Project no: 022936
Project acronym: Beneris
Project title: Benefit-risk assessment for food: an iterative value-of-information approach

Instrument: STP-Specific Targeted Project

Deliverable 39:

Combined database

Due date of deliverable:	**1 Jan, 2009**
Actual submission date:	**15 Nov, 2009**
Dissemination level:	**PU**
Start date of project:	**April, 1st 2006**
Duration:	**3,5 years**

Organisation name of the lead contractor for this deliverable: **THL**

Combined database
(D39)

This documentation of this deliverable consists of three sections:

1) General description of the Opasnet base
2) Description of the Opasnet Base structure
3) Connection from Opasnet variables to the Opasnet Base: An example

The first section introduces the Opasnet Base and its functionalities The wikipage from which section 1 has been printed out below can be accessed at http://en.opasnet.org/w/Opasnet_Base using the following credentials: **username=**bioher, **password=**qADaC4h

The second section describes the interface and the functionalities of the Opasnet Base in further depth. The wikipage from which section 2 has been extracted can be accessed (using the credentials above) at http://en.opasnet.org/w/Opasnet_Base_structure

The third section presents an example illustrating how Opasnet variables are connected to the Opasnet Base, from where one can retrieve the stored results of each variable.

1 Description of the Opasnet base

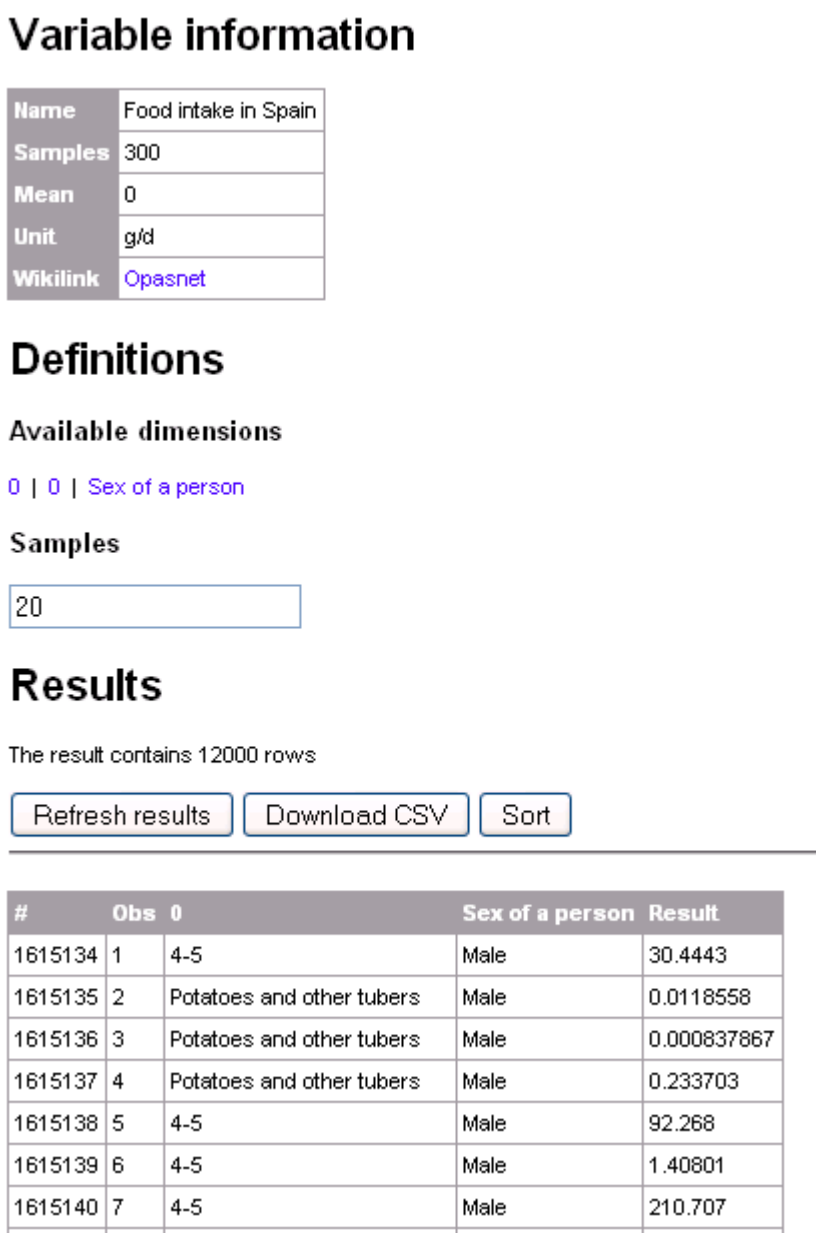
This section is a general description about the **Opasnet Base**. For a detailed description of its structure, see Opasnet Base structure.

> **Opasnet Base** is a part of Opasnet, constituting a storage and retrieval system for the results of variables and the data from studies. It is designed to be flexible enough to store information in almost any format: probability distributions or deterministic point estimates; spatially or temporally distributed data; or data with multiple dimensions. It can be used as a direct source of input data for quantitative models, thus making it possible to share multi-purpose information such as population data, climate scenarios, or the dose-responses of pollutants. The Opasnet Base can be accessed via links in "metadata boxes" on "variable" and "study" pages , or via a dedicated web interface, and also via the Analytica-based model Opasnet base connection.ANA.

Content

1 User interface for downloading data

The easiest way to use Opasnet Base is to find the variable or study of interest from Opasnet, and then click links either on the meta data box in the top right corner, or on the Result section. They will lead you to a page where you can select which details of the variable you want to see.



Variable information

Name	Food intake in Spain
Samples	300
Mean	0
Unit	g/d
Wikilink	Opasnet

Definitions

Available dimensions

0 | 0 | Sex of a person

Samples

20

Results

The result contains 12000 rows

Refresh results | Download CSV | Sort

#	Obs	0	Sex of a person	Result
1615134	1	4-5	Male	30.4443
1615135	2	Potatoes and other tubers	Male	0.0118558
1615136	3	Potatoes and other tubers	Male	0.000837867
1615137	4	Potatoes and other tubers	Male	0.233703
1615138	5	4-5	Male	92.268
1615139	6	4-5	Male	1.40801
1615140	7	4-5	Male	210.707

Figure 1. User interface of Opasnet Base.

The results in Opasnet Base are typically not single numbers but tables with several indices (also called dimensions or determinants). For example, food intake results may be given for several age groups, and both sexes separately. In this case, sex and age are used as indices. Because many variables may have a lot of indices, also the results may be large multi-dimensional tables. This makes it difficult to find the particular numbers the user is interested in.

To overcome this problem, the user interface makes it possible to select only those rows of each index that are of interest. It is also possible to select the samplesize, i.e. how many iterations are shown for each value. The output of Opasnet Base is a table that first lists technical information such are the row number of the result in the database, the number of the iteration, and the row description for each index. The last column, Result, contains the actual value of the variable.

It is possible to copy and paste the resulting table to another software. It is also possible to save the result first as a comma separated value (CSV) file for further use.

2 User interface for uploading data

There are two ways of uploading data. If you have Analytica Enterprise, you can open the module Opasnet base connection.ANA into your model and use it for uploading. You can even upload a whole model at one time. If you don't have Analytica Enterprise, you can use web interface to upload data. With the web interface, you can only upload one variable or study at a time.

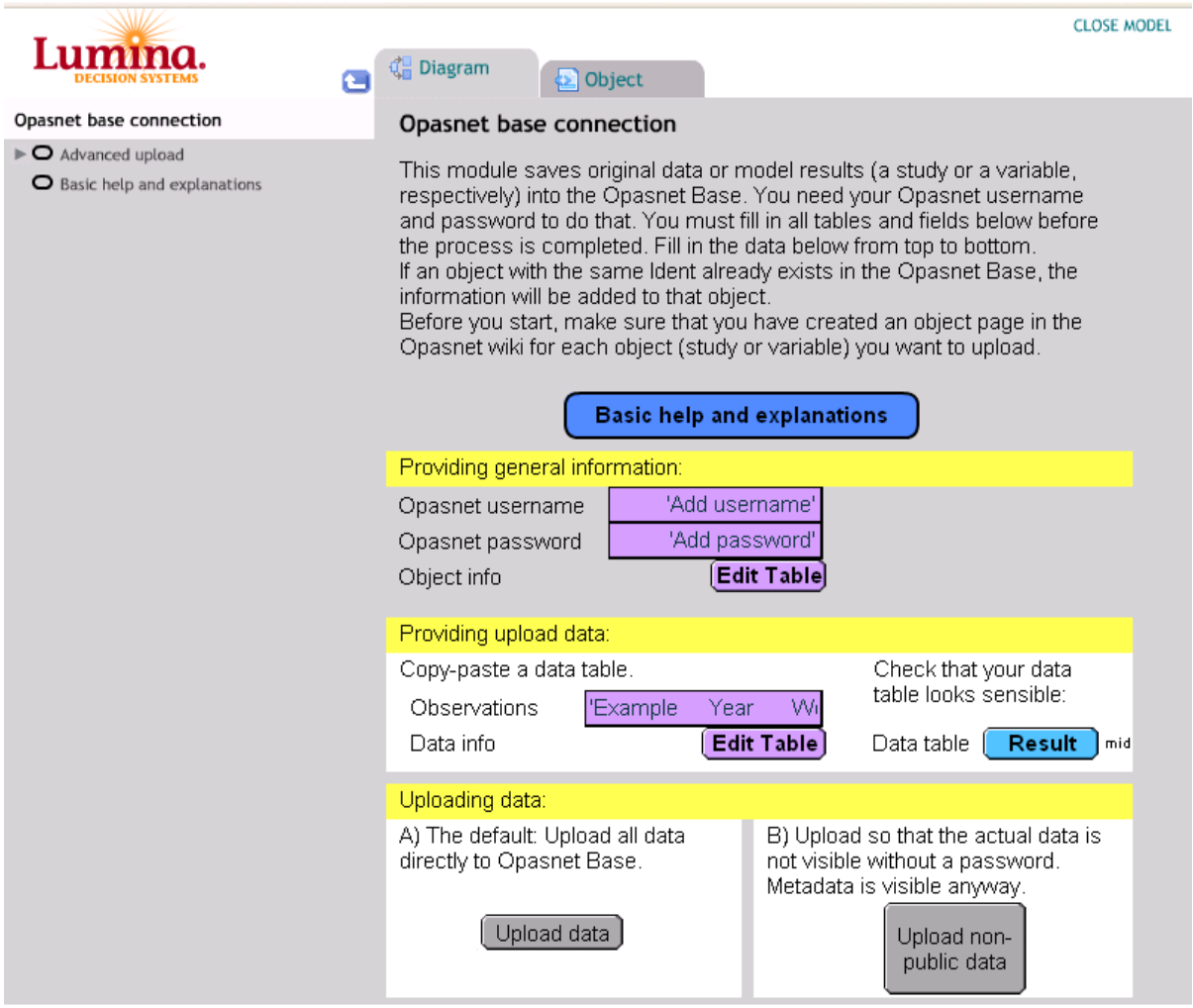


Figure 2. User interface for uploading data to Opasnet Base.

3 Some uses of Opasnet Base

3.1 Storage of interpreted model results

Originally, Opasnet base was designed to serve as the storage for interpreted model results, i.e. variable results. Variables attempt to answer specific real-world questions, and their results are the current best estimates as the answers to these questions. This is different to studies that report the observations from a single study. Variable results are expected to be eternally improving in time, while data from a study is fixed after the study has been done and the observations made.

3.2 Storage of study results

Opasnet Base can be used to collect observation data from studies. A study can be a traditional research study, documented in Opasnet Data afterwards, or it can be an Opasnet study where the data is collected on a particular Opasnet page with a web form. There can be several purposes to using Opasnet Base as a data repository:

- To collect observation data to be directly and conveniently usable in interpretations of variables and other objects.
- To collectively collect information about specific cases for the purpose of conditionalizing generalised assessment models with data specific to particular cases.

However, there are some things about variables and studies that should be understood:

- The object for a collection of observations is called a study, while the object of interpretations is called a variable. As an example, a study can collect information about a population group by a questionnaire and by taking a blood sample. The study identifier is the Obj.id in the Opasnet Base.
- The object may be divided into smaller pieces along one or more indices. For example, the questionnaire may have 30 questions, and therefore the questionnaire data can be indexed by an index with 30 columns (or rows, depending on which way you think), one row for each question. Each column of the study object has one cell, i.e. an answer to one question. For example, if ten blood markers will be studied along with 30 questions,the study object will have 40 cells, and the index has 40 columns (30 from the questionnaire, 10 from the blood sample). The cell identifier is the Cell.id in the Opasnet Base.
- For each individual patient, there is one row of observations (in the example each 40 cells). The observation row identifier is Res.Sample in Opasnet Base.
- The actual result of a particular cell of a particular patient is located in Sam.Result in Opasnet Base (or in Res_info.Description in the case where the result is text, i.e. non-numeric).
- Each study may be multidimensional just like a variable and have indices along e.g. space, time, or sex.
- If the data is collected using an Opasnet web form, then the timestamp and username or IP will be recorded for each entry into Resinfo.When and Resinfo.Who fields, respectively. This is not needed, if the data comes from a previously performed study (which is static data in the eyes of Opasnet).
- In some cases, it might be useful to restrict the number of entries per user to one. However, this is done only at the interpretation phase where only the last entry is counted. There are no restrictions to enter new data, and therefore a user may change his/her previous entry by simply making a new entry.

3.3 Downloading results automatically to Analytica model

It is possible to automatically download values from Opasnet Base into an Analytica model. To do this, you must have Opasnet Base connection.ANA module uploaded into your model, and you must have Analytica Enterprise version. Then you use

```
get_sample(variable_identifier)
```

or other functions to call values from Opasnet Base. In this way it is possible to keep data in Opasnet Base so that every time the model is run, it will use the most up-to-date values from the database. If the results are then uploaded to the database, they are automatically coherent with other results in the database. On the other hand, if you always want to use a particular version of data, that can be defined in the Analytica get_sample function as an additional parameter. In this way, you can always get results that are consistent with your previous results (but not necessarily with the newest information about the topic).

3.4 Making value-of-information analyses in Opasnet base

Value of information (VOI) is a decision analysis tool for estimating the importance of remaining uncertainty for decision-making. Opasnet base can be used to perform a large number of VOI analyses, because all variables are in the right format for that: as random samples from uncertain variables. The analysis is done by optimising an indicator variable by adjusting a decision variable so that the variable under analysis is conditionalised to different values. All this can in theory be done in the Opasnet base by just listing the indicator, the decision variable, and the variable of interest. Practical tools should be developed for this. After that, systematic VOI analyses can be made over a wide range of environmental health issues.

3.5 Analysing the change in the quality of a variable result in Opasnet base

All results that have once been stored in the Opasnet base remain there. Old results can be very interesting for some purposes:

- The time trend of informativeness and calibration (see performance) can be evaluated for a single variable against the newest information.
- Critical pieces of information that had a major impact on the informativeness and calibration can be identified afterwards.
- Large number of variables can be assessed and e.g. following questions can be asked:
    - How much work is needed to make a variable with reasonable performance for practical applications?
    - What are the critical steps after which the variable performance is saturated, i.e., does not improve much despite additional effort?

4 See also

- Open Office document about Opasnet Base
- Opasnet Base structure
- Opasnet
- Open assessment
- Opasnet base connection.ANA
- CSC e-infrastructure strategy

2 Opasnet base structure

This page is about the **structure of the Opasnet Base**. For a general description, see Opasnet base.

1 Scope

Opasnet base is a storage and retrieval system for results of variable and data from studies. What is the structure of Opasnet base such that it enables the following functionalities?

1. Storage of results of variables with uncertainties when necessary, and as multidimensional arrays when necessary.
2. Automatic retrieval of results when called from Opasnet wiki or other platforms or modelling systems.
3. Description and handling of the dimensions that a variable may take.
4. It is possible to protect some results and data from reading by unauthorised persons.
5. If is possible to build user interfaces for easily entering observations into the Base.

2 Definition

2.1 Data

2.1.1 Software

Because Opasnet base will contain very large amounts of mostly numerical information, the state-of-the-art structure is a SQL database. Because of its flexibility, ease of use, and cost, MySQL is an optimal choice among SQL software. In addition to the database software, a

variable transfer protocol is needed on top of that so that the results of variables can be retrieved and new results stored either automatically by a calculating software, or manually by the user. Fancy presenting software can be built on top of the database, but that is not the topic of this page.

2.1.2 Storage and retrieval of results of variables

The most important functionality is to store and retrieve the results of variables. Because variables may take very different forms (from a single value such as natural constant to an uncertain spatio-temporal concentration field over the whole Europe), the database must be very flexible. The basic solution is described in the variable page, and it is only briefly summarised here. The result is described as

$P(R|x_1,x_2,...)$

where P(R) is the probability distribution of the result and x_1 and x_2 are defining locations of a dimension where a particular P(R) applies. Typically locations are operationalised as discrete indices. A variable must have at least one dimension. Uncertainty about the true value of the variable is operationalised as a random sample from the probability distribution, in such a way that the samples are located along an index *Sample*, which is a list of integers 1,2,3...n, where n=number of samples.

- Old description of the structure

2.2 Dependencies

- Opasnet structure
- Open assessment

3 Result

Opasnet base is a MySQL database located at http://base.opasnet.org.

3.1 Table structure

3.1.1 Objinfo: new structure

The structure of Objinfo should be changed. The original plan was that there is at most one row of Objinfo per Object. Now it is clear that this does not have all functionalities we need. Instead, there should be a possibility to add any number of actions per object. Therefore, even the name of the table should be changed to Act. The structure should be changed accordingly:

- The field id is the primary field for the table. It is NOT the Obj.id any longer.
- A new field Obj_id should be added. This is the old field id.
- End field should be removed, it is not used.
- Url should be changed to Comment, as it may contain also other info.
- The length of Comment should be 250 characters (at least).

- Begin should be replaced by When, which is the current timestamp of the row addition.
- A new field Act_id should be added.
- A new table Acttype for actions should be added. It would contain only fields id and Act, and the following rows:
    1. Start the object
    2. Finish the assessment
    3. Add a reference
    4. Add an URL
    5. Peer review the object definition: accept based on the discussion
    6. Peer review the object definition: reject based on the discussion
    7. Peer review the object definition: accept (personal opinion)
    8. Peer review the object definition: reject (personal opinion)
    9. Clairvoyant test for the scope: pass
    10. Clairvoyant test for the scope: fail
    11. Save a run of the object.

4 See also

4.1 Some useful syntax

- http://www.baycongroup.com/sql_join.htm
- Opasnet base connection.ANA for Analytica: for writing and reading variable results into and from the database. Writing requires a password. For SQL used in the model, see the model page.
- Some historical queries

```
SELECT Obj.id, Obj.Ident, Obj.Name, Obj.Typ_id, Sty_id, Itemm.Ident as
Iident, Itemm.Name as Iname
FROM Obj
LEFT JOIN Sett ON Obj.id = Sett.Obj_id
LEFT JOIN Item ON Sett.id = Item.Sett_id
LEFT JOIN Obj AS Itemm ON Item.Obj_id = Itemm.id
```

NOTE! The queries below work in the new database "opasnet_base", not "resultdb" as the old versions.

List all dimensions that have indices, and the indices concatenated:

```
SELECT Dim.Ident, Dim.Name, Dim.Unit, Group_concat(Ind.Ident
ORDER BY Ind.Name SEPARATOR ', ') as Indices
```

```
FROM Obj AS Dim, Obj as Ind, Sett, Item
WHERE Dim.id = Sett.Obj_id
AND Sett.Settype_id=1
AND Sett.id = Item.Sett_id
AND Item.Obj_id = Ind.id
GROUP BY Dim.Name
ORDER BY Dim.id
```

List all indices, and their locations concatenated:

```
SELECT Ident, Name, Unit, GROUP_CONCAT(Location ORDER BY Roww SEPARATOR ',
') AS Locations
FROM Obj AS Ind, Loc
WHERE Ind.id = Loc.Obj_id_i
GROUP BY Name
ORDER BY Name
```

List all variables and their runs, and also list all indices (concatenated) used for each variable for each run.

```
SELECT Var_id, Run_id, Ident, Name, GROUP_CONCAT(Indic SEPARATOR ', ') AS
Indices, N, Method
FROM
   (SELECT Var.id as Var_id, Run.id as Run_id, Var.Ident AS Ident, Var.Name
as Name, Ind.Ident AS Indic, N, Run.Name AS Method
   FROM Obj AS Var, Obj AS Run, Obj AS Ind, Loccell, Loc, Cell
   WHERE Var.id = Cell.Obj_id_v
   AND Run.id = Cell.Obj_id_r
   AND Cell.id = Loccell.Cell_id
   AND Loc.id = Loccell.Loc_id
   AND Ind.id = Loc.Obj_id_i
   GROUP BY Var_id, Run_id, Ind.Ident ) AS Temp1
GROUP BY Var_id, Run_id
```

3.1.2 All tables: Overview

- We need **Ressec** (Result secure) and **Resinfosec** (Result info secure) tables for secure information. All other tables are openly readable except these two. They have the same structure as Res and Resinfo tables, respectively.

Obj		
Describes all objects		
FIELD	**TYPE**	**EXTRA**
id	int(10)	primary
Ident	varchar(20)	unique
Name	varchar(200)	
Unit	varchar(16)	
Objtype_id	tinyint(3)	
Page	int(10)	
Wiki_id	tinyint(3)	

Cell (previously Res)		
Cells of an object		
FIELD	**TYPE**	**EXTRA**
id	int(12)	primary
Obj_id_v (variable id)	int(10)	
Obj_id_r (run id)	int(10)	
Mean (mean of the cell)	float	
N (samplesize)	int(10)	

Loc		
Location information		
FIELD	**TYPE**	**EXTRA**
id	int(10)	primary
Obj_id_i (index id)	int(10)	
Location	varchar(1000)	
Roww (row # of index)	Mediumint(8)	
Description	varchar(150)	

--**1**: Unit should have at least 32 characters. --**Jouni** 19:29, 17 September 2009 (EEST)

--**2:** We can increase it to 64 at once. --**Juha Villman** 07:52, 18 September 2009 (EEST)

Item		
Items of a set		
FIELD	**TYPE**	**EXTRA**
id	int(10)	primary
Sett_id (set to which the item belongs)	int(10)	
Obj_id (item id)	int(10)	
Fail (membership not valid?)	tinyint(1)	

Loccell (previously Locres)		
Locations of a cell		
FIELD	**TYPE**	**EXTRA**
id	int(10)	primary
Cell_id	int(10)	
Loc_id	int(10)	

Res (previously Sam)		
Result distribution (actual values)		
FIELD	**TYPE**	**EXTRA**
id	bigint(20)	primary
Cell_id	int(12)	
Obs (previously Sample)	int(10)	
Result	float	

Sett		
List of sets		
FIELD	**TYPE**	**EXTRA**
id	int(10)	primary
Obj_id	int(10)	
Settype_id	tinyint(3)	

Settype (previously Sty)		
Types of set-item memberships		
FIELD	**TYPE**	**EXTRA**
id	tinyint(3)	
Settype (previously Stype)	varchar(30)	

Objtype (previously Typ)		
Types of objects		
FIELD	**TYPE**	**EXTRA**
id	tinyint(3)	primary
Objtype (previously Type)	varchar(30)	

Wiki (previously Wik)		
Wiki information		
FIELD	**TYPE**	**EXTRA**
id	tinyint(3)	primary
Url	varchar(255)	
Wname	varchar(20)	

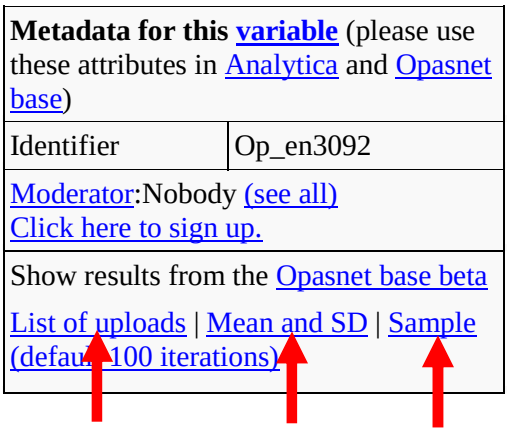
Resinfo (previously Descr)		
Additional description of the result		
FIELD	**TYPE**	**EXTRA**
id	bigint(20)	primary
Restext (previously Description)	varchar(250)	
Who	varchar(50)	
When	timestamp	

Objinfo (previously Inf)		
Additional information about the object		
FIELD	**TYPE**	**EXTRA**
id	int(10)	primary
Begin	date	
End	date	
Who	varchar(50)	
Url	varchar(250)	

3 Connection from Opasnet variables to the Opasnet Base: An example

Chapter 3 demonstrates the linkage between Opasnet and the Opaset Base in light of the example variable "*Food intake in Spain*".

In Opasnet, real-world phenomena of interest are analyzed by means of their dissection into "variables" that collectively represent the relevant set of interconnected elements, processes, and features of those phenomena which is necessary for analyzing, understanding, and mitigating the problem. Each variable is defined and described using a "variable page" in Opasnet. In a small window located in the top right corner of each variable page, there is a "metadata box" that among other information contains "result links" to the Opasnet Base, including the links *"List of uploads"*, *"Mean and SD"*, and *"Sample"*. This is shown in picture 1, where red arrows indicate the result links for the variable *"Food intake in Spain"*. In addition, one can embed such "results links" anywhere within a variable page.



Picture 1. "Metadata box" can be used to access the results of the Opasnet variable as stored in the Opasnet Base.

Other variables connected to the Opasnet Base are exemplified by the variables *"Food intake in Finland"* and *"Food intake in Denmark"*, which are linked to in the "References" section of the variable "*Food intake in Spain*" as printed out below. The original variable can be accessed at http://en.opasnet.org/w/Food_intake_in_Spain using the credentials **username=**bioher, **password=**qADaC4h

Food intake in Spain

Contents

1 Scope

Food intake in Spain for age groups 0-65+ years

2 Definition

2.1 Data

The data contains:

a) Main food groups (ingredient level, classification used in EPIC)

b) Nutrients

c) Nutrient supplements

d) Fish and other seafood by species

Age groups are: 0, 1, 2, 3, 4-5, 6-9, 10-13, 14-17, 18-24, 25-34, 35-44, 45-54, 55-64

This data is generated from the ENKID Study, which is the largest nutrition survey on the child and adolescent Spanish population to date.[1][2]. ENKID is a cross-sectional epidemiological study of a representative sample of the Spanish population aged 2 to 24 years (n=3534). Height and weight measurements were carried out using standard procedures and equipment. Obesity and overweight were defined according to BMI values for the 97th and 85th percentiles, respectively using Spanish reference data provided by Hernández et al (1998).

2.2 Causality

No upstream variables

2.3 Unit

g/day (fish consumers)

2.4 Formula

```
var a:=Spain_fi[Percentiles=Cumprob];
a:= if a = null then 0 else a;
{var b:=array(percentiles,[0,0.00001,0.00002,0.00003,0.0004,0.00005]);
if a=0 then b else a;}
cumdist(Cumprob,a,Cumprob)
```

3 Result

4 See also

[3]

[4]

5 References

1. Aranceta Bartrina J, Serra-Majem L, Pérez-Rodrigo C, Ribas-Barba L, Delgado-Rubio A. 2006. Nutrition risk in the child and adolescent population of the Basque country: the enKid Study. Br J Nutr. 2006 Aug;96 Suppl 1:S58-66.

2. Serra Majem L, Ribas Barba L, Aranceta Bartrina J, Pérez Rodrigo C, Saavedra Santana P, Peña Quintana L. 2003. [Childhood and adolescent obesity in Spain. Results of the enKid study (1998-2000)]. Med Clin (Barc). 2003 Nov 29;121(19):725-32.

3. Finnish data

4. Danish data