



*"This project has received funding
from the European Union's Seventh Programme for
research, technological development and demonstration
under grant agreement N°603946*



HEALS

**Health and Environment-wide Associations
based on Large population Surveys**

FP7-ENV-2013- 603946

<http://www.heals-eu.eu/>

D12.2 The HEALS GeoDatabase Platform

**WP12 Exposure and Health Data management
Version 0.1**

Lead beneficiary: 11 - VTT

Date: 30.09.2016

Nature: O - Other

Dissemination level: PU - Public




 HEALS FP7-ENV-2013-603946	D12.2 – The HEALS GeoDatabase Platform		
	WP12: Exposure and Health Data Management		Security: Public
	Author(s): Luc Cluitmans, Eija Parmes		Version: 2/28

TABLE OF CONTENTS

1	INTRODUCTION	5
2	ARCHITECTURE OF THE HEALS GEODATABASE.....	6
2.1	The original design.....	6
2.2	Changes in the design	7
2.3	Updated user management and access control.....	8
2.4	The file storage system.....	10
2.5	Database access.....	12
2.6	MySQL spatial extension	14
2.7	Environmental grid and raster data	14
2.7.1	Installation of Geoserver for raster data	14
2.7.2	Export of raster data by WMS protocol	14
3	AVAILABLE VARIABLES.....	16
4	HARMONIZATION OF THE DATASETS	23
4.1	Georeference and spatial coverage.....	23
4.2	Time stamp, temporal reference and coverage.....	23
4.3	Country names and codes.....	24
4.4	Grid systems	24
4.5	Variable names	24
4.6	NULL value	24
4.7	Link to external databases	25
5	ACCESS TO HEALS GEODATABASE	26
6	ISSUES.....	27

 HEALS FP7-ENV-2013-603946	D12.2 – The HEALS GeoDatabase Platform		
	WP12: Exposure and Health Data Management	Security: Public	
	Author(s): Luc Cluitmans, Eija Parmes	Version:	3/28

 HEALS FP7-ENV-2013-603946	D12.2 – The HEALS GeoDatabase Platform		
	WP12: Exposure and Health Data Management		Security: Public
	Author(s): Luc Cluitmans, Eija Parmes		Version: 4/28

Document Information

Grant Agreement Number	ENV-603946	Acronym	HEALS
Full title	Health and Environment-wide Associations based on Large population Surveys		
Project URL	http://www.heals-eu.eu/		
EU Project Officer	Tuomo Karjalainen, - Tuomo.KARJALAINEN@ec.europa.eu		


Deliverable	Number	12.2	Title	The HEALS GeoDatabase Platform
Work Package	Number	12	Title	Exposure and Health Data Management

Delivery date	Contractual	M36	Actual	dd/mm/yyyy
Status	Draft <input type="checkbox"/>		Final <input type="checkbox"/>	
Nature	Demonstrator <input type="checkbox"/>	Report <input type="checkbox"/>	Prototype <input type="checkbox"/>	Other X
Dissemination level	Confidential <input type="checkbox"/> Public X			

Author (Partners)	Luc Cluitmans (VTT), Eija Parmes (VTT)		
Responsible Author	Luc Cluitmans		Email luc.cluitmans@vtt.fi
	Partner	VTT	Phone +358 40 5492156

Document History

Name	Date	Version	Description

 HEALS FP7-ENV-2013-603946	D12.2 – The HEALS GeoDatabase Platform		
	WP12: Exposure and Health Data Management		Security: Public
	Author(s): Luc Cluitmans, Eija Parmes		Version: 5/28

1 Introduction

This is the report to accompany deliverable D12.2 of the HEALS project. This deliverable is of nature “O” (“other”) and described in the DoW as:


D12.2) HEALS GeoDatabase platform fully developed and populated: HEALS GeoDatabase platform fully developed and populated [month 36]

This report describes datasets available via the platform and provides updates on the architecture that was described in deliverable D12.1, its implementation, and its current status.

This deliverable describes the structure and architecture of the HEALS geodatabase. In addition, it lists the datasets that have been 1) imported to the database, 2) prepared for the import to the database, and 3) are available for such preparation. This content however evolves continuously with already available and totally new datasets (for instance results from analysis), and the new situation on datasets is updated and will be shown in the UI. The datasets and files from other work packages were provided to WP12 during summer 2016, which caused delay in populating the database.

Based on feedback from project partners it was decided to leave the analysis functions and visualization outside the HEALS database for now. The feedback indicated that analysis of the data will be done by each HEALS partner’s local software and it is unnecessary to copy the analysis functions to geodatabase. These applications also contain the needed visualization, so the visualization components are now also left outside the geodatabase. Later, simple visualization and analysis for the data that can be exposed publicly is to be planned.

Summa summarum, the geodatabase acts as a data archive, from where observations and measurements can be queried in synchronized format: variable, timespan and resolution, area and resolution, for input to applications software.

 HEALS FP7-ENV-2013-603946	D12.2 – The HEALS GeoDatabase Platform		
	WP12: Exposure and Health Data Management		Security: Public
	Author(s): Luc Cluitmans, Eija Parmes		Version: 6/28

2 Architecture of the HEALS GeoDatabase

This chapter provides background on the system, explaining restrictions, and explaining the different ways data is exposed. It does not go into details of what datasets are present.

In work package 12 of the HEALS project we implemented a “GeoDatabase platform” to act as data repository and information source on where to find data to be used within the project. The architecture and specifications were already described in deliverable D12.1 “*Report on Prototype Design of Geospatial Platform*”.

In the HEALS database the requirement of confidentiality on data on individual health and exposure is a crucial requirement and directed the implementation. Besides this data part of the environmental data resides in EDMS system. This is described in chapter 2.7. The system is now up and running and can be accessed at the URL <https://heals.vtt.fi/>. This chapter provides background information on how the system is implemented.

2.1 The original design

An overview of the architecture as designed in D12.1 is given in Figure 1.

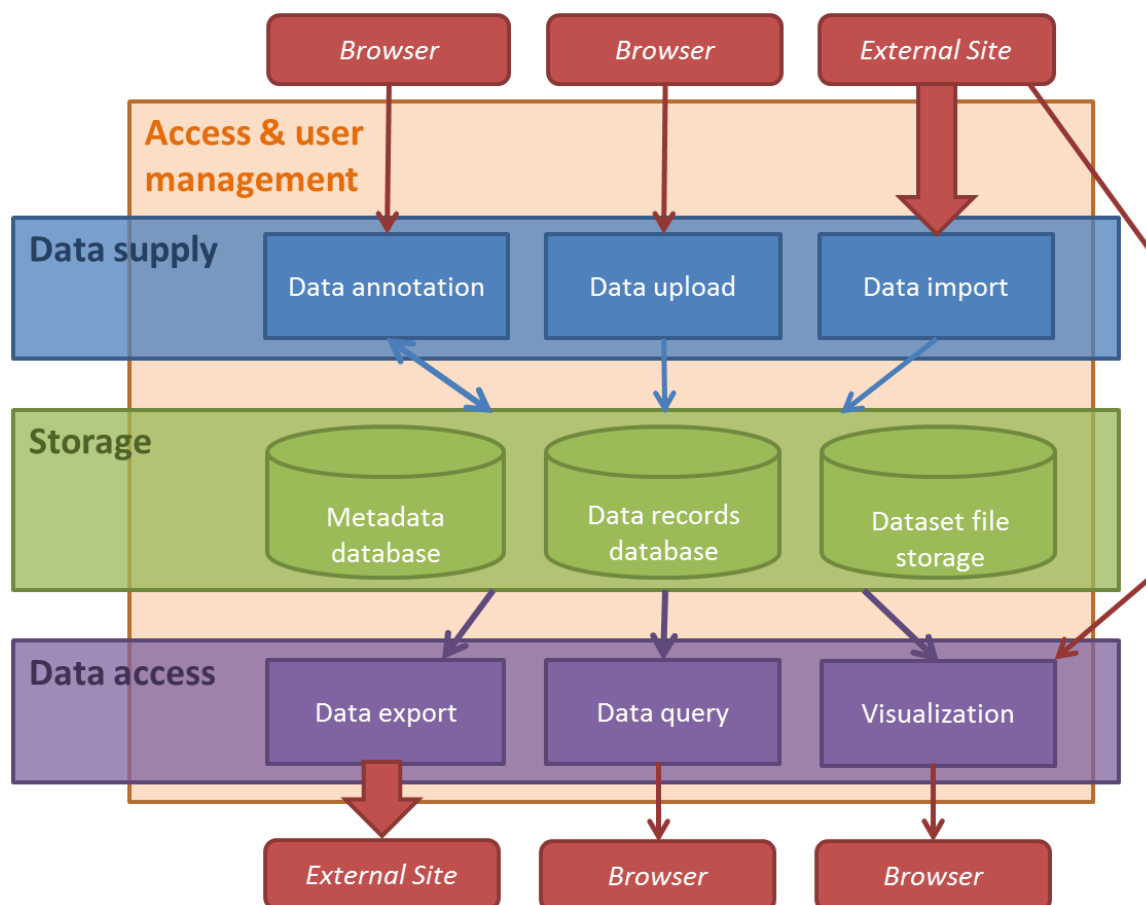




Figure 1: the architecture of the platform as designed in deliverable D12.1

 HEALS FP7-ENV-2013-603946	D12.2 – The HEALS GeoDatabase Platform		
	WP12: Exposure and Health Data Management		Security: Public
	Author(s): Luc Cluitmans, Eija Parmes		Version: 7/28

2.2 Changes in the design

While the overall design of the system is largely similar to what was described in D12.1 there are several updates to mention, both at a high level and at a lower level:

- The site now fully supports secure connections (HTTPS). It can be accessed as <https://heals.vtt.fi/>.
- Data storage is at the moment of writing largely file based.
- Database-based storage is available, but work to expose that in the user interface has only just started.
- Database storage is implemented in 3 different databases. This fact is not directly visible to end users, since database access is only indirectly through the web site.
 - A MongoDB NoSQL database can store “NoSQL” style data, such as JSON objects.
 - A Microsoft SQL Server Express database stores system information, such as the metadata used to provide the “file system like” data store: file names, descriptions, folder names, what files are in what folders, who can access what data, file versioning and history information, etc.
 - A MySQL database can store any tabular data. Initially it was intended to store this type of data in the SQL Server database as well, but once data started to become available it became clear that the capacity limit to that database would be hit. MySQL is more flexible in this aspect. Separating this data to a separate database also simplified the implementation a bit, since the technology used access to this type of data was already designed to be different from the technology used for accessing the system data: The system data is accessed via “Entity Framework v6” technology, while the “real data” is accessed via a more flexible technology (“Dapper” on top of “ADO.NET”)
- As mentioned above, data visualization implementation is postponed until a later stage.
- Data import is currently a manual process. While much of the data can be used “as is” in file form, trying to store it in databases is a more laborious task, showing that different data sources that should have been compatible in practice often aren’t. Data import would also be a good opportunity to perform anonymization, but that too is a laborious task that isn’t trivially automated.
- Since data is currently mostly file based, the “Data Query” and “Data Export” modules aren’t really at this moment as they were intended to be; they will become more fleshed out as the project goes on and actual requirements become clearer. For now they can be interpreted as “you can find data files and download them”.
- The system for defining access (which user can access which data, and in which way) has been extended beyond the D12.1 design; see the next subsection for a more detailed description
- Based on user feedback the prototype file system storage has been redesigned, and is no longer a large flat listing of files accessed by querying metadata, but appears more like a “classical” folder hierarchy. This hierarchy is now also tightly integrated with the updated access rights system, as will be explained below.
- The file storage subsystem now supports versioning / history: a file can be “overwritten” by a newer version, but the older version is kept available. As a side effect that means that files normally cannot truly be deleted – rather they are moved to a trashcan, with their history

 HEALS FP7-ENV-2013-603946	D12.2 – The HEALS GeoDatabase Platform		
	WP12: Exposure and Health Data Management		Security: Public
	Author(s): Luc Cluitmans, Eija Parmes		Version: 8/28

preserved. In case of emergencies (such as accidental leaking of personal data) the site administrator can truly delete the offending file.

2.3 Updated user management and access control

The “role” system described in section 4.2.1 of D12.1 has been extended a bit. While each user still has one of the “basic roles” described in that section (“Public”, “User”, “Contributor”, “Partner”, “Administrator”), a user can have multiple additional “roles” that allow for a more fine-grained access control.


For instance, every user has a set of roles that define the HEALS Work Packages they are involved in. These roles are combined with the new “access grant” system to define in more detail what access limits a user has to a particular resource (such as a folder, file, or database table). For instances, this system allows setting up a folder that is writable to users involved in a specific work package, read-only to other HEALS partners, and invisible when not logged in as any user. This system allows for “non-linear” access restrictions: it is possible to have a User A that has write access to Folder 1 but only read access to Folder 2, while at the same time the reverse is true for User B. Note that a simple ranked access system (from “Public” to “Administrator”) would not allow that.

The access grant system works as follows: for each “resource” a list of “grants” exist that define an “access level” for a “role”. Here you can think of a “resource” as, for instance, a folder (and all the files and sub-folders in it), but it may also be a data table, or possibly in the future specific columns in a data table. The “access levels” in our system are:

- None: the user cannot access the resource nor even learn that the resource exists. Example: a folder that has access “None” for a particular user isn’t even shown in the user interface. For security reasons attempts to access such a resource appear indistinguishable from attempts to access truly non-existing resources.
- Meta: the user cannot read or write the resource, but can find out that it exists and find out some of its properties. Example: a folder that the user has “meta” access to is shown and can be browsed, but files in it cannot be read. Sometimes it is useful to know that something exists, even without being able to actually access it.
- Read: the user can read the resource (e.g. download a file), but cannot write it.
- Write: the user can read and write a resource, e.g. replace a file with a new version, upload new files to a folder, create new subfolders, rename folders, and “delete” files or folders (move them to a trashcan)

The grant system is “additive”: to determine the access level a particular user has to a particular resource, we start from “no access” and subsequently all grants for that resource for all of the user’s roles are checked and the highest access level found is used as the actual access level. One implication of this is that adding a new role to a user can only *increase* their access level to any resource, never decrease it. The exceptions to this are the special roles “Demo” and “Enabled” mentioned in D12.1: Presence of the “Demo” role means that the resulting access right will be capped at “Read” (that is: if the calculated access level would be “Write”, it is lowered to “Read”); Absence of the “Enabled” role means that the user gains the “Public” role and all other roles are ignored. In other words: the system acts as if the user was not logged in at all. Normally “disabled” users cannot even log in, but this case can play a role when a user is already logged in and then is disabled (think of the scenario of a reported hacked account).

Here are a few screenshots that show the access control system in action.


 HEALS FP7-ENV-2013-603946	D12.2 – The HEALS GeoDatabase Platform		
	WP12: Exposure and Health Data Management		Security: Public
	Author(s): Luc Cluitmans, Eija Parmes		Version: 9/28

HEALS Data Platform Log in

[Documents](#)

Documents

Not finding what you were expecting to find here? Please note that the list of folders shown here depends on the access rights assigned to your user account, so make sure to be logged in! If you cannot see a folder that should be visible to you please contact the site administrator.

 [Public Documents](#) (changed Fri 2016-01-29)


© 2016 VTT - HEALS WP12 www.heals-eu.eu  HEALS














Figure 2: The site’s “Documents & Data Files” page when not logged in. Only the “Public Documents” folder is visible.

HEALS Data Platform Documents & Data Files Example Partner Log out

[Documents](#)

Documents

Not finding what you were expecting to find here? Please note that the list of folders shown here depends on the access rights assigned to your user account, so make sure to be logged in! If you cannot see a folder that should be visible to you please contact the site administrator.

 Data Files - WP05	(changed Tue 2015-10-13)
 Data Files - WP08	(changed Thu 2015-09-10)
 Data Files - WP09 (all HEALS partners)	(changed Fri 2016-07-15)
 Data Files - WP09 (WP09 partners only)	(changed Wed 2016-08-17)
 Data Files - WP10	(changed Thu 2015-09-10)
 Data Files - WP11	(changed Fri 2016-01-08)
 Data Files - WP14-15-16	(changed Wed 2016-03-09)
 Data Files - WP17	(changed Tue 2015-10-20)
 Demonstration Area	(changed Wed 2016-03-09)
 For All Users	(changed Tue 2016-04-19)
 Public Documents	(changed Fri 2016-01-29)
 Test Area	(changed Wed 2015-09-16)
 WP12 Partner Content	(changed Wed 2016-08-17)



© 2016 VTT - HEALS WP12 www.heals-eu.eu  HEALS

Figure 3: The same page when logged in as a HEALS partner. This particular demonstration user does not have special WP09 rights and the “WP09 partners only” folder is in meta-only mode: visible, but files in it are not accessible.


 HEALS FP7-ENV-2013-603946	D12.2 – The HEALS GeoDatabase Platform		
	WP12: Exposure and Health Data Management		Security: Public
	Author(s): Luc Cluitmans, Eija Parmes		Version: 10/28

2.4 The file storage system

As originally implemented, the file storage system provided a way to upload files to one big pool of files, where the user could subsequently add tags as “metadata” that could be used to “query” files. In practice this didn’t prove to be a useful way of organizing data – it is hard to query a set of files if you cannot see what files there are to query, and just querying for “everything” gives too many results to be useful. Additionally, to make this work, users uploading data files would need to carefully tag their data files, which is not a realistic expectation.

The system was replaced with a more conventional “folder” like organization structure. While at first look the system looks like a normal folder based file system, there are some additional features.

- As described in the previous section, an access grant system determines how each user can access each folder, and even determines if a folder shows up at all (compare Figure 2 and Figure 3).
- The access control system interacts with the file storage folder system in a slightly different way than you may think at first: The access control restrictions are only set on the first level folders (the folders that are listed in the “root” folder), and apply recursively to all files and subfolders within those. This has a few effects:
 - Administration is relatively simple: grants only need to be managed for those top level folders
 - There cannot be any files in the root folder, since access rights to those would be ill-defined.
 - The root folder itself is special-cased and has hard-coded access restrictions (readable to anyone, including “not logged in”, writable to administrators only, its restrictions do not apply recursively to its children)
 - It is not possible to have a more restricted “subfolder” within a folder. It is for this reason that you can see two separate folders for WP09 in Figure 3: one that is readable for all HEALS partners and one that is only readable (and writable) to HEALS partners involved in WP09.
 - The previous point is actually related to *why* this implementation was chosen. When desiring to have special permissions for a sub-folder it only makes sense to have those permissions be more restrictive than the parent folder (otherwise you end up with a situation where you may “see” a child folder, but not its parent, making the child folder undiscoverable). However, as mentioned in the previous paragraph, the access grant system can only *extend* grants, not *reduce* them. Which means that to have a more restrictive access to a child folder you would need to start from zero and not be based on the access grants existing for the parent; this “do not inherit” would lead to an unintuitive management experience (prone to make errors)
- The system supports a “history” system: when you attempt to overwrite a file with a new upload, you are creating a new “version” of that file, merely hiding the previous version. Going to a file’s page shows information about that file (size, type, upload date), and also the list of versions. See Figure 4 for an example.
- Files and folders can have a description attached to help understand the purpose or content of the file or folder. The description is a text in “Markdown” format: text with a set of options to tag special formatting; for instance, enclose a phrase in single asterisks


 HEALS FP7-ENV-2013-603946	D12.2 – The HEALS GeoDatabase Platform		
	WP12: Exposure and Health Data Management		Security: Public
	Author(s): Luc Cluitmans, Eija Parmes		Version: 11/28

(*example*) to make it appear italic in the description, or double asterisks (**example**) to make it appear bold. See for instance <http://commonmark.org/help/tutorial/> for an introduction to the formatting options. See Figure 5 for an example.

- There are a set of commands available within each folder or file description page; which commands are available depend on your access rights to the file or folder
- If the file or folder is writable, “Cut and Paste” style buttons are available to move a document to a different folder. To preserve the access rights restrictions a cut document/folder can only be pasted within the same top level folder. The “cut” file/folder is actually still present until you paste it elsewhere; if you forget to paste it anywhere (or cut another one) it just stays where it was.


HEALS Data Platform
Documents & Data Files
Exan

Documents
Demonstration Area
Example
Versioning Samples



VersionDemo.xlsx


Current version (20160309-1155):



VersionDemo.xlsx

content type	application/vnd.openxmlformats-officedocument.spreadsheetml.sheet
size	9714 bytes

Upload new version...


...

Rename...

Edit description...

Delete...

Cut this document

All versions :






Revision	Document	Author	File fingerprint
 20160309-1155	 VersionDemo.xlsx	Luc Cluitmans (VTT)	d6wdbj323ctn3yddtgb4y88f86
 20160309-1152	 VersionDemo.xlsx	Luc Cluitmans (VTT)	d1cd96fqu6y45ypp0lmh8lrnqw

Figure 4: Example file information page, including the list of versions and the actions that can be applied to it

 HEALS FP7-ENV-2013-603946	D12.2 – The HEALS GeoDatabase Platform		
	WP12: Exposure and Health Data Management		Security: Public
	Author(s): Luc Cluitmans, Eija Parmes		Version: 12/28

HEALS Data Platform
Documents & Data Files
Demo User
Log out

Documents
Data Files - WP09 (WP09 partners only)
NRC Portal Processed

Fatsecret.xlsx

Current version (20160927-1138):

Download: [Fatsecret.xlsx](#)

View: [Fatsecret.xlsx](#)

content type	application/vnd.openxmlformats-officedocument.spreadsheetml.sheet
size	8877448 bytes

Description:

This Excel Sheet contains FatSecret data from the NRC portal (based on the data dump extracted on 2016-08-02).

Description of the pre-processing performed, and other notes:

- There are two tabs, splitting the entries for the baby as a separate tab
- The columns are a re-ordered subset of the NRC portal columns; Rows have been reordered to try to get the data for each entry together
- In a few cases there seem to be intentional "duplicate" entries, probably because the user recorded two separate but equivalent food intakes, e.g. two glasses of milk; In these cases the entries may have been mixed up in the row order
- Columns that did not provide information for this dataset were removed
- Note that (apparently) different software versions recorded data in different ways: in some versions there are entries in the "FoodSerie product name" and "FoodSerie product quantity" columns, while the "Extra Info" column is blank or has the fixed text *updated*, and in other versions the FoodSerie columns are blank, but the "Extra Info" column has an entry record ID.

Upload new version... > ...

All versions :


Revision	Document	Author	File fingerprint
20160927-1138	Fatsecret.xlsx	Luc Cluitmans (VTT)	kjsk5rvk04p0tyxjvnhx0xqhqv

Figure 5: Example of a data document annotated with a custom description

2.5 Database access

As explained in deliverable D12.1, all access to the databases of the platform will be indirectly, via the web site or via web services. As such, the databases are not directly visible to end users, but only indirectly. In particular, end users will not be able to submit their own SQL scripts to the databases, but rather use query templates and provide parameters that the server will use to generate the underlying SQL queries (or generate the underlying NoSQL “query”). This mechanism allows tighter control over who can access what data and avoids risks related to “SQL Injection” attacks on the server.

At the moment of writing, this part of the server is being implemented. The intention is that the platform will have a “dataset index” where each entry will provide a (“wiki” style) description of each dataset providing both information about the dataset, possibly external links, and providing link(s) to

 HEALS FP7-ENV-2013-603946	D12.2 – The HEALS GeoDatabase Platform		
	WP12: Exposure and Health Data Management		Security: Public
	Author(s): Luc Cluitmans, Eija Parmes		Version: 13/28

the query UI to access these datasets. The query UI will allow specifying query parameters, perform “preview” queries and to download query results to the user’s local machine.

An example of how a dataset page looks can be found in the screenshot below, in Figure 6. (Data download and parameterized querying is not visible there yet).

HEALS Data Platform
Documents & Data Files

Table 'nrcnetatmo'

General:

Name: nrcnetatmo
Kind: Table
Rows: 5898938
Comment: Reshaped Netatmo data from NRC portal


Columns:

Name	Nullable	Data Type	Column Type	Comment
Subject	NO	varchar	varchar(32)	Original Subject Name
Stamp	NO	char	char(25)	Time stamp
Hsid	NO	int	int(11)	HSID (standardized subject ID)
ModuleChildBedRoom	YES	varchar	varchar(32)	ModuleId in child bed room
Tcr	YES	double	double	Child bed room temperature
Hcr	YES	double	double	Child bed room Humidity
Press	YES	double	double	Pressure
CO2	YES	double	double	CO2
Noise	YES	double	double	Noise
ModuleLivingRoom	YES	varchar	varchar(32)	ModuleId in living roome
Tlr	YES	double	double	Temperature in living room
Hlr	YES	double	double	Humidity in living room

Samples:

- First 10 rows
- First 100 rows
- First 1000 rows

Figure 6: Example dataset info page.

 HEALS FP7-ENV-2013-603946	D12.2 – The HEALS GeoDatabase Platform		
	WP12: Exposure and Health Data Management		Security: Public
	Author(s): Luc Cluitmans, Eija Parmes		Version: 14/28

2.6 MySQL spatial extension

The HEALS geodatabase can utilize MySQL “spatial extensions” to support location based queries. Enabling this functionality requires some extra steps while creating the tables involved and importing data in the spatially indexed columns.

2.7 Environmental grid and raster data

Part of the environmental data resides in EDMS (Work Package 8). This is mostly grid and raster format data on air quality and land cover available from global and European portals.

The intention is to implement a data import module in the HEALS portal that can communicate besides EDMS with external sites using WCS/WMS as provided by GeoServer (the external site has to run GeoServer, see <http://geoserver.org/>). Geoserver is utilized for the database of raster imagery and access by WCS or WMS.

2.7.1 Installation pof Geoserver for raster data

1. Check/Install Java JRE xxx
2. Install Geoserver xxx
3. Add image.tiff to <Your geoserver installation dir>\data_dir\data
4. Start Geoserver with the startup.bat in the bin directory
- (5. Shutdown from bin/shutdown.dat)


2.7.2 Export of raster data by WMS protocol

Raster data will be transferred by Web Map Server (WMS) standard. WMS manages raster data type objects represented as images (digital maps). Operations are executed by JavaScript submitting an URL layer:


- WMS_GetCapabilities.url (info on existing maps)
- WMS_GetMap.url,
 - <http://localhost:8080/geoserver/wms?bbox=60,20,65,25&styles=population&Format=image/tiff&request=GetMap&layers=image&width=1000&height=1000&srs=EPSG:4326>
- WMS_GetMap_multilayer.url
 - <http://localhost:8080/geoserver/wms?bbox=60,20,65,25&styles=population&Format=image/tiff&request=GetMap&layers=bluemarble:red,bluemarble:green,bluemarble:blue&width=1000&height=1000&srs=EPSG:4326>

(here “localhost:8080” is the placeholder for the GeoServer server that exposes its data to the WP12 GeoDatabase)

The metadata to be transferred is:

 HEALS FP7-ENV-2013-603946	D12.2 – The HEALS GeoDatabase Platform		
	WP12: Exposure and Health Data Management		Security: Public
	Author(s): Luc Cluitmans, Eija Parmes		Version: 15/28

	Layers	Char		Name of the image
	Width= NumberOfGridColumns	Integer		10, 10 000
	Height= NumberOfGridLines	Integer		10, 10 000
	Format	Char		GeoTIFF/GeoTIFF32
	bbox: SWLatitude	Float	Degrees	Latitude of the raster area South-west corner
	bbox: SWLongitude	Float	Degrees	Longitude of the raster area South-west corner
	bbox: NELatitude	Float	Degrees	Latitude of the raster area North-East corner
	bbox: NELongitude	Float	Degrees	Longitude of the raster area North-East corner
	EPSG: Coordinate system	Integer		4326 (WGS84 geographical latitude and longitude)

 HEALS	D12.2 – The HEALS GeoDatabase Platform		
	WP12: Exposure and Health Data Management		Security: Public
	Author(s): Luc Cluitmans, Eija Parmes	Version:	16/28

3 Available variables

Table 1 shows the HEALS data variables. The coloring shows the status of the variable regarding the harmonization and import to the HEALS file or geodatabase.

Table 1. Variables in HEALS geodatabase

Grey = in original format


Green = CSV-file prepared for import to GDB (Stamp, Location, Variable, Value, Unit, Source)

Brown= imported to file database


Red= imported to geodatabase

Black= query ready


Source/ Owner	Variable Class	Conf Level *	Variable	Spatial Coverage	Spatial Resolution	Temporal Coverage	Temporal resolution
USTUTT	OutdoorAirHeavyMetal		As	EU-29 countries	Country	2010	Year
USTUTT	OutdoorAirHeavyMetal		Cd	EU-29 countries	Country	2010	Year
USTUTT	OutdoorAirHeavyMetal		Dioxins	EU-29 countries	Country	2010	Year
USTUTT	OutdoorAirHeavyMetal		Ni	EU-29 countries	Country	2010	Year
USTUTT	OutdoorAirHeavyMetal		Pb	EU-29 countries	Country	2010	Year
USTUTT	OutdoorAirHeavyMetal		PCB	EU-29 countries	Country	2010	Year
USTUTT	OutdoorAirHeavyMetal		Hg	EU-29 countries	Country	2010	Year
EMEP	Pollutants		NO2		Degree	1980 1985 1990 1995-2014	Year
EMEP	Pollutants		PM25		Degree	1980 1985 1990 1995-2014	Year
EMEP	Pollutants		PM10		Degree	1980 1985 1990 1995-2014	Year
IER	OutdoorAirHeavyMetal		As		1x1km2	1990 2000	Year
IER	OutdoorAirHeavyMetal		Cd		1x1km2	1990 2000	Year

 HEALS FP7-ENV-2013-603946	D12.2 – The HEALS GeoDatabase Platform			
	WP12: Exposure and Health Data Management		Security: Public	
	Author(s): Luc Cluitmans, Eija Parmes		Version:	17/28


Source/ Owner	Variable Class	Conf Level *	Variable	Spatial Coverage	Spatial Resolution	Temporal Coverage	Temporal resolution
IER	OutdoorAirHeavyMetal		CO		1x1km2	1990 2000	Year
IER	OutdoorAirHeavyMetal		CR		1x1km2	1990 2000	Year
IER	OutdoorAirHeavyMetal		Cu		1x1km2	1990 2000	Year
IER	Pollutants		DioxinsAndFurans		1x1km2	1990 2000	Year
IER	OutdoorAirHeavyMetal		Hg		1x1km2	1990 2000	Year
IER	Pollutants		NMVOX		1x1km2	1990 2000	Year
IER	Pollutants		NOx		1x1km2	1990 2000	Year
IER	OutdoorAirHeavyMetal		Pb		1x1km2	1990 2000	Year
IER	Pollutants		PM10		1x1km2	1990 2000	Year
IER	Pollutants		PM25		1x1km2	1990 2000	Year
IER	Pollutants		SOx		1x1km2	1990 2000	Year
IER	Pollutants		TotalPAH		1x1km2	1990 2000	Year
UNFCCC	EnergyRelatedSubmission		CO2,CH4,N2O,CO,NO2, xx		Country	1990-2012	
Eionet	OutdoorAirPollutants		O3,NO,SO2,PM2.5,PM10		10x10 km2	2004-2013	
NOAA-SoDa	EMF		UV			1979-2016	
CIESIN, CIAT	GriddedPopulation				1x1 km2	1990,1995, 2000	
Eurostat EU-SILC	SES variables list						
Eurostat EU-SILC	Income,LivingConditions	C		31 European countries		2004-2013	
MTUS	Time activity data/variable list?						
ISSDA	Time use data	C	DiaryData	Ireland			
?	Time use data	C	DiaryData	Greece			
EXHES (WP17)	Electronic Checklist	R	Building, dwelling and room parameters	10 countries (Croatia, France, Germany,	NA	Baseline and follow-up of the child, according to	

 HEALS FP7-ENV-2013-603946	D12.2 – The HEALS GeoDatabase Platform		
	WP12: Exposure and Health Data Management		Security: Public
	Author(s): Luc Cluitmans, Eija Parmes		Version: 18/28


Source/ Owner	Variable Class	Conf Level *	Variable	Spatial Coverage	Spatial Resolution	Temporal Coverage	Temporal resolution
				Greece, Italy, Poland, Portugal, Slovenia, Spain and UK).		DOW (specific dates would be defined)	
HEALS Sensor pilot	SES Questionnaires	R	SES		Person		
HEALS Sensor pilot	User feedback questionnaire	R	User feedback		Person		
HEALS Sensor pilot	Moves	R	Activity,Dist,D uration,Place or GPS,Steps		Person		Min
HEALS Sensor pilot	Fatsecret baby	R	Food diary: Product, Field, Quantity, Value		Person	2015:12:09- 2015:12:14	N/day
HEALS Sensor pilot	Fatsecret	R	Food diary: Product, Field, Quantity, Value		Person	2015:12:09- 2015:12:14	N/day
HEALS Sensor pilot	Netatmo	R	Temperature		MainLivingAr ea	2015-06	Min
HEALS Sensor pilot	Netatmo	R	Humidity		MainLivingAr ea	2015-06	Min
HEALS Sensor pilot	Netatmo	R	Pressure		MainLivingAr ea	2015-06	Min
HEALS Sensor pilot	Netatmo	R	CO2		MainLivingAr ea	2015-06	Min
HEALS Sensor pilot	Netatmo	R	Noise		MainLivingAr ea	2015-06	Min
HEALS Sensor pilot	Netatmo	R	Temperature		ChildBedRoo m	2015-06	Min
HEALS Sensor pilot	Netatmo	R	Humidity		ChildBedRoo m	2015-06	Min

 HEALS FP7-ENV-2013-603946	D12.2 – The HEALS GeoDatabase Platform		
	WP12: Exposure and Health Data Management		Security: Public
	Author(s): Luc Cluitmans, Eija Parmes		Version: 19/28


Source/ Owner	Variable Class	Conf Level *	Variable	Spatial Coverage	Spatial Resolution	Temporal Coverage	Temporal resolution
HEALS Sensor pilot	Netatmo	R	Pressure		ChildBedRoom	2015-06	Min
HEALS Sensor pilot	Netatmo	R	CO2		ChildBedRoom	2015-06	Min
HEALS Sensor pilot	Netatmo	R	Noise		ChildBedRoom	2015-06	Min
HEALS Sensor pilot	Netatmo	R	Pressure		OtherRoom	2015-06	Min
HEALS Sensor pilot	Netatmo	R	CO2		OtherRoom	2015-06	Min
HEALS Sensor pilot	Netatmo	R	Noise		OtherRoom	2015-06	Min
HEALS Sensor pilot	Fitbit	R	NumberOfSteps		Person		Min
HEALS Sensor pilot	Passive dust samples	R					
HEALS Sensor pilot	Vacuum dust samples	R	Dust (PM2.5?) from vacuum		Apartment		
HEALS Sensor pilot	Dylos	R	PC0.5,PC2.5		MainLivingArea		Min
HEALS Sensor pilot	Widenoise	R	NoiseLevel				
HEALS Sensor pilot	Activity log	R					
HEALS Sensor pilot	Noise	R	Spectrum				
HEALS Sensor pilot	barcode	R	ScannedBarcodes				
HEALS sensor pilot	HarvardImpactorPM2.5	R					
HEALS	ConsumerProductsL	R					

 HEALS FP7-ENV-2013-603946	D12.2 – The HEALS GeoDatabase Platform		
	WP12: Exposure and Health Data Management		Security: Public
	Author(s): Luc Cluitmans, Eija Parmes		Version: 20/28


Source/ Owner	Variable Class	Conf Level *	Variable	Spatial Coverage	Spatial Resolution	Temporal Coverage	Temporal resolution
Sensor pilot	og						
HEALS sensor pilot	AQ monitor		CO, CO2, NO, NO2, TVOCs, T, RH, P				
HEALS sensor pilot	Grimm						
HEALS sensor pilot	Aerocet						
HEALS sensor pilot	Aeroqual NO2						
HEALS sensor pilot	Aeroqual O3						
HEALS sensor pilot	Radiello passive samplers (BTX, Aldehydes)						
National Registries	Integrated TwinData (no access yet)	Confid ential		10-15 European countries			
	Singleton data						
ACAG	OutdoorAirPollutan ts		O3,PM2.5,PM 10	Global?	1x1 km2	2005-2014?	Year
Meteo?	OutdoorAirParamet ers			FrenchMetS tations	Point	2000-2016?	
Meteo? Chimera	OutdoorAirParamet ers		Estimated	France	1x1 km2	2009-2013?	Day
Cohort data/Fran ce	EDEN	Confid ential	asthma, ...				
EFCA	Food consumption concise	Public	Dietary surveys adults	19 countries	Country		Yearly?
EFCA	Food consumption comprehensive	Public	Dietary surveys infants < 18m	2 countries	Country		Yearly?
EFCA	Food consumption comprehensive	Public	Dietary surveys children < 10y	14 countries	Country		Yearly?
EFCA	Food consumption comprehensive	Public	Dietary surveys adults	20 countries	Country		Yearly?

 HEALS FP7-ENV-2013-603946	D12.2 – The HEALS GeoDatabase Platform		
	WP12: Exposure and Health Data Management		Security: Public
	Author(s): Luc Cluitmans, Eija Parmes		Version: 21/28

Source/ Owner	Variable Class	Conf Level *	Variable	Spatial Coverage	Spatial Resolution	Temporal Coverage	Temporal resolution
			18-64y				
EFCA/me mber states	Food contamination	Restri cted (HEAL S)	Cd,As,Pb,dioxi n,nondioxin, perfluor	19 countries	City,country?	2011-2014	
?	Air quality		Phtalates in dust		Country		
?	Biomonitoring		Phtalates in human		Country		
EEA IpChem	Environmental monitoring						
EEA IpChem	Food and feed						
EEA IpChem	Consumer products						
EEA IpChem	Human Biomonitoring						
EEA IpChem	Metadata template	P					
OMICS	Multigenomics data						
	Air quality		Aerosol optical depth/density ?		30x30 m2		
HEALS Sensor pilot	Questionnaires	C			Person		
	Metadata UI						
	Metadata template						
European met stations			Particles mostly				
			Land cover				
			Toxic metals				
EEA	Green and blue		GreenArea,BI	Europe	250x250 m2		

 HEALS FP7-ENV-2013-603946	D12.2 – The HEALS GeoDatabase Platform		
	WP12: Exposure and Health Data Management		Security: Public
	Author(s): Luc Cluitmans, Eija Parmes		Version: 22/28

Source/ Owner	Variable Class	Conf Level *	Variable	Spatial Coverage	Spatial Resolution	Temporal Coverage	Temporal resolution
Corine	space		ueArea				
EEA Urban Atlas	Green and blue space		GreenArea,Bl ueArea	21 HEALS cities	50x50 m2	2006, 2012	
EEA Urban Atlas	Green and blue space %		GreenArea,Bl ueArea %	21 HEALS cities	City	2006, 2012	

 HEALS FP7-ENV-2013-603946	D12.2 – The HEALS GeoDatabase Platform		
	WP12: Exposure and Health Data Management		Security: Public
	Author(s): Luc Cluitmans, Eija Parmes		Version: 23/28

4 Harmonization of the datasets

4.1 Georeference and spatial coverage

The coordinate system of HEALS GDB is WGS84 based geodetic latitude and longitude coordinates. The name of the fields are Latitude and Longitude. The coordinates standard is, depending on the resolution, DD, DD:MM or DD:MM:SS.

The data is originally georeferenced in different ways: by point location, by GPS track, by grid or pixel, by city, by country, by household, by ChildrenRoom, or by MainLivingArea. Individual data is assigned to point, household or track. To enable the integrated analysis between different datasets, the following geospatial integration is prepared to the variables as shown in Table 2.

Table 2. Geocoding of the datasets

	Point	Track	City	Country	Pixel/Grid	Room	Household
Point	X		X	X	X		
Track		X					
City				X			
Country					X		
Pixel/Grid			X	X	X		
Room			X	X	X	X	
Household			X	X	X		X

4.2 Time stamp, temporal reference and coverage

The timestamp standard is the Universal TU with YYYY:MM:DD:HH:MM:SS. The time zone follows the convention *+hh:mm* or *-hh:mm*.

The time reference varies in original datasets from seconds to years. The interval of the measurements, or the temporal resolution, varies from minutes (for instance Move) to several years (for instance the Green spaces layer) or ten years (for instance emission data from EMEP). The temporal coverage is from days to tens of years. To allow integrated analysis of the datasets, the following integration of the time reference is prepared.


 HEALS FP7-ENV-2013-603946	D12.2 – The HEALS GeoDatabase Platform		
	WP12: Exposure and Health Data Management		Security: Public
	Author(s): Luc Cluitmans, Eija Parmes		Version: 24/28

Table 3. Temporal reference of datasets

	Minute	Hour	Year	10 years
Minute	X	X	X	X
Hour		X	X	X
Year			X	X
10 years				X

4.3 Country names and codes

The names of the countries are written followingly:

EU-29: Austria, Belgium, Bulgaria, Cyprus, CzechRepublic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Ireland, Italy, Latvia, Lithuania, Luxembourg, Malta, Netherlands, Norway, Poland, Portugal, Romania, SlovakRepublic, Slovenia, Spain, Sweden, Switzerland, UnitedKingdom

4.4 Grid systems

1 degree grid

1 x 1 km² grid


4.5 Variable names

Names of the variables from the original dataset have been harmonized as follows. The main reason has been not to create difficulties in interpreting columns and fields by applications.

PM2.5 -> PM25


Dioxins and furans -> DioxinsAndFurans

4.6 NULL value

 HEALS FP7-ENV-2013-603946	D12.2 – The HEALS GeoDatabase Platform		
	WP12: Exposure and Health Data Management		Security: Public
	Author(s): Luc Cluitmans, Eija Parmes		Version: 25/28

4.7 Link to external databases


Conformality and future transport to ipChem database is foreseen by applying the IpChem metadata template.

 HEALS FP7-ENV-2013-603946	D12.2 – The HEALS GeoDatabase Platform		
	WP12: Exposure and Health Data Management		Security: Public
	Author(s): Luc Cluitmans, Eija Parmes		Version: 26/28

5 Access to HEALS geodatabase

As implied in chapter 2, to make full use of the platform users need to log in, using a personal account. To obtain an account send an email to luc.cluitmans@vtt.fi including information about what HEALS partner you represent and what work packages you are involved in.

While towards the end of the HEALS project parts of the data (that are not privacy sensitive) will be made available publicly, currently most data is only available after logging in.

 HEALS FP7-ENV-2013-603946	D12.2 – The HEALS GeoDatabase Platform		
	WP12: Exposure and Health Data Management		Security: Public
	Author(s): Luc Cluitmans, Eija Parmes		Version: 27/28

6 Issues

Most of the data was available to VTT only in summer 2016 and the import to the database is still ongoing in time of delivery 30th September, 2016. The status of the variables available will be updated and shown in in the UI.

The confidentiality of the datasets was and continues to be one of major challenges for the delays. The database contains methodology to restricted usage, but for instance with individual and twins data, food diaries, time use and SILC data there is still administrative work to do to get the data in the HEALS database. With twins data, most probably only statistics and integrated data can be imported to the HEALS geo database.

Also the terms of usage prevent in some cases the full exploitation of the datasets. This is the situation also with EEA and EU datasets, which should be open to distribute and use in databases at least in EU projects (CORINE data, Urban Atlas data).

The functionalities correspond to the requirements and Correspond to the requirements in D12.1 Report on prototype Design of Geospatial platform (March 2015) except in the following points: 7.1: basic data analysis and 8.1: visualization of the data on the web. The analysis and visualization will be kept in minimum because they are made outside the database with applications oriented software. However, it is foreseen to add simple analysis and visualization before the database for its public parts is exposed to public after the project.



*"This project has received funding
from the European Union's Seventh Programme for
research, technological development and demonstration
under grant agreement N°603946*

