Some GIS programs allow automated generalizing of 'Shape' files. Within Nesstar, most detail can be omitted, as only a rough outline of the areas (boundaries) on the map is often sufficient.

A MIF/MID map can be translated into 'Shape' format using the translator tool in MapInfo.

6.1.2 The dbf file

The .dbf file that is part of the 'Shape' file needs to include the same identification codes that match the corresponding areas in the data file. This file can be opened in a spreadsheet or a dBase4 editor but should be saved in 'dBase4' format.

The 'ID' information should be in the first (left-most) column. The second column must be called 'NAME' and should contain the area names. These will then be displayed when hovering over the map within WebView.

The ID column should contain all the IDs in the '.shp' file and the numbers must match the numbers relating to the spatial categories in the dataset. This column should only contain numbers which if using Excel may mean removing the leading ' (quote) in each cell.

The IDs do not need to be consecutive, but do need to match the category values of the areas they relate to in the Nesstar Publisher. If there are areas on the map that are not in the data, these will be presented as 'Missing' when the map is viewed through Nesstar. It is also possible to have regions in the data that are not on the map.

The 'NAME' column (of 'character type') should contain the names of the corresponding areas and must be 32 characters or less in length.

Note: Although the '.dbf' file can contain many columns, only the first two will be used.

6.2 Placing the map files in an accessible area on a Server

The three map files (.shp, .shx, .dbf) should be placed in a zip file which can then be placed on any Web Server. The URL of this file should then be noted as it will be needed by the Publisher. See section 6.3 below.

The map can be placed on the Nesstar Server, e.g. within the **[Nesstar server]/jboss/server/default/deploy/root.war** directory. The link would then be either a relative one, i.e. /nameofmap.zip, or the full URL to this area, i.e. <http://Server-address/nameofmap.zip>

For example: <http://www.nesstar.com/map.zip>.

6.3 Adding the map link to the data file

To add a link to a map within a cube file:

1. Select the required '.NSDstat' file.
2. Select the spatial dimension variable.
3. Go to **Variable Description** in the Publisher and click on one of the categories (areas) at the appropriate level.

4. Add the URI in the '**GeoMap URI**' box below it.
5. Save the file.

Note: If the actual map file (.zip) is updated, but the name stays the same, files referencing this map will need to be re-published so that the updated map information is used.

6.4 Adding more than one map for one dataset

As maps are linked to a 'level' of geography, in some cases you may want to add more than one map to a cube, e.g., one for each level in the spatial dimension. A separate '.zip' file needs to be created for each map. In the Nesstar Publisher, each map URI should be added to the corresponding level within the spatial dimension.

For example: There may be two maps for England - one at Counties level (containing 25 counties), called counties.zip, and one at Town level (containing 500 towns), called towns.zip. The areas for the Town level could be numbered 1-500, whereas the areas for the Counties level could be numbered 501-525. The areas should have unique IDs. In the Publisher, the '**GeoMap URIs**' for the two maps would be added at the correct level of the spatial dimension (Counties and Towns).

7 More Complex Cubes

7.1 Multiple Tables

Often, several tables exist containing related data. For example, there may be a table displaying the number of houses built in Southern England for 1999, and a similar table for 2000 and 2001. Instead of publishing three separate cubes, each containing the data for a specific year, it would be more useful to add the data together to form one cube. In this way, the user has the opportunity to compare data for the different years or to only view the data for one particular year.

The cube can be prepared by combining the tables and adding an extra dimension called YEAR (see Table 6 below). When the cube is published, the user can decide on the default view, e.g., whether YEAR should be in the rows or columns of the table, or not shown at all.

YEAR	PROVIDER /AREA	Private	Council
1999	Essex		
	Sussex		
	Kent		
	Gr. London		
2000	Essex		
	Sussex		
	Kent		
	Gr. London		
2001	Essex		
	Sussex		
	Kent		
	Gr. London		

Table 6. Multiple tables merged together.

7.2 Multiple Measures

In some cases there may be more than one measure available for a particular cube. For instance, data may exist for both the total sales profit for a company and the number of units sold. This means that for each cell in the table, two figures are available.

Two or more measures can be added to the same cube and published together. One of the measures is then shown in the published table by default, and the others are available for selection by the user.

It is also possible to publish more than one type of the same measure. For example, data may exist for the average age of people in a number of towns. If the minimum

and maximum ages for each town are also known, the data for minimum and maximum ages could be published alongside the data for average ages.

The available measure types are:

Mean: For 'Mean' measures, aggregations can only be performed if an additional measure variable is present; either 'Sum' or 'Count'. If one of these is present, the average value for higher levels in a hierarchy can be calculated using the formula $\text{Mean} = \text{Sum} / \text{Count}$. If no other measure type is present, the data cannot be aggregated and will be presented as zeros at the higher level of a hierarchy. For example, the average age for people in a particular industry

Count: Number on which a mean is based, e.g., number of people that were included to calculate the average age

Min: Minimum, e.g. the minimum age of people working in a particular industry.

Max: Maximum, e.g. the maximum age of people working in a particular industry.

Sum: Total, e.g., the total number of people working in a particular industry.

The input file for the Nesstar Publisher should contain all the measures for a cube. The data for each measure should be placed alongside the other information for each cell in a cube, as follows:

Area	Gender	Age (Mean)	Age (Min)	Age (Max)
Newtown	male	46	2	78
Newtown	female	52	1	88
Oldtown	male	45	1	67
Oldtown	female	46	3	78

Note: When 'multiple measures' are published for the same cube, these measures may be additive or non-additive or a mixture of both.

7.3 Incomplete Cubes

In some cases, not all data may be available to form a full cube. For example, in Table 7 a table is presented on "Number of jobs in Southern England" (fictional) where data on the percentage change from the previous year is not available for 2000.

JOBS	Number of jobs			% change	
	2000	2001	2002	2001	2002
Essex	120,564	125,644	163,257	4.21	29.94
Sussex	156,232	195,471	187,463	25.12	-4.1
Kent	325,567	287,457	303,911	-11.71	5.72
Gr. London	447,691	489,253	463,745	9.28	-5.21

Table 7. An incomplete cube.

There are three possible ways in which to deal with an incomplete cube:

1. Publish the table as it is, and the missing category will be added automatically, with no data present. Thus, for Table 7, the category **2000** will be added under **% change**. If you wish to present both **Number of jobs** and **% change** in the same table, the table as in the example should be created with the dimensions JOBS, YEAR and AREA.

A variation on this is to present the data as multiple measures (see [Section 7.2](#)) such that either the **Number of jobs**, or **% change** is displayed. The advantage of using multiple measures, rather than a dimension with two categories, as in Table 7 (above), is that you can publish the table, if suitable, as an additive cube for one or all measures (see [Section 2.3](#)).

2. Alternatively the table can be split into two tables, from which two cubes are published. In Table 7, a cube entitled '**Number of jobs**' could be made, along with a cube called '**% change**'. However, '*number of jobs*' and '*% change*', are related data and the user may prefer to see this data in the same cube. Therefore, if only one or a few extra categories need to be added to make a full cube, adding categories is preferred. It is advisable to split the table into two cubes only when a number of empty cells would be created through the addition of extra (empty) categories.
3. A third option is to create one (hierarchical) dimension under which the other categories are placed. Thus, a dimension **JOB** would have two categories, **Number of jobs** and **% change**, which would have lower-level categories **2000**, **2001** and **2002**, and **2001** and **2002**, respectively. For Table 7 this is unnecessary but for more incomplete cubes, this is a useful option, as demonstrated in Table 8 below.

Region / Population	Dudley	Sandwell	Walsall	W'hampton
Population 30/06/01	305100	282800	253300	236400
Live births in 2001				
Males	1724	1854	1621	1457
Females	1589	1840	1578	1443
Totals	3313	3694	3199	2900
Under 2500 grams (%)	8	11	10	10
Conceptions in 2000				
All ages	4271	4845	3926	3788
Under 18	272	346	306	286
Rate per 1,000 women aged 15-44	71	82	77	77
Rate per 1,000 women aged 15-17	51	63	63	64
Deaths in 2001				
Male	1568	1609	1225	1285
Female	1610	1716	1314	1245
Total	3178	3325	2539	2530
Standardised Mortality Rates (UK=100)	101	116	102	102
Deaths under 1 year				
Stillbirths	20	23	17	16
Under 1 week	7	9	11	18
Under 4 weeks	10	13	12	20
Under 1 year	16	25	18	26
Stillbirth rate	6	6	5	6
Perinatal mortality rate	8	9	9	12
Infant mortality rate	5	7	6	9

Table 8: An incomplete cube.

In Table 8, above, a variety of information, all related to population data, has been brought together. The only dimension that can be created here (other than REGION) is POPULATION, which contains all the categories in the left-hand column. No other dimensions can be created, as the data points do not have anything in common, other than that they are population data, e.g., there is no 'gender' associated with all data points and the data points are not all about birth or all about death.

Rather than splitting the table into several smaller cubes, the POPULATION dimension can be turned into a hierarchical dimension, preserving the structure of the table. The categories in bold would be at one level, and the remaining categories at the lower level. Data missing at the higher levels is not problematic, but when viewed through Nesstar WebView these cells will contain a dash ('-') as the information is not available.

8 Hints and Tips

8.1 Nesstar Exporter Excel Macro

8.1.1 Security settings

The 'NesstarExporter' macro will not work if the security settings in Excel are set to 'High'. To change the security setting within Excel, go to **Tools | Macro | Security** and set to 'Medium'. Then close and reopen Excel.

8.1.2 Category values and labels in Excel

In the Excel file, only category labels should be used, e.g., do not use "1 Male", "2 Female" spread over two columns (i.e. one for the value and one for the label). Use only the labels, e.g., "Male", "Female". Use the values rather than the labels only if the values are strings or numerical values that are not consecutive and need to stay as they are. Otherwise, the Publisher will add numerical values (1-n) when it imports the labels.

8.2 Additive Cubes

8.2.1 Higher-level data are calculated automatically

For additive cubes, higher level data in a hierarchical dimension will automatically be calculated. For example, if the dimension REGION, contains 'Towns' at the lowest level, and 'Counties' at the higher level, the data for 'Towns' will be used to calculate the data at the 'Counties' level.

It is important that these higher-level data are NOT present in the original data, as the calculations may not be performed correctly. If higher-level data is present in the original table, these need to be deleted, either in the original file, the Publisher input file, or within the Publisher itself.

8.2.2 Missing data

Data can be missing in two ways: the cell for a particular data point is empty (e.g., there are no population data available for a certain year for the category Colchester) or not all categories for a certain high-level category are in the table (e.g., for the County Essex, the town Colchester has not been included). In the latter case there will be no actual empty cells in the table.

If there are missing data it is up to the data publisher to decide whether or not to publish the table as additive. In the strictest sense, one cannot create higher-level data if not all lower-level data are available. For example, it is not possible to calculate the total population of a County if data for one or more of the underlying Towns are missing. However, with data such sales, it may be preferable to present some approximation of the total sales, rather than have no data for the (higher-level) total.

8.3 Access to Maps when running the Nesstar Server off-line

In some cases, maps may not be local to the server, i.e., they are stored elsewhere. When using Nesstar WebView off-line, you then need to make sure the map is added to the Server before your machine is taken off-line, or the map functionality will not work.

1. Open a dataset on Nesstar WebView and click on the map icon.
2. This action will save the map in C:\Nesstar-Server\jboss\server\default\deploy\root.war\nesstar\temp\maps\nesstarweb\kart.

Note: This is only a temporary folder which will be deleted as soon as the server is stopped.

3. This is only a temporary folder which will be deleted as soon as the server is stopped. Copy this map into C:\Nesstar-Server\jboss\server\default\deploy\root.war.
4. The map will now be available when you use the Nesstar system off-line.

Appendix A

A.1 The Nesstar Exporter Excel Macro

To facilitate the process of transforming a table into a structure suitable for input to the Publisher, Nesstar can provide a Nesstar Exporter Excel Macro. The Exporter macro should not be used if the data are already in the form described in [Section 3.1](#). It is only suitable for data that are in Excel format, or that can be copied into Excel.

A.1.1 Prepare the data for the Exporter

1. Copy the data table into an Excel spreadsheet, or to a new worksheet if the table is already in an Excel file.
2. Delete empty rows and columns to avoid the creation of empty variables in the publisher.
3. Do not use asterisks [*] or other characters to indicate missing values. Leave cells with missing data blank.
4. Add rows and columns in which to write names for the header and stub dimensions. The header consists of the names of the dimensions at the top of the table in the columns. The stub consists of the names of the dimensions (or variables) on the left-hand side of the table in the rows. See Figure 4 (below).

Figure 4. Nesstar Exporter Excel Macro

8							
9	STUB		Sex	Year	Male	Male	Male
10			Year		1993	1994	1995
11	Hair Colour	Blond	Isle of Anglesey		458	466	471
12	Region	Blond	Gwynedd		933	965	1,000
13		Blond	Conwy and Denbighshire	CATEGORY NAMES	1,357	1,357	1,361
14		Blond	South West Wales		2,406	2,530	2,691
15		Blond	Central Valleys		1,972	2,089	2,221
16		Blond	Gwent Valleys		2,082	2,246	2,441
17		Blond	Bridgend and Neath Port Talbot		1,966	2,111	2,261

5. Add category names for each dimension. Note: the 'Sample' sheet in the Nesstar Exporter Excel Macro is an example of a table ready to be exported. Make sure when category names are repeated (such as 'Blond' in Figure 4) they are written identical every time. The Publisher imports any different category names as separate categories.
6. Ensure that there are no unusual characters left in the data (e.g., superscripts which refer to footnotes). Data must be in 'number' format, i.e. no percentages or pound signs should be present.
7. For hierarchical dimensions (non-additive data only), the categories should all be at the same level, i.e., all categories should be in the same row (for a header dimension) or column (for a stub dimension).

Note: The order in which the categories are in the table, from top to bottom and left to right, is the order in which the categories will appear when the table is published. That is, the publisher uses the order in which the categories first appear in the input file. Therefore, if the categories are not in the required order (e.g., for GENDER the order is Male, Total, Female), then this needs to be changed at this stage. There is no opportunity to change it later.

A.1.2 Using the Nesstar Exporter Excel Macro

1. Open the Nesstar Exporter [NesstarExporter.v4.xls] and choose to 'Enable Macros', if prompted.

The Exporter has three worksheets: 'Sheet1', 'Sample', and the 'Export macro'. The Export macro creates the file needed for the Publisher. The 'Sample' sheet is for illustration purposes only.

2. Copy the newly created table into the worksheet called 'Sheet1'.
3. Highlight the entire table including the stubs and headers.
4. Go to the **Export Macro** sheet.
5. Press 'OK' and fill out the fields in the new window that appears:

Source Worksheet Name	The worksheet with the table (This is usually sheet1 unless a new name has been assigned)
Export File Name	Enter a name for the new file
Number of Stubs	Enter the number of different dimensions (variables) on the left hand side of the table
Number of Headers	Enter the number of different dimensions (variables) on the top of the table
Measure Label	Enter a name for the actual data points – e.g., number of jobs

6. Click 'OK'.
7. Close the file.

A new 'Excel' file will then be displayed containing:

- The original table on 'Sheet1',
- A comma-separated version of that table,
- The Export Macro, and
- The sample sheet.

A.1.3 Check your data

1. Check that all of the data and category names are correctly displayed in the new version of the file (this is the sheet that has the Export file name you provided).
2. This sheet will have been automatically saved as a '.csv' (comma-separated) file, but it is advisable to save this as a '.txt' (tab-delimited) file. If your data contains commas in the category names you will encounter problems when the file is imported into the Publisher.
3. Exit Excel. This file can be saved as all the information added will then be preserved along with the Excel macro.

The data are now ready for 'import' into the Publisher.

A.1.3 Multiple measures

When there is more than one measure or measure type, the process above needs to be repeated for each measure, i.e., a full table is exported for each measure. From the resulting delimited files, one file then needs to be created, that contains the columns for all dimensions, followed by the columns for each of the measures (see [Section 5.1.2](#)).

A.2 Using Data from a Database (e.g., MS Access)

1. Create a query that will produce a table with columns in the form as in [Section 3.1](#).
2. Choose "Text Files (*.txt; *.cvs; *.tab; *.asc)" as the Export format.
3. Choose to use tab-delimited files.
4. Column titles are included at the first row. Text Qualifier is selected as 'none'.

Disclaimer:

While the University of Essex and the Norwegian Social Science Data Services (NSD) have taken all reasonable efforts to compile accurate documentation, we cannot accept liability for any loss or damage consequential or otherwise suffered by the client or any third party arising as a result of its use.