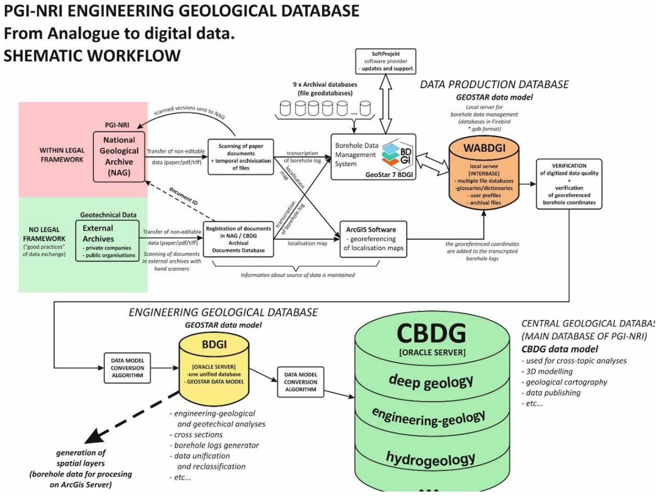
COST is supportrd by the
EU Framework Programme
Horizon 2020

TU1206-WG2.2-003

1970s

2010s


PGI-NRI ENGINEERING GEOLOGICAL DATABASE
From Analogue to digital data.
SHEMATIC WORKFLOW
WITHIN LEGAL FRAMEWORK
PGI-NRI
National Geological Archive (NAG)
NO LEGAL FRAMEWORK
("good practices" of data exchange)
Geotechnical Data
External Archives
- private companies
- public organisations
scanned versions sent to NAG
Transfer of non-editable data (paper/pdf/tiff)
Scanning of paper documents + temporal archivisation of files
document ID
Transfer of non-editable data (paper/pdf/tiff)
Scanning of documents in external archives with hand scanners
Registration of documents in NAG / CBDG Archival Documents Database
Information about source of data is maintained
9 x Archival databases (file geodatabases)
SoftProjekt software provider - updates and support
transcription of borehole log
localisation map
transcription of borehole log
localisation map
Borehole Data Management System
BD GI
GeoStar 7 BDGI
ArcGIS Software - georeferencing of localisation maps
DATA PRODUCTION DATABASE
GEOSTAR data model
Local server for borehole data management (databases in Firebird *.gdb format)
WABDGI
local server [INTERBASE]
- multiple file databases
-glossaries/dictionaries
- user profiles
- archival files
the georeferenced coordinates are added to the transcripted borehole logs
VERIFICATION of digitized data quality + verification of georeferenced borehole coordinates
ENGINEERING GEOLOGICAL DATABASE
GEOSTAR data model
DATA MODEL CONVERSION ALGORITHM
BDGI
[ORACLE SERVER]
-une unified database
- GEOSTAR DATA MODEL
DATA MODEL CONVERSION ALGORITHM
- engineering-geological and geotechical analyses
- cross sections
- borehole logs generator
- data unification and reclassification
- etc...
generation of spatial layers (borehole data for procesing on ArcGis Server)
CBDG
[ORACLE SERVER]
deep geology
engineering-geology
hydrogeology
CENTRAL GEOLOGICAL DATABAS
(MAIN DATABASE OF PGI-NRI)
CBDG data model
- used for cross-topic analyses
- 3D modelling
- geological cartography
- data publishing
- etc...

Data Acquisition & Management

TU1206 COST Sub-Urban WG2 Report

Carl Watson, Niels-Peter Jensen, Grzegorz Ryżyński, Krzysztof Majer and Martin Hansen

COST is supportrd by the
EU Framework Programme
Horizon 2020

TU1206 COST Sub-Urban Report
TU1206-WG2.2-003

Published February 2017

Authors: Carl Watson (BGS), Niels-Peter Jensen (I-GIS A/S), Grzegorz Ryżyński (PGI), Krzysztof Majer (PGI) and Martin Hansen (GEUS)

Contributions from: David Entwisle (BGS), Ane Bang-Kittilsen (NGU) and Gerold Diepolder (LFU), Ingelöv Eriksson (Oslo City) and Susie Mielby (GEUS)

Editor and layout : Guri V. Ganerød (NGU)

www.sub-urban.eu
www.cost.eu

COST is supportrd by the
EU Framework Programme
Horizon 2020

Acknowledgements

This report is based on work from COST Action TU1206 Sub-Urban, supported by COST (European Cooperation in Sciences and Technology), and itself, a component of the European Union', itself a component of the European Union's Horizon2020 programme. Sub-Urban is a network to improve understanding and the use of the ground beneath our cities (www.Sub-Urban.eu).

Many participants of the Sub-Urban COST Action have contributed to this report and to the Action as a whole, each with their own experience and expertise. This evaluation would not have been the same without their presentations, comments, and discussions; they are all thanked for their contributions.

Content

1. Introduction to WG2.2 Data Acquisition and Management

1.1 Rationale

City authorities and other stakeholders in urban environments produce and have access to a greater density of data than is often the case in lesser populated areas, however, it is often very difficult to collate all relevant information together in a useful and easily communicated manner. With such a wide spectrum of stakeholder groups, each with specialist requirements and differing levels of knowledge, it is extremely challenging to provide effective communication tools that disseminate geoscience data and models as useable information. Information about the subsurface needs to be made available in ways which are appropriate to each type of consumer, from a geotechnical engineer carrying out a site investigation to a member of the public wanting to know if their house is at risk of flooding.

Arguably the biggest challenges facing those who attempt to understand urban subsurface environments is developing a reliable and affordable strategy for data acquisition, storage, management and communication. Relationships between geological properties and human processes need to be better understood, this requires a greater understanding of interdisciplinary relationships. Geological Survey Organisations (GSOs), and other public bodies, need to incorporate data from external, sometimes commercial, sources in order to see the whole picture and despite advances in technology which have resulted in more data being made available in digital formats, there remains a large body of analogue data sources which are expensive to digitize. Financial constraints on public authorities and the increasing volumes and variability of data generated means that the current labour intensive processes for acquiring subsurface data are unsustainable. In order to minimize manual processing it is necessary for newly acquired data to be captured and communicated between stakeholders using standardized digital formats that support automated processing.

1.2 Knowledge base

The WG2.2 Data Acquisition and Management group were pulled together to discuss good practice, unresolved issues and strategies for improving data management amongst the Sub-Urban community.

The group is made up of the following members

- Ane Bang-Kittilsen: Norwegian Geological Survey (NGU)
- Anna Wilimowska: City of Warsaw
- Carl Watson: British Geological Survey (BGS)
- Gerold Diepolder: Bavarian Environment Agency (LFU)
- Grzegorz Ryzynski: Polish Geological Institute (PGI)
- Krzysztof Majer: PGI
- Martin Hansen: Geological Survey of Denmark (GEUS)
- Niels-Peter Jensen: I●GIS

With contributions from:

- David Entwisle (BGS), Ingelöv Eriksson (Oslo) and Susie Mielby (GEUS)

Priorities and hot topics

Over a period of 10 months the group discussed the issues surrounding geoscience data acquisition and management in urban environments by meeting for two dedicated workshops, numerous remote communications and through attending a number of collaborative meetings with other Sub-Urban project groups. Legislation and urban planning priorities differ significantly across Europe but there are a number of common priorities with regards to the types of datasets and technologies that could be used to enhance our understanding of urban subsurface properties and processes.

Priority datasets

The datasets which are of most value to a city will depend upon local environmental conditions, historical developments and future requirements. There are however a number of common priorities across most European cities and for the purpose of this report the needs of five cities were reviewed and recurring themes were identified. The cities considered were Oslo in Norway; Odense, Denmark; Glasgow, Scotland; London, England and Warsaw in Poland. The analysis was carried out by review project documentation and by contacting city partners that have recently requested subsurface data from geological survey organisations, we looked at what information and services they requested from the geoscience community in order to satisfy immediate development queries or inform decisions about longer term sustainability.

The following datasets featured prominently in the requirements identified by all 5 cities:

- Geotechnical properties (often in the form of borehole analysis)
- Groundwater data (modelled and observed)
- Tunnel locations (in three dimensions)
- Utilities (e.g. fresh and waste water piping)
- Pollution information (location, type, history, geochemical properties)
- Land use (historical, current and in some cases future plans)
- Surface water features

A selection of datasets were only requested by stakeholder in a few of the cities but anecdotally appear to be gaining significance across Europe, and beyond, therefore we expect them to become higher priority to many more cities in the near future:

- Live, and near live, environmental monitoring sensor data (helping researches to identify significant events in real time as well as model dynamic processes)
- Detailed information about anthropogenic deposits (ideally categorised using communally agreed standards)
- Geothermal /energy well locations and details
- Integrated 3D building and subsurface models (BIM)

Technological priorities

In addition to the identification of high priority datasets our analysis revealed a number of technology requirements that were repeatedly identified by city partners, these included:

- Decision Support Systems (DSS) that integrate all key datasets and models for a city in a single tool
- Efficient digitisation of analogue data, which is something that all cities would benefit from, traditionally this is a very labour intensive process so there is a big appetite for automated and semi-automated processes that reduce the costs of manual data conversion
- Distributed /federated data architectures that allow support the combining or related datasets from multiple sources, securely managed using appropriate authentication and permissions tools.
- Calculating and communicating uncertainty

The so called challenge of BIG data was raised by several stakeholders and working group members although the understanding of what this term meant varied significantly. In many cases, working with BIG data means being able to transfer and store of large data volumes whilst others were more concerned about the variability of data gathered from many different sources or the rapidly changing nature of temporal data (e.g. sensor data). Regardless of what stakeholders consider to be BIG data the amount of data available to decision makers is increasing and they need tools to help manage, describe and combine these datasets in meaningful ways.

1.3 Report structure

The WG2.2 Data Acquisition and Management group have investigated and documented four of the high priority topics, described in the following four chapters.

Chapter 2: Integrating urban datasets

Using the software tool GeoScene3D as an exemplar: How to capture, in a single model, multi scaled data covering the key sub urban datasets such as geology, anthropogenic deposits, infrastructure. Incorporate comments on the range of data source formats and ways in which the data can be structured and displayed.

Lead author: Niels-Peter Jensen

Chapter 3: From analogue to digital data

Using examples from the Polish Geological Institute: How to develop a set of procedures and systems that will enable the migration from paper and PDF documents towards well-structured datasets. Covering how the work was planned, systems developed and the quality assurance processes refined.

Lead author: Grzegorz Ryzynski

Chapter 4: Commercial data and public data centre services

Using the example of the geotechnical data format AGS and how the British Geological Survey are developing workflows and systems to enable data sharing between commercial organisations and public sector data centres for the benefit of the city of Glasgow, UK.

Lead Author: Carl Watson

Chapter 5: Managing permissions and roles

Using the experience of the GEUS distributed database systems this topic will describe the technical architecture and constraints which are required to administer a system that involved many users of different roles across a range of organisations throughout Denmark.

Lead Author: Martin Hansen

2: Integrating urban datasets

2.1 Introduction

Until now, many geological or geological- geotechnical models have been built primarily in the non-populated area outside the urban areas. Thus, many modelling tools supports this kind of modelling, and the data related to this. When moving to urban environments, a lot of other data matters, such as man-made structures, infrastructure, buildings, and the “geology” present may not be related to geological processes, but to a large degree the result of human activities.

The level of detail that the models should deal with is also quite different from the more regional (outside urban) models, in the way, that even quite small features might have a huge impact on what the model can predict. For instance, a relative thin coarse water bearing gravel layer, may give problems constructing a metro, or a tunnel, and have a serious impact on the economy in the construction project. Therefore the model software should be able to model these relative small features, and enable the modeller to maintain overview of the model and all the data related to it.

2.2 Good practices

Urban modelling differs from other modelling, in the sense that the area covered with the model is quite large, ranging from one kilometre, to several tens of kilometres. The data types are very inhomogeneous and that there are many of them, coming from many different sources adding complexity in the data setup for the model.

An important task is therefore to try to homogenize these data into a structure that is similar across the different datasets. An example of this, is to assign the same attribute types on fundamentally different datasets as the having an attribute describing geotechnical strengths on both road and a geological formation and a sewer section, so that these can be incorporated in a final model.

2.3 Workflows

Here, in brief, is a step by step outline of the current modelling workflow.

Step 1: Regional geology model

The basis of the model is an underlying regional geological model. The detail of this model is normally considerably lower than the detail in the urban model, as data is sparser. This is the basic geological setting of the area, incorporating regional geological knowledge.

Step 2: Digitizing at bottom of fill/anthropogenic layer

Next step is figuring out the bottom of the anthropogenic layer. This is done by going through available information, primarily boreholes and logs, but also other data, such as other known information from excavations old maps from before landfill etc.

Based on these data, a point is digitized, marking the border between the untouched geology and the man-influenced overburden. These digitized points are then interpolated into a surface that is used in the modelling process.

Step 3: Evaluating and digitize wells and other information creating a 3D point dataset

By examine the wells a geological or hydrogeological value can be determined and digitized as point information. A point is digitized in the borehole (or other kind of information) and a value is assigned to this point. This is done, preferably at uniform depth for example every 0.25 meter. The resulting points can be seen in Fig 1.

Step 4: 3D interpolation point dataset

A 3D interpolation of the point dataset follows. A search radius must be defined, and this will vary on the inhomogeneity anticipated in the model area. The interpolation results in a 3D voxel grid partly filled with values.

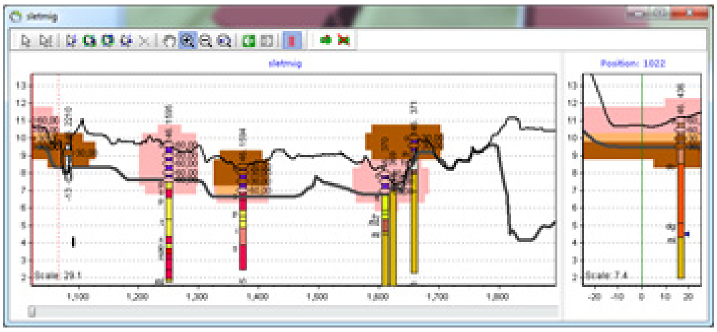


Fig 1: 3D interpolation of the data points digitized in Step 3.

Step 5: Fill the rest of the anthropogenic with a reasonable value

If no other data is present, the blank values of the model must be filled with a “best guess” value. The terrain and bottom of the anthropogenic layer is used as cut off layers, so that no values are set above terrain and below the bottom of the anthropogenic layer.

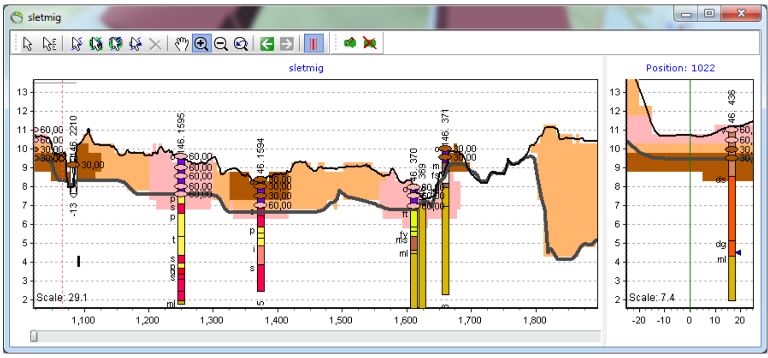


Fig 2: Infill with an anticipated value of the surrounding fill.

Step 6: Voxelation of infrastructure elements

The final steps includes assigning values from different datasets, especially infrastructure and building etc. ending up with a final model that includes all relevant datasets.

2D features, can be rendered in 3D according to an attribute, a building footprint, as an example, can be placed on the terrain, and extruded below terrain based on an attribute telling that basement depth is 4 meters.

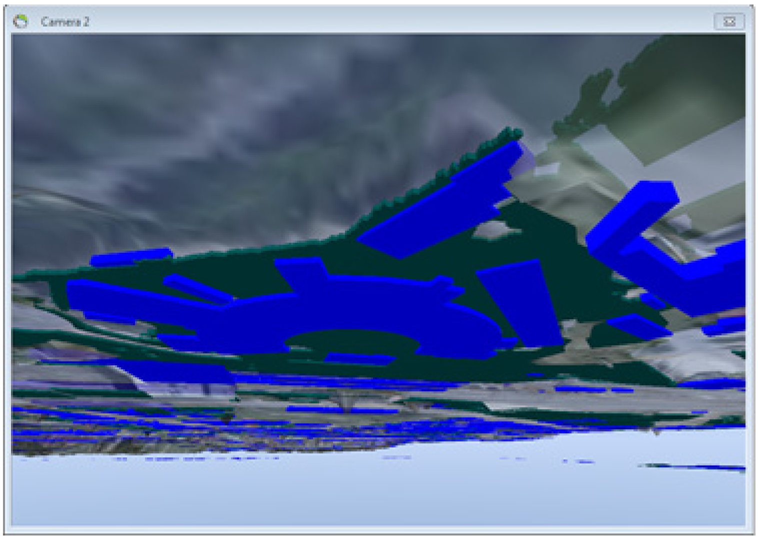


Fig 3: Assignment of infrastructure and buildings. Looking from below on basements and parking areas.

Voxels inside this basement will then get a value telling that indicating that is a basement, or, if the voxel model should be used as base for a flowmodel, a value representing that no flow can occur in this voxel.

This is done for all structures resulting in a final model, incorporating all available information. The uncertainty in the different steps can be estimated and recorded in an uncertainty attribute. It is obvious, the uncertainty should be smaller, where actual measurements or borehole information are present, as well, as where foundations or infrastructure elements are located, compared to areas where voxel values are a more or less qualified guess, as in step 5

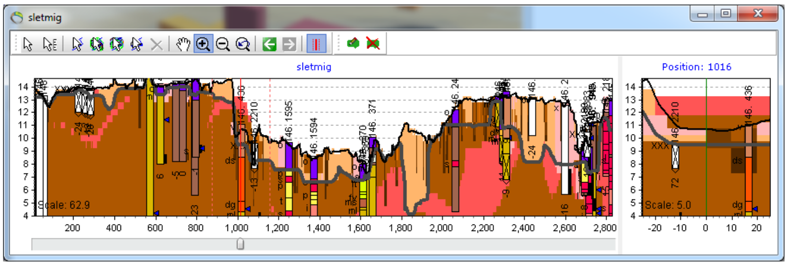


Fig 4: Cross section of the final model, with all data integrated

2.4 Examples

GeoScene3D is a 3D modelling tool, mainly developed to build geological modelling of groundwater, polluted areas and geotechnical projects. This software has now been revised to be used in urban environments and to build urban geological models and this report is built on experiences with developing and using this software.

A workflow has also be introduced, which has been applied in an urban modelling project in the city centre of Odense, a Danish town of 175.000 inhabitants, on the island Fyn.

For a detailed description of the application of GeoScene3D in Odense see Appendix 2.

2.5 Knowledge gaps

ID	Current State	Desired State	Gap Description	Gap Reason	Remedies
1	A common problem with 3D models is that a lot of effort is spent on their creation but updates are often hard to perform and track, this results in static/frozen models.	Versioned and audited 3D models which are easily updated to reflect new and updated data or understanding.	Most updates performed on the model tend to involve significant manual effort and there are only a few examples of truly 'live models' that actively evolve over time.	Urban models are often created for a general purpose and the route between "raw" data and model data is too long.	Develop 3D modelling software functionality to automate (and semi-automate) some of the steps involved in processing raw data. As regards audited versioned models, the database powered systems such as GiGa Systems GST or BGS Geological Object Store show promise.
2	Updating 3D models often requires a significant amount of time and money to manually convert new data into modelling software compliant formats and introduces the risk of errors due to human handling.	Data used in 3D models should be kept as close to their native formats as possible and automated workflows should perform conversions when needed.	Source data is often provided in formats that are not automatically supported by the modelling software or lack key information required by the modellers.	Source data has been created for a single purpose and that objective does not often include re-use within 3D software, e.g. attributes such as the thickness of road beds is often not present in the native data.	3D modelling software vendors could develop additional support for the high value native data formats that incorporate the logical steps currently performed during manual data conversion.
3	Many 3D models contain a snapshot view of the	Time plays a significant role in urban	Most elements within a 3D geological model	The majority of urban 3D models have been	Incorporate time varying observational data, such as sensor values, process

	subsurface environment without much, if any, time varying elements.	environments, as features can be replaced by other features over time. Therefore a more or less automated workflow that incorporates age of features and chooses the most recent features for assigning values in a voxel would be a good enhancement.	lack the concept of time and therefore time varying properties are not well represented.	developed to describe the current state of the subsurface rather than as an ongoing resource that is regularly updated with new data and understanding of processes.	model data such as predictions of time varying properties. And include the concept of time in structural elements of the model such as buildings and subsurface material whether that is anthropogenic or naturally deposited.
4	Many data, especially geotechnical data, only exist in paper formats. Some are scanned, few are georeferenced. Full digital versions of geotechnical logs are often quite expensive to produce (Not to mention the task of acquiring the data).	Ideally the modeller would have access to digital log data, such as geotechnical data in AGS format (see chapter 4). When only analogue data is available modellers should be able to digitise efficiently within the modelling software.	The process of digitization typically involves three steps: 1: Identify paper records 2: Scan images 3: Manually digitize layers and log curves. The last step can be especially time consuming.	Historically, the majority of borehole data is held and supplied in a wide range of analogue formats. A multitude of methods have been developed to process such data.	Develop modelling software functionality to incorporate digital formats, where available, or digitize on scanned images of analogue data which has undergone basic geographical registration.

3: From analogue to digital data

3.1 Introduction

Well-structured and interoperable databases or data stores are becoming a more and more common way of storing geological and geotechnical data, however, a lot of geological information remains in paper format within archives of private companies and public organisations. These archives consist of non-editable data, which is kept only for reference, as a hard copies of documents or as scans in pdf/tiff format. The digitisation of this data from analogue to digital format allows processing this data and generating many useful maps/models and analyses necessary for municipalities and geological surveys for managing the subsurface space of city areas.

Non editable geological data cannot be used for quick analyses. However, municipalities often need quick access to geological data for purposes of spatial planning, management of city subsurface space and for crisis management. To allow geological data processing within well-structures databases, efficient procedures for digitising analogue data need to be developed and implemented.

3.2 Good practices

The organisations which hold archives of geological and geotechnical data in paper formats are mostly private companies and their data comes from their commercial projects. Other archives of non-editable data are often in public companies, especially those which manage roads railways and metro or manage underground infrastructure, like water and sewage system, gas piping and telecommunication and electric network. Geological information resources are often digitized for certain projects (mostly at site scale, on selected areas or new construction projects). There are also situations, that the archival data is shared with geological surveys as a good practice of data exchange. This is mostly for city scale local authorities, public organisations can then use the geological databases run by geological surveys.

With development of procedures and workflows for migration form paper and pdf documents to geological databases several problems and challenges should be addressed.

- Legal regulations / ownership rights.
- Full identification of archival data sources.
- Temporal repository for external archives data.
- Verification of archival data quality.
- Data management interface.
- Harmonized data using (international or industry) standards and controlled vocabularies

- Managing spatial coordinate specifications in a city relevant scale

3.3 Workflows

The following steps form the core workflow developed by the Polish Geological Institute for a project (BDGI) described in Appendix 3 that involved digitising analogue borehole data:

- Data coming within legal framework is scanned as a part of general procedures whereas that coming from external sources is registered and have their document IDs added to archival documents database.
- Scanned versions of paper documents are temporarily held on server and are processed to the Borehole Data Management System
- When the both types of data are scanned, all necessary ID codes needed to maintain the link to original data source documents are attributed the data is processed to the Data Management System.
- Profiles of boreholes are introduced in database with the help of special creators/wizards. Thanks to that, possible errors are minimized during entering data to database. The application creator allows typing data coherent with the relevant dictionaries.
    - Dictionaries used for both national and ISO standards as well as geological dictionaries used in local geological survey.
    - Dictionaries are controlled and managed by administrators, which significantly improves the quality of data entry.
- Localization of boreholes is a problem of two types, the first concerns the quality and accuracy of information about the position on the different- scale maps (or the quality of written location data in the borehole sheet) and the second is related to the use of different coordinate systems.
    - If the sources documentation has information about localization, then it is put directly to the database after verification.
    - Otherwise, each map with documentation points is calibrated / geo-referenced.
- At this stage entered archival data by many users have to be verified. It is very important because you should be sure that quality of digitized boreholes' profile and their coordinates are correct.
- The next step is to automate any transformations of data from one database to other more specialized databases such as spatially enabled databases that can auto generate spatial layers for GIS analysis. Thanks of that you can use information included in database for creating maps or for quick solving critical issues/ resolution of crises.
- Last stage is conversion to a Central Geological Database (CBDG) which is main database which provides access to internal and public users.

One of the main problems and tasks within the BDGI project is putting the analogue geological data into the database. Due to large amount of data to be digitized (more than 60 000 boreholes) the dedicated Data Management System – Geostar7 BDGI was developed and workflow of geological and geotechnical data migration to database was developed (see figure 5).

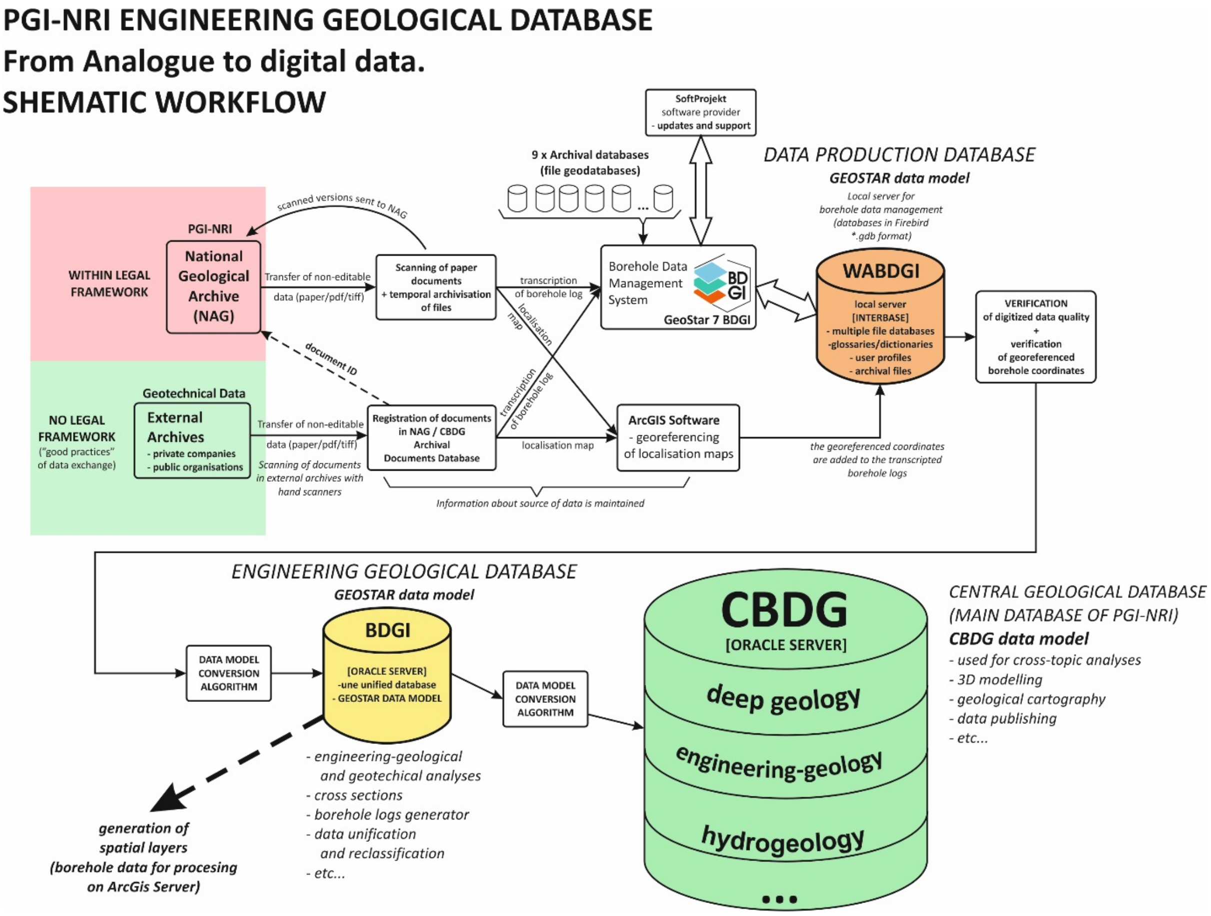
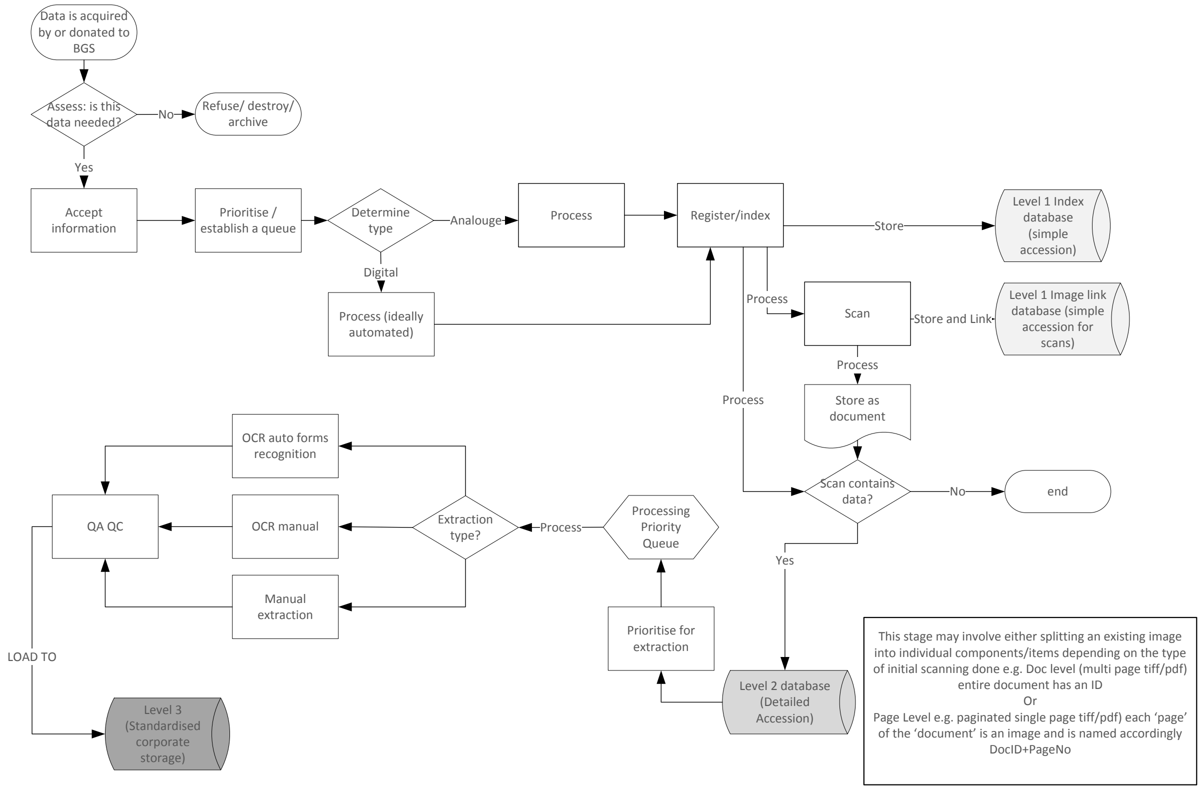

PGI-NRI ENGINEERING GEOLOGICAL DATABASE
From Analogue to digital data.
SHEMATIC WORKFLOW
SoftProjekt
software provider
- updates and support
9 x Archival databases
(file geodatabases)
DATA PRODUCTION DATABASE
GEOSTAR data model
Local server for
borehole data management
(databases in Firebird
*.gdb format)
scanned versions sent to NAG
PGI-NRI
WITHIN LEGAL FRAMEWORK
National Geological Archive (NAG)
Transfer of non-editable data (paper/pdf/tiff)
Scanning of paper documents + temporal archivisation of files
transcription of borehole log
Borehole Data Management System
BD GI
GeoStar 7 BDGI
WABDGI
local server [INTERBASE]
- multiple file databases
-glossaries/dictionaries
- user profiles
- archival files
VERIFICATION of digitized data quality + verification of georeferenced borehole coordinates
localisation map
document ID
transcription of borehole log
Geotechnical Data
NO LEGAL FRAMEWORK
("good practices" of data exchange)
External Archives
- private companies
- public organisations
Transfer of non-editable data (paper/pdf/tiff)
Scanning of documents in external archives with hand scanners
Registration of documents in NAG / CBDG Archival Documents Database
localisation map
ArcGIS Software
- georeferencing of localisation maps
the georeferenced coordinates are added to the transcripted borehole logs
Information about source of data is maintained
ENGINEERING GEOLOGICAL DATABASE
GEOSTAR data model
DATA MODEL CONVERSION ALGORITHM
BDGI
[ORACLE SERVER]
-une unified database
- GEOSTAR DATA MODEL
DATA MODEL CONVERSION ALGORITHM
CBDG
[ORACLE SERVER]
deep geology
engineering-geology
hydrogeology
...
CENTRAL GEOLOGICAL DATABASE
(MAIN DATABASE OF PGI-NRI)
CBDG data model
- used for cross-topic analyses
- 3D modelling
- geological cartography
- data publishing
- etc...
- engineering-geological and geotechical analyses
- cross sections
- borehole logs generator
- data unification and reclassification
- etc...
generation of spatial layers
(borehole data for procesing on ArcGis Server)

Fig. 5. Schematic workflow for PGI – NRI Engineering Geological Database.

Figure 6 shows an alternative, generic, data acquisition and processing workflow based upon the experiences of the British Geological Survey. In this workflow the analogue records are prioritised according to their apparent value and data is entered into a series of database structures. Initially records are entered into a Level 1 database containing simple accession details, links to the scanned images and physical records. Level 2, detailed accessions, involves splitting up large accessions into sensible component parts, capturing spatial and other key information needed to help with subsequent discovery and prioritisation. Level 3 is the most manually intensive stage and involves full digitisation and often requires data to be extracted from specialist formats into the standardised data structures used by the BGS corporate systems.



Data is acquired by or donated to BGS
Assess: is this data needed?
No
Refuse/ destroy/ archive
Yes
Accept information
Prioritise / establish a queue
Determine type
Analouge
Process
Register/index
Store
Level 1 Index database (simple accession)
Digital
Process (ideally automated)
Process
Scan
Store and Link
Level 1 Image link database (simple accession for scans)
Process
Store as document
Process
Scan contains data?
No
end
Yes
OCR auto forms recognition
QA QC
OCR manual
Extraction type?
Process
Processing Priority Queue
Manual extraction
Prioritise for extraction
LOAD TO
Level 2 database (Detailed Accession)
Level 3 (Standardised corporate storage)
This stage may involve either splitting an existing image into individual components/items depending on the type of initial scanning done e.g. Doc level (multi page tiff/pdf) entire document has an ID
Or
Page Level e.g. paginated single page tiff/pdf) each 'page' of the 'document' is an image and is named accordingly DocID+PageNo

Figure 6 High level data acquisition workflow (OCR: Optical Character Recognition. QA QC: Quality assurance and quality checking).

3.4 Examples

To describe the process of migration of analogue data to geological database, the example form Polish Geological Survey Project "Engineering Geological Database" (BDGI) is used in Appendix 3. The BDGI project (duration 2013 -2016) is aimed at unification of 9 separate databases of largest agglomerations of Poland (total more than 260 000 boreholes) and creation of one unified database BDGI compatible with Central Geological Database of PGI-NRI with new extra 66 000 boreholes added to the database till end of the project.

The data which is gathered into BDGI database comes mainly from two sources. One is National Geological Archive (NAG), held by Polish Geological Institute, where all boreholes are kept for reference in paper or scanned (pdf/tiff) and they are archived within legal framework of Geological Law. The second source of data is from external archives of public organisations and private companies and covers mostly geotechnical data. This data is not covered by legal regulations of Geological Law, so therefore is not archived within National Geological Archive. The two sources of data require different procedures of migration into the database.

3.5 Knowledge gaps

ID	Current State	Desired State	Gap Description	Gap Reason	Remedies
1	Most digitisation workflows used by Sub-Urban stakeholders are very manual, slow processes, prone to human error.	Automation or at least semi-automation of the routine digitisation tasks.	Many digitisation workflows involve indexing, metadata capture, scanning, transcription and verification. Some also include standardisation of the data. Some of these steps, like verification, should probably remain manual but many of the other steps are very repetitive in nature and could potentially be carried out using appropriate technology.	Technologies such as Optical Character Recognition (OCR) have not proved to be effective within institutions such as national geological surveys, yet. The formats of the data sources highly variable and simply storing digital versions of text in an unstructured form has not been desirable.	Potential, non-mutually exclusive, remedies include: 1. Continue to monitor improvements made in OCR and related technologies 2. Investigate the potential of 'Data Science' / 'Data Analytics' techniques for deriving meaning from unstructured datasets 3. Encourage the use of digital formats which could remove the need for digitisation 4. If digital formats are not possible, encourage the use of proscribed formats (forms) which would be easier for OCR software to understand
2	Valuable analogue data is not being identified or digitised efficiently and the backlogs of analogue yet to be processed are growing.	Develop triage workflows to prioritise particularly valuable records within backlogs of analogue data.	Many institutions attempt to process all incoming analogue data as it comes in and others digitise a mass of records on a project basis. In both these cases the approach has been to process the data from original format to fully digital, with minimal prioritisation taking place.	It is common for digitisation to be carried out in a very linear workflow, acquire data, index it and digitise from start to finish using human eyes and knowledge to identify the meaning of the analogue information.	Focus on digitising only crucial information, for example - the triage step should focus on capturing index level metadata such as identifying information, locality, high level descriptions and conditions of deposit (e.g. access restrictions, confidentiality and ownership)

4: Commercial data and public data centre services

4.1 Introduction

Efficient management of Europe's urban environments requires an efficient means of communicating existing information amongst stakeholders and effective systems that support the capture and storage of newly created data. The production and management of data can be expensive and all too frequently the information contained within the data is used only within the project it was produced for. Recycling this data for use by both public and private organisations could provide the basis of future desk studies, ground models and resources for planning and regeneration. However, the data must be managed and in a form that is readily available.

Whilst advances in technology mean more of the data, which is important to city management, is increasingly digital there remains a large body of analogue data sources that are expensive to convert into usable digital formats for current and future projects. Financial constraints on public bodies have led to the need to increasingly automate the digitisation of analogue datasets rather than rely on manual checking and conversion.

4.2 Good practices

There are a number of technical challenges which need to be overcome by communities looking to integrate the data and information gathered by a range of organisations regardless of the discipline involved, good practices include implementation of:

- Standard exchange formats
- Automated and semi-automated systems to deliver:
  - Data validation (against the agreed data & exchange formats)
  - Data verification (to ensure that the data is valid and valuable)
- Centralised/communal data storage (includes standardisation of data structures and dictionaries used)
- Tools that enable efficient data discovery, data visualisation and data access

4.3 Workflows

The data flow shown in Fig. 7, below, was developed for the Glasgow case study described in Appendix 4. It centres around an online website through which data donors are authenticated and files submitted, once a file has been submitted it triggers an automated validation and ant-virus check process, if valid, files are transferred to internal BGS servers where the data verification and data storage processes take place. This information workflow was initially designed in 2013 and the solution was launched early 2014 and closely resembles the draft design.

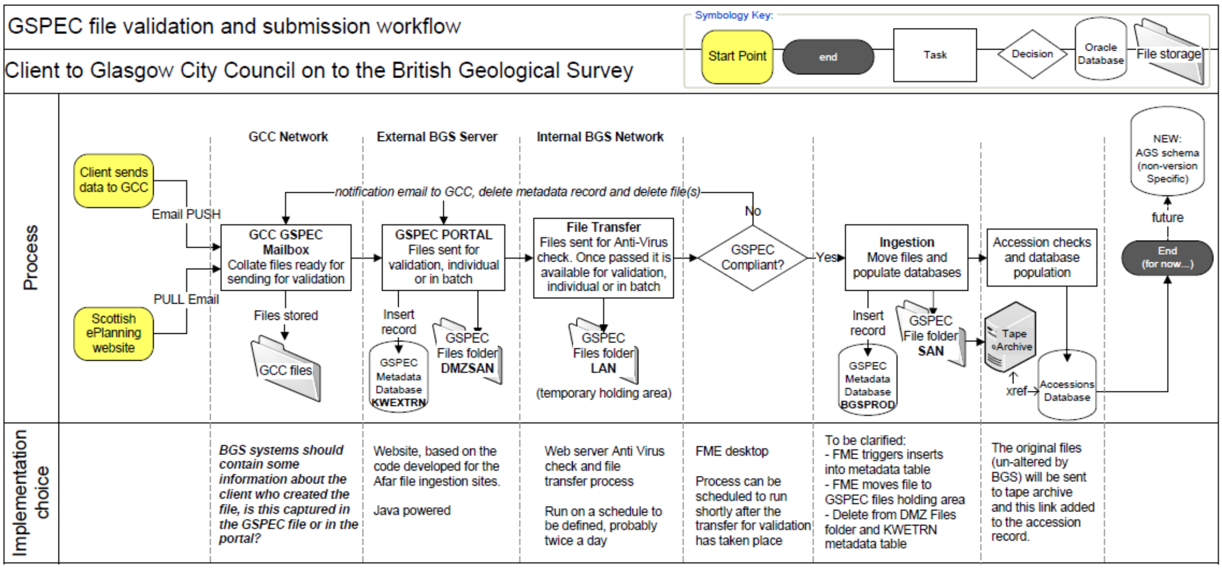

GSPEC file validation and submission workflow
Client to Glasgow City Council on to the British Geological Survey
Symbology Key:
Start Point
end
Task
Decision
Oracle Database
File storage
Process
Implementation choice
GCC Network
External BGS Server
Internal BGS Network
Client sends data to GCC
Email PUSH
PULL Email
Scottish ePlanning website
notification email to GCC, delete metadata record and delete file(s)
GCC GSPEC Mailbox
Collate files ready for sending for validation
Files stored
GCC files
GSPEC PORTAL
Files sent for validation, individual or in batch
Insert record
GSPEC Metadata Database
KWEXTRN
GSPEC Files folder
DMZSAN
File Transfer
Files sent for Anti-Virus check. Once passed it is available for validation, individual or in batch
GSPEC Files folder
LAN
(temporary holding area)
No
GSPEC Compliant?
Yes
Ingestion
Move files and populate databases
Insert record
GSPEC Metadata Database
BGSPROD
GSPEC File folder
SAN
Tape Archive
xref
Accession checks and database population
Accessions Database
NEW: AGS schema (non-version Specific)
future
End (for now...)
BGS systems should contain some information about the client who created the file, is this captured in the GSPEC file or in the portal?
Website, based on the code developed for the Afar file ingestion sites.
Java powered
Web server Anti Virus check and file transfer process
Run on a schedule to be defined, probably twice a day
FME desktop
Process can be scheduled to run shortly after the transfer for validation has taken place
To be clarified:
- FME triggers inserts into metadata table
- FME moves file to GSPEC files holding area
- Delete from DMZ Files folder and KWETRN metadata table
The original files (un-altered by BGS) will be sent to tape archive and this link added to the accession record.

Fig. 7: Draft workflow for the ASK Network data acquisition workflow

4.4 Examples

Appendix 4 describes the experiences of Glasgow City Council and the British Geological Survey in developing the data management workflows required to share geotechnical data between all stakeholders in the Glasgow / Clyde urban area brought together as part of the Accessing Subsurface Knowledge (ASK) Network. In order to deliver the technical aspects of the ASK Network vision required the development of solutions to all of the key issues raised in the previous section, namely:

- Standard data exchange format
- Data validation
- Assisted data verification
- Communal data store
- Data discovery
- Data visualisation
- Data access

In addition to these general requirements Glasgow City Council required a GCC branded web interface that allowed them to provide a file submission and file validation process which would trigger emails to confirm acceptance or a failure report if invalid. The BGS also required the solution to be secure, restrict submission functionality to authorised users only and support the capture of appropriate discovery and data accession metadata.

4.5 Knowledge gaps

ID	Current State	Desired State	Gap Description	Gap Reason	Remedies
1	Standard exchange formats for geotechnical data in use in a small number of cities	Common standards used across all dataset themes identified as high priority by city partners	Many cities that could benefit from AGS for geotechnical data are not using it.	Some cities are simply not aware of the standard, whilst others may consider it an unnecessary expense.	Provide free and open case studies which illustrate the cost-benefit of implementing such a standard and provide guidance for interested parties as part of WG3 toolkit.
2	Most Sub-Urban community stakeholders store and exchange data in a wide range of bespoke and proprietary formats	All Sub-Urban community stakeholders have access to and make appropriate use of standard exchange formats for the data which are seen as high priority by city partners	Some standards for data storage and data exchange formats exist for groundwater data, tunnels, utilities, pollution data, land use and surface features .	The standards which do exist have not become common practice, yet. This is partly due to a lack of evidence justifying the cost of adopting such standards.	By highlighting the need for greater collaboration and documenting potential efficiency gains could provide the justification for greater adoption of standards.
3	Automated validation of AGS data is performed by tools develop by commercial software vendors or the BGS	A free validation tool which is developed and maintained by the AGS standards authority or wider community	Commercial companies have developed tools which interpret the standard differently. The BGS have also developed a validation tool but it is not as robust as it could be & is not available externally	The AGS standards committee cannot or will not fund the development of such a tool	Either: i. AGS committee to develop such a tool, or ii. AGS committee could formally approve a tool developed by another body
4	Many of the GSOs and other community data centres involved in Sub-Urban activities ingest their data in labour intensive processes.	Seamless integration of distributed data generation and management data repositories.	Analogue data donations require manual processing and there is no wide spread automated validation and verification tools for digital data.	Many data donations are received in analogue formats or if digital the systems used are owned by different organisations leading to data silos.	Data centres and commercial software vendors should work together to develop data exchange interfaces such as the development of APIs to integrate BGS data with geotechnical software HoleBase SI.

5: Managing permissions and roles

5.1 Introduction

Sharing of information between different public units can seriously increase the amount of data available to each of the units, however, sharing of data between different levels of administration as well between different administrative units does not seem to be very common. This is often due to a history of data collection by different units that have been focusing on different parts of the data and have established different local data models making data sharing difficult.

In some countries the legislation also hinders the sharing of data as data owners are often the organisation generating the data rather than the one ordering the data. In this chapter the Danish model, where part of the data sharing is enforced by legislation and others through voluntary activities, is described as an example of good practice, for more details please see Appendix 5.

The solution described involves the development of a system where a central database is made available through services allowing all stakeholders to maintain their part of the system. A system like this requires a clear definition of responsibilities and an effective administrator to manage permissions and roles.

5.2 Good practices

In order to make a communal system truly accessible and usable by all parties it is necessary to support an architecture that is sympathetic to local organisation requirements. The Danish system supports this by making data available through SOAP web services that allow constructions of systems targeting specific user tasks.

A system using a central database, as well as local databases on a data collection, where different parts of data are managed by different stakeholders is strongly discouraged as setting up a two way synchronisation of data in a complex database is very difficult. In cases where a local database is required this can be done by downloading a subset from the central database or maintaining a read only copy of data locally.

In addition, if the creation of users and management of their privileges is maintained locally this removes a very significant overhead from the organisation responsible for the management of the system whilst empowering local users.

5.3 Examples

Appendix 5 describes the key aspects of the Danish Environmental Portal system, used by the municipalities within Denmark. It covers the technical architecture and constraints which are required to administer a system that involves many users of different roles across a range of organisations and administrative levels.

5.4 Knowledge gaps

It is not common for geological surveys, city authorities, utility companies and commercial enterprises to develop or use common information management systems. There are a few exceptions, such as the Danish example, but in general organisations tend to store their own data and retain control over the administration of related systems.

The use of cloud technologies is growing and there are a growing number of organisations who are opening up, internally held, data through web services. There is a very good chance that the integrated data systems which Sub-Urban stakeholders will require in the future will be powered by the loose coupling of web services from a number of organisations. Many of the web services developed by the Sub-Urban community are un-secured read-only mechanisms for sharing public data, few have attempted to develop services that authorise and authenticate data access and data editing functions. The use of such uncontrolled public services has grown rapidly and the rate of growth seems to be increasing, this is starting to impact on the systems which power these services and results in the need for new rules to regulate their use.

Going forward we will need to develop secure web services that support the definition of rights and responsibilities based upon legislation and commercial considerations as well as ensure data integrity, i.e. messages must remain unaltered in transit.

All of these issues are solvable through a variety of technologies, but it is not clear who should be responsible for ensuring the right technology is used and that interfaces between each stakeholder group is appropriately maintained.

References

Open BIM http://www.designingbuildings.co.uk/wiki/Open_BIM Chadwick, N Pickles, A and Sekulski. E. 20016. Data transfer and the Practical Application of Geotechnical Databases. GeoCongress 2006, 1-6. doi: 10.1061/40803(187)110

Bland, J, Walthall and Toll, D. The development and governance of AGS Format for geotechnical Data. Proceedings of the 2nd International Conference (ICITG). Information Technology In Geo-engineering, editors D G Toll, H Zhu, A Osman, W Coobs, X Li and M Rousainia. 67 to 74. IOS Press BV, Netherlands.

Campbell, Seumas, and Helen Bonsor. "The ASK Network: Glasgow Specification of Data Capture." (2013). http://nora.nerc.ac.uk/506350/

Kingdon, Andrew, Martin Nayembil, Keith Holmes, and Graham Smith. "PropBase QueryLayer: a single portal to UK physical property databases." (2010): 1-2.

Self, Suzanne, David Entwisle, and Kevin Northmore. "The structure and operation of the BGS National Geotechnical Properties Database. Version 2." (2012). http://nora.nerc.ac.uk/20815/

Power, C M, Patterson, D A, Rudrum, D M and Wright, D J. 2012. Geotechnical asset management for the UK Highways Agency. In Earthworks in Europe, editor T Radford, Engineering Group of the Geological Society Special Publication 26, 33-39.

Walthall, S and Palmer M. 2006 The development, implementation and future of the AGS data formats for the transfer of Geotechnical and Geoenvironmental data by electronic means. GeoCongress 2006: pp 1-4 doi: 10.1061/40803(187)109

Berg, Richard C.; Mathers, Stephen J.; Kessler, Holger; Keefer, Donald A., eds., 2011. Synopsis of current three-dimensional geological mapping and modeling in Geological Survey organizations. Ilinois State Geological Survey Circular, 578. 104, pp. http://nora.nerc.ac.uk/17095/

Appendix 1: High level recommendations

In summary, the WG2.2 Data Acquisition and Management group have identified the following key recommendations to develop efficient and effective data management systems and workflows.

- Clarify unclear legislation related to data acquisition and management policies as soon as possible
- Adopt standard naming conventions and use of controlled glossaries
- Develop data validation tools which are independent of software vendors
- Maximise use of open data discovery and data access platforms, with low financial and security costs
- More metadata is needed, especially within the commercial stakeholders in urban developments. Metadata should encompass data discovery, how to use the data, tailored to each audience and finally it should capture terms and conditions of use.

There is a wide held belief amongst those who have adopted these recommendations that the development costs are outweighed by the benefits, however, there is a lack of hard evidence to support this belief.

Remaining priority issues

There are number of topics that the WG2.2 Data Acquisition and Management group could not investigate in detail yet recognise as relevant areas that are worthy of mentioning, they include:

- Lack of standard terms for manmade deposits
- Coordinate reference systems at city wide scale
  - Transforming up from site specific or down from region, national or international
  - Deriving coordinates from names, maps and other ‘relative’ location descriptions
- Decision Support Systems tailored for decision makers
  - GIS style tools, GIS + 3D synthetic borehole and cross sections such as the BGS porcupine tool used in Glasgow

More scenario based tools such as the Decision Support Environment (DSE) which was developed by Accenture, Macomi and AIT as part of Transform smart cities program (http://urbantransform.eu/decisionsupportenvironment/), the DSE is designed to supports cities by providing quantitative insights on possible sustainability measures which users can adjust to see the impact of such changes.

Appendix 2: Integrating urban datasets

Introduction

Until now, many geological or geological- geotechnical models have been built primarily in the non-populated area outside the urban areas. Thus, many modelling tools supports this kind of modelling, and the data related to this, mainly borehole data, different kinds of geophysics like well logs, and, predominant in the oil industry, seismic data, as well as electrical and electromagnetic data for models related to groundwater investigations and mineral exploration.

When moving to urban environments, the modellers challenges the fact that a lot of other data matters, such as man-made structures, infrastructure, houses, and the “geology” present, are not related to geological processes, but to a large degree the result of human activities.

The model detail that the models should deal with is also quite different from the more regional models, in the way, that even quite small features might have a huge impact on what the model can predict. For instance, a relative thin coarse water bearing gravel layer, may give problems constructing a metro, or a tunnel, and have a serious impact on the economy in the construction project. Therefore the model software should be able to model these relative small features, and enable the modeller to maintain overview of the model and all the data related to it.

This report will focus on challenges in building such a model tool and some of the specific developments done to make this tool productive.

GeoScene3D is a 3D modelling tool, mainly developed to build geological modelling of groundwater, polluted areas and geotechnical projects. This software has now been revised to be used in urban environments and to build urban geological models and this report is built on experiences with developing and using this software.

A workflow will also be introduced, which has been applied in an urban modelling project in the city centre of Odense, a Danish town of 175.000 inhabitants, on the island Fyn.

Finally, some further developments are suggested.

Choosing the type of model

In the process of developing and using the tool to create models, a lot of requirements both on the modelling tool, and for the data to be used for modelling have been revealed.

Variety and quality is very different from data being used in other aspects of geological modelling.

The following table compares different model situations:

	Oil	Mining	Groundwater	Polluted sites	Geotechnical	Geothermal	GeoEnergy	Urban	Colour coding		
Scale of work involved	1	2	2	3	3	2	3	2	*Large*		*Small*
# Data types		2	2	2	2	3	3	1	*Many*		*Few*
Area size	1	2	2	3	2	2	3	1	*Large*		*Small*
Data inhomogeneity	3	3	2	2	3	3	3	1	*High*		*Low*
Update demands	2	3	2	3	3	3	3	1	*High*		*Low*
									Easy		*Hard*

Table App2-1: Data requirements for Geological models. Comparisons between industries building 3D geological models. The precise nature of specific models varies greatly so this is from an average point of view

What is clear is that urban modelling **differs from other** modelling, in the sense that the area covered with the model is quite large, and the data types are very inhomogeneous and that there are many of them, coming from many different sources.

To create a valid urban model, as well as for other models, it is basically a matter of having enough data, both in terms of quality and quantity.

There are several different types of model types that can be used for an urban model, a few mentioned here:

Layered models

Or layer cake models, consisting of surfaces of 2D grids with an elevation. Surfaces in the model should not cross each other. These models are normally well suited for sedimentary areas.

GeoScene3D in the layered model mode is an example of this

TIN based models

Here the modelling process is manipulating several TIN (Triangulated Irregular Network) There are several software that can be used for this. More complex structures can be built with this data type, as faults and over shooting blocks etc.

Examples are GSI3D and GoCAD.

Voxel models

These models consist of square boxes in which a value can be assigned. Every voxel of a voxel model can have a distinct value. By using voxels, it is possible to model abrupt changes in model properties. The challenge is the size of the models, i.e. a detailed model can consist of millions of voxels, which can be difficult to handle in in terms of memory and hardware demands.

GeoScene3D in the Voxel mode is an example, as well as Petrel.

As urban areas, from a geological point of view, are very inhomogeneous, the practical solution is to use voxels as the resulting modelling data set. However this voxel model can be combined with a layered model that can act as the surrounding limits of the model, for example an underlying bedrock.

Building urban models in GeoScene3D

To explain the challenges met, an explanation of the way models are built in GeoScene3D is necessary.

Voxels has normally, for practical reasons, the same dimensions in X and Y direction and another in the Z direction throughout the model.

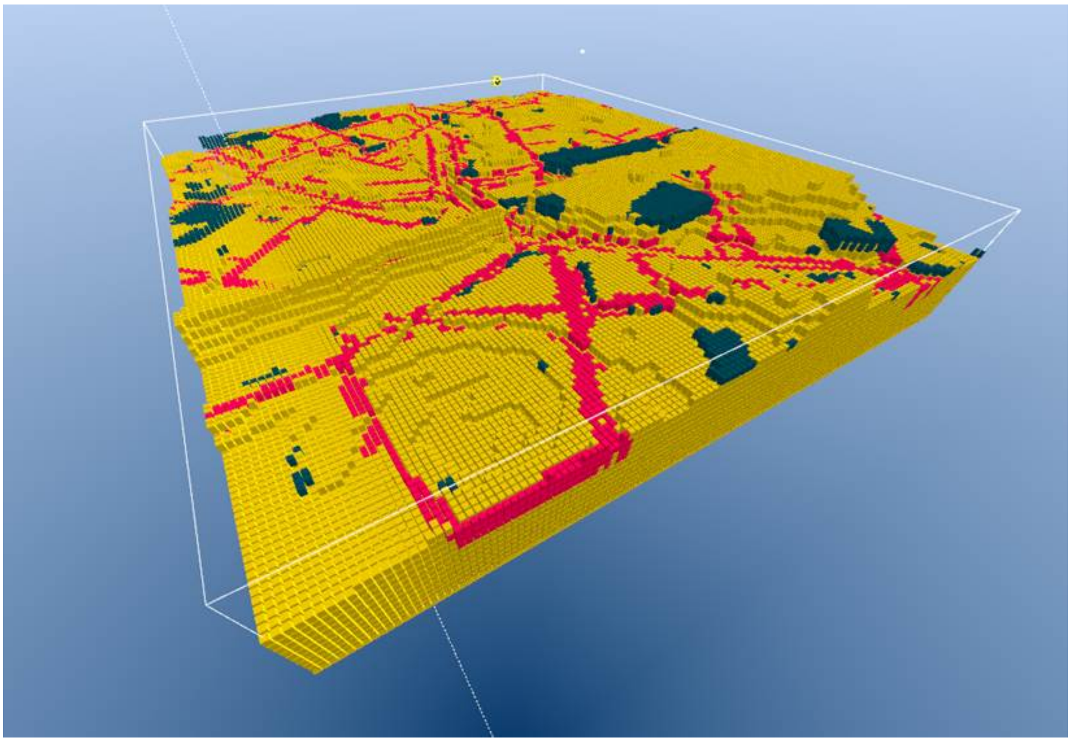


Fig App2-1: Urban voxel model, where different features has been assigned into the voxel dataset forming internal structures.

So the goal of the modelling process is to create a voxel dataset based on a valid geological basic setting and supplemented with manmade “geology” based on information from many different data sources.

One task, as an example, is to assign specific values from a sewer tube line network, which is show in the following figure.

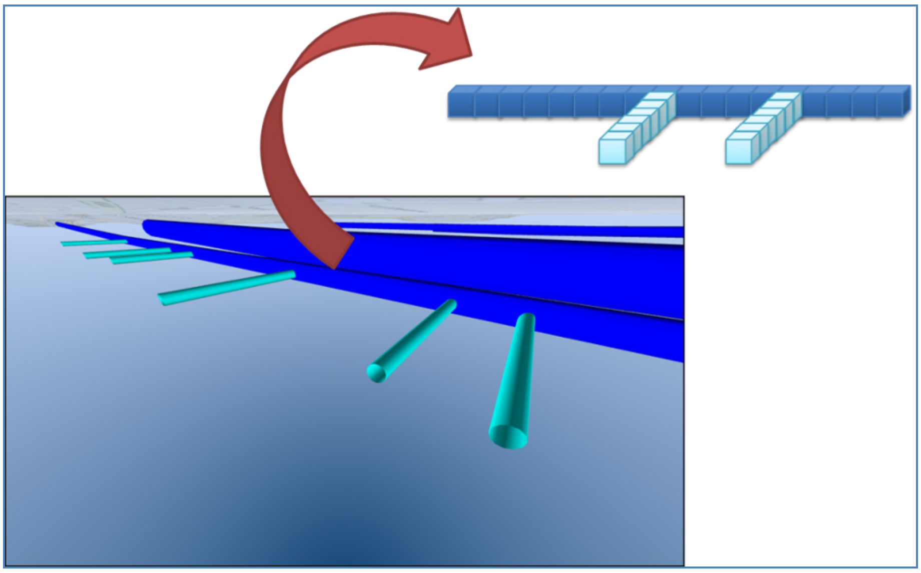


Fig App2-2: Modelling tubes in voxels.

The process should lead to a final model incorporating the different data.

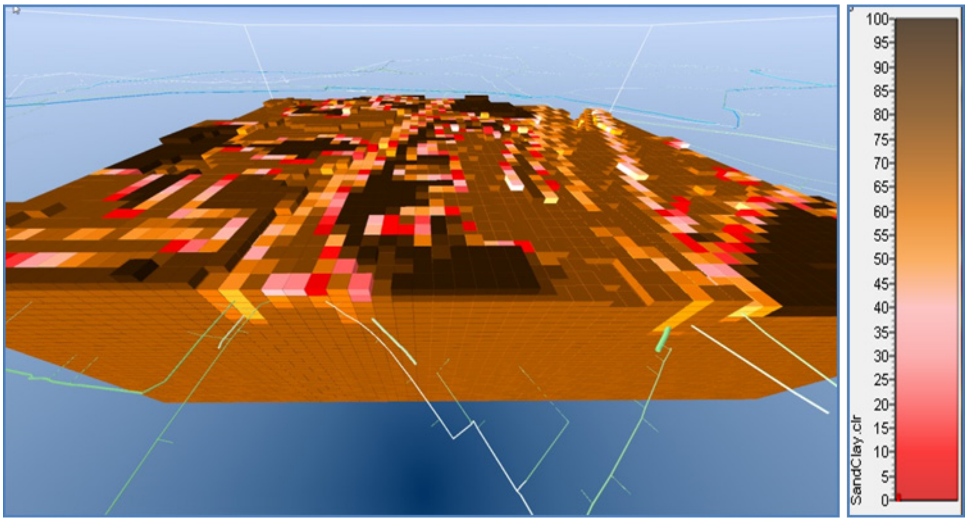


Fig App2-3: Example of urban voxel model.

As the voxels normally never can be small enough some kind of blend or mix is required, because each voxel can be influenced by more than one feature. This can be accomplished by calculating a value, based on the existing value in the voxel blended with the value from the feature, taking the volume of the feature into account.

As an extra attribute, the uncertainty of the model could be estimated.

The modelling workflow

Here, in brief, is a step by step outline of the current modelling workflow.

Step 1: Regional geology model

The basis of the model is an underlying regional geological model. In the case of GeoScene3D, this is normally a layer based model. The detail of this model is normally a lot lower than the detail in the urban model, as data is sparser, than will be the case in the urban model. But this is the basic geological setting of the area, incorporating regional geological knowledge.

Step 2: Digitizing at bottom of fill/anthropogenic layer

Next step is figuring out the bottom of the anthropogenic layer. This is done by going through available information, primarily boreholes and logs, but also other data, such as other known information from excavations old maps from before landfill etc.

Based on these data, a point is digitized, marking the border between the untouched geology and the man-influenced overburden. These digitized points are then interpolated into a surface that is used in the modelling process.

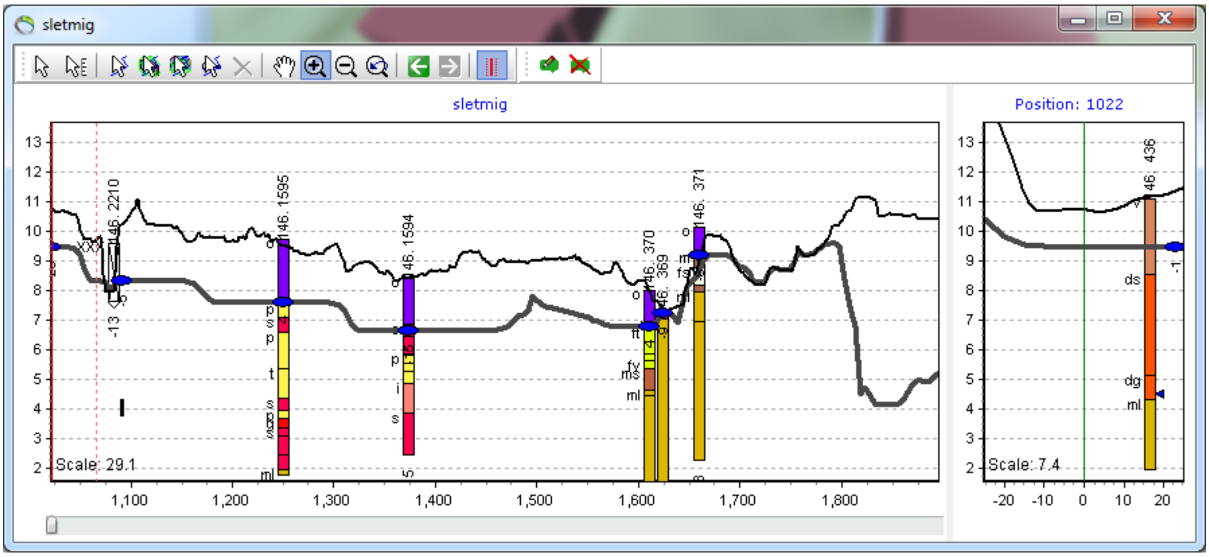


Fig App2-4: A cross section of the model, showing well information, terrain (upper surface, thin black line) and depth to base of anthropogenic layer (lower surface, thick grey line). Remark: The small section on the right, is a perpendicular cross section of the main cross section to the left.

Step 3: Evaluating and digitize wells and other information creating a 3D point dataset

By examine the wells a geological or hydrogeological value can be determined and digitized as point information. A point is digitized in the borehole (or other kind of information) and a value is assigned to this point. This is done, preferably at uniform depth for example every 0.25 meter. The resulting points can be seen in Fig App2-5.

Step 4: 3D interpolation point dataset

A 3D interpolation of the point dataset follows. A search radius must be defined, and this will vary on the inhomogeneity anticipated in the model area. The interpolation results in a 3D voxel grid partly filled with values.

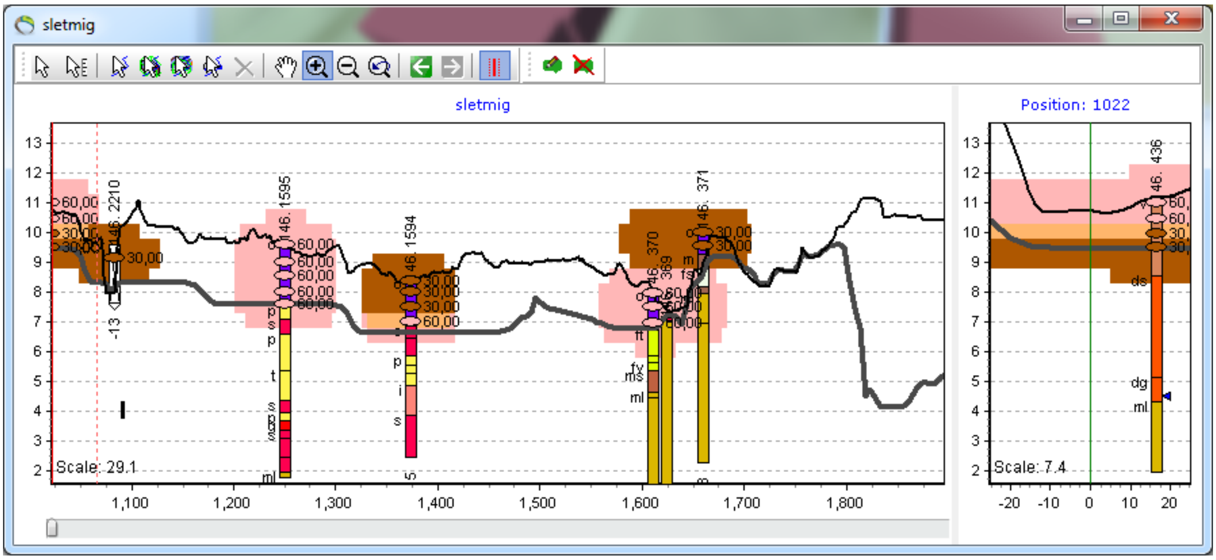


Fig App2-5: 3D interpolation of the data points digitized in Step 3.

5: Fill the rest of the anthropogenic with a reasonable value

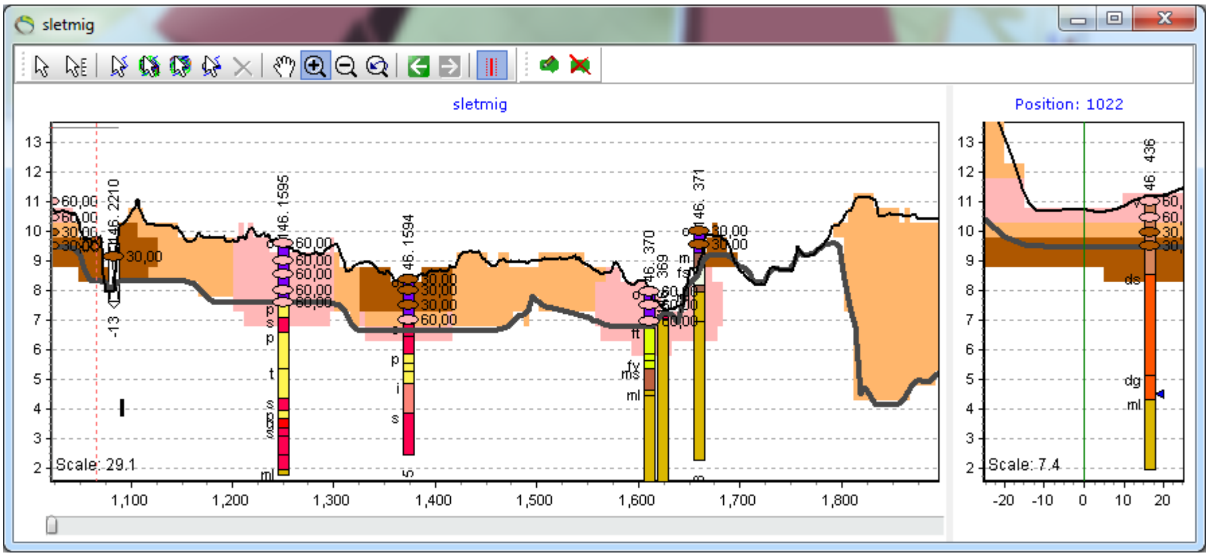


Fig App2-6: Infill with an anticipated value of the surrounding fill.

If no other data is present, the blank values of the model must be filled with a “best guess” value. The terrain and bottom of the anthropogenic layer is used as cut off layers, so that no values are set above terrain and below the bottom of the anthropogenic layer.

6: Voxelation of infrastructure elements

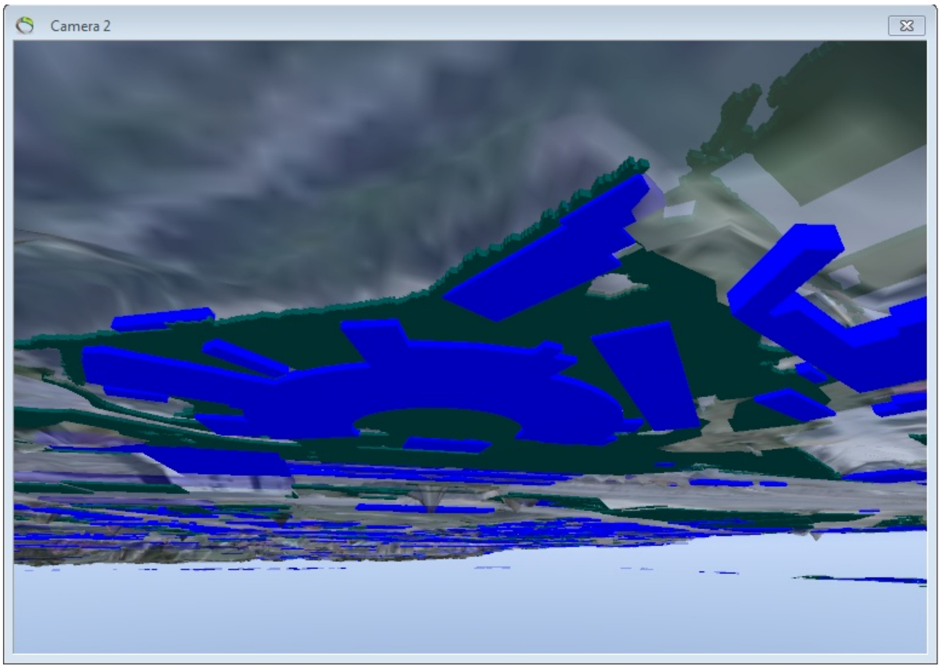


Fig App2-7: Assignment of infrastructure and buildings. Looking from below on basements and parking areas.

The final steps includes assigning values from different datasets, especially infrastructure and building etc. ending up with a final model that includes all relevant datasets.

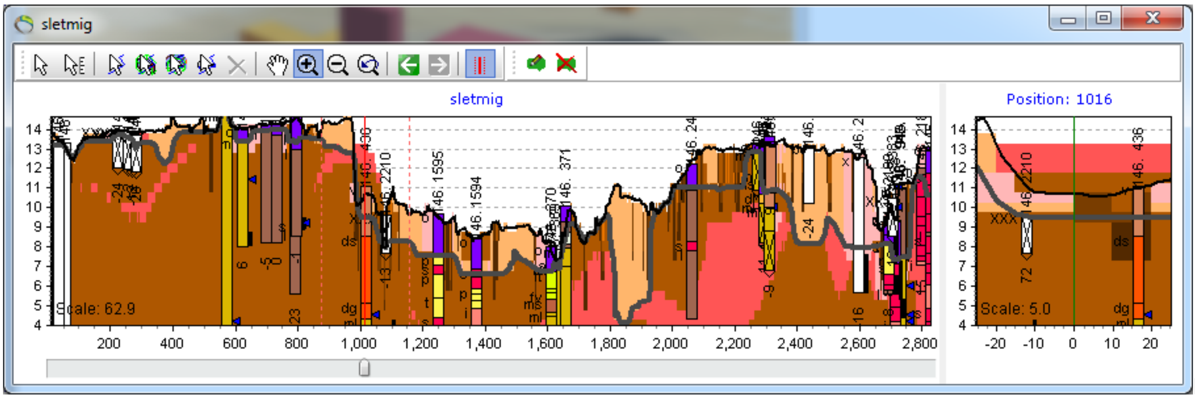


Fig App2-8: Cross section of the final model, with all data integrated

In GeoScene3D 2D features, can be rendered in 3D according to an attribute.

A building footprint, as an example, can be placed on the terrain, and extruded below terrain based on an attribute telling that basement depth is 4 meters.

Voxels inside this basement will then get a value telling that this is basement, or, if the voxel model should be used as base for a flowmodel, a value telling that no flow can occur in this voxel.

This is done for all structures resulting in a final model, incorporating all available information. The uncertainty in the different steps can be estimated and recorded in an uncertainty attribute. It is obvious, the uncertainty should be smaller, where actual measurements or borehole information are present, as well, as where foundations or infrastructure elements are located, compared to areas where voxel values are a more or less qualified guess, as in step 5

Datasets for urban geological models

Most wanted data for modelling, are boreholes and geotechnical information and if present existing geological maps and soil maps , as these are the most “geological” datasets, see Berg at all, 2011.

In addition, geophysical measurements can be used to expand geological information for boreholes and geophysical logs, and thus make better interpretations. Geophysical measurements in urban environments are often influenced by noise due to installations and other disturbances.

Although many datasets of the afore mentioned data are preferable, it is seldom the situation. But as a positive thing, in many cases, the “geology” is defined by manmade objects, which are mapped to some degree. This is for example the case with roads and utility lines, where information related to building the road or utility lines - the road bed and excavations - are available somehow.

Essential datasets are:

- Geological data
- Geotechnical data
- Geophysical data
- Digital terrain models
- Infrastructures: roads, pavement, railroad tracks….
    - Derived structures: Sand/gavel beds around/beneath these structures
- Buildings and foundations, underground parking areas, tunnels…
    - Derived structures: Sand/gavel beds around/beneath these structures
- Utility networks
    - Derived structures: Sand/gavel beds around/beneath these structures

I addition all other available geological data, such as:

- maps on surface geology/soil,
- existing geological models,
- morphological maps,
- agricultural soil maps,
- maps on streams and lakes, moors etc.
- historical maps, which can often reveal drained and filled areas

And other datasets:

- Archaeological data are also valuable, although the archaeological community has only recently started to use GIS and digital formats.
- Maps on groundwater tables can also help, especially if the urban model should target water issues.

In searching datasets for an urban model of a test area centred on the Danish City of Odense more than 80 datasets have been located, that contain some kind of information about the underground.

Data in general

The process of finding and acquiring the datasets can be quite elaborate. But in addition, many of the datasets are not in a state for direct use in 3D models, as they have never been intended for that. Many datasets, for instance, only exists as 2D datasets (Polygons and even just Lines or Points).

There are a few utility datasets can be rendered directly as 3D objects – sewers are an example of this, as the Z level is important in most circumstances and therefore documented, whereas water pipes often lack a Z level, as they operate under pressure and therefore the Z level is not considered as that important.

None of these datasets have any geology related information directly in the original datasets, and must therefore be enriched with these parameters.

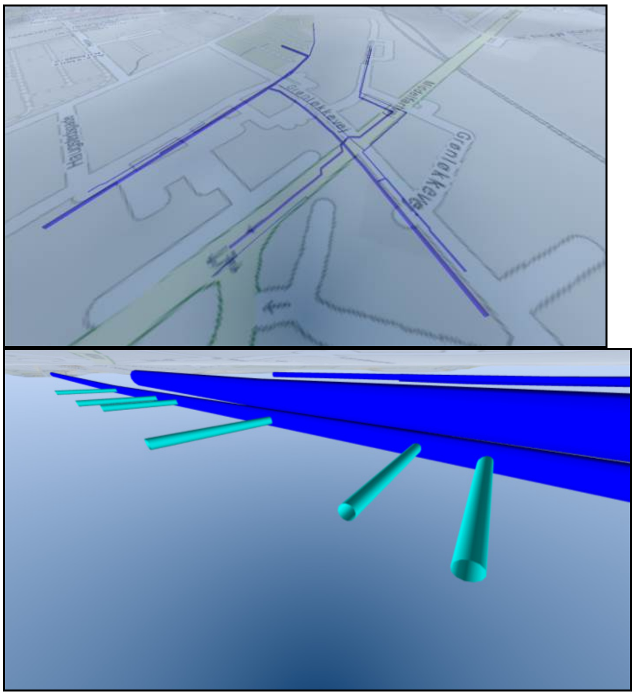


Figure App2-9: 2D Polylines rendered as 3D objects, based upon tube diameter and material.

So from a modelling software point of view, the software must be able to both render and interact with these datasets.

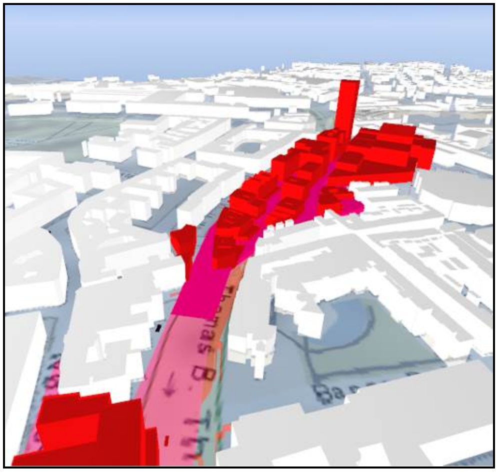
Extra attributes on GIS data

The GIS data often needs to be enriched in some way.

An example could be a road dataset. Not all roads are polygons; some might just be a polyline dataset that has some information on the road type, and maybe the year of construction. By examining this, the road dataset can be buffered to a common width depending on the type of road. Likewise, the road bed - the sand and gravel used for the road construction – can be estimated and assigned as an attribute to the dataset. With these attributes, the road and roadbed can be visualized and used for voxelation.

Basic data formats

- Boreholes with lithological and other information. Normally non spatial database tables, but in more simple cases ASCII files of different formats
- Logs, geotechnical and geophysical. Normally non spatial database tables, but in more simple cases ASCII files of different formats
- GIS formats. The most common formats are:
    - ESRI shape files
    - MapInfo Tab files
    - Spatial databases are also gaining popularity, such as Oracle Spatial and Postgres SQL, SQL Server Spatial.
    - Raster, e.g. aerial photos, topographical maps. Ecw, Jpeg, Tiff etc.
    - CAD formats are used to display buildings and projects



    - Web services, WMS and WFS are constantly growing in numbers and should be included for modelling.
- Image data, e.g. photos of archaeological sections, scanned logs. The most common formats are Jpeg, Tiff or PDF. These images should contain some kind of metadata that enables placement of them in 3D space.

Conversion of data

Data should be kept as close to their native formats as possible due to the task of updating, as this reduces the time and cost spend on conversions and the risk of introducing errors due to human handling. If data can be kept in their original format, it’s the task of the data provider to do updates, and this means that less workload is placed on urban modeller resources. This puts some requirements on the software that should be able to handle such data.

Although this is preferable, alterations and additions to data is often required. An example could be adding attributes describing the thickness of road beds as this is often not present in the native data.

This is especially of great importance, if the model later needs updating, as the modeller will then have to do the same tasks all over again.

As an alternative, an automated workflow can be created to do the conversions when needed. This can reduce the time when updates has to be done.

Digitization problem

Many data, especially geotechnical data, only exists in paper version. Some are scanned, fewer has an x,y coordinate associated. A full digital version of the geotechnical log is often quite expensive to produce, as this calls for a time consuming digitization. In this context, an Urban modelling project can be overwhelmed in the bare effort of the digitization of data (Not to mention the task of acquiring the data).

The process of digitization involves typically three steps:

1: Paper -> 2: Scanned images -> 3: Manual digitizing process of layers and log curves.

The last step can be especially time consuming

So unless this step is considered as a vital part of the process and has become the sufficient funding, it is crucial that less digitized data can be useable in the modelling process.

A way of accomplishing this is being able to digitize on scanned image material in the modelling software, when the material has undergone a basic geographical registration in space. This registration is done by registering the X and Y coordinates of the log, as well as the top and bottom.

An example is shown in Fig App2-10. With this simple registration, the image is present in the modelling environment, and any digitization on the image, is positioned in the right place in the modelling space.

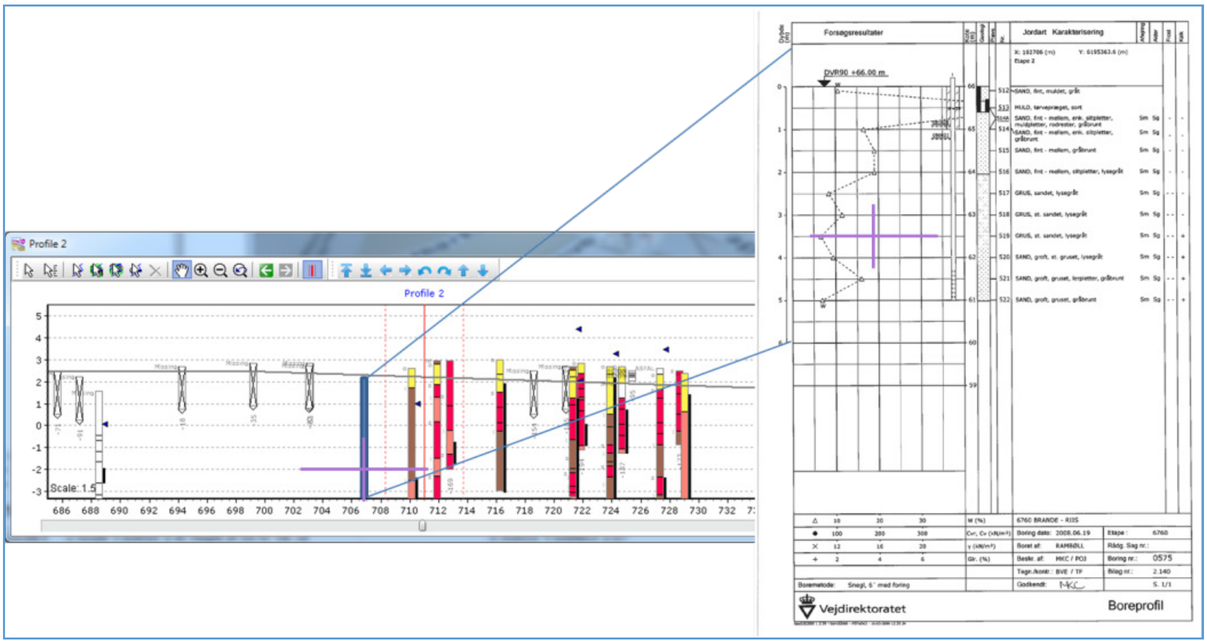

Profile 2
Profile 2
5
4
3
2
1
0
-1
-2
-3
Scale: 1.5
686 688 690 692 694 696 698 700 702 704 706 708 710 712 714 716 718 720 722 724 726 728 730 732
Dybde (m)
Forsøgsresultater
Kote (m)
Geologi
Jordart Karakterisering
X: 182706 (m) Y: 6195363.6 (m)
Etape 2
DVR90 +66.00 m
512 SAND, fint, muldet, gråt
513 MULD, tørvepræget, sort
514A SAND, fint - mellem, enk. siltpletter, muldpletter, rodrester, gråbrunt
514 SAND, fint - mellem, enk. siltpletter, gråbrunt
515 SAND, fint - mellem, gråbrunt
516 SAND, fint - mellem, siltpletter, lysegråt
517 GRUS, sandet, lysegråt
518 GRUS, st. sandet, lysegråt
519 GRUS, st. sandet, lysegråt
520 SAND, groft, st. gruset, lysegråt
521 SAND, groft, gruset, lerpletter, gråbrunt
522 SAND, groft, gruset, gråbrunt
Sm Sg
6760 BRANDE - RIIS
Boring dato: 2008.06.19
Etape: 6760
Boret af: RAMBØLL
Rådg. Sag nr.:
Beskr. af: MKC / POJ
Boring nr.: 0575
Tegn./kontr.: BVE / TF
Bilag nr.: 2.140
Godkendt: MKC
S. 1/1
Boremetode: Snegl, 6" med foring
w (%)
Cvr, Cv (kN/m²)
γ (kN/m³)
Gl. (%)
Vejdirektoratet
Boreprofil

Fig App2-10: Digitizing directly on scanned log image in the modelling environment. The Geotechnical log is shown in a large scale, and modelling points can be put directly into both the cross-section and the log, with the right position in model space.

Updating models

A common problem with models can be that models are built with big efforts, and delivered, but updates are often hard to do, because the road between "raw" data and model data is too long.

Urban models are often created for a general purpose with a demand for being able to be updated.

It is therefore critical that the work in an update is manageable, and the number of processes involved in coming from the "Raw" data to model data is low or automated.

- Estimating properties – problem. Clay-Sand value. Lots of guessing
- What is the purpose of modelling? Water issues, geothermal, geotechnical.
- Can we model geology as such – or some kind of sand/clay factor?

Coordinate systems

A careful examination of coordinate systems of the different datasets needs to be done as well as the choice of the best one.

It is often a problem that data is in different coordinate systems. And this is specially a problem, in urban modelling, as the data providers are very inhomogeneous, coming from different organizations and traditions.

Cross Sections – with all elements

A 3D environment makes it possible to see all elements in the model. The 3D environment can be confusing to look in because of the sometimes overwhelming amounts of information. Many users prefer to look at data in cross sections, as this limits data and makes it more tangible. So it should be possible to see all elements in cross sections, and to interact with both the model and the elements in this manner.

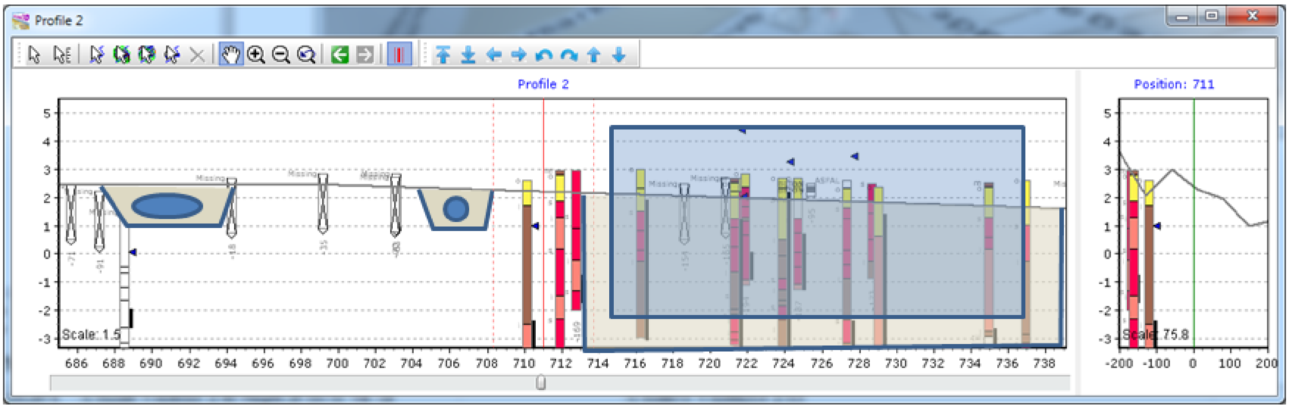


Fig App2-11: Visualizing all elements present in 3D in a cross-section

Further developments

Time plays a significant role in urban environments, as features can be replaced by other features over time. Therefore a more or less automated workflow that incorporates age of features and chooses the most recent features for assigning values in a voxel would be a good enhancement.

This could also support the task of updating models, as older features can be left in the model, and new features then just will overwrite the old features.

References

Berg, Richard C.; Mathers, Stephen J.; Kessler, Holger; Keefer, Donald A., eds., 2011. Synopsis of current three-dimensional geological mapping and modeling in Geological Survey organizations. Ilinois State Geological Survey Circular, 578. 104, pp. http://nora.nerc.ac.uk/17095/

Appendix 3: From analogue to digital data

Problem introduction

Well-structured and interoperable databases are becoming more and more common way of storage for geological and geotechnical data, however still a lot of geological information remains in paper format in archives of private companies and public organisations. These archives consist of non-editable data, which is kept only for reference, as a hard copies of documents or their scans in pdf/tiff format. The transcription of this data from analogue to digital format allows processing this data and generating many useful maps/models and analyses necessary for municipalities and geological surveys for managing the subsurface space of city areas.

The organisations which hold archives of geological and geotechnical data in paper formats are mostly private companies and their data comes from their commercial projects. Other archives of no-editable data are often in public companies, especially those which manage roads railways and metro or manage underground infrastructure, like water and sewage system, gas piping and telecommunication and electric network. Geological information resources are often digitized for certain projects (mostly site scale, on selected areas or new construction projects). There are also situations, that the archival data is shared with geological surveys as a good practice of data exchange. This is mostly for city scale, public organisations can then use the geological databases run by geological surveys.

Non editable geological data cannot be used for quick analyses. However municipalities often need the quick access to geological data for purposes of spatial planning, management of city subsurface space and for crisis management. To allow geological data processing within well-structures databases, the efficient procedures of analogue data digitalisation need to be developed and implemented.

With development of procedures and workflows for migration form paper and pdf documents to geological databases several problems and challenges should be addressed.

- **Legal regulations**. The information about the legal status and regulations considering the geological data should be maintained concerning the archival data being transferred to database. Often the archival data from private companies and public organizations have legal limitations for publishing raw data, it can be presented only in processed form of maps, models or analyses.
- **Full identification of archival data sources**. The analogue (paper) data are kept in many separate places / organisations/ companies. The working with paper documents brings logistical challenges. The identification of all sources of archival data is necessary to have an efficient and optimal planning of paper documents transport and scanning.
- **Temporal repository for external archives data**. if the data is already scanned (pdf/tiff) or has some form of a local database the temporary servers / ftp, with

necessary amount of disk space, need to be established. The local databases should be analysed and possible migration algorithms should be developed.

- **Verification of archival data quality**. In the process of migration to database very important step is to have the information on boreholes purpose (resources, geotechnics, heat pumps, wells, etc…) and date of completion. Low quality data can be skipped.
- **Borehole data management interface**. To minimize the transcription errors, to manage boreholes and the schedule of boreholes digitization the data management system is necessary. The developed data management system should also allow the presentation of final data from database in one of interoperable standards specific for certain country / region / city (like AGS standard in UK or firebird file database in Poland).
- **The need for harmonized data**. The controlled glossaries should have a connection with standards, like Eurocodes or National Standards, especially for soil and rock classifications. This would allow better interoperability of datasets and also make geological data more usable by their final users, for example geotechnical engineers or construction industry people.
- **Managing spatial coordinate specifications** in a city relevant scale
  - Transforming from site specific reference systems to urban scale as well as from regional/national or international scales.
  - Driving coordinates from positions marked on a map, free text descriptions and other relative or non-absolute position data
- **Managing controlled glossaries**. Urban data management requires the use of many diverse datasets which in turn requires the use of well documented terms for categories and units of measure. Without such standards it is extremely difficult to combine datasets in a meaningful way.

Suggestions for new technologies & good practices

PGI-NRI Engineering Geological Database. From analogue to digital data workflow.

To describe the process of migration of analogue data to geological database, the example form Polish Geological Survey Project "Engineering Geological Database" (acronym BDGI) was used. The BDGI project (duration 2013 -2016) is aimed at unification of 9 separate databases of largest agglomerations of Poland (total more than 260 000 boreholes) and creation of one unified database BDGI compatible with Central Geological Database of PGI-NRI with new extra 66 000 boreholes added to the database till end of the project.

Project BDGI description

Engineering-geological digital databases are the base for the preparation of engineering-geological atlases of large agglomerations of Poland. Engineering geological atlases of urban agglomerations are the largest and unique digital collection of such data in Poland. They include detailed information obtained from an engineering-geological, geotechnical, hydrogeological documentations and borehole profiles. Current state of the project can be seen on *atlasy.pgi.gov.pl* website.

They are created on the basis of engineering-geological, geotechnical, hydrogeological documentation and borehole profiles. Not only documents archived in a National Geological Archive (NAG) PGI- NRI, but also materials stored in the external archives of state-owned enterprises, municipal offices, private companies as well as data obtained from field mapping are used for this purpose.

A prepared BDGI database is used as a rich reference material for analyses in GIS technology. Precise quantitative and qualitative geostatistical analyses are carried out together with defining the relations between the data sets. The creation and connecting of different digital layers prepared using GIS methods allows performing so-called maps generation, depicting and synthesizing information contained in the database. This enables the presentation of the factors influencing the construction conditions in the ground. The Engineering Geological Database is used to generate numerous engineering-geological and thematic maps and spatial layers.

One of the main problems and tasks within the BDGI project is putting the analogue geological data into the database. Due to large amount of data to be digitized (more than 60 000 boreholes) the dedicated Data Management System – Geostar7 BDGI was developed and workflow of geological and geotechnical data migration to database was developed (see figure App3-1).

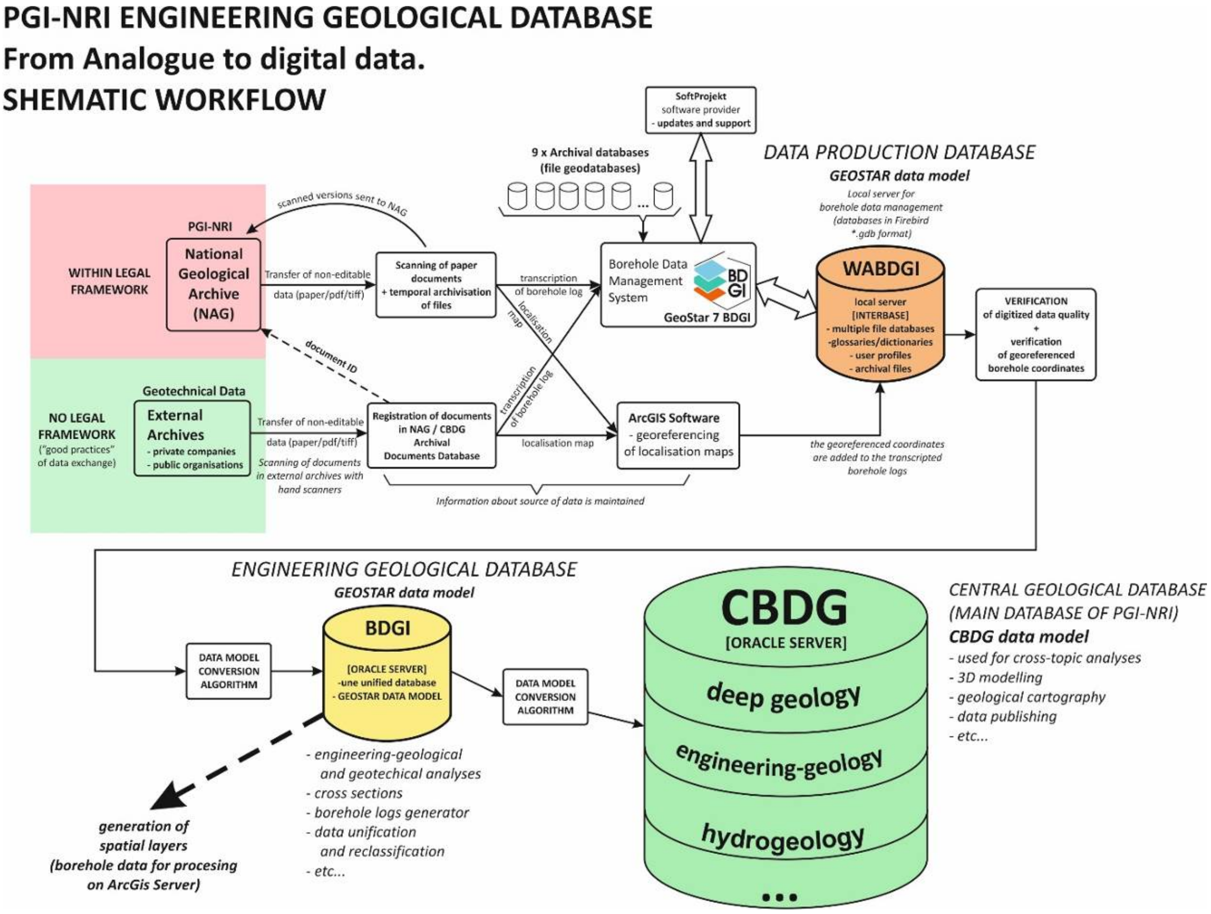

PGI-NRI ENGINEERING GEOLOGICAL DATABASE
From Analogue to digital data.
SHEMATIC WORKFLOW
WITHIN LEGAL FRAMEWORK
PGI-NRI
National Geological Archive (NAG)
scanned versions sent to NAG
Transfer of non-editable data (paper/pdf/tiff)
Scanning of paper documents + temporal archivisation of files
document ID
NO LEGAL FRAMEWORK ("good practices" of data exchange)
Geotechnical Data
External Archives
- private companies
- public organisations
Transfer of non-editable data (paper/pdf/tiff)
Scanning of documents in external archives with hand scanners
Registration of documents in NAG / CBDG Archival Documents Database
transcription of borehole log
localisation map
transcription of borehole log
localisation map
Information about source of data is maintained
9 x Archival databases (file geodatabases)
SoftProjekt
software provider
- updates and support
Borehole Data Management System
BD GI
GeoStar 7 BDGI
ArcGIS Software
- georeferencing of localisation maps
DATA PRODUCTION DATABASE
GEOSTAR data model
Local server for borehole data management (databases in Firebird *.gdb format)
WABDGI
local server
[INTERBASE]
- multiple file databases
-glossaries/dictionaries
- user profiles
- archival files
the georeferenced coordinates are added to the transcripted borehole logs
VERIFICATION of digitized data quality
+
verification of georeferenced borehole coordinates
ENGINEERING GEOLOGICAL DATABASE
GEOSTAR data model
DATA MODEL CONVERSION ALGORITHM
BDGI
[ORACLE SERVER]
-une unified database
- GEOSTAR DATA MODEL
- engineering-geological and geotechical analyses
- cross sections
- borehole logs generator
- data unification and reclassification
- etc...
generation of spatial layers (borehole data for procesing on ArcGis Server)
DATA MODEL CONVERSION ALGORITHM
CBDG
[ORACLE SERVER]
deep geology
engineering-geology
hydrogeology
...
CENTRAL GEOLOGICAL DATABASE
(MAIN DATABASE OF PGI-NRI)
CBDG data model
- used for cross-topic analyses
- 3D modelling
- geological cartography
- data publishing
- etc...

Fig. App3-1. Schematic workflow for PGI – NRI Engineering Geological Database.

The data which is gathered into BDGI database comes mainly from two sources. One is National Geological Archive (NAG), held by Polish Geological Institute, where all boreholes are kept for reference in paper or scanned (pdf/tiff) and they are archived within legal framework of Geological Law. The second source of data is from external archives of public organisations and private companies and covers mostly geotechnical data. This data is not covered by legal regulations of Geological Law, so therefore is not archived within National Geological Archive. The two sources of data require different procedures of migration into the database.

The data coming within legal framework, taken form the resources of National Geological Archive (NAG) is scanned as a part of general procedures within Polish Geological Survey of digitization of analogue resources from NAG. The scanned versions of paper documents are temporarily held on server and are processed to the Borehole Data Management System - Geostar7 BDGI.

The data, which is coming from external sources needs to be at first registered. All incoming paper documents need to have their document ID added in the NAG/CBDG archival documents database to maintain the information about source of geological data. The paper documents are scanned in the external archives with the use of mobile scanners.

Then the scans in pdf/tiff files from original documents are registered with unique source document ID.

When the both types of date are scanned and all necessary ID codes needed to maintain the link to original data source documents are attributed the data is processed to the Borehole Data Management System. Thanks to that all borehole data can be uploaded to the BDGI database in relation to source documents/reports etc. in National Geological Archive (NAG).

Profiles of borehole are introduced in database by GeoStar program. It is standard geological database program in Poland and it is used by many companies and institutions. Data are written with the help of special creators/wizards. Thanks to that, possible errors are minimized during entering data to database. The application creator allows typing data coherent with the relevant dictionaries. When creating main form of dictionaries used for both standards: national and ISO, in accordance with Eurocode. In addition, while working on the dictionaries were related Polish and ISO standards (commonly used in geotechnical engineering) with geological dictionaries used in PGI-NRI, in CBDG database.

Dictionaries are managed by Administrators of BDGI. They can change and actualize dictionaries on server if necessary. It allows controlling and managing dictionaries, which significantly improves the quality of data entry.

Very important problem is localization of boreholes. This is a problem of two types. The first concerns the quality and accuracy of information about the position on the different- scales maps or written by inaccurate data about the coordinates included in the borehole's sheet. The second is related with used different coordinate systems so far. Consequently, the two ways selected. If the sources documentation has information about localization, then it is put directly to the database after prior verification. Otherwise, each map with documentation points is calibrated / geo-refereed. Next every point can automatically obtain the exact coordinates, which are associated with the boreholes already included in the database. In this way we are sure good localization all documents points in the sense of boreholes.

Data with the coordinate entered into the database form database 'WABDGI'. It is exist as *the interbase* (**.gdb format*) on the local Firebird server. At this point the data can be further improved and modified if it becomes a reason. At this stage entered archival data by many users have to be verified. It is very imported because you should be sure that quality of digitized boreholes' profile and their coordinates are correct.

The next step is conversion Firebird's GeoStar data model to Oracle's GeoStar data model. It is very significant the selected algorithm to the data conversion, there were no mistakes. At this point, to properly constructed databases are possible engineering- geological (geotechnical) analyses, creating cross-sections, data unification and reclassification, generating borehole's sheets and logs and other. On the basis of data from the database, you can already generate spatial layers for GIS analysis. Thanks of that you can use information included in database for creating maps or for quick solving critical issues/ resolution of crises.

Last stage is conversion GeoStar data model to PGI-NRI standard data model. It is Central Geological Database (CBDG) which is main database of Polish Geological Survey. This database is largely public and contains many standardized data related to geology like: hydrogeology, engineering-geology, deep geology, deposits, mines, caves and other. Therefore the placement engineering-geological and geotechnical data in CBDG allows you to show and access to data for all the needy. Thanks of that data may be used for different purposes by different users, and the various programs.

Good example(s) / evaluation of techniques

BDGI – From analogue to digital data

The task with digitisation of analogue data is a big part of BDGI (Engineering-Geological Database) project of Polish Geological Survey. The task is to digitize 66 000 archival engineering-geological and geotechnical boreholes during the project duration 2013 – 2016. To fulfil this task the special data interface was developed, made especially for the BDGI project. The interface is GeoStar 7 BDGI and was developed by company Soft-Projekt Jan Szymanski. The presented example of analogue data digitisation is based on experiences from the BDGI project gained in year 2014. First year of the project took the development of GeoStar 7 BDGI tool. Then first 3 months of 2014 was the beta-testing of software. The other 9 months of digitizing in 2014 brought the effect of over 20 000 digitized boreholes up to the end of 2014. During this period the team working on the BDGI project, on the digitisation task, consisted of average 25 people. About half of them was working on the project full time, others were working remotely on their computers in PGI in LAN network, using their extra time for boreholes digitisation.

The composition of digitisation task team was as follows:

- 2 database administrators
- 5 people working on geo-referencing od localisation maps
- 2 people on scanning of borehole logs, localisation maps and reports
- 7 coordinators of regional subsets of archival data (working both as data collectors and data reviewers and quality managers)
- 7 ÷ 14 data collectors (varied number in time)

The equipment used for digitisation was:

- 10 laptop computers for data collection (see figure 2)

- 10 desktop computers data reviewers and quality managers and for BDGI database administrators.
- 1 large scale scanner
- 2 office scanners
- 6 hand held scanners for use in external archives, where is no possibility to move the documents to PGI offices
- 5 desktop computers (graphic workstation) with ArcGIS software for geo-referencing of localisation maps

The figures App3-2 to App3-10 give an overview of the performed workflow and organisational challenges connected with the digitisation task.

Fig. App3-2. Borehole data digitization.

The data collectors use terminals of Geostar7 BDGI (laptop and desktop PC) working in LAN network of PGI. The source data (scans of original borehole logs and reports are stored on a network drive folder). Each time the user (data collector) starts working with GeoStar7 BDGI the program automatically updates the glossaries from server which is crucial for maintaining the data quality and to minimize the transcription errors.

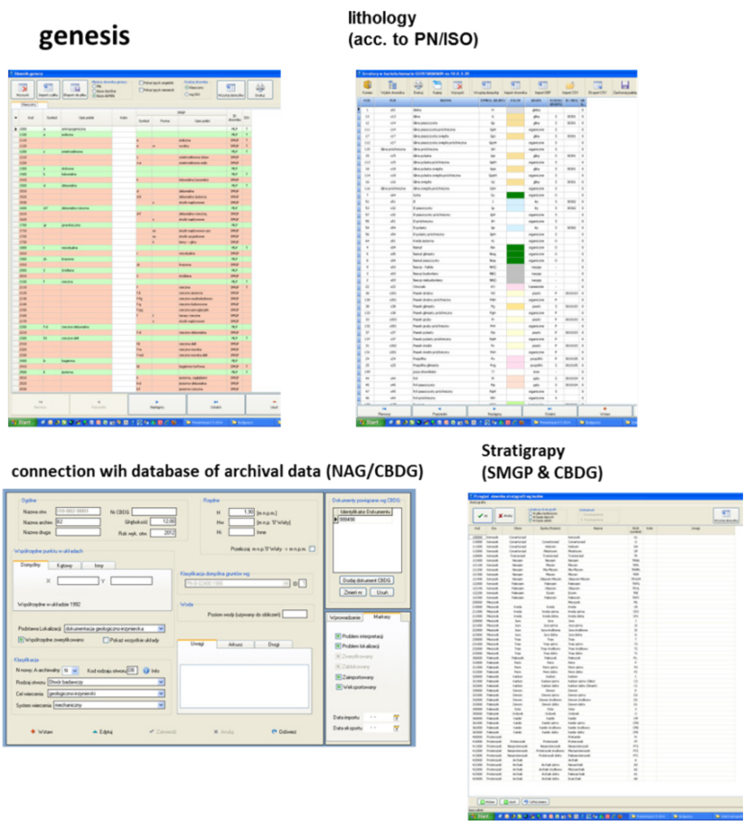
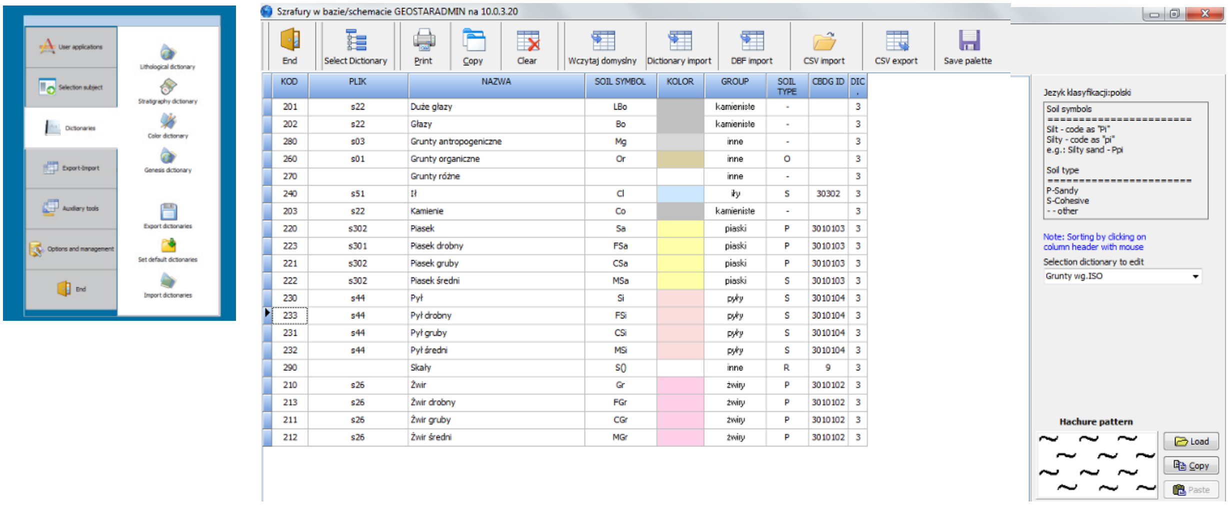


Fig. App3-3. The glossaries manager in Geostar7 BDGI borehole data input interface.

GeoStar7 BDGI uses glossaries stored on a local server in PGI. BDGI Administrator updates the glossaries during the project, so the transcription errors are minimized and the quality of the digitized borehole data can be maintained.



KOD	PLIK	NAZWA	SOIL SYMBOL	KOLOR	GROUP	SOIL TYPE	CBDG ID	DIC.
201	s22	Duże głazy	LBo	■	kamieniste	-		3
202	s22	Głazy	Bo	■	kamieniste	-		3
280	s03	Grunty antropogeniczne	Mg	■	inne	-		3
260	s01	Grunty organiczne	Or	■	inne	O		3
270		Grunty różne			inne	-		3
240	s51	Ił	Cl	■	iły	S	30302	3
203	s22	Kamienie	Co	■	kamieniste	-		3
220	s302	Piasek	Sa	■	piaski	P	3010103	3
223	s301	Piasek drobny	FSa	■	piaski	P	3010103	3
221	s302	Piasek gruby	CSa	■	piaski	P	3010103	3
222	s302	Piasek średni	MSa	■	piaski	S	3010103	3
230	s44	Pył	Si	■	pyły	S	3010104	3
233	s44	Pył drobny	FSi	■	pyły	S	3010104	3
231	s44	Pył gruby	CSi	■	pyły	S	3010104	3
232	s44	Pył średni	MSi	■	pyły	S	3010104	3
290		Skały	S()		inne	R	9	3
210	s26	Żwir	Gr	■	żwiry	P	3010102	3
213	s26	Żwir drobny	FGr	■	żwiry	P	3010102	3
211	s26	Żwir gruby	CGr	■	żwiry	P	3010102	3
212	s26	Żwir średni	MGr	■	żwiry	P	3010102	3

Jezyk klasyfikacji:polski

Soil symbols
========================
Silt - code as "Pi"
Silty - code as "pi"
e.g.: Silty sand - Ppi

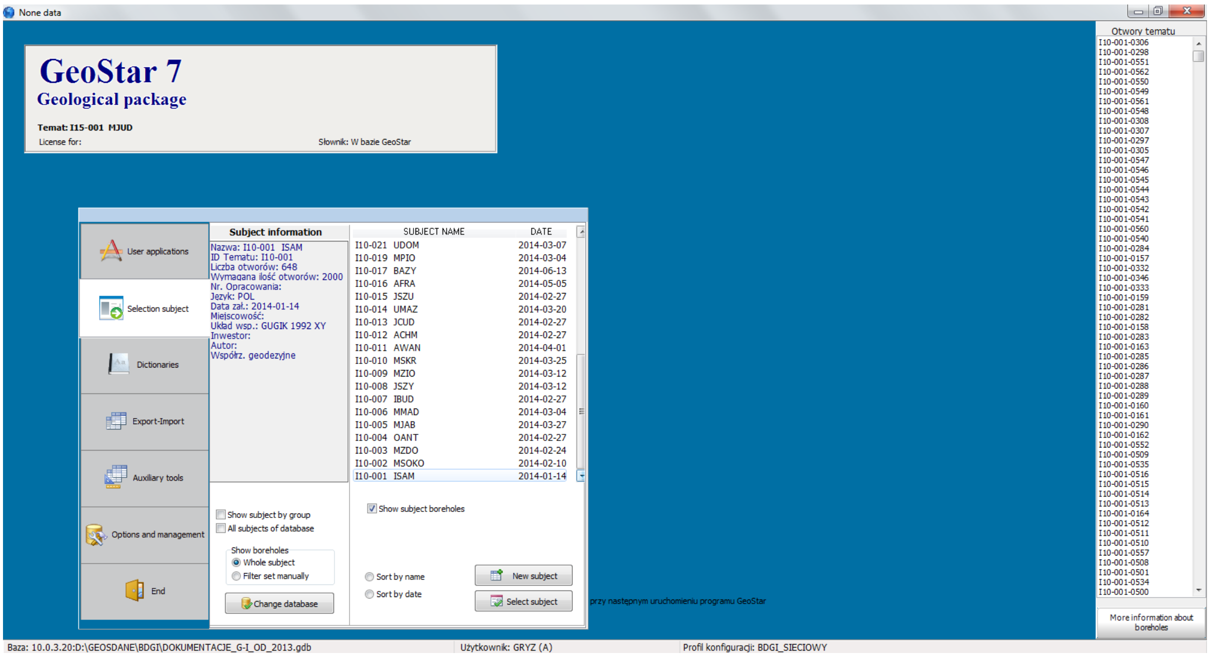
Soil type
========================
P-Sandy
S-Cohesive
- - other

Note: Sorting by clicking on column header with mouse

Selection dictionary to edit: Grunty wg.ISO

Fig. App3-4. Tools for managing the glossaries

The tools for managing the glossaries were developed in GeoStar7 BDGI software. The glossaries can be easily exported to CSV or XLSX files. Also patterns for profiles can be managed on the basis of unique codes in the glossaries.



Temat: I15-001 MJUD

License for:

Słownik: W bazie GeoStar

Subject information

Nazwa: I10-001 ISAM
ID Tematu: I10-001
Liczba otworów: 648
Wymagana ilość otworów: 2000
Nr. Opracowania:
Język: POL
Data zał.: 2014-01-14
Miejscowość:
Układ wsp.: GUGIK 1992 XY
Inwestor:
Autor:
Współrz. geodezyjne

SUBJECT NAME	DATE
I10-021 UDOM	2014-03-07
I10-019 MPIO	2014-03-04
I10-017 BAZY	2014-06-13
I10-016 AFRA	2014-05-05
I10-015 JSZU	2014-02-27
I10-014 UMAZ	2014-03-20
I10-013 JCUD	2014-02-27
I10-012 ACHM	2014-02-27
I10-011 AWAN	2014-04-01
I10-010 MSKR	2014-03-25
I10-009 MZIO	2014-03-12
I10-008 JSZY	2014-03-12
I10-007 IBUD	2014-02-27
I10-006 MMAD	2014-03-04
I10-005 MJAB	2014-03-27
I10-004 OANT	2014-02-27
I10-003 MZDO	2014-02-24
I10-002 MSOKO	2014-02-10
I10-001 ISAM	2014-01-14

☐ Show subject by group
☐ All subjects of database

Show boreholes
● Whole subject
○ Filter set manually

☑ Show subject boreholes

○ Sort by name
○ Sort by date

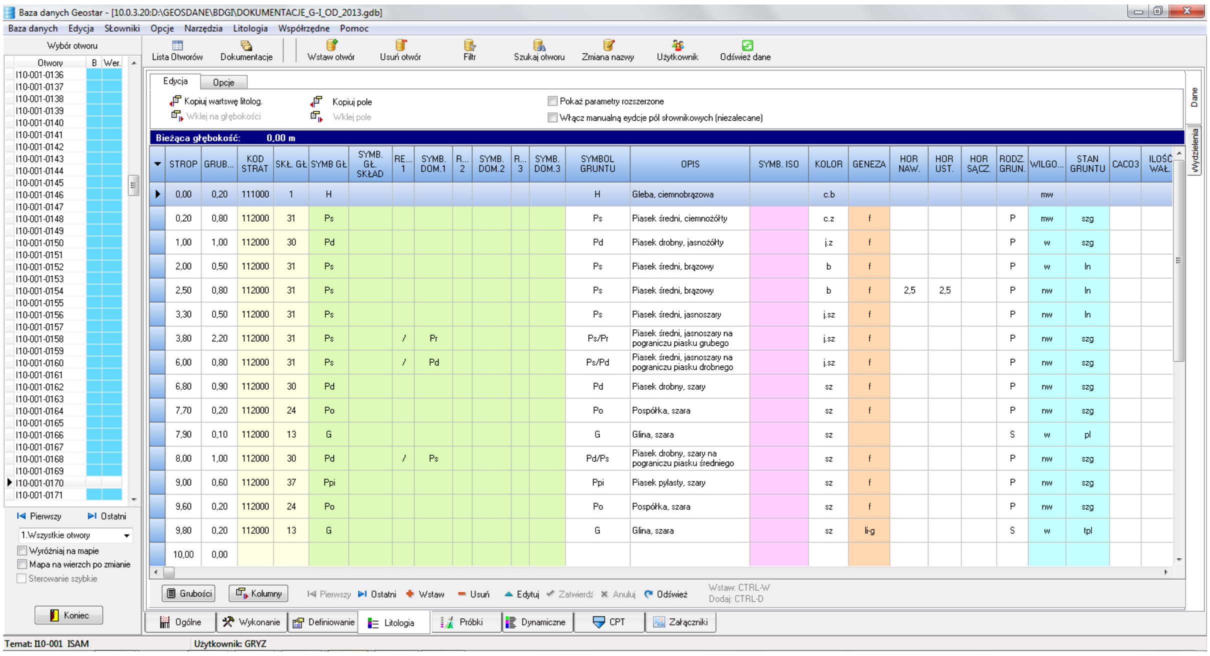
przy następnym uruchomieniu programu GeoStar

Otwory tematu
I10-001-0306
I10-001-0298
I10-001-0551
I10-001-0562
I10-001-0550
I10-001-0549
I10-001-0561
I10-001-0548
I10-001-0308
I10-001-0307
I10-001-0297
I10-001-0305
I10-001-0547
I10-001-0546
I10-001-0545
I10-001-0544
I10-001-0543
I10-001-0542
I10-001-0541
I10-001-0560
I10-001-0540
I10-001-0284
I10-001-0157
I10-001-0332
I10-001-0346
I10-001-0333
I10-001-0159
I10-001-0281
I10-001-0282
I10-001-0158
I10-001-0283
I10-001-0163
I10-001-0285
I10-001-0286
I10-001-0287
I10-001-0288
I10-001-0289
I10-001-0160
I10-001-0161
I10-001-0290
I10-001-0162
I10-001-0552
I10-001-0509
I10-001-0535
I10-001-0516
I10-001-0515
I10-001-0514
I10-001-0513
I10-001-0164
I10-001-0512
I10-001-0511
I10-001-0510
I10-001-0557
I10-001-0508
I10-001-0501
I10-001-0534
I10-001-0500

Baza: 10.0.3.20:D:\GEOSDANE\BDGI\DOKUMENTACJE_G-I_OD_2013.gdb
Użytkownik: GRYZ (A)
Profil konfiguracji: BDGI_SIECIOWY

Fig. App3-6. The data management tools.

To have control of digitization progress the data management tools were developed. To do so, the roles were designated as follows: data collector, coordinator and administrator. Data collectors digitize the scanned borehole logs, Coordinators verify and check how many boreholes were loaded into database by each data collector, verify and review the quality of digitized boreholes. Boreholes verified and accepted by coordinators are then uploaded by Administrator from Interbase (Firebird) local server WABDGI (production server) to Oracle server BDGI used for analytical purposes and data publishing.



Otwory	B	Wer.
I10-001-0136		
I10-001-0137		
I10-001-0138		
I10-001-0139		
I10-001-0140		
I10-001-0141		
I10-001-0142		
I10-001-0143		
I10-001-0144		
I10-001-0145		
I10-001-0146		
I10-001-0147		
I10-001-0148		
I10-001-0149		
I10-001-0150		
I10-001-0151		
I10-001-0152		
I10-001-0153		
I10-001-0154		
I10-001-0155		
I10-001-0156		
I10-001-0157		
I10-001-0158		
I10-001-0159		
I10-001-0160		
I10-001-0161		
I10-001-0162		
I10-001-0163		
I10-001-0164		
I10-001-0165		
I10-001-0166		
I10-001-0167		
I10-001-0168		
I10-001-0169		
I10-001-0170		
I10-001-0171		

1.Wszystkie otwory

□ Wyróżniaj na mapie
□ Mapa na wierzch po zmianie
□ Sterowanie szybkie

□ Pokaż parametry rozszerzone
□ Włącz manualną eydcje pól słownikowych (niezalecane)

Bieżąca głębokość: 0,00 m

STROP	GRUB...	KOD STRAT	SKŁ. GŁ	SYMB GŁ	SYMB. GŁ. SKŁAD	RE... 1	SYMB. DOM.1	R... 2	SYMB. DOM.2	R... 3	SYMB. DOM.3	SYMBOL GRUNTU	OPIS	SYMB. ISO	KOLOR	GENEZA	HOR NAW.	HOR UST.	HOR SĄCZ.	RODZ. GRUN.	WILGO...	STAN GRUNTU	CACO3	ILOŚĆ WAŁ.
0,00	0,20	111000	1	H								H	Gleba, ciemnobrązowa		c.b						mw			
0,20	0,80	112000	31	Ps								Ps	Piasek średni, ciemnożółty		c.z	f				P	mw	szg		
1,00	1,00	112000	30	Pd								Pd	Piasek drobny, jasnożółty		j.z	f				P	w	szg		
2,00	0,50	112000	31	Ps								Ps	Piasek średni, brązowy		b	f				P	w	ln		
2,50	0,80	112000	31	Ps								Ps	Piasek średni, brązowy		b	f	2,5	2,5		P	nw	ln		
3,30	0,50	112000	31	Ps								Ps	Piasek średni, jasnoszary		j.sz	f				P	nw	ln		
3,80	2,20	112000	31	Ps		/	Pr					Ps/Pr	Piasek średni, jasnoszary na pograniczu piasku grubego		j.sz	f				P	nw	szg		
6,00	0,80	112000	31	Ps		/	Pd					Ps/Pd	Piasek średni, jasnoszary na pograniczu piasku drobnego		j.sz	f				P	nw	szg		
6,80	0,90	112000	30	Pd								Pd	Piasek drobny, szary		sz	f				P	nw	szg		
7,70	0,20	112000	24	Po								Po	Pospółka, szara		sz	f				P	nw	szg		
7,90	0,10	112000	13	G								G	Glina, szara		sz					S	w	pl		
8,00	1,00	112000	30	Pd		/	Ps					Pd/Ps	Piasek drobny, szary na pograniczu piasku średniego		sz	f				P	nw	szg		
9,00	0,60	112000	37	Ppi								Ppi	Piasek pylasty, szary		sz	f				P	nw	szg		
9,60	0,20	112000	24	Po								Po	Pospółka, szara		sz	f				P	nw	szg		
9,80	0,20	112000	13	G								G	Glina, szara		sz	li-g				S	w	tpl		
10,00	0,00																							

Temat: I10-001 ISAM   Użytkownik: GRYZ

Fig. App3-7. Interface for data input.

The process of borehole data input is performed with the use of special wizard. This “wizard” allows only selection of lithology, genesis, stratigraphy and other values from the list based on latest updated version of glossaries from the server. The transcription errors are minimized then. The data collectors can also use checkbox “problem” if they are not sure what field from the list they should match with the original geological descriptions from the archival borehole logs and reports. This is a very useful tool for coordinators, who can verify such records in borehole profiles much easier. This also improves the digitized data quality.

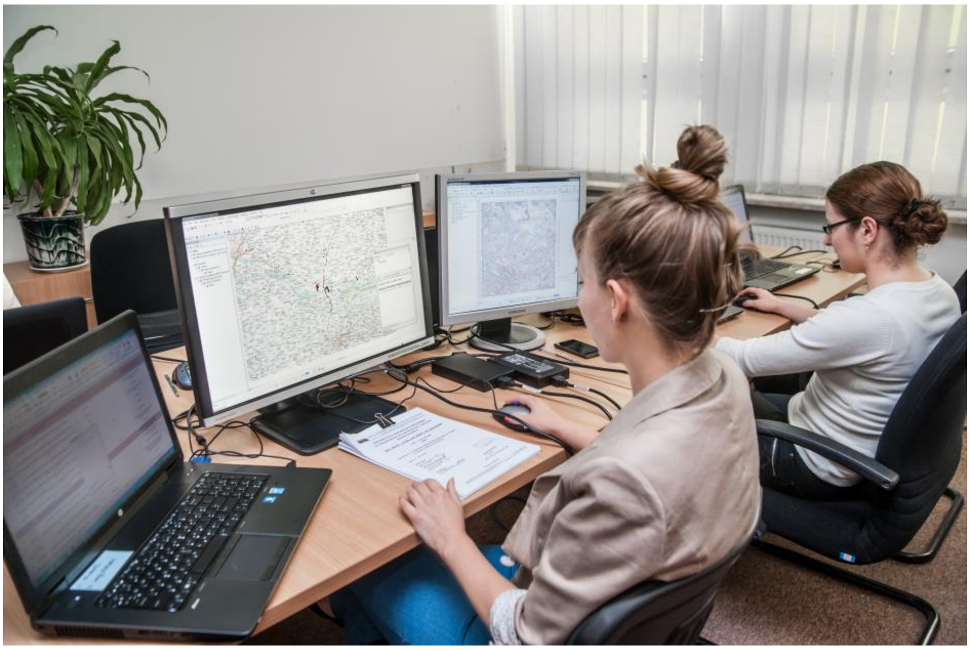
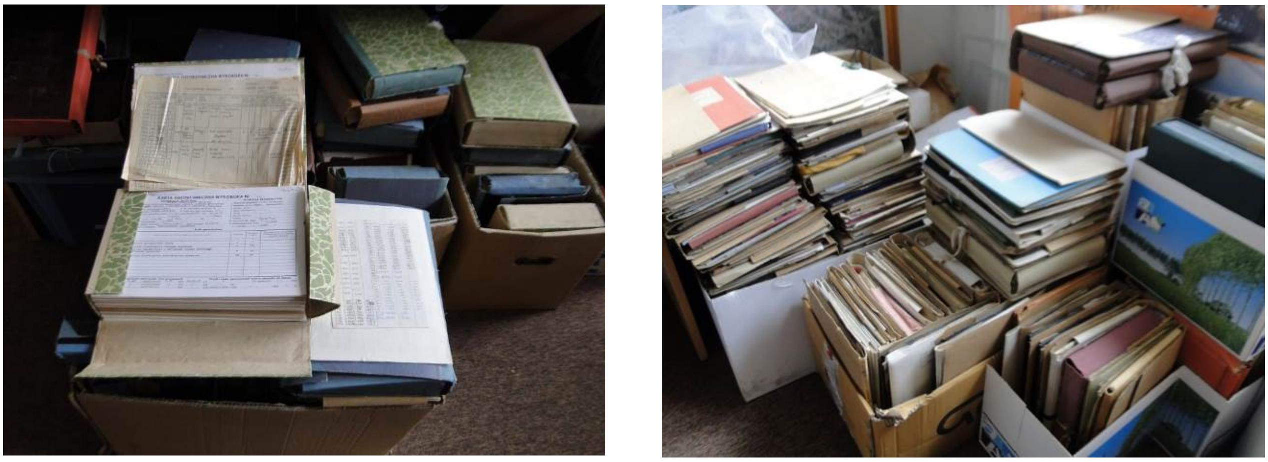


Fig. App3-8. The geo-referencing of localisation maps.

The geo-referencing of localisation maps is performed by different persons than those who digitize boreholes. Many of archival ground investigation reports made before year 1990 have no information about coordinates of boreholes. The only source of this data are scanned localisation maps. On the geo-referenced scans of maps (geo-tiff format) the boreholes localisation points are digitized as a feature class in Arc GIS geodatabase. The x and y coordinates are then recalculated in ArcGIS in predefined coordinate system for whole database (in Poland we use the “1992” projected coordinate system). The information of archival borehole name and source documentation ID code are the fields used to match the digitized boreholes in GeoStar 7 BDGI database with their geo-referenced coordinates from ArcGIS geodatabase tables.

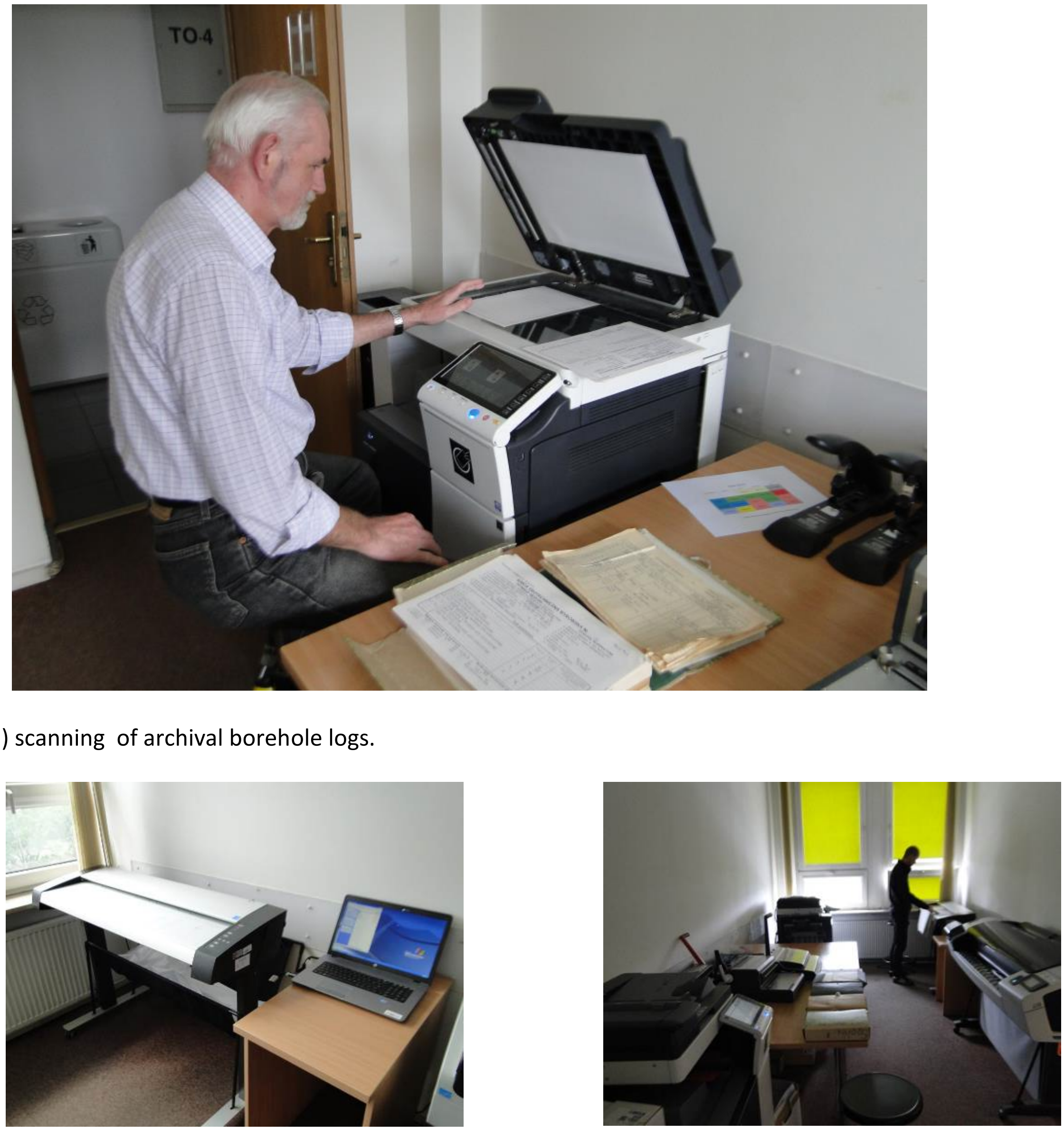


a) borehole logs from the 1970’s,

b) geological reports

Fig. App3-9 a, b. Source data in analogue format; very time consuming scanning process.

The archival borehole logs, localisation maps and reports are brought from National Geological Archive in portions, 1 or 2 times a week. The form of archival documents is varied, some of them can be scanned quickly with the use of office scanner with automation for multiple pages scanning (like for borehole logs from the 1970, see figure a). Other reports require more effort in scanning, due to non-standard formats or the need to separate the split pages. In BDGI project the two persons are working their full time only on scanning of documents and uploading them to server.



a) scanning of archival borehole logs.

b) large scale scanner for localisation maps

c) scanning room setting; two office scanner units were used and a large scale scanner

Fig. App3-10. Scanning room setting.

Special room was prepared to allow performing all scanning activities in one place, to avoid the turning of archival documents into separate parts (borehole logs, maps, etc.). All scanned documents are on the run uploaded to server, where data collectors use scans for digitisation into Geostar7BDGI software.

Appendix 4: Commercial data and public data centre services

Current State

Efficient management of Europe’s urban environments requires an efficient means of communicating existing information amongst stakeholders and effective systems that support the capture and storage of newly created data. The production and management of data can be expensive and all too frequently the information contained within the data is used only within the project it was produced for. Recycling this data for use by both public and private organisations could provide the basis of future desk studies, ground models and resources for planning and regeneration. However, the data must be managed and in a form that is readily available.

Whilst advances in technology mean more of the data, which is important to city management, is increasingly digital there remains a large body of analogue data sources that are expensive to convert into usable digital formats for current and future projects. Financial constraints on public bodies have led to the need to increasingly automate the digitisation of analogue datasets rather than rely on manual checking and conversion.

Databases are a key element of modern organisations, whether they are publicly funded or private commercial ventures. Numerous databases are designed to hold organisation or project specific data, this results in the data being locked in many isolated, and possibly, incompatible databases making integration difficult or impossible.

The geotechnical and geoenvironmental industry has produced a data transfer format for geotechnical and geoenvironmental site investigation data (Association of Geotechnical and Geoenvironmental Specialists AGS data transfer format) (Bland et al. 2014, Walthall and Palmer 2006 Chadwick et al. 2006).The data is entered once at source by the field or laboratory contractor and then transferred and used many times within the project, i.e. contractors, consultants and clients as required. Ideally the data is also made available to the construction contractors and sub-contractors as required. This makes the whole process of site investigation, design and construction more efficient and as all the data is available during the site investigation and when it is completed. It is also more likely that risks will be identified thereby reducing the chances of unforeseen ground conditions.

This data can also be recycled by others including local, regional and national authorities to inform future development planning and construction if it is made available in a database. A part of this includes the interpretation and the production of themed ground models and maps to aid this process.

Much of the data is produced by the geotechnical community for public and private development. For the efficient recycling of the data it needs to be made available much more widely.

The Highways Agency (the major roads) provides digital data transfer format files to consultants, contractors and others via Highways Agency Geotechnical Data Management System (HA GDMS) (Power et al. 2012). Whilst this is not a database it does make the relatively small data format files available for easy and rapid download for those who have access. However, some of the data transfer files do contain errors that make accessing the data more time consuming than needs be. This system is efficient for the client, the Highways Agency as it is a managed store of files with the metadata on each file. Those that download the files then use it as they require. The data might be recycled and reused a number of times at the desk study stage to aid the design of future site investigations within the area covered by previous site investigation. There is no attempt to produce a database of the data and it is up to the different consultant, contractors and others how they use the data they download, usually using commercially available software.

The geotechnical engineering community provide one example of public and commercial organisations coming together to develop standardised digital data formats with the aim of transferring data between different organisations within a civil engineering project reducing uncertainty in ground conditions through greater data access and re-cycling and re-use (Bonsor_etal_2013). The geotechnical community in the UK aim to reduce costs, lower transcription errors and speed up the construction process by transferring data through electronic means using the format developed by the Association of Geotechnical and Geoenvironmental Specialists (AGS). A number of pilot projects such as the Accessing Subsurface Knowledge (ASK) network operating in the greater Glasgow region are actively developing the communities, workflows and digital tools needed to improve data access and re-use. This report will focus on the technical challenges faced by initiatives such as the ASK network.

There are a number of technical challenges which need to be overcome by communities looking to integrate the data and information gathered by a range of organisations regardless of the discipline involved, the key challenges considered in this report are:

- Standard exchange formats
- Producing automated and semi-automated systems to deliver:
    - Data validation (against the agreed data & exchange formats)
    - Data verification (to ensure that the data is valid and valuable)
    - Centralised/communal data storage (includes standardisation of data structures and dictionaries used)
    - Data discovery
    - Data visualisation

- Data access (to enable use and re-use)

Suggested good practice

Standard Exchange Formats – Geotechnical example

In order to deliver on the aims of the ASK network, and similar initiatives, it is necessary to develop a community of users who will adopt common standards, tools and techniques. One way to develop such a community is through voluntary initiatives that recognise the mutual benefits of working together, for example the coming together of stakeholders in the Glasgow region. Another approach is to introduce legislation which stipulates the use of certain standards, for example the UK Cabinet office published their Government Construction Strategy in 2011. "The report announced the Governments intention to require: collaborative 3D BIM (with all project and asset information, documentation and data being electronic) on its projects by 2016." source: www.bimtaskgroup.org (accessed 21/01/2015).

Building Information Modelling (BIM) has focussed on information and communication between the different organisations involved in construction, it has not explicitly addressed geological or geotechnical data requirements, indeed geology, geotechnics and ground engineering are not really referred to within BIM. However, that is no reason for the providers of geotechnical data to ignore such initiatives, geotechnical data providers should ensure that their data can be directly incorporated into BIM tools or, if this is not possible, devise clear and simple methods for the conversion of their data into a suitable format. BIM level 2 is 'a managed 3D environment held in separate discipline 'BIM' tools with data attached. The AGS digital transfer format fulfils BIM level 2 requirements for geotechnical site investigation data.

Automated Data Validation

Where users of a particular data exchange format have different interpretations of the standard it is necessary to develop methods to address any incompatibilities which may arise as a consequence. One option is to produce community approved format validation tools which can scrutinise the format of the data produced by each user, ideally this would be in the form of a freely available tool that was maintained by an organisation which is independent of the software vendors who produce the tools which generate the data.

Assisted Data Verification

In this report the act of verification is considered to be the evaluation of whether or not a data submission is suitable for ingestion into a communal data store. This is difficult to automate, in most cases there will need to be some level of human intervention to identify whether the data is relevant, a duplicate of previously acquired data and if possible the correctness of the data and metadata is checked, for example does the metadata accurately describe the spatial or temporal contents of the submission.

Although it may be possible to assist manual processing of the data, with systematic reports and automated notifications, it is difficult to envisage a fully automated system which could cope with the wide variety of possible data submission types which could be received.

Communal data storage

Communal data stores provide a means of pooling information together for a common purpose, the users of communal data stores are often organisations or institutions.

There are two main options for implementing a communal data store, the most common and arguably the simplest option is a centralised data store, this is a facility which is located, stored and maintained at a single physical location. The alternative approach is a distributed data store, where data is held in multiple physical locations and are logically combined to provide a single view of all the data.

The following table shows some of the advantages and disadvantages of each approach:

	General advantages of each implementation
Centralised	- Easier to secure - All data accessible at once - Single master copy of the data - Updates to data are immediately available to all users - Single data model is easier for users to comprehend - Lower costs for power and staff maintenance
Distributed	- Greater redundancy - Less prone to bottlenecks due to high traffic - Can accommodate multiple data model structures - Greater ability for multiple users to access, create and update data

The way in which data is grouped for communal use is evolving, from data created and used by an individual to an intra-organisation project bases to organisation wide data stores. There is an increasing trend towards inter-organisation data sharing, either for specific purposes, such as a large collaborative project, or more generally as a strategic decision to share data with specific external organisations.

It seems appropriate that communal data stores were designed to pool the key datasets relevant to city management, such a store could be implemented as a centralised or distributed system.

In order to develop a communal data store it is crucial that the terminology and units of measure used are commonly understood and their meaning clearly communicated. Ideally the data store should be constrained by a number of controlled glossaries (or sometimes referred to as data dictionaries). Gaining agreement between all community members on which terms to use is often a difficult and timely process but if individuals feel the need to use an alternative set of terms it is possible to build mapping processes into the solution so that specific names and units of measure can be transformed to community controlled equivalent.

Data discovery

Regardless of how and where data is stored, users and potential users should be able to discover that data through simple search facilities. As most of the data which has been considered in this report has some spatial context it is possible that a spatial metadata search could provide the basis for data discover portals, within Europe the spatial metadata INSPIRE directive. The directive is *"establishing an infrastructure for spatial information in Europe to support Community environmental policies, and policies or activities which may have an impact on the environment."* Source - http://inspire.ec.europa.eu/ (accessed 22/01/2015)

Data visualisation

One of the quickest ways to assess the likely value of data, especially spatial data, is to visualise it, the primary goal of data visualisation is to communicate information clearly. Seeing data in the correct spatial and temporal dimensions is a particularly powerful way to analyse data allowing users to identify spatiotemporal patterns within the data and recognise relationships between datasets.

Data access

In order to maximise data access and re-use data exchange and discovery standards should be used such as international spatial metadata standard ISO 19115. Data models are also an effective way to describe the meaning, structure and interrelationships between data.

-

  Where possible data should be distributed in open formats and barriers to access such as expensive proprietary software should be avoided. When distributing 3D geological models to stakeholders in the commercial sector the BGS have been asked to provide access via web based tools that do not require applications to be installed on networked computers due to security concerns.

Example of good practice

GSPEC – Glasgow SPecification of AGS

In order to deliver the technical aspects of the ASK Network vision required the development of solutions to all of the key issues raised in the previous section, namely:

- Standard data exchange format
- Data validation
- Assisted data verification
- Communal data store
- Data discovery
- Data visualisation
- Data access

In addition to these general requirements Glasgow City Council required a GCC branded web interface that allowed them to provide a file submission and file validation process which would trigger emails to confirm acceptance or a failure report if invalid.

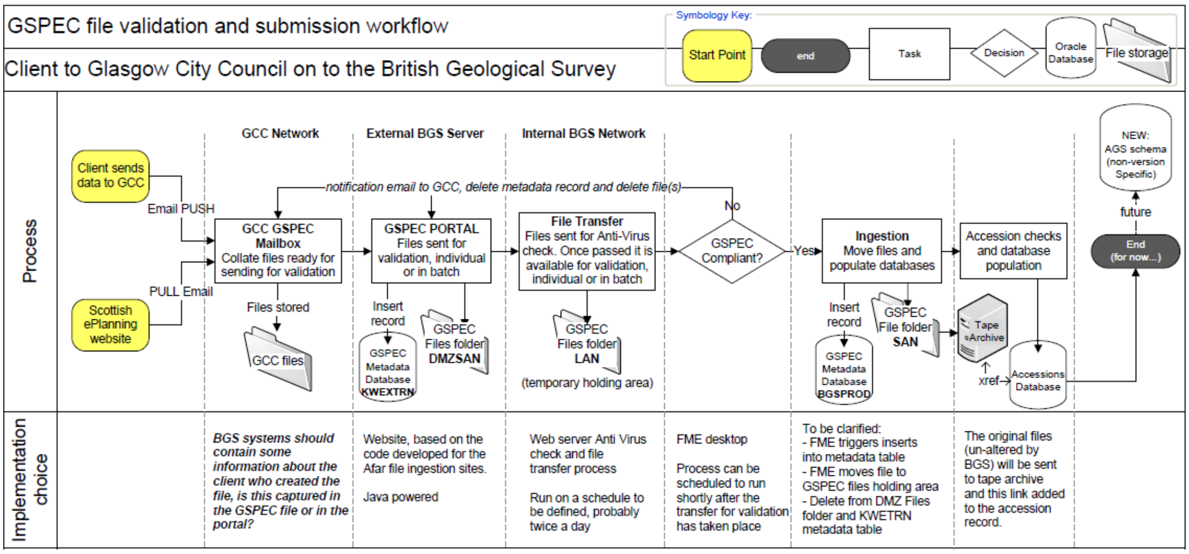
The BGS also required the solution to be secure, restrict submission functionality to authorised users only and support the capture of appropriate discovery and data accession metadata.

The responsibility for developing the solution fell largely on the BGS and required a small team of experts, the team contained individuals with the following skills and experience:

- Knowledge of the AGS data format
- An ability to develop the programmatic logic needed to check text files for conformance to validation rules
    - This could have been achieved using a programming language such as Java, C# etcetera but was implemented using an expert in the software FME (Feature Manipulation Engine) by Safe Software Inc: http://www.safe.com/fme/
- Web programming
- Database development
- Report generation
- Configuration of anti-virus checkers
- Advanced visualisation programming

Illustration of suggested workflow

The data flow which was developed for the ASK Network centres around an online website through which data donors are authenticated and files submitted, once a file has been submitted it triggers an automated validation and ant-virus check process, if valid, files are transferred to internal BGS servers where the data verification and data storage processes take place. This information workflow was initially designed in 2013 and the solution was launched early 2014 and closely resembles the draft design shown in Figure App4-1.


GSPEC file validation and submission workflow
Client to Glasgow City Council on to the British Geological Survey
Symbology Key:
Start Point
end
Task
Decision
Oracle Database
File storage
Process
Implementation choice
GCC Network
External BGS Server
Internal BGS Network
Client sends data to GCC
Email PUSH
Scottish ePlanning website
PULL Email
notification email to GCC, delete metadata record and delete file(s)
No
GCC GSPEC Mailbox
Collate files ready for sending for validation
Files stored
GCC files
GSPEC PORTAL
Files sent for validation, individual or in batch
Insert record
GSPEC Metadata Database KWEXTRN
GSPEC Files folder DMZSAN
File Transfer
Files sent for Anti-Virus check. Once passed it is available for validation, individual or in batch
GSPEC Files folder LAN
(temporary holding area)
GSPEC Compliant?
Yes
Ingestion
Move files and populate databases
Insert record
GSPEC Metadata Database BGSPROD
GSPEC File folder SAN
Accession checks and database population
Tape Archive
xref
Accessions Database
NEW: AGS schema (non-version Specific)
future
End (for now...)

BGS systems should contain some information about the client who created the file, is this captured in the GSPEC file or in the portal?

Website, based on the code developed for the Afar file ingestion sites.

Java powered

Web server Anti Virus check and file transfer process

Run on a schedule to be defined, probably twice a day

FME desktop

Process can be scheduled to run shortly after the transfer for validation has taken place

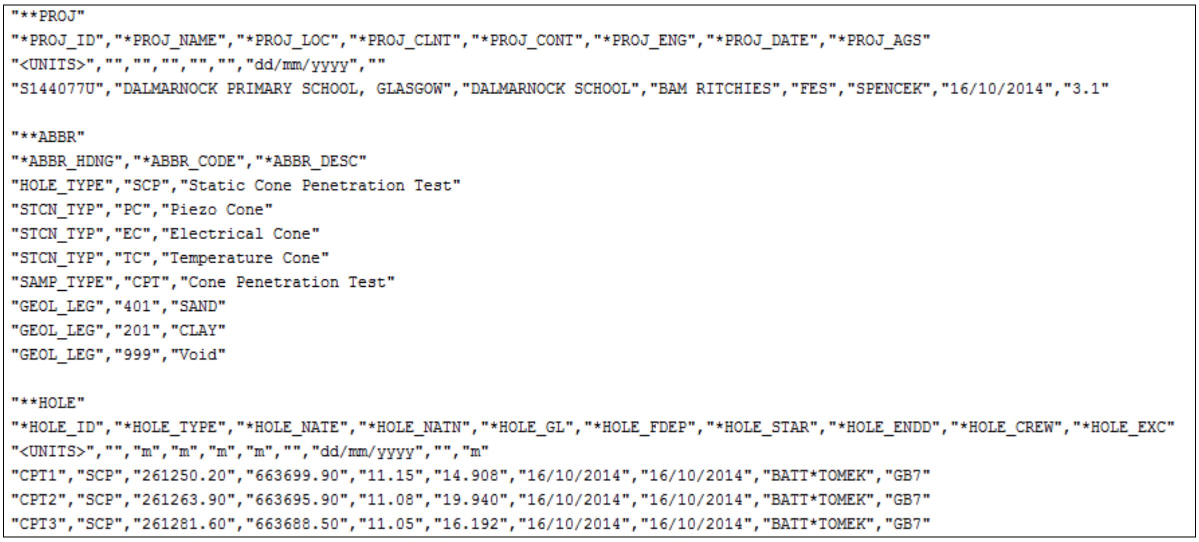
To be clarified:
- FME triggers inserts into metadata table
- FME moves file to GSPEC files holding area
- Delete from DMZ Files folder and KWETRN metadata table

The original files (un-altered by BGS) will be sent to tape archive and this link added to the accession record.

Figure App4-1: Draft workflow for the ASK Network data acquisition workflow

The **G**lasgow **SPE**cification for data **C**apture (GPEC) was developed as part of the project and was used as the standard data exchange format (Campbell & Bonsor, 2013). GSPEC is essentially AGS version 3.1 with additional rules, most notably all point data (trial pits, boreholes and sample) should have British National Grid Reference (x and y) and Ordnance Datum (z).

GSPEC files are human readable ASCII files containing comma separated labels and values, see Figure App4-2.


```
"**PROJ"
"*PROJ_ID","*PROJ_NAME","*PROJ_LOC","*PROJ_CLNT","*PROJ_CONT","*PROJ_ENG","*PROJ_DATE","*PROJ_AGS"
"<UNITS>","","","","","","dd/mm/yyyy",""
"S144077U","DALMARNOCK PRIMARY SCHOOL, GLASGOW","DALMARNOCK SCHOOL","BAM RITCHIES","FES","SPENCEK","16/10/2014","3.1"

"**ABBR"
"*ABBR_HDNG","*ABBR_CODE","*ABBR_DESC"
"HOLE_TYPE","SCP","Static Cone Penetration Test"
"STCN_TYP","PC","Piezo Cone"
"STCN_TYP","EC","Electrical Cone"
"STCN_TYP","TC","Temperature Cone"
"SAMP_TYPE","CPT","Cone Penetration Test"
"GEOL_LEG","401","SAND"
"GEOL_LEG","201","CLAY"
"GEOL_LEG","999","Void"

"**HOLE"
"*HOLE_ID","*HOLE_TYPE","*HOLE_NATE","*HOLE_NATN","*HOLE_GL","*HOLE_FDEP","*HOLE_STAR","*HOLE_ENDD","*HOLE_CREW","*HOLE_EXC"
"<UNITS>","","m","m","m","m","","dd/mm/yyyy","","m"
"CPT1","SCP","261250.20","663699.90","11.15","14.908","16/10/2014","16/10/2014","BATT*TOMEK","GB7"
"CPT2","SCP","261263.90","663695.90","11.08","19.940","16/10/2014","16/10/2014","BATT*TOMEK","GB7"
"CPT3","SCP","261281.60","663688.50","11.05","16.192","16/10/2014","16/10/2014","BATT*TOMEK","GB7"
```

Figure App4-2: Selected example of a GSPEC compliant AGS file

The online portal was developed using Java and Restlet (http://restlet.com/) technologies, these provide a scalable solution for delivering user authentication and file transfer functionality. User account details are encrypted and stored in a simple Oracle database.

The decision to use these technologies was influenced by BGS developers having experience of similar projects to ingest digital files over the web and therefore had a template system in place which could be re-engineered to satisfy ASK Network requirements.

In order to protect BGS internal systems from potential security risks associated with incoming files from external donors the key elements of the solution were split across externally facing servers in a DMZ (Demilitarised Zone) and internal servers which reside on the BGS network and are protected by the corporate firewall. Files are replicated and moved between the DMZ and the internal network by the Windows file replication command Robocopy and antivirus checked using Sophos software.

To date no files have failed the antivirus check but this is probably due, in part, to the limited number of individuals who have permissions to submit files, if and when the service is opened up to a wider audience this may become a more significant step in the process.

Files which have passed the antivirus check are transferred by Robocopy into the BGS internal network for validation. There are currently at least three AGS validation checkers, each with their advantages and disadvantages

The software vendors Keynetix and Bentley both provide free AGS checking tools, which support multiple versions of the format. These tools are extremely helpful and provide users with a simple way of ensuring their AGS transfer format files are generally well structured. Unfortunately, these tools do differ in their interpretation of the AGS rules and a file checked in both tools could generate differing results.

The BGS developed an AGS validation checker using the spatial data transformation software FME, by Safe Software Inc., it checks AGS files for compliance to the GSPEC variation of the format. If the ASK Network initiative was rolled out to a wider user base it is likely that there would need to be validation of other versions of the AGS format. The BGS validation checker is currently only accessible to external users when they donate a file to the BGS, this functionality could be expanded to allow pre-submission checking.

The BGS validation checker parses each files to establish whether it contains all of the mandatory AGS 3.1 content, it also checks that there are XYZ coordinates for every point location (specified in the AGS group called HOLE). Where location details exist the FME workflow produces a Google Earth viewable KMZ file containing the borehole positions and interpreted lithology. Due to Google Earths limited ability to display subsurface features the locations of such boreholes have to be projected above surface as shown in Figure App4-3. Although the visualisation of the borehole locations in software such as Google Earth is not strictly necessary for the validation it is very useful in the subsequent verification step and is routinely used to check the quality of submissions. If the file passes validation it is moved to a dedicated folder containing valid but yet to be verified files, if it fails validation an email is sent to the file donor containing details of why the file failed and if possible it includes the KMZ file is attached.

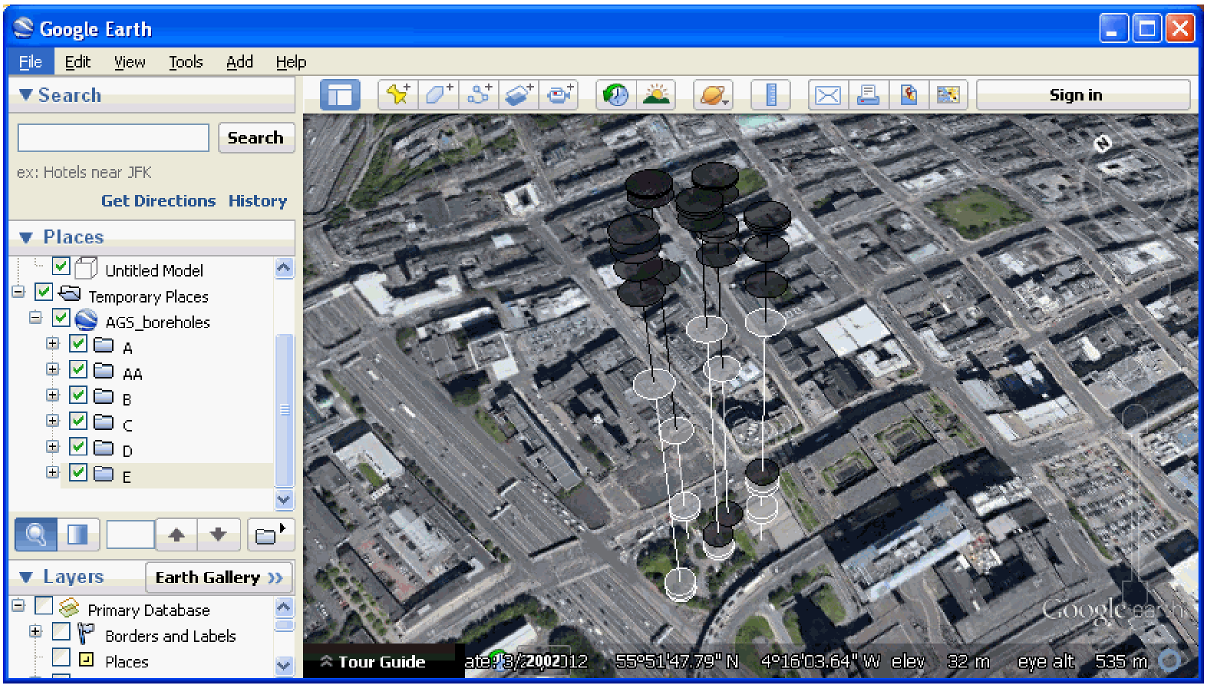


Figure App4-3: Position of boreholes derived from an AGS file using a FME workflow

Data verification is the next step in the process to take place and requires a suitably knowledgeable individual who is able to assess the quality of the data contents within the valid files.

Verification involves checking:

- Is the data is a duplicate of a previous submission?
- Are locations correctly positioned as expected?
- Is the data appropriate for the BGS to store?

If a file fails the verification step it may be possible to rectify the issue by contacting either the file donor (GCC) the client who paid for the data to be collected (also GCC in most ASK network examples) or the contractor who carried out the work to collect the data. If necessary though the file may simply be deemed removed as valueless and deleted.

Valid and verified files are stored in network file store with a cross reference made in the metadata record held in a relational database, where possible metadata information is derived automatically from file contents, donor account information (identified through the web portal authentication function), however there are a number of attributes which are filled in manually.

All GSPEC submissions are associated with a pre-defined high level spatial metadata record, this describes the geographical bounding box for the Glasgow area, describes the file format used and identifies the main contact points for the datasets. This enables all GSPEC records to be easily identified in a number of spatial data discovery portals

The BGS spatial metadata database has been developed over a number of years and is based upon international standards such as ISO 19115 and the Inspire directive, extra attributes have been added over the years to satisfy local requirements but the core information is mapped to the international standards. A freely available version of the BGS metadata database has been published on the Data Models Knowledge Exchange website EarthDataModels.org website:
http://www.earthdatamodels.org/designs/metadata_BGS.html

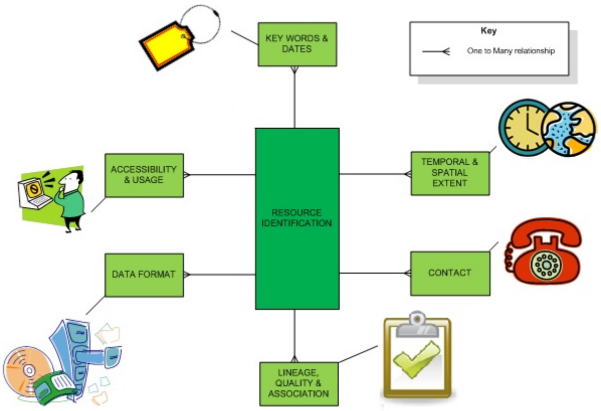
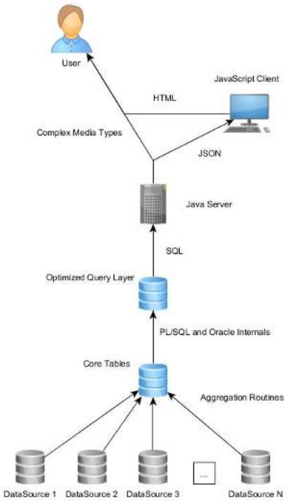


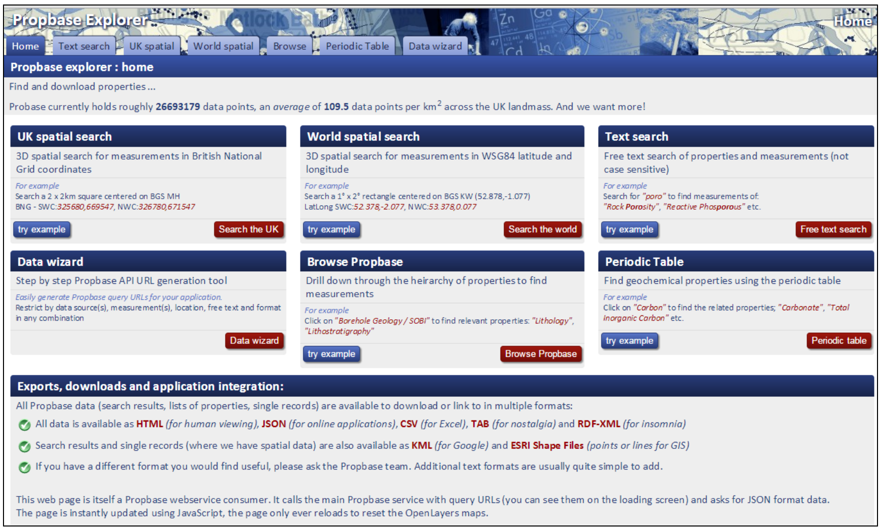
Figure App4-4: High level data model for Metadata, designed to meet international spatial metadata standards such as the European INSPIRE Directive and ISO 19115.

The BGS use a centralised Oracle relational database and a networked file storage facility as the official data store.

The Geotechnical database (Self et al., 2012) is compatible with the data model for AGS and is linked into the BGS database for all onshore boreholes. Like the metadata database a public version of the BGS borehole data model is available on EarthDataModels.org and the Geotechnical database will be released mid to late 2015.

By developing the appropriate infrastructure, the BGS have been able to disseminate the metadata through to many data discovery portals, for example the British Geological Survey maintain a spatial metadata database which feeds into the BGS discovery metadata website (http://www.bgs.ac.uk/discoverymetadata/home.html), NERC Data Catalogue Service (http://data-search.nerc.ac.uk/search/full), the UK government open data search facility (http://data.gov.uk/data/map-based-search) and the Europe wide INSPIRE GEOPORTAL (http://inspire-geoportal.ec.europa.eu/). In addition to relatively basic metadata discovery portals it has been necessary to develop ways to discover and visualise the data itself, one system which has been developed for internal BGS use, but which might be adapted for a community such as the ASK Network, is the Propbase Query Layer (Kingdon et al., 2010). The Propbase architecture is designed to pull together all subsurface property data, their locations, property type and property values into a single query layer, this query layer is then exposed in a variety of search facilities.





Propbase explorer : home

Find and download properties ...

Probase currently holds roughly **26693179** data points, an *average* of **109.5** data points per km^2 across the UK landmass. And we want more!

UK spatial search

3D spatial search for measurements in British National Grid coordinates

For example
Search a 2 x 2km square centered on BGS MH
BNG - SWC:*325680,669547*, NWC:*326780,671547*

try example | Search the UK

World spatial search

3D spatial search for measurements in WSG84 latitude and longitude

For example
Search a 1° x 2° rectangle centered on BGS KW (52.878,-1.077)
LatLong SWC:*52.378,-2.077*, NWC:*53.378,0.077*

try example | Search the world

Text search

Free text search of properties and measurements (not case sensitive)

For example
Search for *"poro"* to find measurements of:
"Rock Porosity", *"Reactive Phosporous"* etc.

try example | Free text search

Data wizard

Step by step Propbase API URL generation tool

Easily generate Propbase query URLs for your application.
Restrict by data source(s), measurement(s), location, free text and format in any combination

Data wizard

Browse Propbase

Drill down through the heirarchy of properties to find measurements

For example
Click on *"Borehole Geology / SOBI"* to find relevant properties: *"Lithology"*, *"Lithostratigraphy"*

try example | Browse Propbase

Periodic Table

Find geochemical properties using the periodic table

For example
Click on *"Carbon"* to find the related properties; *"Carbonate"*, *"Total Inorganic Carbon"* etc.

try example | Periodic table

Exports, downloads and application integration:

All Propbase data (search results, lists of properties, single records) are available to download or link to in multiple formats:

✓ All data is available as **HTML** *(for human viewing)*, **JSON** *(for online applications)*, **CSV** *(for Excel)*, **TAB** *(for nostalgia)* and **RDF-XML** *(for insomnia)*

✓ Search results and single records (where we have spatial data) are also available as **KML** *(for Google)* and **ESRI Shape Files** *(points or lines for GIS)*

✓ If you have a different format you would find useful, please ask the Propbase team. Additional text formats are usually quite simple to add.

This web page is itself a Propbase webservice consumer. It calls the main Propbase service with query URLs (you can see them on the loading screen) and asks for JSON format data.
The page is instantly updated using JavaScript, the page only ever reloads to reset the OpenLayers maps.

Figure App4-5: Propbase architecture and Propbase Explorer search facility homepage

Well structured, fully described and discoverable data still requires additional processing before it is readily usable for purposes such as spatial analysis in GIS software or the construction of 3D models.

Data has to be transformed into those required by the GIS and 3D modelling community from that held in the Oracle relation database. A number of procedures have been developed over the years which involve the use of FME, bespoke Java applications and manipulation of csv files (comma separated values). These procedures continue to develop as the requirements of the BGS and ASK Network community evolve.

The BGS has developed a Java software prototype which proves that it is possible to generate AGS compliant files from the Geotechnical database, regardless of the format the data was originally provided in. The intention is to develop a public webservice that would enable individuals and software vendors to use this service to search for geotechnical data, view PDF scans of any drill logs and download the data in AGS format.

Common 3D modelling practices tend to consume data, such as that held in the Geotechnical and Borehole database, without feeding back any information. This is changing and organisations like the BGS and GiGa infosystems (http://giga-infosystems.com/) are developing relational database systems which can store and manage data produced as part of the 3D modelling process. Modelled geological data such as cross sections or 3D surfaces can be versioned and associated with all the key source data held in the communal data store along with the relevant metadata. It is anticipated that by integrating the source and interpreted data it will be possible to ensure consistency between these datasets. When a borehole record is exported from the database and the data contained in the file is altered in response to new insights gained during 3D modelling it is possible that the database is not updated as it should be. Through the new integrated database designs a more iterative approach to data capture > data store > data use and data update can be achieved.

3D models derived mostly from the geotechnical data (primarily borehole log descriptions) are freely available to community members using an online tool that users could use to generate synthetic boreholes and cross sections via a map interface, opening the data and derived information to a wider, non-expert, audience.

The feedback from stakeholders to the workflow and systems put in place has been very encouraging as illustrated in the following testimonials:

Jane Morgan, Deputy Director, Digital Public Services at The Scottish Government stated at the ASK Network workshop, Glasgow City Chambers, 4th March 2014:

"The ASK Network and GSPEC data formats fit well with the Scottish Government's spatial information priorities, namely promoting the wider exploitation of spatial data, efficient hosting and data sharing to encourage use and collaborative projects across the public, private and academic sectors."

Iain Hall, Technical Manager of Grontmij Edinburgh, stated at the same event:

"For modest initial investment in training and communication with contractors on site, applying GSPEC has led to significant benefits downstream including improved ability to handle large data sets, reduced financial risk to design and ability for rapid 3D visualisation of data. These have given improved confidence of ground conditions and enabled closer attainment of optimum design."

Whilst Jackie Bland, Geotechnics Ltd and Chair of the AGS Data Management Working Group provided the following confirmation of how well the GSPEC scheme aligned with the AGS vision:

"GSPEC is AGS format really being used in the way it was originally intended to be."

Garry R. Baker (Head of National Geoscience Data Centre, UK) provided the authors of this report with the following comment in February 2015 *"The National Geoscience Data Centre holds geoscience data for professional long-term management, storage and future re-use. In recent years we have been streamlining the ingestion processes into a common workflow to best utilise our resources and more strongly align to the requirements for digital data ingestion to support future multi-disciplinary science. The ASK network (with its GSPEC/AGS standard) has provided a user community ready to fully engage with digital ingestion, both to provide appropriate geoscience data and to help work through workflows, stages and processes."* .

Identification of knowledge gaps

This section compares the gap between the Sub-Urban community's actual performance against the potential performance with a focus on digital data acquisition, validation and verification.

- Lack of automated ingestion processes for almost all of the priority datasets
    - Standard exchange formats
    - Validation of exchange formats
- Lack of knowledge about the structure and meaning of datasets from 'external organizations'
    - Data models
- Available or potentially available datasets are sometimes hidden
- Lack of common terminologies

ID	Current State	Desired State	Gap Description	Gap Reason	Remedies
1	Standard exchange formats for geotechnical data in use in a small number of cities	Common standards used across all dataset themes identified as high priority by city partners	Many cities that could benefit from AGS for geotechnical data are not using it.	Some cities are simply not aware of the standard, whilst others may consider it an unnecessary expense.	Provide free and open case studies which illustrate the cost-benefit of implementing such a standard and provide guidance for interested parties as part of WG3 toolkit.
2	Most Sub-Urban community stakeholders store and exchange data in a wide range of bespoke and proprietary formats	All Sub-Urban community stakeholders have access to and make appropriate use of standard exchange formats for the data which are seen as high priority by city partners	Some standards for data storage and data exchange formats exist for groundwater data, tunnels, utilities, pollution data, land use and surface features .	The standards which do exist have not become common practice, yet. This is partly due to a lack of evidence justifying the cost of adopting such standards.	By highlighting the need for greater collaboration and documenting potential efficiency gains could provide the justification for greater adoption of standards.
3	Automated validation of AGS data is performed by tools develop by commercial software vendors or the BGS	A free validation tool which is developed and maintained by the AGS	Commercial companies have developed tools which interpret the standard differently. The BGS have also	The AGS standards committee cannot or will not fund the development	Either: i. AGS committee to develop such a tool, or ii. AGS committee

		standards authority or wider community	developed a validation tool but it is not as robust as it could be & is not available externally	of such a tool	could formally approve a tool developed by another body
4	Many of the GSOs and other community data centres involved in Sub-Urban activities ingest their data in labour intensive processes.	Seamless integration of distributed data generation and management data repositories.	Analogue data donations require manual processing and there is no wide spread automated validation and verification tools for digital data.	Many data donations are received in analogue formats or if digital the systems used are owned by different organisations leading to data silos.	Data centres and commercial software vendors should work together to develop data exchange interfaces such as the development of APIs to integrate BGS data with geotechnical software HoleBase.

Suggestions for research and development

The ASK network / GSPEC workflow and systems architecture described in the previous section illustrate how multiple organisations can work together to develop a bespoke solution to satisfy communal requirements on a city wide scale. If such as solution were to be re-applied to another city there are a number of city based as well as universal considerations which need to be explored to incorporate local requirements into tailored implementations.

Highlights of key technical requirements for the future

One of the most useful tools that could be developed for the community would be a freely available AGS committee approved validation checker, if such a tool could be integrated into digital acquisition workflows and support the integration of geotechnical data from a range of commercial and public sector sources.

Initiatives such as the BIM for subsurface project should be expanded to provide open APIs that enable the linking of systems and datasets from communal data centres and industry stakeholders.

Topic 3 References

Open BIM http://www.designingbuildings.co.uk/wiki/Open_BIM Chadwick, N Pickles, A and Sekulski. E. 20016. Data transfer and the Practical Application of Geotechnical Databases. GeoCongress 2006, 1-6. doi: 10.1061/40803(187)110

Bland, J, Walthall and Toll, D. The development and governance of AGS Format for geotechnical Data. Proceedings of the 2nd International Conference (ICITG). Information Technology In Geo-engineering, editors D G Toll, H Zhu, A Osman, W Coobs, X Li and M Rousainia. 67 to 74. IOS Press BV, Netherlands.

Campbell, Seumas, and Helen Bonsor. "The ASK Network: Glasgow Specification of Data Capture." (2013). http://nora.nerc.ac.uk/506350/

Kingdon, Andrew, Martin Nayembil, Keith Holmes, and Graham Smith. "PropBase QueryLayer: a single portal to UK physical property databases." (2010): 1-2.

Self, Suzanne, David Entwisle, and Kevin Northmore. "The structure and operation of the BGS National Geotechnical Properties Database. Version 2." (2012). http://nora.nerc.ac.uk/20815/

Power, C M, Patterson, D A, Rudrum, D M and Wright, D J. 2012. Geotechnical asset management for the UK Highways Agency. In Earthworks in Europe, editor T Radford, Engineering Group of the Geological Society Special Publication 26, 33-39.

Walthall, S and Palmer M. 2006 The development, implementation and future of the AGS data formats for the transfer of Geotechnical and Geoenvironmental data by electronic means. GeoCongress 2006: pp 1-4 doi: 10.1061/40803(187)109

Appendix 5: Managing permissions and roles

Managing permissions and roles across shared distributed database systems:

Introduction

Using the experience of the Danish systems which are used by the municipalities within Denmark this topic will describe the technical architecture and constraints which are required to administer a system that involved many users of different roles across a range of organisations and administrative levels.

Current state

Sharing of data between different levels administration of as well of administrative units on the same levels does not seem to very common. This is often due to history of data collection where different units have been focusing on different parts of de data and have established different local data models making data sharing difficult. In some countries the legislation also hinder sharing of data as the data owner often is the company generating the data the one ordering the data. In this document the Danish model where part of the data sharing is enforced by legislation other parts voluntary is described as an example of god practice.

Good practice

Danish shared public database

Sharing of information's between different public units can seriously make the amount of data available to each of the units much larger.

During a restructuring of the public administration in Denmark in 2007 the handling of groundwater, drinking water and soil pollution data was transferred from the counties (which were closed) to municipalities, regions and state agencies. In this process data concerning groundwater and drinking water was moved to the Geological Survey of Denmark and Greenland (GEUS) where a national database for geology, groundwater and drinking water were established by expanding the surveys database Jupiter. The data were made freely available for all actors in the field of groundwater and drinking water. At the same time all data (except a few tables containing personal information) was made publicly available.

Access to data

The intensions were to establish a central primary data storage where all actors working with groundwater and drinking water could access updated information about environmental boreholes, groundwater and drinking water data. When the system was designed it was decided to make the central database available through SOAP web services and leave it to the market to develop the systems needed in the public administration. Besides the web services a download system was also developed to allow download of data in the same format as made available as web services. This was done to allow complex queries not possible through web services.

Data content

In advance of the restricting process the data content and dataflow requirements were outlined, the initial scope of the data content was completed by January 1st 2007 but has been extended several times, it now covers:

- Laboratories send chemical data (water, air and soil) to the central database. Data must be validated by the data owner (municipalities, regions and state agencies) before coming public available
- Municipalities update the database with drinking water data, including:
    - Plants and well fields
    - Permits for water extraction
    - Water level measurements
    - Permissions for private (1 to 10 users) water plants
    - Report yearly yields, exchange of water between plants
    - QC of chemical data from drinking water, soil pollution and monitoring data for drinking water wells
    - Entering confidential boreholes
- State agencies – taking acre of groundwater mapping
    - QC of chemical data from groundwater monitoring, soil pollution
    - Water level measurements
    - Entering confidential boreholes
- Regions – taking care of soil pollution mapping and clean up
    - Remediation plants
    - QC of chemical data related to soil pollution
    - Water level measurement
    - Entering confidential boreholes
- The Geological survey update the well part of the database
    - Administrative data
    - Technical data
    - Geological description
    - Water level measurements
    - Entering non-confidential boreholes
- Advising companies including drilling companies
    - Entering boreholes, both confidential and non-confidential

In the initial plans it should be possible for all public authorities to update coordinates and the use of the different wells in the database. This was changed at the last minute as some of the developers decided to use local databases. It was decided to allow only one administrative unit at the time to be allowed to update single cords in the database.

In the time up to and after January 1st 2007 three systems were developed. It quickly became obvious that not all of the systems developed would conform to the plans for a primary central database. One of the systems was made to run on local databases, one to work on a semi central database located at system developers and one system working entirely online (but capable to run on downloaded data for advanced data queries).

To manage this situation is was necessary to change user permissions to ensure that each post in the data model could be changed by only one administrative unit at the time. This was necessary for the synchronisation of the different databases to function.

Data ownership

Data are owned by the authority entering or responsible for the data the data.

Permissions and roles

The user management used for the system is an ADSF2 system made available by the Danish Natural Environment Portal (http://www.miljoeportal.dk/English/Sider/default.aspx) which also pays GEUS to keep the system running, maintain the database and help users and system developers. The Environmental Portal is owned by the Danish state 45%, the municipalities 45% and the regions 10 %. The Environmental Portal manages a lot of system used by public authorities in Denmark. The creation of users and management of their privileges are maintained locally. As a system provider we at GEUS don’t know the users before we receive a request with a token that can be validated against the Environmental Portal. The token contains information’s about the user, such as:

1. User name
2. Name
3. E-mail
4. Authority id (identifying the authority the user belongs to)
5. A list of roles

If the request is allowed, (if the user comes from an administrative unit allowed to enter this type of data and if the user have been granted the needed privilege) the request is executed, if not the user receives an error. All inserts and updates are marked with administrative unit and username.

Types of users and roles currently implemented in the system:

	Municipalities	Regions	State agencies	Advisers	GEUS
Water resources (maintain own data concerning drinking water)	X				
Water level soundings (maintain own water level measurement data)	X	X	X	X	X
Groundwater chemistry validation (release own groundwater chemical data)	X	X	X		
Drinking water chemistry validation (release own drinking water chemical data)	X				
Borehole read (read all borehole information)	X	X	X	X	X
Borehole (maintain own borehole information)	X	X	X	X	X

To access the common database a suite of SOAP web services was developed. Each of the requests was then mapped to one or more of the roles. As an example the insertAirSample request can be used by anyone having either Borehole or the Laboratory permission. The requests are used to insert an air sample from a borehole, a surface sample or a plant (water works / remediation plants). For administrative units covering only a part of Denmark the permissions are restricted to samples within the borders of the administrative unit.

Legislation

When the system was made available, January 1st 2007, it was preceded and followed by several legislative acts. These stated which data should be delivered to the public database by whom. The change compared to earlier legislation was mainly that the report should be done continuously as opposed to once a year. In later changes to the legislations more and more time limits were defined for data entry.

Data agreement form

In the data agreement form all the data in the common database was listed with reference to the responsible administrative unit(s) and legislation and a list of users needing the type of data. As not all types of data are covered by legislation, some of the data are entered voluntarily by the administrative units.

Clients for the public administration

When the system was laid out it was decided that the Environmental Portal should only deliver services to read, edit and delete data and leave it to the private sector to develop the systems needed by the different sectors of the public administration. Now more than 8 years after the reform there are 4 systems used in the municipalities. One of these is also used in the regions in the state agencies. With 98 municipalities of which about 5 does not have a system at all the marked for this kind of systems it is rather small. As a resulted and new features are often introduced slowly if at all. The state agencies and regions have had to pay the developer of their system to have get a system with the needed features..

This decision only to deliver a web service interface to the central databases has later been changed. For new systems a simple web interface must be created with the services to ensure that all public administrations can do what they are mend to do.

Other types of access to data

Besides making data available through SOAP web services for professional users the data are made available in the following ways:

- Web pages - Form the surveys homepage, most of the data is available for lookup
- Web Map Services and Web Feature Services - Different thematic maps are exposed as WMS and WFS allowing users to show data in local GIS
- Download - For advanced use as for example data geological modelling data are available for download as a relational database
- Synchronisation of local databases (one way survey –> user) - GEUS has recently been paid by one of the larger engineering companies to develop services to keep an external database updated with changes made to the central database. This service will be public available and allowing private companies and local authorities to have an updated local (read only) database for advanced use.

Experience with the system

Easy and public access to data for all users including public administration, water works, advisory companies and all other stakeholders working in the geology, hydrology drinker water.

Updated data becomes publically available as soon as they are entered into the database.

As data are available on a common format (defined by the municipalities, regions, state agencies and GEUS in cooperation) it is cheaper conduct studies of new areas.

The public defined formats are widely used in both the public and private sector. Public authorities as well as private companies now pay for further development of the system for the benefit of all users

Danish Railroads have decided to pay for an extension of the system including extension of one of the systems used by private companies and administrative units with geotechnical data and advisor access to. This extension will be available for all users.

The process of changing the drinking water administrative system was only a very small part in the restructuring and received rather little political interest. Also the change was done in a hurry resulting in rather late change in legislation and very late description of how people working with ground- and drinking water should behave. In one case the description of how to handle the data was made public two years after the new structure was in effect.

The decision to leave it to the marked to develop the needed systems has only been partly successful. Some of the developed systems do not meet the needs of the users. Now more than 8 years after the system were introduced not all systems allows the users to do what they must in order to comply the legislation.

The use of local databases in one of the system makes data corrections and synchronisation hard.

Unclear and complex legislation have resulted in many discussions about who – if any – are responsible for different types of data.

Identification of knowledge gaps

The Danish system rely to a high extend on voluntary reporting of data to the local databases. The experience with the system is that not all sectors of the public administration put the same efforts in data sharing and a large number of data does never reach the central databases. Where the reporting of data is based on legislative act the system works much better but not perfect. This is partly due to the fragmentation of knowledge about drinking water and groundwater as a result of the restructuring. The peoples working with these data were transferred from 14 counties to 98 municipalities, 7 state agencies and 5 regions. This have resulted in – in places – a rather thin coverage of the subject area.

Highlights of key technical requirements for the future

For the Danish system most of the needed functionality is available for the public authorities. The data model developed very fast in 2006, had to include data from three different systems and where decisions about who should be responsible for the different parts of the data set needs a clean-up. Part of the data model allows the users to do the same thing in different ways. Also the need at state level for data e.g. to report to the EU system is impossible as this needs was not considered during the establishment of the data model.

Where the data flow could be optimized is the flow of data from e.g. water works to the central databases.

Suggestions for research and development

As the system is built on Danish standards it is not compliant with INSPIRE. Users from other countries will have a hard time using the data. All the code lists translates to Danish and the comments are added solely in Danish. The model with a central national database will however make this huge task of harmonizing the data model and code lists easier as the import export functionality only have to be developed at one place. That the Inspire model is achievable have, for the export functionality, been demonstrated in small scale in the eEarth and eWater EU projects.