European Journal of Geography Volume 10, Number 3:6-23, September 2019
© Association of European Geographers

A CONCEPT OF AN INTEGRATED GEODATABASE FOR SURFACE WATER, SOIL AND GROUNDWATER POLLUTION WITH ARSENIC IN THE UPPER PART OF OGOSTA VALLEY, NORTHWESTERN BULGARIA

Emilia TCHERKEZOVA
National Institute of Geophysics, Geodesy and Geography, Bulgarian Academy of Sciences
eti2015@abv.bg / et@geophys.bas.bg

Velimira STOYANOVA
National Institute of Geophysics, Geodesy and Geography, Bulgarian Academy of Sciences
stoyanovavelimira@gmail.com

Tsvetan KOTSEV
National Institute of Geophysics, Geodesy and Geography, Bulgarian Academy of Sciences
tsvetankotsev@mail.bg

Abstract

This paper presents a concept of an integrated geodatabase for surface water, soil and groundwater pollution with arsenic in the upper part of the Ogosta Valley (Northwestern Bulgaria) using ArcGIS™ (ESRI®) file geodatabase. A geodatabase model diagram is created to enable data storage and to meet research requirements ensuring clear geodatabase structure. Further research and activities in this area will focus on the extension and upgrade of the geodatabase to a web format with different levels of data access to support the long-term monitoring and analysis of surface water and groundwater contamination with heavy metals in the investigated area.

Keywords: *Geodatabase, Metadata, Surface water, Soil and Groundwater pollution, Arsenic, Ogosta*

1. INTRODUCTION

Heavy metal and metalloid in surface water, soil and groundwater is a serious risk for humans through contamination of drinking water, crop plants and food (Byrne et al., 2012, Jordanova et al., 2013). According to Tchounwou et al. (2012) and Kim et al. (2015) many trace elements, e.g. arsenic (As), cadmium (Cd), chromium (Cr), and nickel (Ni) are carcinogens and can increase the risk of cancer and cancer-related diseases. Arsenic is a metalloid, which occurs in river channel sediment, river floodplains and adjacent wetlands affected by mining, ore smelting and industrial accidents (Kim et al., 2015; Mandaliev et al., 2013; Mikutta et al., 2014), and because of that can impact negatively the environment (e.g. Byrne et al., 2012).

The contaminated soil-water system in the upper part of the Ogosta River Valley, Northwestern Bulgaria, is an object of long-term investigation mostly in the frame of the research projects ASCOR (2012-2016) and ARSENT (2016-2020), as well as in previous studies, e.g. Kotsev (2003), Dimitrova et al. (2007), Mandaliev et al. (2013), Jordanova et al. (2013), Mikutta et al. (2014), Tcherkezova (2015), Kotsev et al. (2015), Antonov et al. (2015), Benderev et al. (2015), Stoyanova and Kotsev (2016), Simmler et al. (2016), Simmler et al. (2017), Gerginov et al. (2017), Antonov et al. (2018). During the ASCOR project different thematic spatial and non-spatial data and information is gathered, e.g. topographic and hydrographic data from Airborne Laser Scanning (ALS), Digital Elevation Model (DEM), Digital Surface Model (DSM), orthoimages, hydrogeological, hydrological, geophysical, geomorphic, soil, and land use data for the upper part of the Ogosta Valley. This digital data and the obtained results such as hydrological, hydraulic and As-prediction models should be integrated into an appropriate spatial geodatabase together with the field sampling data, e.g. data from piezometric wells, which are collected during the project ARSENT.

In the study area, a long-term monitoring of surface water and groundwater pollution with As is also carried out. This monitoring provides permanent retrieving, processing, analysis, and updating a variety of thematic geospatial data regarding the chemical composition of water, physicochemical indicators, surface water level, and meteorological data, e.g. temperature, precipitation, atmospheric pressure, wind velocity and direction. In this regard, a creation, implementation and maintenance of an integrated spatial database in the near future can ensure and facilitate an efficient spatial data manipulation and management, offering geometrically correct and easy accessible basic and thematic spatial and data and information, cartographic and graphic representations of the geodatabase contents, as well as the geodatabase extent and upgrade to a web format. The benefits of using the spatial database technology for long-term environmental monitoring of soil, surface and groundwater pollution with As in the study area are related to its capability to store, manipulate, query, and manage the basic and thematic spatial data and attribute tables using Geographic Information System (GIS). GIS provides a framework and tools for integration, storage, spatial analysis, spatial data retrieval and combination to create new geographic geometries and information, spatial modelling, visualization, and exchange of all types of spatial data (Bill and Fritsch, 1994; Rigaux et al., 2003). It helps a deeper understanding of a variety of natural, socio-economic and economic processes, market analysis, and supports expert users, e.g. public administrative bodies, decision and policymakers.

Data storage and management play an important role in any GIS (Bill and Fritsch, 1994: 263). Many GIS, e.g. ArcGIS™ (ESRI®) have their own geodatabases managed by DBMS (Database management systems) (Bill and Fritsch, 1994; Rigaux et al., 2003; Watt and Eng, 2014). A DBMS ensures the interrelation between database, stored database metadata and application programs (Rigaux et al., 2003: Figure 1.1: 5).

Such as classical database, a geodatabase describes the representation of an object from the real world additionally on the basis on two components for its modelling: spatial component and its attributes (Samson et al., 2017: 112). According to Rigaux et al. (2003), the spatial component refers to spatial or geometric objects, their attributes, or spatial extent describing location, shape, orientation, size of the object in 2D and 3D space, while the non-spatial attributes are thematic or descriptive attributes. In this regard, similar to database design, the geodatabase (GDB) design involves data modeling at conceptual and logical levels, and physical implementation of the logical schema (Rigaux et al., 2003).

A conceptual model describes the objects (entities) and the relationships among them (Bill and Fritsch, 1994; Zehnder, 1998; Rigaux et al., 2003).

Logical modelling is a process to define the data schema (Zehnder, 1998; Rigaux et al., 2003). According to Bill and Fritsch (1994), a logical data model represents the spatial and conceptual model within a GIS. The logical structure of a geodatabase and the geodatabase itself can be defined using entity-relationship model (ER), hierarchical, network model, relational, object-oriented, object-relational, and hybrid data models (Bill and Fritsch, 1994; Watt and Eng, 2014).

At the physical level, similar to the classical database development described by Graves (2002), the logical schema for a concrete application is then converted into a DBMS (Bill and Fritsch, 1994). Any DBMS manages the data storage, the access to data and resources from several users, evaluates a query, guarantees the security, consistency, as well as the recovery of the database (Rigaux et al., 2003: 8).

The main objective of this work is to create a concept for building up and maintenance of an integrated upgradable geodatabase for surface water, soil and groundwater pollution with As in the upper part of the Ogosta Valley. It is planned for internal use to support the analysis and monitoring of the surface water, soil and groundwater contamination with As in the investigated area. Nevertheless, the building up and implementation of a comprehensive and very well structured geodatabase can ensure data exchange, as well as data export to external users and connection to other applications.

The methodological approach is outlined in Figures 2 and 3 and the following paper's sections. It is based on application of a GIS-based spatial data processing workflow and the geodatabase design methodological framework. A simple spatial database schema is to be developed to accommodate the necessary data while using all necessary geodatabase functionality.

In the first section, the main physiographic characteristics of the study area are presented. In the following section, we introduce a selected collection of spatial and non-spatial data, the methodology, as well as the used software in this work. The highlights are discussed in the final section.

2. STUDY AREA

The study area is located in the valley of the Ogosta River between the settlement Martinovo and the Ogosta Reservoir and is approximately bounded by the coordinates: 43°21′42″N − 43°26′17″N and 22°48′15″E − 23°07′40″E (Figure 1). The upper part of the Ogosta River basin till Ogosta Reservoir covers an area of around 800 km2 in the Western Balkan Mountain and Western Fore- Balkan. In this part, the area is drained by the Ogosta River and some tributaries, such as the rivers Prevalska Ogosta, Chiprovska, Dalgodelska, and Zlatitsa.

Iron, silver and gold mining activities in NW Bulgaria dates since ancient times due to complex geological setting and polymetallic ores (Benderev et al., 2015). Gold mining in the Ogosta River basin are known since Roman times, while Pb, Ag and Fe mining in the river basin of Chiprovska dates since Thracian time (II century B.C.) (Jordanova et al., 2013; Benderev et al., 2015). Intensive mining is documented in the period XIV to XVII centuries (Jordanova et al., 2013). Recent industrial ore production (Fe, Au, Pb, Ag) is very intensive from 1959-1975 and continues till the year 1999. The ore extraction, flotation, operation of several tailing ponds, and a tailings dam failure west from settlement Martinovo in 1964, have caused the surface water, soil and groundwater pollution of the Ogosta River floodplain with As and heavy metals at some areas (Kotsev, 2003; Mladenova et al., 2008; Mandaliev et al., 2013; Jordanova et al., 2013; Mikutta et al., 2014; Benderev et al., 2015; Kotsev et al., 2015; Simmler et al. 2016; Stoyanova and Kotsev, 2016; Simmler et al., 2017; Antonov et al. 2018).

According to Benderev et al. (2015: 198) the Neozoic and Mesozoic underlying rocks of the Ogosta River floodplain in the study area are covered by alluvial deposits with two-layer structure: a lower layer built of gravels and boulders with sand, and an upper layer of sandy-clayey deposits (sandy clays, clayey sands and clays).

The soil types in the valley are determined according to FAO (Food and Agriculture Organization of the United Nations) classification (IUSS Working Group WRB, 2015). The eutric and dystric Fluvisols, and Gleysols are the most widespread soil types within the bottom of the valley. Usually, gleyic and dystric Colluvisols, rendzic Leptosols, dystric Cambisols, and luvic Phaeozems are found in the adjacent areas.

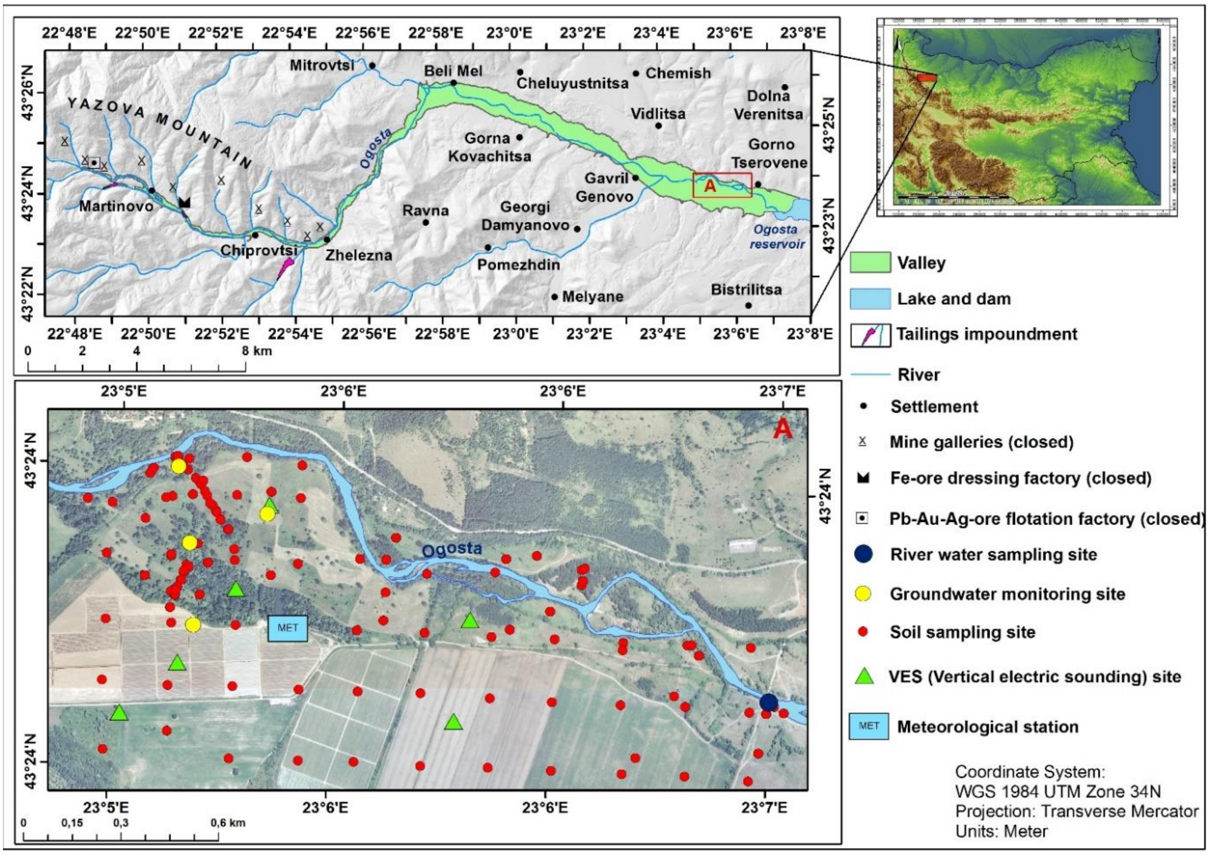


Figure 1. Schematic map of the upper part of Ogosta Valley, NW Bulgaria and site location of field collected data near the Ogosta Reservoir

The upper part of the Ogosta Valley is shared between three municipalities: Chiprovtsi, Georgi Damyanovo and Montana.

3. MATERIALS AND METHODS

3.1. Data collection

During the projects of ASCOR and ARSENT numerous field campaigns for samplings of soil, surface water and groundwater were conducted, and geophysical and hydrogeological data for site locations in the Ogosta River floodplain is collected and compiled for a geodatabase (Table 1). The total number of the measured soil samples is as follows: 1388 soil samples from the depth intervals of 0-20, 20-40, 40-60, 60-80, 80-100,100-150, 150-200, 200-300, 300-400, and 400-500 cm; 1223 groundwater samples; 524 river water samples. Physical and physico-chemical parameters of river water, groundwater and soil pore-water such as water level, redox potential, pH, electrical conductivity and dissolved oxygen are measured permanently using automatic telemetric systems LogCom-2 (SEBA Hydrometrie) and Stylitis-10 (Symmetron), which are installed at four sites in the Ogosta Valley near the village of Gavril Genovo. An automatic weather station (LaCrosse WS2355) is mounted 4 km away in the valley to collect continuous data on air temperature, atmospheric pressure, precipitation, and wind velocity and direction.

Table 1. A part of the collected field sampling data for the floodplain of the Ogosta River

Theme	File format	Measurement	Year
Meteorology	Shape (Point) / Excel (.xlsx)	Air temperature - T_out; T_out_min; T_out_max; Precipitation - P_1h; P_1h_max; P_24h; P_24h_max; Wind – velocity and direction; Atmospheric pressure.	since 2013
Hydrology	Shape (Point) / Excel (.xlsx)	River water level; Trace and macro elements in river water (48) – total dissolved As (ppb), As^{III} (ppb), As^{V} (ppb), etc; Physicochemical indicators – temperature (°C), EC (mS/cm), water level (m), salinity (mg/l), TDS (g/l), pH, O_2 (mg/l), O_2 (%), ORP (mV).	since 2014
Hydrogeology	Shape (Point) / Excel (.xlsx)	Groundwater level; Trace and macro elements in groundwater (48) – total dissolved As (ppb), As^{III} (ppb), As^{V} (ppb), etc; Trace and macro elements in soil pore water (48) – total dissolved As (ppb), As^{III} (ppb), As^{V} (ppb), etc; Physicochemical indicators – temperature (°C), EC (mS/cm), water level (m), salinity (mg/l), TDS (g/l), pH, O_2 (mg/l), O_2 (%), ORP (mV), soil moisture (dielectric permittivity).	since 2014
Geophysics	Shape (Point) / Excel (.xlsx)	Measurement of the magnetic susceptibility of soils (MSS); Vertical Electrical Sounding (VES)	2013 2014
Soil physics	Shape (Point) / Excel (.xlsx)	Hydraulic parameters - θr, θs, α, n, Ks	2015
Soil chemistry	Shape (Point) / Excel (.xlsx)	Trace and macro elements in soils – As (ppm), Na (pct), Mg (pct), Al (pct), Si (pct), P (pct), S (ppm), Cl (ppm), K (pct), Ca (pct) and others	1994 2010 2012 2013 2014 2017

Table 2. Airborne Laser Scanning data for the study area

Data/ Area of each tile - 1 km²	File format	Airborne Laser Scanner	Point density	Relative accuracy of the ALS	Year
Topographic data (elevation)/ 252	LAS 1.2 files	*RIEGL™* LMS Q680i topographic scanner	8 points per square meter at the "corridor" area	15 cm relative horizontal accuracy	2013
Topo-hydrographic data)/ 205	LAS 1.2 files	*RIEGL™* VQ-820-G Topo-hydrographic scanner	8 points per square meter at the "corridor" area	9 cm relative vertical accuracy	2013
DSM data/ 252	LAS 1.2 files	*RIEGL™* LMS Q680i topographic scanner	8 points per square meter at the "corridor" area	15 cm relative horizontal accuracy	2013
DSM data/ 205	LAS 1.2 files	*RIEGL™* VQ-820-G Topo-hydrographic scanner	8 points per square meter at the "corridor" area	9 cm relative vertical accuracy	2013
Orthoimages/ 252	.tif	RGB Camera - IGI Digicam H-39 using a Hasselblad RGB objective HC 3,5/35	Array size - 39 Megapixels; 36.80 x 49.07 mm Pixel size - 6.8 µm Image size - 7216 x 5412 pixels Max. exposure rate - 1.9 sec Data storage with HDD or SSD - External storage units for 8000 images (2 x 4000)		

Source: Zaff, M. (2013). Interim report Ogosta River - T182T. Airborne Technologies GmbH, Wiener Neustadt Austria

The study area is scanned using Airborne Laser Scanning (ALS) technology in 2013 by Airborne Technologies GmbH, Austria in contract with National Institute of Geophysics, Geodesy and Geography at Bulgarian Academy of Sciences (Table 2). The ALS data is used for the creation of precise DEM, DSM and normalized DSM (nDSM) with pixel size 1 m, as well as for digital terrain analysis to calculate relevant local and compound terrain variables and to recognise landform units in the investigated part of the Ogosta Valley (Tcherkezova, 2015).

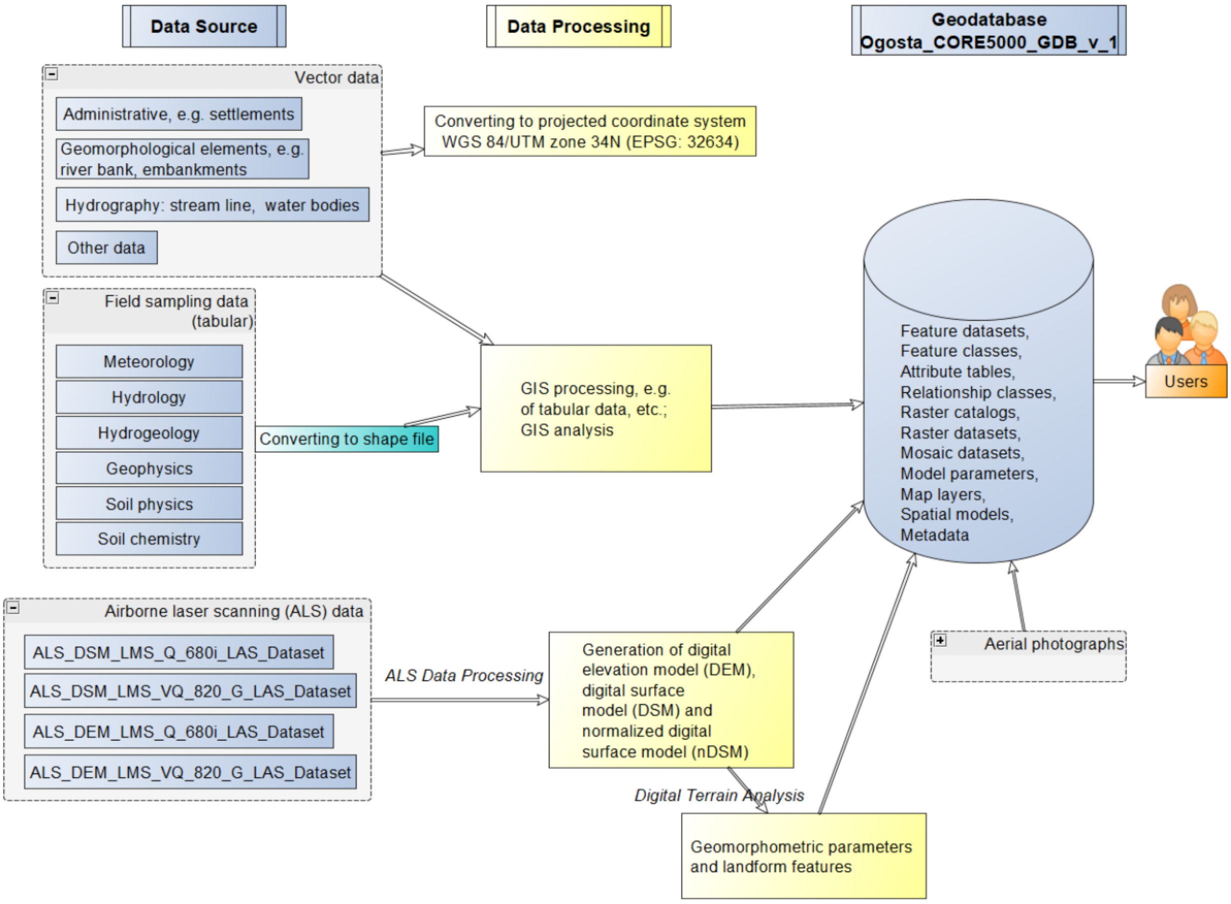


Figure 2. Data flowchart of the development of the geodatabase Ogosta_CORE5000_GDB_v_1

Additionally, the following information is collected: analogue topographic maps on scale 1:5000; aerial photographs (years 1961-1962, 1970-1972, 1978, 1985, and 1993); land use digital data; basic digital data on the hydrographic network, administrative boundaries, settlements, and roads. Most of the digital data resources come with different quality, scale and spatial reference system. For this reason, the compilation process included a qualitative and quantitative analysis of each data source. Figure 2 shows the data flowchart in the development of the spatial database. The spatial database should ensure consistency and reliability of the information, avoids data redundancy, and identifies relevant data for the investigation. The field data on soil, river water and groundwater, as well as geophysical data (VES – vertical electrical sounding and data from measurements of magnetic *susceptibility of soils*) is stored in separate tables (see Table 1). This data is also compiled in vector format (shape layers).

3.2. Methodology and Software

The geodatabase design is an important step in the geodatabase development. The methodological framework for designing a geodatabase is based on the concept of thematic layers according to Zeiler (1999) and ESRI (2018a) and integrates key database development, as well as application tasks and their dependencies. It involves three main steps: requirement analysis, conceptual and logical modelling, and physical implementation following the main geodatabase design steps described by ESRI (2018a), namely: 1) identifying the key thematic layers; 2) specifying the scale ranges for data use and spatial representations for each scale range; 3) decomposing each representation into data sets: feature classes, raster data sets, tables, etc.; 4) identifying the attribute fields; 5) specifying any valid values and relationships; 6) identifying subtypes for each feature class; 7) defining spatial relationships, integrity rules, and behaviours (i.e., topologies and networks); 8) designing the geodatabase; 9) implementing, review, and refine the geodatabase design; 10) designing workflows for building and maintaining each layer; 11) documenting the design using appropriate methods (e.g. database schema, data dictionary, UML diagrams) (Figure 3).

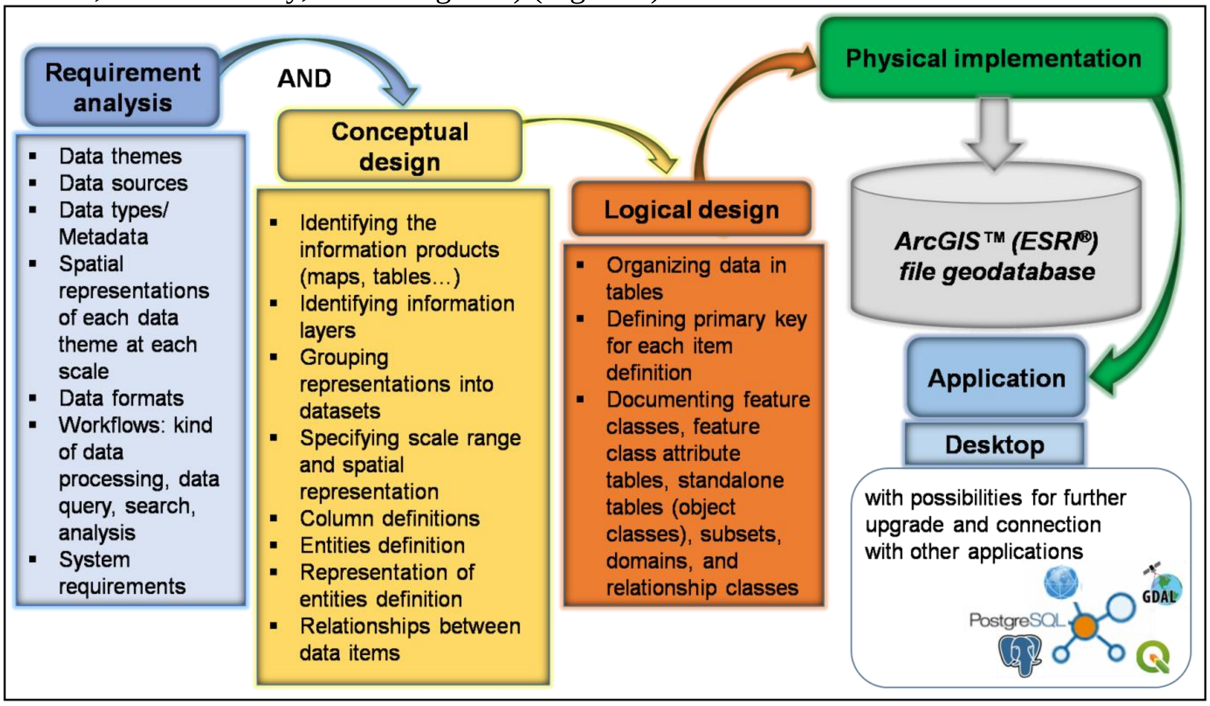


Figure 3. Methodological framework of the geodatabase design process in this work

The investigation of surface water, soil and groundwater pollution with As is organised in two different hierarchical levels – the upper part of the Ogosta River floodplain between the village of Martinovo and Ogosta Reservoir, and the upper part of the Ogosta River basin. Each level needs an appropriate scale range and spatial resolution. The different spatial data should be harmonized and linked to each other. This paper focuses on the creation of a geodatabase concept for the first level, namely the bottom of the Ogosta Valley between the village of Martinovo and the Ogosta Reservoir, because the long-term monitoring of surface water and groundwater pollution with heavy metals is located in this area (Figure 4).

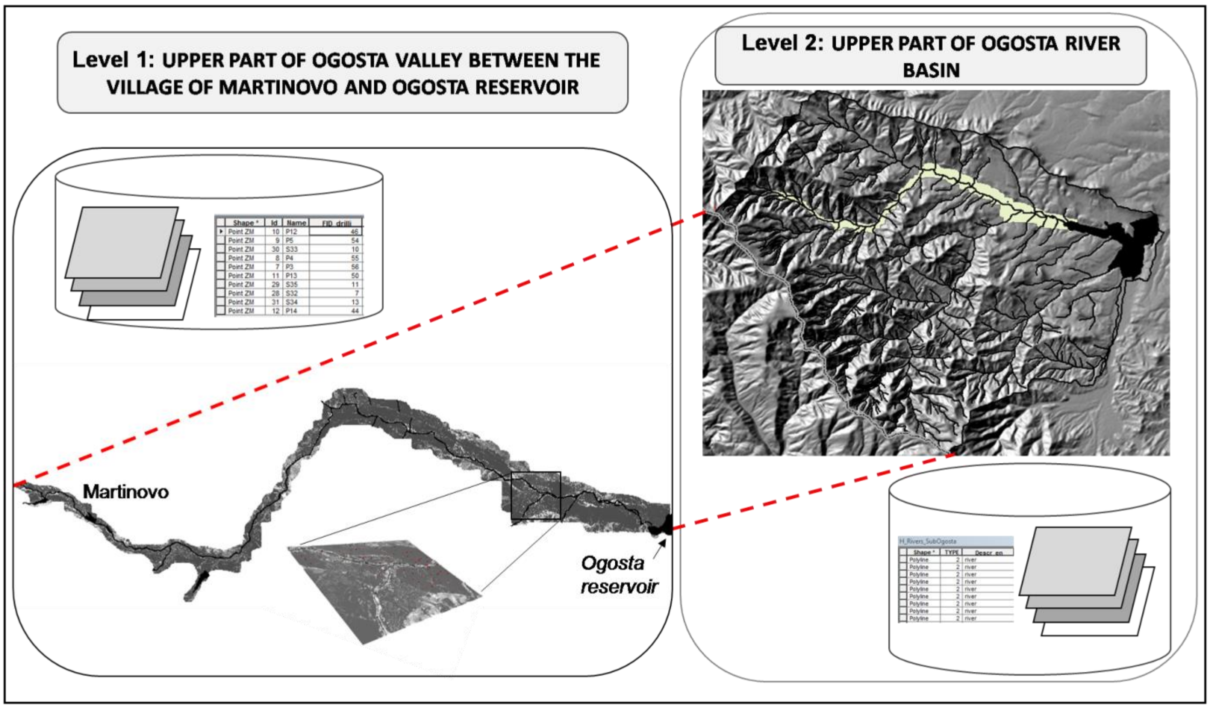


Figure 4. Investigation and geodatabase hierarchical levels

Requirement analysis is based on the objectives of the projects ASCOR and ARSENT with focus on interdisciplinary investigation and long-term monitoring of surface water, soil and groundwater pollution with As and heavy metals in the upper part of the Ogosta Valley. The necessary datasets are analysed according to their content, type and digital format. Additionally, workflows were created to define various types of data processing, data search, query, and analysis for different research tasks.

The conceptual model produces the primary geodatabase datasets: feature class (point, line, or polygon), object class (tables) or continuous spatial information (e.g. DEM), the data themes and the relationships between the objects. For instance, field sampling data is important elements of the spatial geodatabase due to their association with the measurement of trace and macro elements, as well as with the physicochemical indicators (see Table 1).

The logical phase includes identification, documentation and defining the geodatabase infrastructure, e.g. organization of vector data in tables, defining data types, attribute fields, primary keys, field description, associated domains, designing subs-types, and relationships.

The physical implementation in this work will be done using ArcGIS™ file geodatabase.

A common projected coordinate system was defined: WGS 84/UTM zone 34N (EPSG: 32634).

Geodatabase documentation will be created for documenting the geodatabase data model and description of the spatial database.

In this study we use the ArcGIS™ version 10.1 desktop platform by (ESRI® – Environmental System Research Institute's) to create a file geodatabase (ESRI 2018b), and ArcGIS Diagrammer 10.0.1.0 to edit, create and analyse the geodatabase schema. A **file geodatabase** consists of a ***.gdb** folder enabling storage of feature datasets, feature classes, relationship class definition to relate one feature class or geodatabase table to another, as well as raster catalogues and raster classes, LAS data, geodatabase tables and images (ESRI 2018c). ArcGIS Diagrammer is software to build, edit, analyse and illustrate geodatabase schema (ESRI 2018d). It includes legacy schema and data reporting tools. ArcGIS Diagrammer is a visual editor for ESRI's xml workspace documents that can be created in ArcCatalog or ArcMap and vice versa. Any created diagram with this tool can be saved as xml workspace documents, and respectively can be open in ArcCatalog in ArcGIS 10.1, as well as in the current ArcGIS version (ESRI 2018d). This software is chosen due to the ArcGIS™ use in the previous project ASCOR. It uses a geographic data model named ArcGIS data model that utilizes a relational database management model (RDBMS) and supports an object-oriented data model ESRI (2018e). Additionally, recent versions of Geospatial Data Abstraction Library (GDAL) enable efficiently reading and extracting information from ESRI's file geodatabases, e.g. GDAL-utilizing software like QGIS (https://qgis.org/en/site/).

4. RESULTS

4.1. Data processing and modelling

Various thematic geographic datasets should be integrated for creating thematic maps and models as a part of a geographic database for analysis, assessment and modelling surface water, soil and groundwater pollution with As in the upper part of the Ogosta Valley. The data integration includes several GIS procedures for providing data consistency, elimination of data redundancy, and ensuring data interoperability.

The digital data gathered from various sources are non-homogeneous, in different coordinate systems and formats, so the final database schema is created to capture only essential information needed for the research. Tabular information is proofed to identify and eliminate duplicate data. All data is projected in the common projected coordinate system WGS 84/UTM zone 34 N (EPSG: 32634).

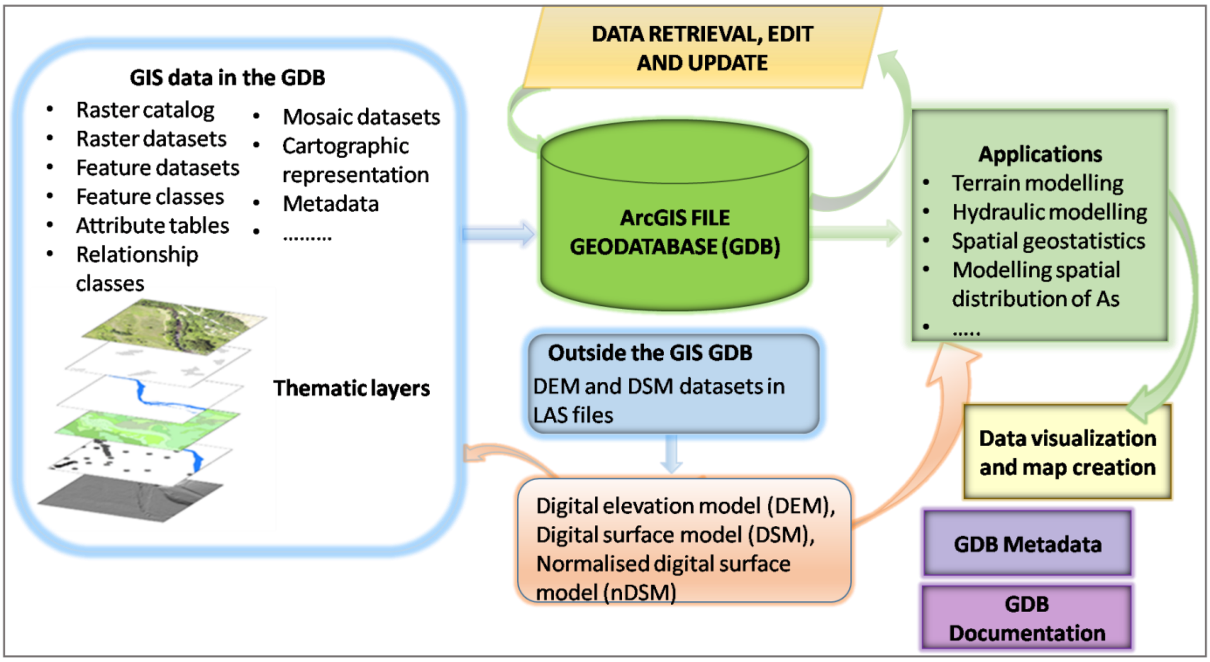


Figure 5. Conceptual flowchart of the geodatabase

A conceptual flowchart of the ArcGIS file geodatabase is created to ensure the basis of an efficient and a comprehensive data structure in order to develop an integrated interdisciplinary GIS database to provide solutions in assessment of surface water, soil and groundwater pollutions with As in the upper part of the Ogosta Valley and similar areas (Figure 5).

The monitoring, analysis and modelling the spatial distribution of arsenic in contaminated river water, soil and groundwater require interdisciplinary research involving hydrogeologists, geophysicists, geomorphologists, hydrologists, physiogeographers, ecologists, GIS experts, and IT specialists.

The hydrogeological (e.g. piezometric), geophysical (e.g. vertical electrical sounding – VES), river water, soil, and groundwater sampling data, as well as meteorological data in table format is transformed as point feature datasets using their coordinates.

This data, as well as the available digital data such as hydrographic network, settlements, roads, administrative and study area boundaries at the same spatial scale, are assimilated into the following categories identifying their geographic component: object classes (tables) and type of feature classes (points, lines, polygons). Relationship classes are defined to link geospatial data stored in the feature classes with related tabular data and information stored in attribute tables. Each spatial and lookup table has a field "primary key", containing a unique identifier for each entry within the table. Primary keys are specified as "Short integer" field type. For example, the soil sampling sites can be joined with the magnetic data and grain size data through the primary key "Key_code".

The data from Airborne Laser Scanning (in LAS 1.2. files) is used to create precise DEM, DSM and normalized DSM (nDSM) with 1 m grid size (Figure 6).

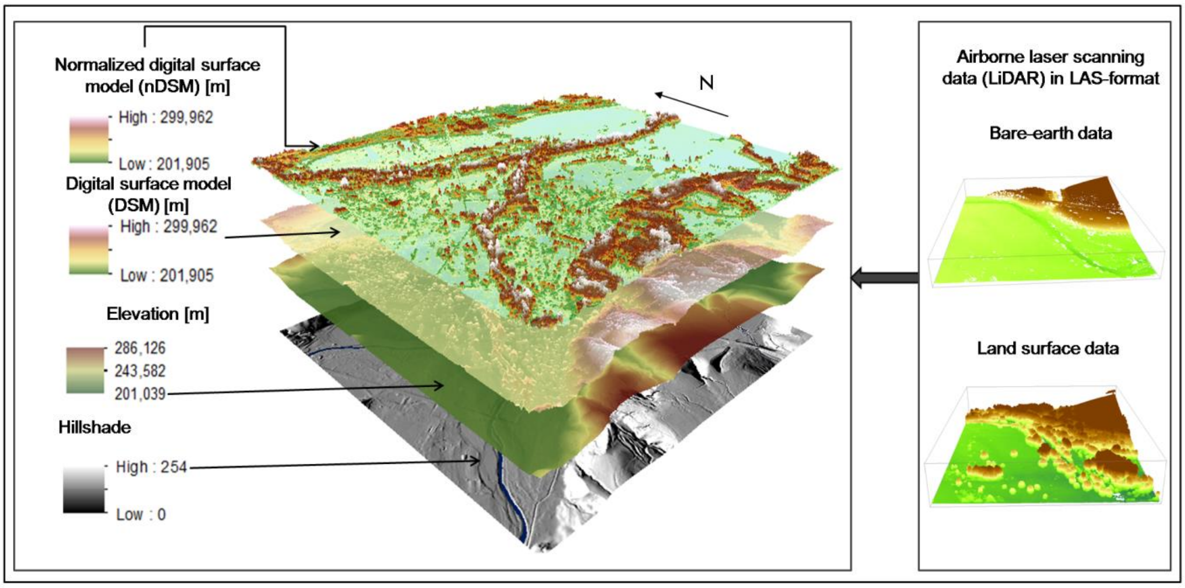


Figure 6. Basic digital models for terrain and land surface analysis on the example of an area near the settlement Gavril Genovo

Orthoimages (in .tif format) are georeferenced, and grouped in mosaic datasets for analysis of stream network changes in the period between 1961 and 2013.

Advanced digital terrain analysis is done to calculate relevant local and compound terrain variables, e.g. slope, aspect, altitude above the stream network, modified topographic wetness index (mTWI), topographic classification index for lowlands (TCIlow), multiresolution index of valley bottom flatness (MRVBF) using SAGA GIS (System for Automated Geoscientific Analysis) (Conrad et al., 2015) for the floodplain description and landform recognition (Tcherkezova, 2015) (Figure 7). These raster data layers are stored as raster datasets stored in raster catalogues.

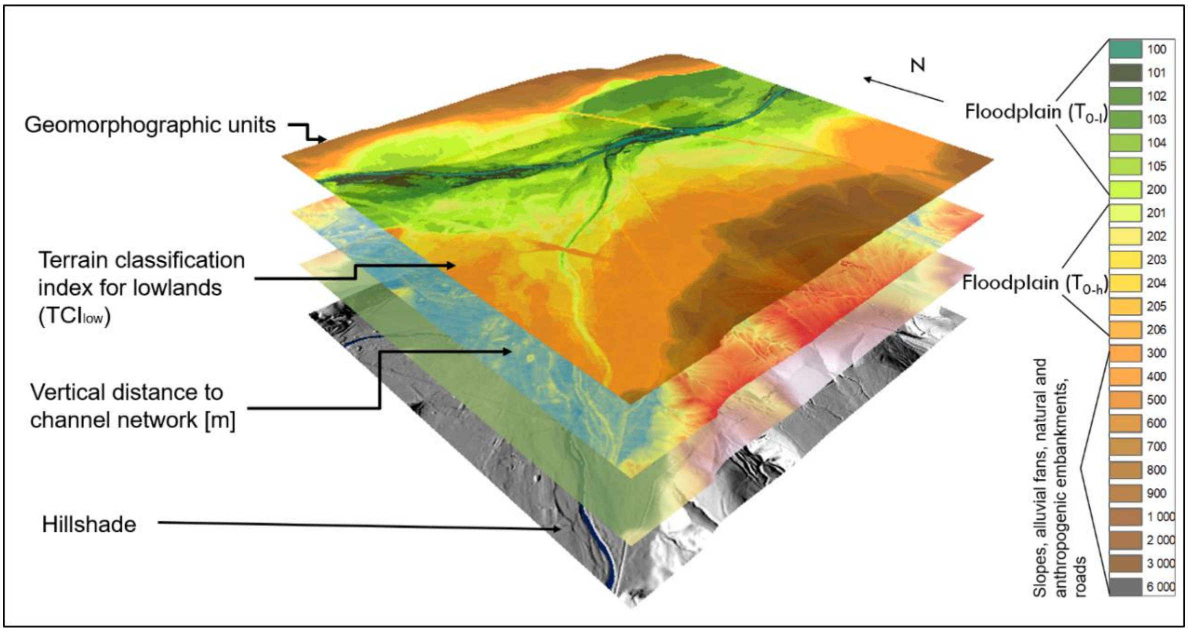


Figure 7. Main geomorphographic floodplain units on the example of an area near the settlement Gavril Genovo (Tcherkezova, 2015)

4.2. Metadata

The creation of metadata for each feature class and dataset, as well as for raster catalogues follows the requirements of the Directive 2007/2/EO INSPIRE (Infrastructure for Spatial Information in Europe) (2007) using the INSPIRE-compliant metadata standard *INSPIRE Metadata Directive* in ArcCatalog (ESRI). The main goals of INSPIRE are to ensure adequate infrastructure for spatial information in Europe, as well as a combination of seamless spatial information from different sources (DIRECTIVE 2007/2/EC…., 2007; Oterro and Torres, 2017). According to INSPIRE, the following metadata standards are developed by the International Organization for Standardization (ISO) and Open Geospatial Consortium (OGC): EN ISO 19115, EN ISO 19119 and ISO 15836 (Dublin Core).

In this work, metadata is created for each feature class, feature dataset, raster dataset, raster catalogues, as well as for each layer and the whole geodatabase. The metadata documentation includes item description such as title, summary, abstract, date of creation, keywords, data sets, geographic location, scale range, constraints related to access and use, publication date, as well as resource and metadata information. Also, the metadata record describes the geographic setting for each spatial data set, including the projected coordinate system and vertical/horizontal datum.

4.3. Geodatabase modelling

Following the main goal in this work to draft a concept for an integrated geodatabase for surface water, soil and groundwater pollution with arsenic in the upper part of Ogosta Valley, a simple geodatabase schema is developed to compile the data needs, while using all necessary geodatabase functionality.

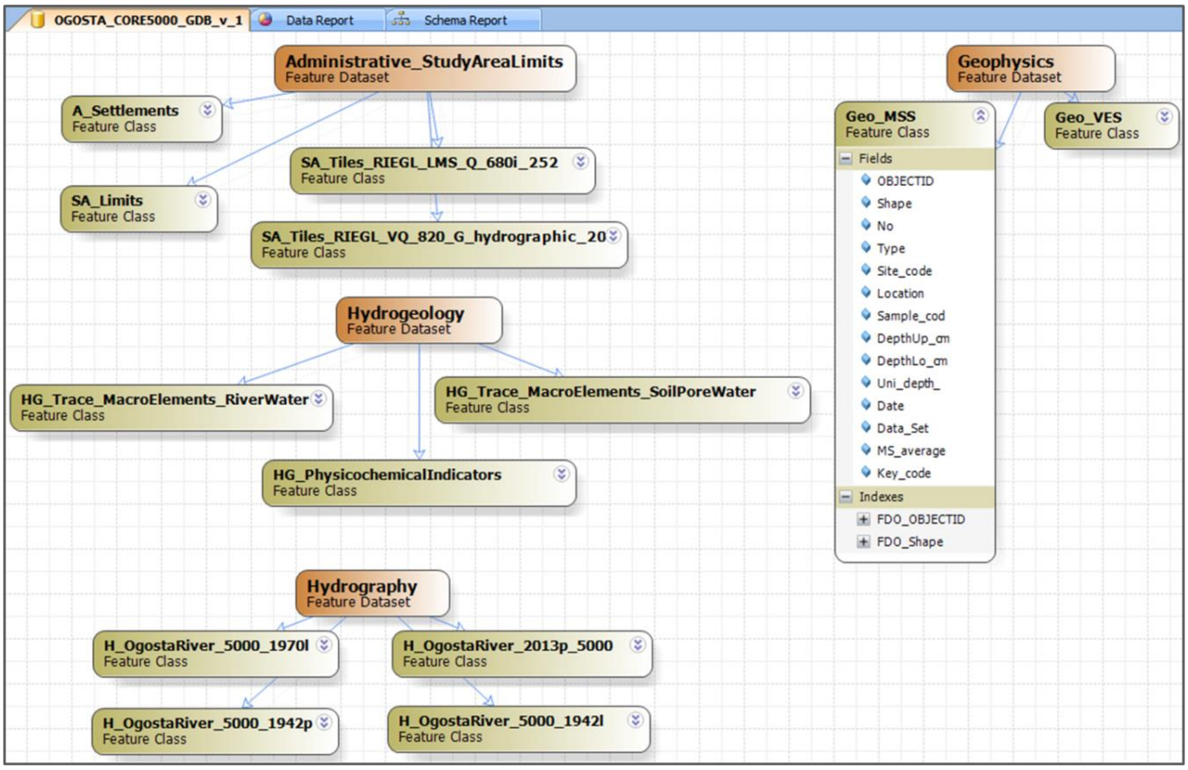


Figure 8. A part of the geodatabase diagram with an example of the feature dataset Geophysics (vertical electrical sounding and geophysical data from measurements of magnetic *susceptibility of soils*)

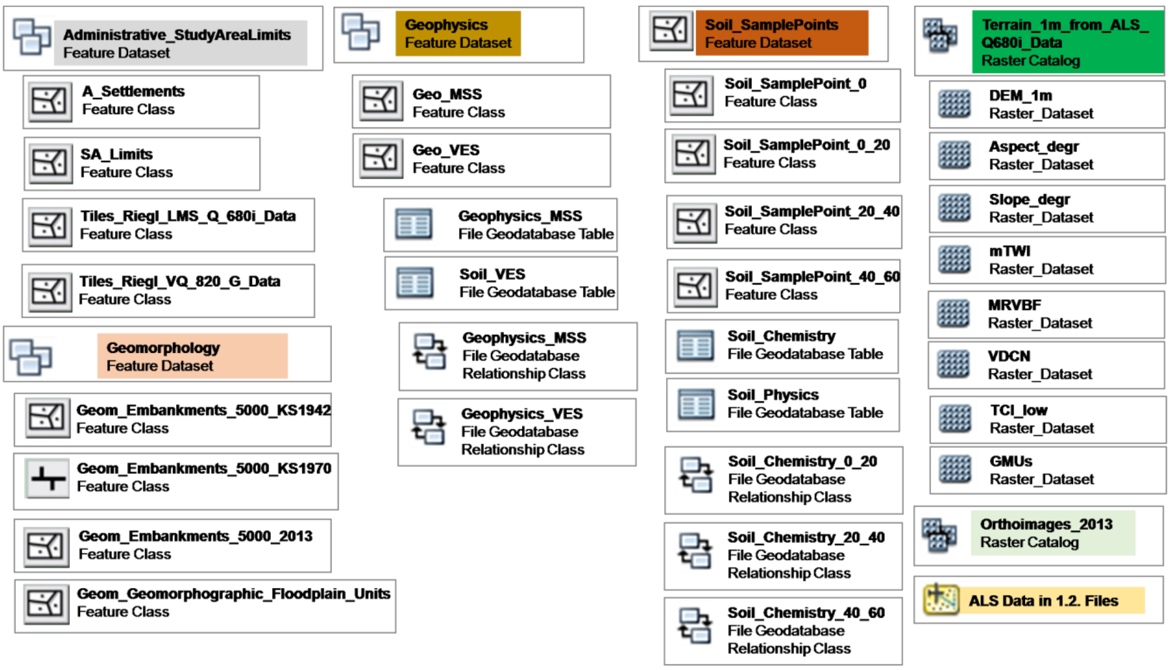


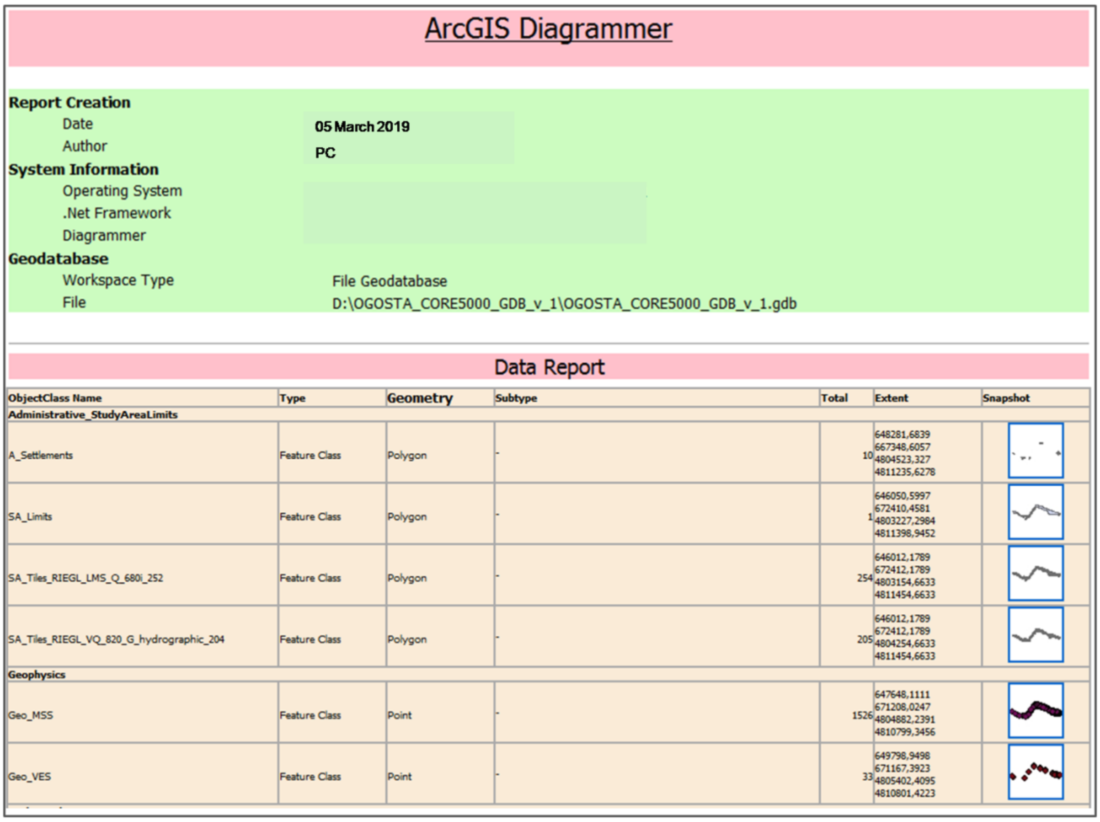
Figure 9. A part of the geodatabase model of the geodatabase Ogosta_Core5000_GDB_v_1. SA – study area, *VES – vertical electric sounding, MSS – magnetic susceptibility of soils, DEM – digital elevation model, degr – degree, mTWI – modified topographic wetness index, MRVBF – multiresolution index of bottom flatness, VDCN – vertical distance to channel network, GMUs – geomorphographic units, ALS – airborne laser scanning*

Figure 8 shows a summary of a part of the entire geodatabase schema diagram. The schema includes only the necessary basic and thematic vector, raster, tabular and ALS data at level 1. The used ESRI's GIS approach for geodatabase design customizes the design for each theme, e.g. meteorology, hydrogeology, hydrology, soil, and others. All the field-collected data (see Tables 1 and 2) is included together with the results from the digital terrain analysis such as locale and compound geomorphometric variables and recognised floodplain units in the investigated area, some of which are described above. Object classes are utilized to their full potential in the geodatabase and are separated into thematic categories. Relationship classes are devised to connect appropriate data elements, making it easier to manipulate and manage the data. For example, the following relationships are created: *soil sample points - soil chemistry*, *soil sample points - soil physics*, *soil sample points - magnetic susceptibility of soils*, *soil sample points - land use*, and others. In this case, relationships allowed effective monitoring storage using the one-to-many cardinality for data management.

As mentioned in subsection 3.2, the geodatabase model diagram is created using ArcGIS Diagrammer 10.0.1.0. for ArcGIS 10.1. (Figure 8). A simplified picture is given in figure 9. The feature datasets and raster catalogues are coloured, while their feature classes, object classes, and relationship classes are displayed under each feature dataset and raster catalogues.

4.4. Documenting the geodatabase design

According to ESRI (2018e), the geodatabase design can be documented using appropriate methods, including database schema, data dictionary and UML diagrams. Figures 10 and 11 show a part of the diagram data report, as well as a part of the geodatabase schema data report, that are documented using ArcGIS Diagrammer.


ArcGIS Diagrammer

Report Creation
Date 05 March 2019
Author PC
System Information
Operating System
.Net Framework
Diagrammer
Geodatabase
Workspace Type File Geodatabase
File D:\OGOSTA_CORE5000_GDB_v_1\OGOSTA_CORE5000_GDB_v_1.gdb

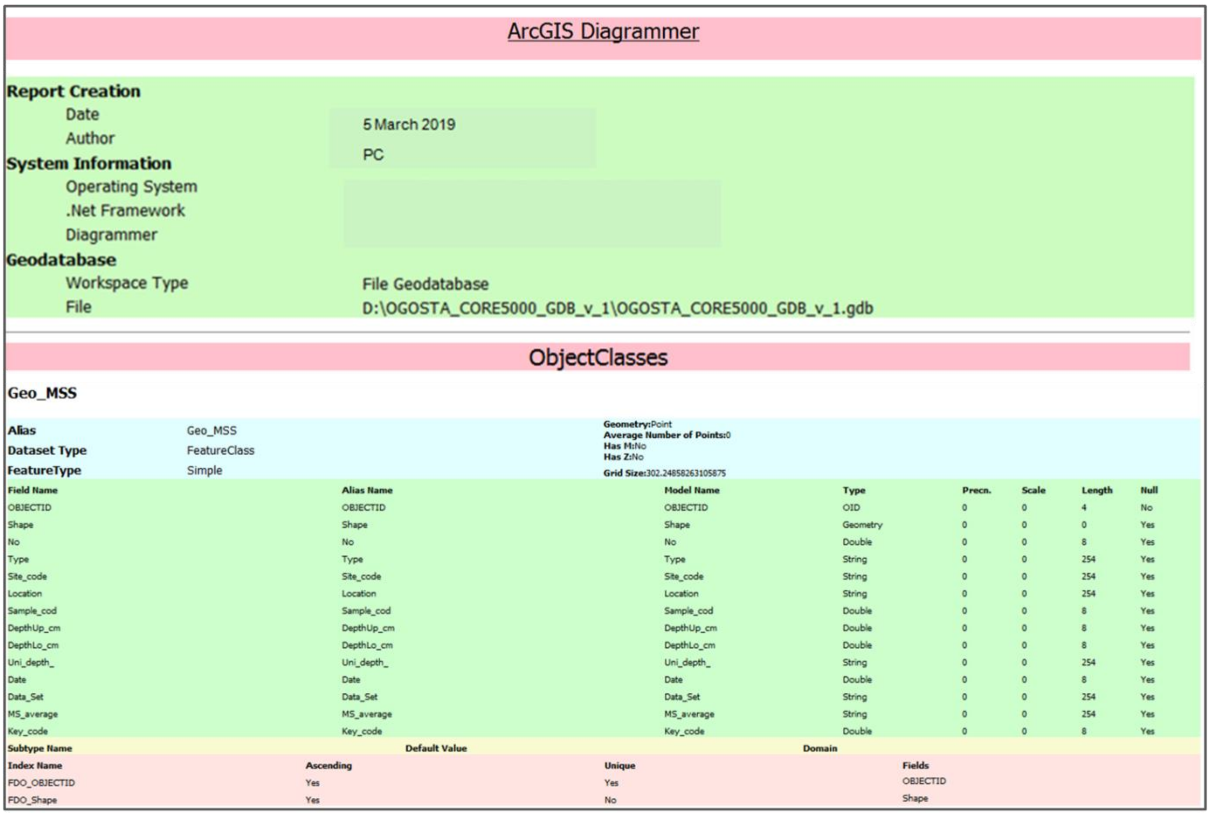
Data Report

ObjectClass Name	Type	Geometry	Subtype	Total	Extent	Snapshot
Administrative_StudyAreaLimits						
A_Settlements	Feature Class	Polygon	-	10	648281,6839 667348,6057 4804523,327 4811235,6278	
SA_Limits	Feature Class	Polygon	-	1	646050,5997 672410,4581 4803227,2984 4811398,9452	
SA_Tiles_RIEGL_LMS_Q_680i_252	Feature Class	Polygon	-	254	646012,1789 672412,1789 4803154,6633 4811454,6633	
SA_Tiles_RIEGL_VQ_820_G_hydrographic_204	Feature Class	Polygon	-	205	646012,1789 672412,1789 4804254,6633 4811454,6633	
Geophysics						
Geo_MSS	Feature Class	Point	-	1526	647648,1111 671208,0247 4804882,2391 4810799,3456	
Geo_VES	Feature Class	Point	-	33	649798,9498 671167,3923 4805402,4095 4810801,4223	

Figure 10. A part of the diagram data report on the example of the feature datasets "Administrative_StudyAreaLimits" and "Geophysics" (data gathered from the measurement of the magnetic susceptibility of soils)

The data and schema report are exported in Web page format – OGOSTA_CORE5000_GDB_v_1.html and in Web archive single file – OGOSTA_CORE5000_GDB_v_1.mht.

A data dictionary is a tool for efficient providing of information that supports the geodatabase use also from scientists outside the ASCOR and ARSENT research teams. In this research it will describe the nature of all the data that have been input into the geodatabase containing a full information about the collected data sets, e.g. object types, data features, attribute data, GIS layers, and so forth.



ArcGIS Diagrammer

Report Creation

Date	5 March 2019
Author	PC

System Information

Operating System	
.Net Framework	
Diagrammer	

Geodatabase

Workspace Type	File Geodatabase
File	D:\OGOSTA_CORE5000_GDB_v_1\OGOSTA_CORE5000_GDB_v_1.gdb

ObjectClasses

Geo_MSS

Alias	Geo_MSS	**Geometry:**Point **Average Number of Points:**0 **Has M:**No **Has Z:**No
Dataset Type	FeatureClass	
FeatureType	Simple	**Grid Size:**302.24858263105875

Field Name	Alias Name	Model Name	Type	Precn.	Scale	Length	Null
OBJECTID	OBJECTID	OBJECTID	OID	0	0	4	No
Shape	Shape	Shape	Geometry	0	0	0	Yes
No	No	No	Double	0	0	8	Yes
Type	Type	Type	String	0	0	254	Yes
Site_code	Site_code	Site_code	String	0	0	254	Yes
Location	Location	Location	String	0	0	254	Yes
Sample_cod	Sample_cod	Sample_cod	Double	0	0	8	Yes
DepthUp_cm	DepthUp_cm	DepthUp_cm	Double	0	0	8	Yes
DepthLo_cm	DepthLo_cm	DepthLo_cm	Double	0	0	8	Yes
Uni_depth_	Uni_depth_	Uni_depth_	String	0	0	254	Yes
Date	Date	Date	Double	0	0	8	Yes
Data_Set	Data_Set	Data_Set	String	0	0	254	Yes
MS_average	MS_average	MS_average	String	0	0	254	Yes
Key_code	Key_code	Key_code	Double	0	0	8	Yes

Subtype Name	Default Value	Domain

Index Name	Ascending	Unique	Fields
FDO_OBJECTID	Yes	Yes	OBJECTID
FDO_Shape	Yes	No	Shape

Figure 11. A part of the geodatabase schema data report on the example of the feature class *Geo_MSS* (magnetic susceptibility of soils)

4.5. The geodatabase prototype

The geodatabase diagram is published as XML workspace document and then implemented in a file geodatabase using the created diagram Ogosta_CORE5000_GDB_v_1. The goal of the constructed prototype is to prove its facility to offer accessible, integrative, updatable, and documentable data sets (Figure 12). The ALS data are stored in LAS datasets outside the geodatabase.

This physical implementation can be exported back as XML workspace document and edited and upgraded again in ArcGIS Diagrammer. Additionally, the ESRI's file geodatabase can be open in desktop QGIS. Using QGIS, the data layers can be exported in various types of data formats, e.g. GeoPackage (an open, standards-based, platform-independent, portable, self-describing, compact format for transferring geospatial information, https://www.geopackage.org/) GML (Geography Markup Language), KML (*Keyhole Markup Language), which improve the data interoperability by users of the free and open source software QGIS.*

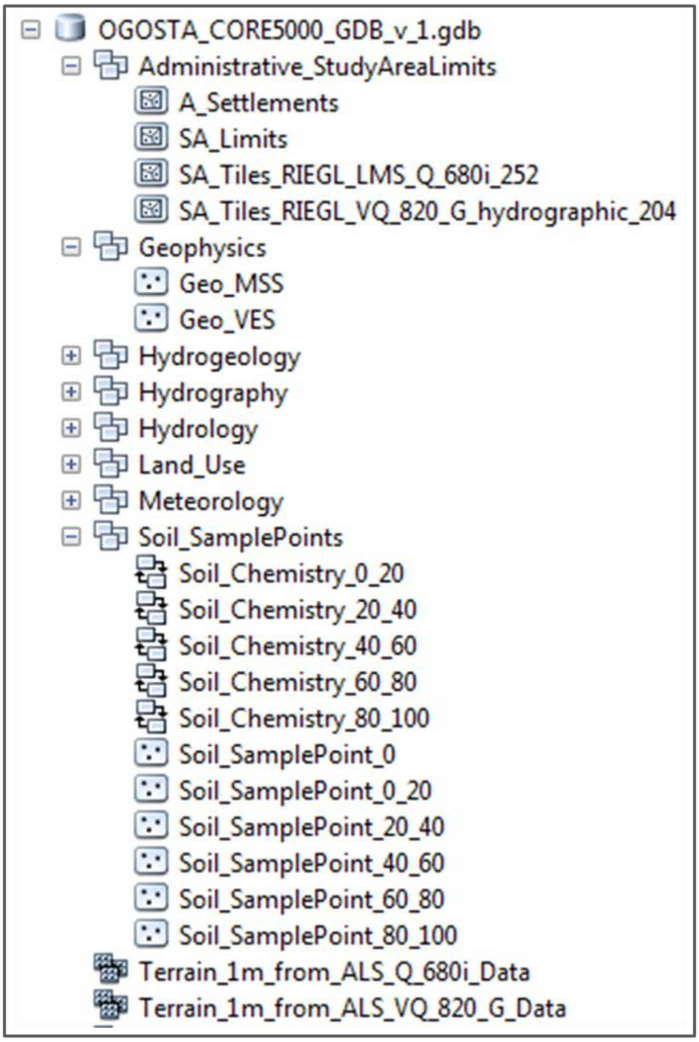


Figure 12. A part of the file geodatabase prototype

This physical implementation can be exported back as XML workspace document, as well as edited and upgraded again in ArcGIS Diagrammer (ESRI 2018d). Additionally, the ESRI's file geodatabase can be open in desktop QGIS. Using QGIS, the data layers can be exported in various types of data formats, e.g. GeoPackage (an open, standards-based, platform-independent, portable, self-describing, compact format for transferring geospatial information, https://www.geopackage.org/) GML (Geography Markup Language), KML (*Keyhole Markup Language), which improve the data interoperability by users of the free and open source software QGIS.*

5. CONCLUSIONS

This paper presents a concept for an integrated geodatabase for surface water, soil and groundwater pollution with arsenic in the upper part of the Ogosta Valley (NW Bulgaria). Relevant available digital data are obtained and proofed according to their quality and accuracy. A total number of 1223 groundwater samples, 524 river water samples, and 1388 soil samples from the depth intervals of 0-20, 20-40, 40-60, 60-80, 80-100, 100-150, 150-200, 200-300, 300-400 and 400-500 cm are collected during two research projects: ASCOR (2012-2016) and ARSENT (2016-2020). Additionally, long-term monitoring of river water and groundwater quality is organized in the study area, producing a continuous time series of data (see Table 1).

The main focus of this work is the development of a geodatabase concept starting from the phase of requirement analysis and concept design, through logical modelling, and ending with prototype creation within the ArcGIS environment. The created geodatabase prototype facilitates integration of the spatial and descriptive data into an integrated geodatabase. The benefits of using spatial database technology for long-term environmental monitoring in the study area are related to the capability of storage, manipulation, querying, and management of the basic and thematic spatial data and attribute tables using a Geographic Information System (GIS).

A constructed prototype of an integrated geodatabase based on this concept offers accessible, integrative, updatable, and documentable data sets to support the analysis of the collected information. Further development of the geodatabase will focus on the extension of the OGOSTA_CORE5000_GDB_v_1 to support the long-term water quality monitoring in the valley and upgrading in the future to a web format with different levels of data access.

ACKNOWLEDGEMENTS

This work is supported by the National Science Fund of Bulgaria, Grant No ДН 04/3, project ARSENT (2016-2020), as well as by the Bulgarian-Swiss Research Programme 2011-2016, Grant IZEBZO-142978, project ASCOR (2012-2016). It is presented during the international conference 100 years Bulgarian Geographical Society, Smart Geography, 2-4 November 2018, Sofia, Bulgaria.

The authors would like to thank Antonia Mokreva from the section Earth magnetism at the National Institute of Geophysics, Geodesy and Geography, Bulgarian Academy of Sciences for metadata creation of the orthoimages and their compilation in the database.

REFERENCES

Antonov D., Vl. Hristov, Al. Benderev, and T. Kotsev. (2015). Comparing the parameters from pedotransfer functions and in situ permeability tests in the vadose zone of the Ogosta River floodplain in connection with validation procedures of contaminant migration modelling. *Proceedings of National Conference with international participation Geosciences `2015, Bulgaria*, 139-140.

Antonov, D., K. Nakamura, T. Kotsev, V. Stoyanova, and R. Kretzschmar. (2018). Application of HYDRUS-1D for evaluation of the vadose zone saturation state in connection with arsenic mobilization and transport in contaminated river floodplain − Ogosta Valley case study, NW Bulgaria. *Proceedings of the 18th International Multidisciplinary Scientific GeoConference SGEM`2018*, 83-90, doi: 10.5593/sgem2018/1.2.

Benderev, A., P. Gerginov, D. Antonov, N. Van Meier, and R. Kretzschmar. (2015). Conceptual hydrogeological model of the Ogosta River floodplain (Western Balkan, Bulgaria) and its application for predicting of groundwater contamination with arsenic. *Science and Technologies in Geology, Exploration and Mining Conference Proceedings*. Vol. II Hydrogeology, Engineering Geology and Geotechnics. 18-24 June 2015, Albena, Bulgaria, 195-202.

Bill, R. and D. Fritsch. (1994). *Grundlagen der Geo-Informationssysteme. Band 1. Hardware, Software und Daten*. 2. Auflage. Heidelberg: WICHMANN.

Byrne, P., P.J. Wood, and I. Reid. (2012). The impairment of river systems by metal mine contamination: a review including remediation options. *Critical Reviews in Environmental Science and Technology*. 42. 2017–2077.

Carcinogenic_Heavy_Metal_Molecular_Toxicity_Mechanism_and_Prevention.

 © All rights reserved

Conrad, O., Bechtel, B., Bock, M., Dietrich, H., Fischer, E., Gerlitz, L., Wehberg, J., Wichmann, V., and J. Boehner (2015). System for Automated Geoscientific Analyses (SAGA) v. 2.1.4. G*eosci. Model Dev*., 8, 1991-2007, doi:10.5194/gmd-8-1991-2015.

Dimitrova, D., Z. Cholakova, N. Velitchkova, T. Kotsev, V. Mladenova, T. Kerestedjian, and D. Antonov. (2007). Heavy metal and metalloid concentration dynamics in mine and surface waters in the vicinity of Chiprovtsi and Martinovo mines, Northwestern Bulgaria. Bulletin of the Geological Society of Greece. 40 (3). 1397-1408. doi: http://dx.doi.org/10.12681/bgsg.16973

DIRECTIVE 2007/2/EC OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 14 March 2007 establishing an Infrastructure for Spatial Information in the European Community (INSPIRE). 2007. L108/1. https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32007L0002&rid=1 (Accessed 2016-06-28).

ESRI (2018a). ArcGIS Desktop: Release 10.3 Redlands, CA: Environmental Systems Research Institute. Geodatabase design steps. http://desktop.arcgis.com/en/arcmap/10.3/manage-data/geodatabases/geodatabase-design-steps.htm (Accessed, 2018-04-26).

ESRI (2018b). ArcGIS Desktop: Release 10.3 Redlands, CA: Environmental Systems Research Institute. Types of geodatabases. http://desktop.arcgis.com/en/arcmap/10.3/manage-data/geodatabases/types-of-geodatabases.htm (Accessed 2018-04-26).

ESRI (2018c). ArcGIS Desktop: Release 10.3 Redlands, CA: Environmental Systems Research Institute. The architecture of a geodatabase. http://desktop.arcgis.com/en/arcmap/10.3/manage-data/geodatabases/the-architecture-of-a-geodatabase.htm (Accessed 2018-04-26).

ESRI (2018d). ArcGIS Diagrammer for 10.1. https://www.arcgis.com/home/item.html?id=5af1723bab4e413bae63dd89daa87984 (Accessed 2019-02-07).

ESRI (2018e). ArcGIS Desktop: Release 10.3 Redlands, CA: Environmental Systems Research Institute. Documenting your geodatabase design. http://desktop.arcgis.com/en/arcmap/latest/manage-data/geodatabases/documenting-your-geodatabase-design.htm (Accessed 2018-04-26).

Graves, M. *Designing XML Databases*. Wiley, (Ru - ISBN 5-8459-0270-3).

IUSS Working Group WRB. 2015. World Reference Base for Soil Resources 2014, update 2015, International soil classification system for naming soils and creating legends for soil maps. World Soil Resources Reports No. 106. FAO, Rome, 192. E-ISBN 978-92-5-108370-3. http://www.fao.org/3/i3794en/I3794en.pdf (Accessed 2019-02-07).

Jordanova, D., S.R. Goddu, T. Kotsev, and N. Jordanova. (2013). Industrial contamination of alluvial soils near Fe–Pb mining site revealed by magnetic and geochemical studies. *Geoderma*. 192. 237-248.

Kim, H.S., Y.J, Kim, and Y.R. Seo. (2015). An overview of carcinogenic heavy metal: molecular toxicity. Mechanism and prevention. *Journal of Cancer Prevention*. 20. 232-240. https://www.researchgate.net/publication/288872385_An_Overview_of_

Kotsev, Ts. (2003). *Geochemical changes of the landscapes in the drainage basin of the Ogosta Reservoir, induced by the mining activities*. PhD Dissertation, in Bulgarian.

Kotsev, Ts., V. Mladenova, Z. Cholakova, and D. Dimitrova. (2015). Arsenic and heavy metal vertical distribution in soil of the Ogosta River low floodplain, NW Bulgaria. *Proceedings of the International Scientific Conference "Sustainable Mountain Regions: Make Them Work"*, 14-16 May 2015, Borovets, Bulgaria. 78-83.

Mandaliev, P.N., C. Mikutta, K. Barmettler, T. Kotsev, and R. Kretzschmar. (2013). Arsenic species formed from arsenopyrite weathering along a contamination gradient in circumneutral river floodplain soils. *Environmental Science & Technology*. 48. 208-217.

Mikutta, C., P.N. Mandaliev, N. Mahler, T. Kotsev, and R. Kretzschmar. (2014). Bioaccessibility of arsenic in mining-impacted circumneutral river floodplain soils. *Environmental Science & Technology*. 48. 13468–13477.

Mladenova, V., T. Kotsev, I. Ivanova, Z. Cholakova, D. Dimitrova, and R.-T. Schmitt (2008). Mineralogy of the heavy metal and metalloid pollution of the Ogosta River floodplains, NW Bulgaria. *Proceedings of Geosciences* 2008, Sofia, Bulgaria, 121−122.

Oterro, J.Á and M.L. de L. y Torres. (2017). Spatial data infrastructures and geography learning. *European Journal of Geography.* 8 (3).19 - 29. http://www.eurogeographyjournal.eu/articles/2.Spatial%20data%20infrastructures%20and%20Geography%20learning.pdf (Accessed 2019-07-13).

QGIS. A free and open source geographic information system. https://qgis.org/en/site/ (Accessed 2019-03-11).

Rigaux, P., M. Scholl, and A. Voisard. (2002). *Spatial databases with application to GIS*. Elsevier, USA.

Samson, G.L., J. Lu, M.M. Usman, and Q. Xu. (2017). *Spatial databases: An overview*. Lu, J. and Q. Xu. Ontologies and Big Data Consideration for Effective Intelligence. IGI Global book series *Advances in Information Quality and Management* (AIQM). 111-149.

Simmler, M., E. Suess, I. Christl, T. Kotsev, and R. Kretzschmar. (2016). Soil-to-plant transfer of arsenic and phosphorus along a contamination gradient in the mining-impacted Ogosta River floodplain. *Science of Total Environment*, 572. 742-754.

Simmler, M., J. Bommer, S. Frischknecht, I. Christl, T. Kotsev, and R. Kretzschmar. (2017). Reductive solubilization of arsenic in a mining-impacted river floodplain: Influence of soil properties and temperature. *Environmental Pollution*. 231 (1). 722-731.

Stoyanova V., and T. Kotsev. (2016). GIS-based assessment of groundwater vulnerability to arsenic contamination in the floodplain of the Ogosta River, NW Bulgaria. *Proceedings, 6th International Conference on Cartography and GIS*, 13-17 June, Albena, Bulgaria. 668-677.

Tcherkezova, E. (2015). GIS-based delineation and regionalization of geomorphographic units in the floodplain of Ogosta River between the settlement Gavril Genovo and the "Ogosta" Reservoir (NW-Bulgaria). *Problems of Geography*, 1-2, Academic Publishing House "M. Drinov". 114-122.

Tchounwou, P.B., Yedjou, C.G., Patrolla, A.K., Sutton, D.J. (2012). Heavy Metals Toxicity and the Environment. EXS. 2012; 101: 133–164. doi:10.1007/978-3-7643-8340-4_6 (Accessed 2019-03-11).

Watt, A. and N. Eng. (2014). *Database design – 2nd Edition*. Victoria, B.C. https://opentextbc.ca/dbdesign01/ (Accessed June 18, 2018-06-18).

Zehnder, C.A. (1998). *Informationssysteme and Datenbanken.* 6. Auflage. Springer.

Zeiler, M. (1999). *Modeling our world*. The ESRI Guide to Geodatabase Design. Environmental Systems Research Institute, Inc., Redlands, CA. http://downloads2.esri.com/support/documentation/ao_/Modeling_our_World.pdf (Accessed 2018-06-18).