Programme Theme: Preparatory Action on development of prevention activities to halt desertification in Europe

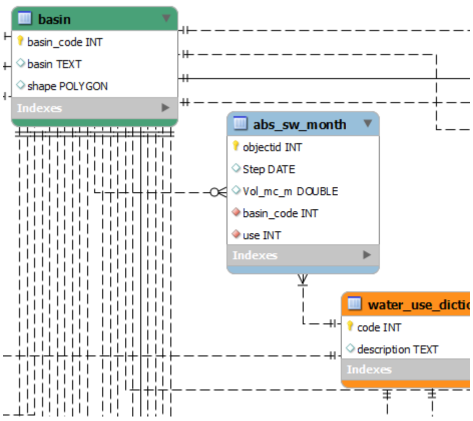
Budget line: 07 03 29

Reference: Desertification 2012–2013 – Call for Proposals "Halting Desertification in Europe"

Grant Agreement No. 07.0329/2013/671279/SUB/ENV.C.1

Duration: January 1, 2014–March 31, 2015

Pilot Arno Water Accounts



Deliverable: D3.2 Geo-referenced database for water accounts

Authors: Bernardo MAZZANTI, Francesco CONSUMI, Giovanni MONTINI and Lucia FIUMI – ARBA
Eric MINO and Carolina CARDETE – EMWIS
Tania LUTI – ISPRA

Due date to deliverable: 31 December 2014

Dissemination Level: Public

Acknowledgments

Project funded by the European Commission, Directorate-General Environment, Directorate D – Water, Marine Environment & Chemicals, ENV.D.1 – Protection of Water Resources European Commission within the Desertification 2012–2013 Call

Table of Contents

List of Figures

List of Tables

1 Introduction

The PAWA project is composed of four successive technical activities and two horizontal activities (**Figure 1**).

During the implementation of Activity 1 "Setting the scene at RBD level", the project partners identified three pilot basins where the calculation of water accounts on a monthly basis would have the best potential by reason of the basin characteristics and main water uses (see **D1.2 Prioritization list of sub-basins**). These sub-basins are:

- Chiana
- Bisenzio
- Pisa

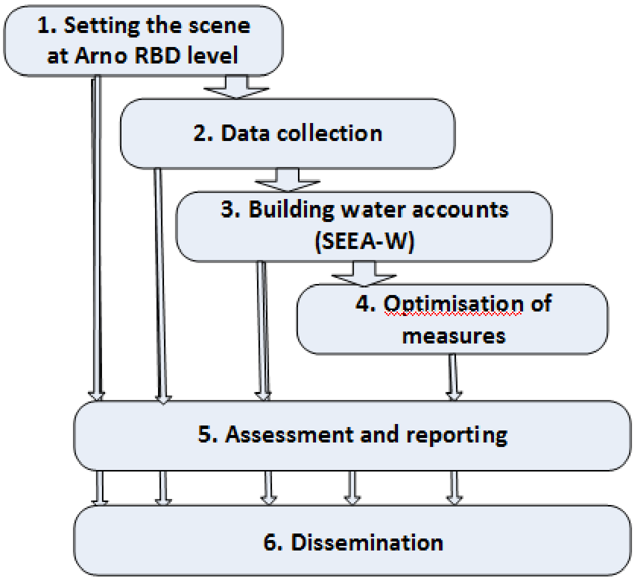


Figure 1 – PAWA project activity chart.

Deliverable D2.2 illustrates the sub-basin water flows used to prioritize data collection. These data flows have been validated with the local stakeholders during the **2nd Stakeholder Workshop**, which took place at the Arno River Basin Authority premises in July 2014.

In **Deliverable D3.1** the "Physical Use and Supply Accounts and Water Asset Accounts" tables for the three pilot territories – Chiana, Bisenzio and Pisa – are presented and discussed.

The present deliverable describes the geo-database which has been built up and contains all data and information necessary to implement the Activity n. 3 "Building Water Accounts", that is, for the production of the SEEA-Water tables. This deliverable describes the geo-database construction process that consisted of three different steps: designing, developing and testing. The attention is focused here on the evaluation of advantages and needs, as well as on the structural and technical choices. The description of the geo-database components and the presentation of the metadata requirements complete the document.

2 Project needs

As illustrated in **Deliverables D2.1** and **D2.2**, the gathered datasets are characterized by a high level of heterogeneity. This degree of complexity is mainly due to the different data sources, since each local data provider has a self-developed, usually non-standard way to store and to elaborate its own datasets. Another aspect that has to be taken into account regards the big differences in dataset sizes. The compilation of the SEEA-Water tables requires the processing of very short time series (a few years in the considered time interval) or long ones, with a very large volume of data (e.g., daily rainfall data).

Hence, the preparatory procedures for the production and testing of SEEA-Water tables have required a significant effort. The following activities have been carried out: data pre-processing; data quality assessment; data coherence appraisal; data layers consolidation; and homogenisation.

All information used is geo-referenced, at different spatial scales. A specific pre-processing phase has required the transformation of the spatial reference system by migrating all layers and each geographical information to a common reference system, the so-called "Monte Mario Italy 1" (also classified in the EPSG Geodetic Parameter Dataset as 3003).

In terms of data typology, the collected datasets are in very different formats: vector format; raster format; tables; and geographical layers. The original digital formats range from pure ASCII table to DBF tables, proprietary worksheet formats, and standard spatial layer (e.g., shape files).

Obviously the compilation of the SEEA-Water tables needs a synthetic view of data availability and consistency. For these needs, the limits of a file system organization in terms of data accessibility, maintenance and update are clear and cannot be underestimated.

Therefore, a relational geodatabase (hereinafter GeoDB) has been designed and implemented, providing a useful tool for any further data integration, with a reliable log procedure. Both data pre-processing and tables preparation can be implemented as coded procedures, storing intermediate datasets in the same database.

3 Choice of database structure

When planning the database structure, the following issues have been taken into consideration:

- Clustering of datasets according their typology
- Predefined relational structure
- Limit to the use of textual fields
- Preference to numerical fields together with related dictionary tables
- Use of pivot tables for 1-to-many relationships

In agreement with the above-mentioned instances, the relational GeoDB has been organized taking into account the following main data categories:

[1] HydroData: hydrological datasets, related to hydrometeorological measurements.

[2] GeoData: geographical (polygons and lines) datasets, like soil use, digital terrain model, slope, river basin. Position of every monitoring station. Position of abstraction/restitution points. It should include all spatial layers, in any representative forms (vector or raster).

[3] Abs_Rest: water abstraction/restitution data.

[4] BalanceData: data derived from the water balance evaluation, according to the water balance carried out by ARBA (first release 2008, updated in 2014).

[5] SocioEconomic: socio-economic data, mainly extracted from the census catalogues provide by the Italian National Institute of Statistics (ISTAT).

[6] SEEA-Water: data directly related to the SEAA-Water tables. The three main tables are

- Assets;
- Supply;
- Use.

The tables have an identical structure, where the variable name is defined in the first field, and the temporal identification is determined by the pair year/month. The SEEA-Water tables are completed by a general "variable" table, describing all used variables.

The HydroData, Abs_rest, BalanceData, and SocioEconomic categories are linked to river basins through the "basin _code" item (INTEGER). As mentioned, a wide use of dictionary tables allows containing textual items; therefore a list of relationship for the linkage between dictionary terms and related codes is applied as reported in **Table 1**.

Table 1 – Relationships between dictionary terms and related codes.

Table A	Item Table A	Relationship	Item Table B	Table B
Soil_use_dictionary	Code	↔	Code_2007 Code_2010	Soil_use
Water_use_dictionary	Code	↔	Use	Abs_sw_month Abs_gw_month Abstraction_point_gw Abstraction_point_sw
Wb_status_dictionary	Code	↔	Ecostatus Chemstatus Quantstatus	River_water_bodies Transitional_water_bodies Lake_water_bodies Groundwater_bodies
Guage_type_dictionary	Code		Gauge_type	Gauge_location

The logical structure of the different data categories is represented in **Figure 2**.

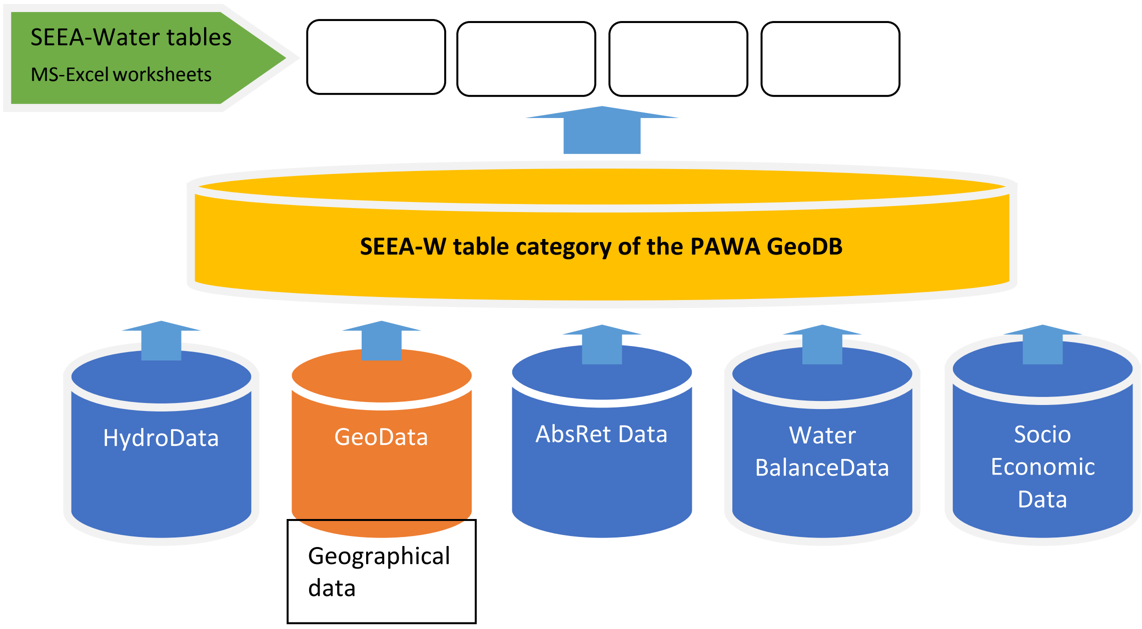


Figure 2 – Hierarchy of datasets in the geodatabase.

4 Choice of database format

The choice of the database format has been determined by the following needs:

- Need to apply easy export procedures for reusing data with different external tools
- Independence from the computer platform and the operating system
- Guarantee of long term support
- Easy data access from desktop application and web applications.

On the basis of the above-listed needs, an open source object-relational database based on PostgreSQL with PostGIS extension has been chosen by the project team.

PostgreSQL is a DB platform with more than 15 years of development and a proven architecture that has earned it a strong reputation for reliability, data integrity, and correctness for all software products that require a capable relational database management system (RDBMS; see **Table 2**). It runs on all major operating systems, including Linux, UNIX, and Windows. It has full support for foreign keys, joins, views, triggers, and stored procedures (in multiple languages). It includes most SQL:2008 data types, including INTEGER, NUMERIC, BOOLEAN, CHAR, VARCHAR, DATE, INTERVAL, and TIMESTAMP. It has native programming interfaces for C/C++, Java, .Net, Perl, Python, PHP, Ruby, Tcl, ODBC, among others. Its SQL implementation strongly conforms to the ANSI-SQL:2008 standard.

PostgreSQL source code is available under the PostgreSQL License, an open source license.

This technical choice has allowed the provision of a list of Open Geospatial Consortium (OGC) open standards (**Figure 3**), such as Web Feature Service (WFS), Web Map Service (WMS), and Web Coverage Service (WCS). The latter are mainly developed by using an open source server like GeoServer (Open Source Geospatial Foundation, 2014), released under the GNU General Public License.

Figure 3 – Logo of OGC standard.

Table 2 – Characteristics of PostgreSQL DB.

Limit	Value
Maximum Database Size	Unlimited
Maximum Table Size	32 TB
Maximum Row Size	1.6 TB
Maximum Field Size	1 GB
Maximum Rows per Table	Unlimited
Maximum Columns per Table	250 – 1600 depending on column types
Maximum Indexes per Table	Unlimited

PostGIS adds support, for geographic objects, to the PostgreSQL object-relational database. In effect, PostGIS "spatially enables" the PostgreSQL server, allowing it to be used as a backend spatial database for geographic information systems (GIS). PostGIS follows the OpenGIS "Simple Features Specification for SQL" and it has been certified as compliant with the "Types and Functions" profile.

5 List of layers and tables

The list of layers and tables used for the project is reported in **Table 3**. It contains information related to *i*) Table name; *ii*) Field name; *iii*) Description; *iv*) Data type; and *v*) Notes.

Table 3 – List of DB tables and related fields.

Table name	Field name	Description	Data type
Assets	idAssets		Integer
Assets	SEEA-w table		Text
Assets	var_name		Text
Assets	id_territory	Linked to "basin" table, "basin_code" item	Integer
Assets	year		Integer
Assets	month		Integer
Assets	id_element		Integer
Assets	value		Double float
Assets	var_type		Text
Assets	var_reference		Text
Assets	var_reliability		Text
Supply	idSupply		Integer

Table name	Field name	Description	Data type
Supply	SEEA-w table		Text
Supply	var_name		Text
Supply	id_territory		Integer
Supply	year		Integer
Supply	month		Integer
Supply	id_element		Integer
Supply	value		Double float
Supply	var_type		Text
Supply	var_reference		Text
Supply	var_reliability		Text
Use	idUse		Integer
Use	SEEA-w table		Text
Use	var_name		Text
Use	id_territory	Linked to “basin” table, “basin_code” item	Integer
Use	year		Integer
Use	month		Integer
Use	id_element		Integer
Use	value		Double float
Use	var_type		Text
Use	var_reference		Text
Use	var_reliability		Text
Variable	var_name		Text
Variable	id_table		Integer
Variable	from		Text
Variable	to		Text
Variable	id_element		Integer
Variable	type		Text
Reference	var_reference		Text
Reference	notes		Text
reference_variable	var_name		Text
reference_variable	var_reference		Text
reliability_literature	reliabilty_coefficient		Text
reliability_literature	notes		Text
reliability_variable	var_name		Text
reliability_variable	id_territory	Linked to “basin” table, “basin_code” item	Integer
reliability_variable	var_reliability		Text
reliability_variable	notes		Text
abs_gw_month	objectid	Sequential number	Integer
abs_gw_month	Step_DATE	Time step	Date
abs_gw_month	Vol_mc_m	Water abstraction from groundwater. Monthly values (for industrial, agricultural, water supply, households uses) in Chiana, Bisenzio and Pisa basins Vol_mc/month	Double float
abs_gw_month	basin_code	Basin code	Integer
abs_gw_month	use	Type of use	Integer
socio_economic_data	objectid	Sequential number	Integer
socio_economic_data	year	Time	Integer
socio_economic_data	population	Number of inhabitants	Integer
socio_economic_data	n_industries	Number of industries	Integer
socio_economic_data	n_employees	Number of employees	Integer
socio_economic_data	cultivated_area_sq_km	Cultivated area expressed in square kilometres	Double float

Table name	**Field name**	**Description**	**Data type**
socio_economic_data	basin_code	Basin code	Integer
sw_bas_03	objectid	Sequential number	Integer
sw_bas_03	Step	Time step	Date
sw_bas_03	Inflow_mc	Inflow volume of Arno river	Double float
sw_bas_03	Arno_Outflow_mc	Outflow volume of Arno river	Double float
sw_bas_03	Scolmatore_Inflow_mc	Inflow water volume related to scolmatore	Double float
sw_bas_03	Scolmatore_Outflow_mc	Inflow water volume related to scolmatore	Double float
sw_bas_03	basin_code	Basin code	Integer
dqu_daily	objectid	Sequential number	Integer
dqu_daily	Date	Time step	Date
dqu_daily	Daily_discharge_mc_s	Daily discharge data in the river stage gauges located in the selected basins. Values in mc/s	Double float
dqu_daily	basin_code	Basin code	Integer
hydrological_data	objectid	Sequential number	Integer
hydrological_data	Step	Time step	Date
hydrological_data	Prec_mm	Precipitation in mm	Double float
hydrological_data	Temp_C	Temperature in °C	Double float
hydrological_data	Evap_mm	Evapotranspiration in mm	Double float
hydrological_data	basin_code	Basin code	Integer
perc_to_sewerage	objectid	Sequential number	Integer
perc_to_sewerage	Step	Time Step	Date
perc_to_sewerage	Vol_mc_m	Portion of precipitation amount directly drained from sewerage in mc/m	Double float
perc_to_sewerage	basin_code	Basin code	Integer
reused_water	objectid	Sequential number	Integer
reused_water	Date	Time Step	Date
reused_water	reused_water_to_ind_mc_y	Reused water for industrial uses (Pisa basin)	Double float
reused_water	basin_code	Basin code	Integer
dqu_month	objectid	Sequential number	Integer
dqu_month	Month	Time	Text
dqu_month	Avg_discharge_mc_s	Monthly average discharge in the river stage gauge (Chiana and Bisenzio basins)	Double float
dqu_month	basin_code	Basin code	Integer
balance_gw	objectid	Sequential number	Integer
balance_gw	Time step	Time Step	Date
balance_gw	From_prec_mc_m	Modeled values for water balance components (Chiana, Bisenzio and Pisa acquifer)	Double float
balance_gw	InUnder_FracPor_mc_m	Modeled values for water balance components (Chiana, Bisenzio and Pisa acquifer)	Double float
balance_gw	InSurf_FromHill_mc_m	Modeled values for water balance components (Chiana, Bisenzio and Pisa acquifer)	Double float
balance_gw	River_feed_gw_mc_m	Modeled values for water balance components (Chiana, Bisenzio and Pisa acquifer)	Double float

Table name	Field name	Description	Data type
balance_gw	GW_feed_river_mc_m	Modeled values for water balance components (Chiana, Bisenzio and Pisa acquifer)	Double float
balance_gw	Inflow_From_upstream_mc_m	Modeled values for water balance components (Chiana, Bisenzio and Pisa acquifer)	Double float
balance_gw	Abstraction_mc_m	Modeled values for water balance components (Chiana, Bisenzio and Pisa acquifer)	Double float
balance_gw	InUnder_FromOtherAq_mc_m	Modeled values for water balance components (Chiana, Bisenzio and Pisa acquifer)	Double float
balance_gw	Delta		Double float
balance_gw	Vol_Mmc	Volume measured in Mmc	Double float
balance_gw	basin_code	Basin code	Integer
water_cost	objectid	Sequential number	Integer
water_cost	water_use	Type of use of water	Integer
water_cost	cost_e_mc	Water cost expressed in Euro/mc	Double float
water_cost	basin_code	Basin code	Integer
ws_leakages	objectid	Sequential number	Integer
ws_leakages	Year	Time	Integer
ws_leakages	leakages_perc	Leakage of water supply system expressed in perc.	Double float
ws_leakages	basin_code	Basin code	Integer
abs_sw_month	objectid	Sequential number	Integer
abs_sw_month	Step	Time	Date
abs_sw_month	Vol_mc_m	Monthly values of water abstraction from surface waterbodies	Double float
abs_sw_month	basin_code	Basin code	Integer
abs_sw_month	use	Type of use of water	Integer
water_supply	objectid	Sequential number	Integer
water_supply	Name	Basin Name	Text
water_supply	basin_code	Basin code	Integer
water_supply	WW_collected_sewer_WU_from_Ind		Double float
water_supply	WW_collected_sewer_WU_from_HH		Double float
water_supply	DWW_discharged_to_ENV_Industry		Double float
water_supply	DWW_discharged_to_ENV_HH		Double float
water_supply	Returns_from_irr_to_GW		Double float
water_supply	Returns_Evap_from_irr		Double float
soil_content	objectid	Sequential number	Integer
soil_content	Step	Time	Date
soil_content	wgrav_mm	Water content by capillarity measured in mm	Double float
soil_content	wcap_mm	Water content by gravityy measured in mm	Double float
soil_content	basin_code	Basin code	Integer
urban_surface	objectid	Sequential number	Integer
urban_surface	Name	Basin Name	Text
urban_surface	Basin_area_kmq	Basin area in Kmq	Double float
urban_surface	Urban_surface_kmq	Urban surface in kmq	Double float
urban_surface	Urban_surface_perc	Urban surface in perc	Double float
urban_surface	basin_code	Basin code	Integer
minimum_vital_flow	objectid	Sequential number	Integer
minimum_vital_flow	eu_cd_rw		Text

Table name	Field name	Description	Data type
minimum_vital_flow	ref_point_Xgb	Section centroids X coordinates of the river	Double float
minimum_vital_flow	ref_point_Ygb	Section centroids Y coordinates of the river	Double float
minimum_vital_flow	name	Name of river	Text
minimum_vital_flow	MVF_ls	minimum vital flow expressed in litre/sec	Double float
minimum_vital_flow	basin_code	Basin code	Integer

6 Relational map of layers and table and general documentation

Taking into consideration the necessary relationships between tables and layers of the implemented GeoDB, it is possible to build a relational map, useful to identify item meaning and function. For an extended view of the graph, the reader can refer to **Annex 1**.

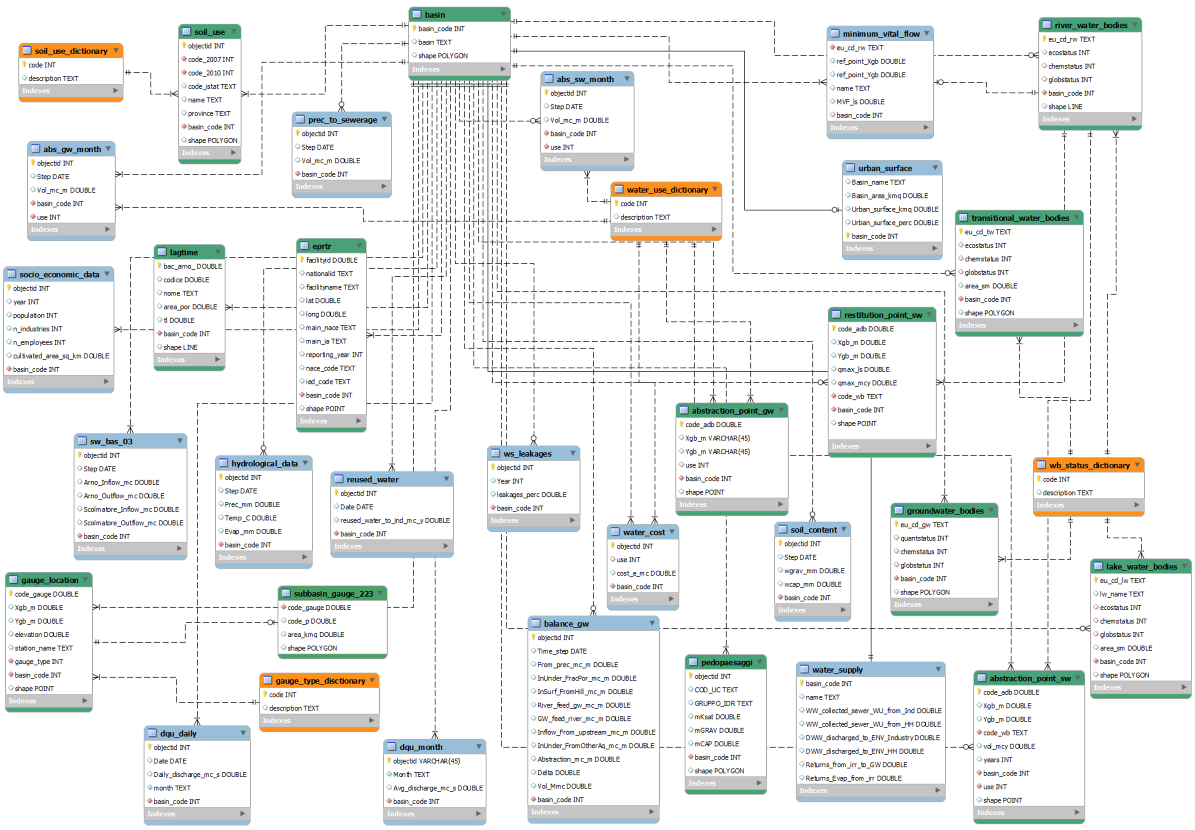


Figure 4 – Relational map of the PAWA GeoDB items.

In the schema reported in **Figure 5**, geographical layers (with attributes) are characterized by a green top label; tables with alphanumeric values by a cyan top label; and dictionary tables by an orange top label.

Deliverable D1.1 describes the metadata catalogue implemented for the PAWA project. The processed GeoDB has been described and documented in the metadata catalogue developed with the GeoNetwork platform (see **Figure 6**). Using the available options, a general metadata item for the GeoDB is available, referred to the hierarchy level "Dataset"; each table or layer in the GeoDB is described with a dedicated metadata item referred to hierarchy level “feature”. The use of “Parent identifier” field allows creating a clear relation between the contents (i.e., layers and tables) and the container (i.e., the GeoDB).

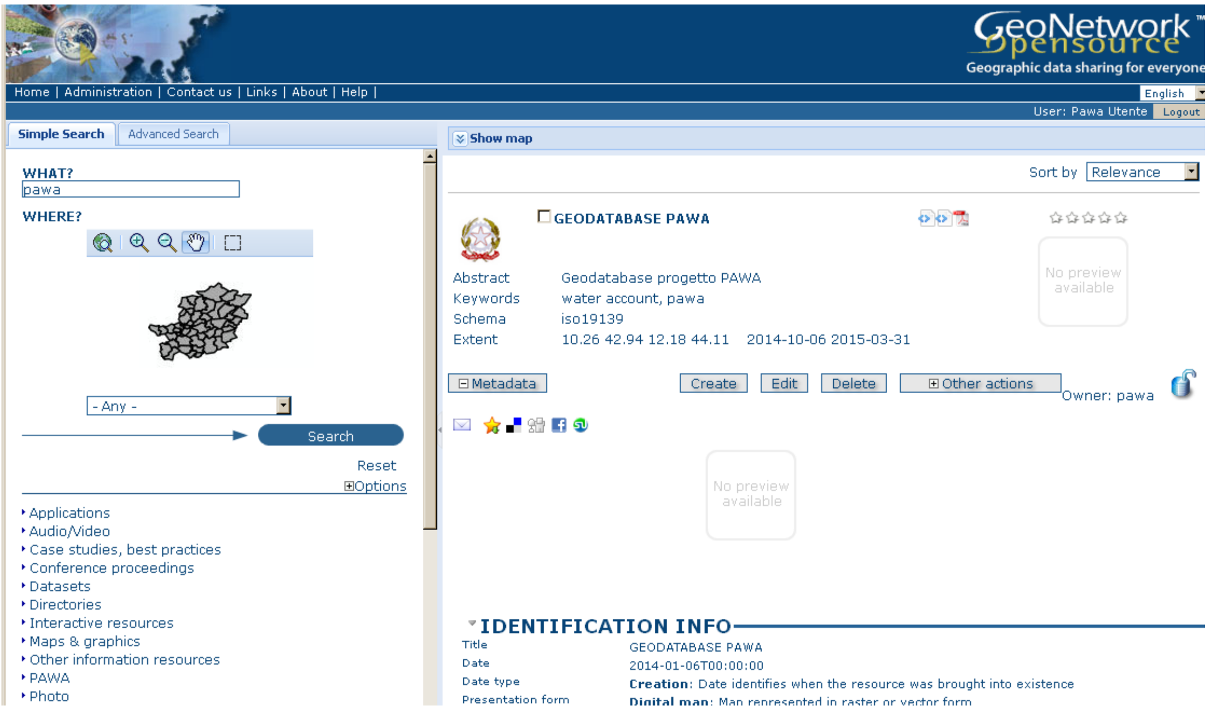


Figure 5 – Extract from metadata item for the PAWA GeoDB.

7 Data elaboration procedures

As mentioned, one of the main reasons to use a geo-referenced database is the possibility to execute a list of queries:

- Check the consistency and coherence of gathered datasets
- Pre-processing gathered datasets for the preparation of intermediate tables (assets, use, supply)
- Procedures to compute the SEEA-Water tables
- Visualization procedures with production of dynamic graphs.

Thanks to the connection with the database and the use of standard queries (SQL), the above listed set of procedures has been coded, by mainly developing PHP scripts, in order to create the intermediate and final products, such as worksheets, documents, graphs.

```
<?php
$dbconn = pg_connect("host='localhost' port=5432 dbname='pawa'
user='pawauser' password='pawa2014'");
?>
```

Example of PHP code for GeoDB connection.

```
SELECT m.eu_cd_rw, m.ref_point_xgb, m.ref_point_ygb, m.name, m.mvf_ls,
m.basin_code, b.basin FROM minimum_vital_flow m, basin b WHERE
m.basin_code=b.basin_code AND m.basin_code=2;

"IT09CI_N002AR083fi3";1668881;4848871;"FIUME BISENZIO";0.709;2;"bisenzio"
"IT09CI_N002AR083fi2";1671309;4856528;"FIUME BISENZIO";0.330;2;"bisenzio"
"IT09CI_N002AR083fi1";1670617;4871655;"FIUME BISENZIO";0.209;2;"bisenzio"
"IT09CI_N002AR579fi2";1671979;4855026;"TORRENTE MARINA";0.150;2;"bisenzio"
"IT09CI_N002AR302ca";1671851;4851510;"COLLETTORE DELLE ACQUE ALTE";0.110;2;"bisenzio"
"IT09CI_N002AR579fi1";1675142;4861316;"TORRENTE MARINA";0.060;2;"bisenzio"
"IT09CI_N002AR580fi";1675150;4861231;"TORRENTE MARINELLA DI LEGRI";0.050;2;"bisenzio"
"IT09CI_N002AR070ca";1671875;4851703;"COLLETTORE ACQUE BASSE";0.0400;2;"bisenzio"
"IT09CI_N002AR488fi";1672385;4878977;"TORRENTE DI FIUMENTA";0.0400;2;"bisenzio"
"IT09CI_N002AR581fi";1671348;4856578;"TORRENTE MARINELLA";0.030;2;"bisenzio"
"IT09CI_N002AR051ca";1676449;4854557;"CANALE DI CINTA
OCCIDENTALE";0.0300;2;"bisenzio"
"IT09CI_N002AR450fi";1670932;4878544;"TORRENTE CARIGIOLA";0.030;2;"bisenzio"
"IT09CI_N002AR537fi";1674858;4858824;"TORRENTE GARILLE";0.020;2;"bisenzio"
```

Example of a SQL query and corresponding results.

References

The PostgreSQL Global Development Group, 2014. PostgreSQL – http://postgresql.org/.

Open Source Geospatial Foundation, 2014. PostGIS web site – http://postigs.refractions.net/.

Open Source Geospatial Foundation, 2014. Geoserver website – http://geoserver.org/.

Annex 1 – Relational map of the PAWA GeoDB items