Project Number: 032518

QVIZ

Query and context based visualization of time-spatial cultural dynamics
Specific Targeted Research Project
Information Society Technologies

Administrative Unit Ontology Report and Schema D3.2

Due date of deliverable: 29/02/2008
Actual submission date: 04/04/2008

Start date of project: 01/05/2006
Duration: 24 month
Umeå University
Final
Abstract

Number and name  Project Number: 032518
QVIZ, Query and context based visualization of time-spatial cultural dynamics

Work Package  WP3

Task  D3.2

Date of delivery  Contractual: 29/02/2008    Actual: 04/04/2008

Code name  032518    Version  draft ☐  final ☑

Nature  Report and schema for the administrative unit ontology

Distribution Type  Public

Authors (Partner)  Paula Aucott, Paula.Aucott@port.ac.uk
Vojtech Kupca, Vojtech.Kupca@humlab.umu.se
Johan Lagrelius, Johan.Lagrelius@humlab.umu.se
Alexander von Lünen, Alex.von-Luenen@port.ac.uk
Fredrik Palm, Fredrik.Palm@humlab.umu.se
Humphrey Southall, Humphrey.Southall@port.ac.uk

Contact Person  Paula Aucott, Paula.Aucott@port.ac.uk

Abstract  The administrative unit ontology for the States of Europe, including detailed histories of administrative units in Great Britain, Sweden and Estonia, which acts as the framework behind the QVIZ user interface.

Keywords List  Administrative units, ontology, database, framework, history, vectors, mapping, domain-ontology, Europe, Estonia, Great Britain, Sweden
# Table of contents

Summary .......................................................................................................................... 5
Objectives ......................................................................................................................... 6
1. Requirements .................................................................................................................. 7
2. Use of the administrative ontology ................................................................................ 8
3. Schema Section ............................................................................................................... 9
   3.1 Background ................................................................................................................ 9
   3.2 Incorporation of recognized standards ....................................................................... 9
   3.3 Entity-Relationship Diagram ................................................................................... 11
   3.5 Administrative Unit Type Relationships Diagram for isPart of relations ... 13
   3.6 Table Description ...................................................................................................... 14
      3.6.1 General Comments ............................................................................................ 14
      3.6.2 General Principles ............................................................................................ 14
   3.7 Framework structure ................................................................................................. 15
      3.7.1 Levels ................................................................................................................ 15
      3.7.2 Types .................................................................................................................. 15
      3.7.3 Statuses .............................................................................................................. 15
      3.7.4 Relationships .................................................................................................... 16
      3.7.5 Boundaries ....................................................................................................... 16
      3.7.6 Jurisdiction ....................................................................................................... 17
      3.7.7 Function ............................................................................................................. 18
      3.7.8 Multiple Languages .......................................................................................... 19
      3.7.9 Authorities ......................................................................................................... 19
4. Data Section .................................................................................................................... 21
   4.1 Introduction to the data .............................................................................................. 21
   4.2 The World data ......................................................................................................... 21
   4.3 European states data .................................................................................................. 22
   4.4 The Estonian data ...................................................................................................... 22
      4.4.1 Dates .................................................................................................................. 23
      4.4.2 Units from Vectors ............................................................................................ 23
      4.4.3 Multiple units with the same name and type ....................................................... 23
      4.4.4 A single root ...................................................................................................... 24
   4.5 The Swedish data ...................................................................................................... 24
   4.6 The British data ................................................................................................-------- 24
4.7 Relationship to Archival Resources

5. Toolkits for validation

5.1 Methodologies employed

5.1.1 Prototype 1

5.1.2 Prototype 2a

5.1.3 Prototype 2b

5.1.4 Text-only Interface

5.2 Administrative unit creation

5.3 Vector creation

5.4 Scanned mapping

5.5 Quality Assurance

6. Conclusion

Appendix 1

A1.1 Administrative Unit Levels

Appendix 2

A2.1 Representation of some administrative units levels 1-5 in the ontology

A2.2 Example of Estonian unit type levels 4-12 showing relationships of Tartu and Kursi

A2.3 Example of British unit type levels 5-12 showing relationships of Hampshire and Newport

A2.4 Example of Swedish unit type levels 4-12 showing relationships of Uppsala län and Uppsala

Appendix 3

A3.1 Statistics relating to Administrative unit content

A3.1.1 Count of units and polygons per unit type where some units exist

A3.1.2 Histogram showing the numbers of polygons over time

A3.1.3 Count of number of names per language

A3.1.4 Count of units and names per jurisdiction

A3.1.5 Count of units with names in two or more languages by authority

A3.1.6 Count of relationship types

4
Summary

The QVIZ European administrative unit ontology was constructed by the Great Britain Historical GIS team based at the University of Portsmouth (UoP), and one of the main starting points was the record of British administrative units they had previously constructed. However, the QVIZ ontology also integrates data from the Swedish and Estonian National Archives recording those countries’ evolving administrative systems, and a newly constructed record of the changing states of Europe, and their membership of international organizations such as the European Union. To hold information covering the whole of Europe, to varying levels of detail, the earlier British system was extended to cope with multiple languages, different coordinate systems and with different historically significant time periods. Once defined and populated the administrative units ontology database acts as the integrating component beneath the interface that connects the administrative unit history, the raster and vector mapping and the archival documents from disparate sources into an integrated whole.
Objectives

To design an ontology of European Administrative Units defining temporal and spatial relationships, and so support dynamic map-based visualization.

The ontology was designed to record complex relations in time and space and the interdependencies between administrative units at different geographical levels and over time. This structure enables users to access archival resources by place name or location, without having to know the detailed administrative history of individual places. The ontology identifies the evolving states and nations of Europe during peacetime over nearly two hundred years. There is much finer granularity for Sweden, Estonia and Great Britain, incorporating data down to parish/village level data created by QVIZ partners either directly for the QVIZ project or in the past. Where older data have been used, significant further work has been done to fit the data to the ontology structure.

Recorded within the ontology are the variant names, sources and (poly-)hierarchical relationships for all units, together with dates and spatial data where they exist. The ontology is also designed to allow for multi-lingual unit typologies and relationships. GIS content is integrated into the main body of the system to contextualize and locate the data in a geographical sense, rather than separately as in conventional GIS software packages. However, the flexibility of the framework is such that actual geographical co-ordinates are not necessary for the administrative unit to exist within the framework.

The use of the ontology as the core element also allows the user to see the exact contents of the archives, which are related to their place or area of interest within the involved Archival Institutions. These links to documents reduce research time significantly for the user, as instead of having to search through all the archives manually to identify potentially relevant resources, the pertinence to the area of investigation is immediately obvious. Users can communicate with one another via the ontology to further their individual projects and extend their knowledge. This is achieved through interaction with other users using the Communities of Practice. Not only can the users bookmark and repeatedly refer to the same administrative units and archival documents, they can liaise with like-minded individuals about the context and historical data associated with their area(s) of interest.
1. Requirements

The administrative unit ontology acts as the core of the system, providing a framework through which the diverse sources of data can be displayed together in a single visualization interface.

An administrative unit is, for the purposes of this project, a public sector corporate body with an area of operation defined by law. The vast majority of administrative units can be grouped into unit types, such as forms of parish or district, which are themselves defined by law. To be held within the QVIZ ontology, an administrative unit must be assigned to a unit type; it must have at least one officially preferred name, and it must have at least one relationship of type IsPartOf with a higher-level unit (the sole exception to this last requirement being the root unit, representing the world); and each item of information must be drawn from a recognized authority. All information about a particular administrative unit within the ontology is linked together by a unit ID number, not by names. Where further information is available, the ontology can hold any number of additional names, many different relationships, which can include textual information about boundary changes, and boundary polygons.

Developed specifically for QVIZ the administrative unit ontology identifies all European States, which have existed in peacetime Europe since The Congress of Vienna in 1814-15. This list was compiled using a combination of authoritative sources to create a definitive list of states for use within the project.

The ontology is designed to be able to grow into a definitive (place) name authority for Europe. A schema for adding additional data from hypothetical new partner countries has also been developed. This exists in both a minimum requirements format and an extended version where fuller data is available. There is no simple way to insert new data so original partners would have to be involved in the insertion of any new partner data as the process is too complicated to implement as an automated procedure.

The ontology is also designed to support a range of evolving data standards, including the Alexandria Digital Library Gazetteer Content Standard ADL GCS) and Encoded Archival Context (EAC).
2. Use of the administrative ontology

The administrative unit ontology is designed to act as the fundamental core of the QVIZ framework. The regimented structure acts as a support mechanism for dynamic spatial-temporal visualization operations within a map interface and for the Faceted Browser search functionality. The ontology helps to contextualize the information held within the social knowledge content environments as well as holding textual content describing the history of individual administrative units.

Within the faceted browser the administrative unit ontology allows for certain kinds of functionality. Firstly the system allows the QVIZ user to make Administrative Unit selections using the faceted browser facility. Secondly the system permits the QVIZ user to select administrative units through the map. Thirdly the system permits the QVIZ user to swap facets within the faceted browser, as well as adding new facets and removing existing ones. Fifthly the system allows the QVIZ user to select a Community of Practice facet. Each of these separate functions requires a standardized internal system to be in place behind the external interface in order for the operations to be successfully completed.

Users can define different filters in their queries using the Faceted Browser, which makes it possible for them to arrange their own search hierarchy and narrow down searches by combing administrative units’ relations with information from social bookmarks of archival resources. Archival resources can be accessed by redirects to archival institutions user interfaces.
3. Schema Section

3.1 Background

Although the administrative unit ontology can support much conventional GIS functionality, the underlying data architecture is very different and could not be implemented using GIS software. Because every administrative unit is required to have hierarchical relationships with other units, but can optionally have associated locational data, it must be implemented in an object-relational database such as Oracle or, for QVIZ, Postgres: hierarchical relationships are held in classic relational structures while the locational data require a spatial extender, such as Oracle Spatial or PostGIS, implemented using the database system’s object extensions.

Providing detailed boundary coordinates for every unit would be very expensive, and the necessary historical maps and other sources may not even exist. However, the minimum data requirements for an administrative unit in our ontology – a type, a name, and a relationship with a higher-level unit – can be met from the most basic taxation listing. Linking units together via explicit semantic relationships creates a web of interconnections which works very well with search engines, so leading users to whatever digital resources cultural heritage institutions link into the structure. Further, this hybrid data structure makes the best possible use of available locational data, as most units without digital boundaries can be assigned approximate locations based on their relationships with units that have boundaries. During the QVIZ project, we have integrated together spatial data from diverse sources, both static snap-shots from Estonia and time-variant vector files from Sweden, into a single dynamic cartographic interface. This system makes much more effective use of the digital resources that already exist than a GIS.

In the GB Historical GIS entities were given a fixed unit type. However within that unit type a unit could gain and lose statuses. Using only status or deciding entity characteristics from current relationships would be more flexible, allowing provinces to become independent states and vice-versa, but despite the clear advantages of employing this approach the complexities of integrating the geospatial content proved particularly intellectually challenging and the relatively short time-span of the QVIZ project and the high concentration of effort and man-hours this would require to be realized meant that this aspect of the project was left for consideration at a later date.

Although the framework for QVIZ is based on the pre-existing Vision of Britain system there are also significant differences. The ability to create a multi-lingual user interface, a process to select the preferred language of the names to suit the user, and the combination of data from many different kinds of administrative geographies into a single consolidated whole. It was designed to solve problems with displaying and mapping information in a user interface from diverse administrative unit types even when no actual polygon data exists.

3.2 Incorporation of recognized standards

Prior to the development of the QVIZ administrative units ontology there existed many on-line gazetteers, but most are simple place-name lists, without hierarchy or variant names, and are not based on recognized standards. The few systems designed by information scientists as (place-) name authorities are still weak in
their inclusion of geo-spatial data. The Fifth Framework LEAF project implemented EAC (Encoded Archival Context), designed for places as well as persons but excluding spatial coordinates. Getty’s Thesaurus of Geographical Names is rigidly hierarchic, and did not try to set or follow standards.

Two previous attempts have been made to define gazetteers standards. The Open Geo-spatial Consortium’s (OGC) work lacks strong support and is arguably too GIS oriented. The Alexandria Digital Library (ADL) Gazetteer Content Standard, Feature Type Thesaurus and Gazetteer Service Protocol together form the most fully developed standard set, incorporating geo-spatial data via OGC elements, and complex relationship typologies, chronologies and variant names. The ADL’s own gazetteer mostly contains US public data with point co-ordinates, no hierarchy, no chronology and one name per feature. It also does not distinguish between officially recognized administrative units and the much looser concept of a place. The British system more fully supports the ADL content standard: locations recorded via multiple date-stamped polygons; poly-hierarchical relationships; names with status, language and dates. It has been developed in consultation with the creators of the ADL gazetteer in an attempt to extend the standard in a methodical way. It also makes a clear differentiation between an administrative unit and a place and systematically maintains the association between them.

The QVIZ research built further on ADL standards as well as the earlier work of the Swedish National Topographic Database upon which the British system was itself based. Each of these systems dealt only with the intricacies of the existence and relationships of units within individual states in isolation. For QVIZ the changing international map, in particular the appearance and disappearance of states and the dilemma of accommodating and displaying multiple languages for labels as well as names presented a new challenge. It also dictated the need of a more generalized lists of typologies and the broadening of legal entity type relationships within that list.
3.4 Ontology Diagram
3.5 Administrative Unit Type Relationships Diagram for isPart of relations
3.6 Table Description

3.6.1 General Comments

The administrative unit ontology consists of six main data tables, plus a larger number of metadata tables. The data tables contain tens of thousands of rows, whilst the metadata tables contain tens or hundreds. The structure implements a highly abstract data model, which can accommodate an infinite range of entities and their relationships. The most important data tables are \texttt{g\_unit}, which defines all the entities in the ontology, and \texttt{g\_rel} that holds all the relationships between them. The most important metadata tables are \texttt{g\_unit\_type}, which defines what kinds of entities can exist in the system, \texttt{g\_rel\_type}, which defines what kind of relationships can exist, and \texttt{g\_legal\_rel}, which defines what kinds of relationships can exist between which kinds of unit. This is an object-orientated database, although there are only two kinds of object both of which are integral to how the structure meets its requirements. They are the spatial objects, which contain co-ordinate data and are located in the \texttt{g\_foot}, \texttt{g\_unit} and \texttt{g\_place} tables, and the date objects, which contain dates of varying precision. The inclusion of spatial objects makes this structure a kind of GIS. The \texttt{g\_place} table is included in the diagram, although little use is made of this within the QVIZ structure. It is used to group together units into a smaller number of \textit{places}, and it is constructed by a piece of software, which looks for units with \textit{more or less} the same name and \textit{more or less} the same location. The current software for generating the \texttt{g\_place} table was designed to work on British data, and has not been modified for use with European data, as it was not deemed necessary for the QVIZ interface.

3.6.2 General Principles

While in other situations one would have had to impose a typology, administrative units are usually defined by law to have types, so the job therefore was to record and arrange those existing legally-defined types, not invent a new typology.

Within QVIZ there were two problems with language, which made this topic especially hard to discuss:

1. This project is working across countries, which use many different languages, although the project works entirely in English.
2. The second problem is that within the existing British system, the words \textit{level}, \textit{type} and \textit{status} have precise meanings, so the word \textit{kind} is used here when not describing one of those precise terms.

What was done for QVIZ is not so very different from what was done for Britain, where four different existing resources were taken together and merged into a single database. The British data sources included a very detailed book that covered England, a less detailed book for Wales, a database created by the Scottish national archives for Scotland, and an additional database created by the main national archives in London listing just manors for England and Wales. The result was a typology which is presented as a thesaurus, but which has just four fixed tiers, which behave in different ways.
3.7 Framework structure

3.7.1 Levels

There are a fixed number of levels (1-13) to which all unit types are assigned, and these levels can be documented in any number of languages. It is important that there be enough different levels to make sense of the most complicated administrative geographies. Many states may need only three levels: Counties (level 7) and Parishes (level 11) with districts or municipalities in between (level 9). Assuming all partners donate data, which can be ascribed to unit types on levels 7 and 11, it is hoped in future to make it possible to create maps that cover more than one state at these levels (see Appendix 1 for full description of levels). Originally there were 12 levels defined and all partners seemed happy with this, but during the final database build for prototype 2b it was discovered the Estonian units actually required a level beneath sub-parish level to cope with some smaller twentieth century units. This level (13) has now been added and the ontology amended accordingly.

3.7.2 Types

The key element in this typology is types. They are what are recorded in the unit records: every unit has a single type, which they cannot change. Types are defined in GIS terms: a type is like an ArcInfo coverage and selecting all the units of one type, or at least all those which existed at a given date, should give you a map of the whole country. They are defined this way whether or not actual boundary polygons for them are held.

The g_unit_type table defines what types exist. If polygons are available for some or all of the units in a type, the type definition includes a bounding box, so the system records the geographical area the units exist in. Some types do not cover a whole country, and that is in fact true of the modern system of counties in Great Britain; there is a lower level of districts, but also unitary authorities, which are basically districts that also have the powers of counties.

3.7.3 Statuses

The 28 British unit types are still not nearly enough to cover all the different kinds of unit that existed, so we have one further tier in the typology: status. Status values are held in a child table, so the rules are much more flexible: some types do not have any associated status values; within some types some units will have associated status values while others do not; some types can have several status values at the same time, whilst others have only one status value but this can change over time. A good example is the system of local government districts, which existed in England and Wales between 1894 and 1974. Local government district (LG_DIST) is defined as a type having the geographical level lower-level district. This type then has 7 associated status values; Metropolitan Borough, Municipal Borough, Urban District, Rural District, County Borough, London City Corporate, London Borough.

This detailed level of data is important because you need to know the status to identify the correct unit. For almost every Rural District there was an Urban District, Municipal Borough or County Borough in the same county with the same name: the system was created out of an earlier system in which the country was divided up into districts centred on a market town, and named after it. In the system created in 1894, the districts were divided into an urban core and a surrounding Rural District, still usually named after the market town but not
including it. Urban Districts, Municipal Boroughs and County Borough all covered towns, but their powers varied; for example, only County Boroughs ran their own schools. As a result, towns tried to raise their status, so over time you see them getting raised from Urban District to Municipal Borough and from Municipal Borough to County Borough. London had a separate system, which was revised in 1965: before you have Metropolitan Boroughs; after you have London Boroughs, which were much bigger. Finally, the City of London – the square mile in the middle – has a unique system of government, which is why it is the only unit with “London City Corporate” status. This system of status values has proved able to record many different historical intricacies that archivists and historians care about, while presentation software can focus on the conceptually simpler types.

### 3.7.4 Relationships

Although each country has its own set of unit types, fitting into the overall system of levels, the system uses a fixed set of relationship types, which can be added to only by general agreement. This is the current set of agreed types of relationship:

- IsPartOf
- SucceededBy
- Administered By
- ReducedToEnlarge
- ReducedToCreate
- AbolishedToEnlarge
- AbolishedToCreate

Although these relationship types are based on English words, the system can hold labels for all relationship types in many different languages. These labels should always be used in public.

### 3.7.5 Boundaries

When considering the issue of describing the history and mapping the polygons of state boundaries two important points emerged:

(a) Historians disagree about the correct boundaries that used to exist, and have arguments about it at conferences.

(b) States themselves disagreed at the time about the boundaries, and had arguments called “wars”, sometimes even genocidal in character. Those create situations where the real areas controlled by particular governments change daily, or are quite fuzzy.

For the QVIZ project the issue of complex relationship typologies was addressed. As already flagged this issue needed to be implemented within the ontology/polyhierarchic thesaurus which acts as the core of the system, whilst the issue around contested polygons was left for another time and a funding source which is more about recording history and less about technology. Trying to build a system capturing the break-up of Yugoslavia would be nightmarish in more than one sense. One of the nice things about this kind of architecture, as compared to a conventional GIS, is that entities can exist without associated polygons.

With the exception of (a) very high level units like continents and states, and (b) the NUTS units currently used by the European Union, the unit types in the administrative ontology are each used only within a particular country. However,
there was no mechanism within the first or second prototypes for enforcing this. There are several detailed reasons why most unit **types** need to be associated with countries:

- When creating maps of a unit or set of units, the system needs to know the wider area that should be shown. This was the only reason the GB Historical GIS needed to know the country, which explains the mechanism that was used.

- There are now so many kinds of unit in the typology that viewing the whole structure can be confusing; we need to be able to show Estonian users only those kinds of unit likely to be of interest to them, i.e. the unit **types** used for Estonia.

- The most important reason for associating unit **types** with countries is to help manage who can edit what, in a decentralized system.

There are, however, two reasons why it was difficult to add a mechanism of this kind in the QVIZ project. Firstly it was only evaluated and considered applicable within the last six months of the project. In practice, this gave enough time to add some kind of mechanism, but not to fully evaluate it, or build a really polished editing system around it. Secondly the current set of QVIZ partners does not allow a full evaluation, however much time is left. This is because the real test of the mechanism will come when more than one organization wants to edit information about the same unit, or at least the same place. This is a real requirement across Europe, but within QVIZ the Swedish National Archives do not want to edit information about Estonia, and the Estonians do not want to add Swedish information. It is, however, quite easy to imagine problems if the Hungarians and the Romanians were both in QVIZ, or the Germans and the Poles. These factors mean that all we can aim for is a first attempt.

### 3.7.6 Jurisdiction

Whatever the mechanism is, it clearly involves holding some information in the **g_unit_type** table, which associates the type with an entity that somehow represents a country. There are three existing kinds of entity in the system we could use:

- **Bounding Boxes**: This mechanism already exists. In the British system, the **g_unit_type** table has a column holding a geographical object, and this is used to store a bounding box for the type as a whole. Only three bounding boxes are in use, covering (i) Great Britain, (ii) England and Wales, and (iii) Scotland, but these bounding boxes are held repeatedly (i.e. there is no mechanism to ensure that all the bounding boxes for Scotland are the same. The absence of a bounding box in the **g_unit_type** table tells the system the unit type cannot be mapped. This mechanism was obviously designed only to support mapping software, and is clearly unsuited to controlling who can edit what.

- **Authority Identifiers**: “Estonian Units” are currently often identifiable because the immediate authority (**im_auth**) value identifies the Estonian National Archives as the authority. This is true not just of the units table, but also of names, relationships, etc. However, this is not how authority
identifiers were meant to be used, as they were designed to identify sources (and the columns in the \textit{g\_authority} table are designed to support the Dublin Core standard). The \textit{im\_auth} values should identify reference books, and the \textit{ul\_auth} values should identify original historical documents referenced by the books, if they are known.

- **Unit Identifiers:** The system already identifies Sweden, Estonia and so on, so why do we need to do it again? Bounding boxes and authority IDs are clearly not the answer, but a case can be made for holding the unit IDs for Sweden, for Estonia and so on within the \textit{g\_unit\_type} table, to indicate which country a particular type belongs to. However, there are several good reasons why not:

  o The \textit{g\_unit} table obviously has to contain a foreign key constraint referencing the \textit{g\_unit\_type} table; it would be a truly horrible database design to have the \textit{g\_unit\_type} table referencing the \textit{g\_unit} table, which in itself suggests this is not the answer.

  o The unit table was not designed for this purpose, and would need many extra columns added; as this is a big table, this would be wasteful.

  o Allowing such a vast collection of units to be used in this way is obvious nonsense, so if we allow unit IDs to appear in the \textit{g\_unit\_type} table we need to constrain which unit IDs can appear.

  o At least in Britain, knowing what geographical area is covered by a \textit{unit type} does not tell you who is responsible for maintaining the list. For example, most of our Scottish \textit{unit types} were defined by the Scottish National Archives, but the Poor Law Combinations were defined by an academic researcher.

For these reasons, the QVIZ partners agreed instead to create a new entity, defined in a new table within the administrative ontology. The preferred name for this is \textit{JURISDICTION} as it explains what the administrative units really do, so the table is accordingly called \textit{g\_jurisdiction}. Within this table are all physical references to actual jurisdiction authorities including name, postal and email address, telephone number and preferred language.

### 3.7.7 Function

This is a new implementation in the system for QVIZ. It was inserted after discussion between project partners about the best way to identify unit types with similar functions. This characterization of unit type functions in the \textit{g\_type\_function} table was mainly to aid the mapping interface, but also to help the search parameters of the faceted browser. A list detailing different kinds of function for the unit was identified and unit types assigned to each function. The following is a list of the defined functions and their purpose:

- **GENERAL** = Unit types which did not have a specific purpose, but like parishes were used for many aspects of local administration.
• CIVIL = Covers all administrative functions including tax and police related activities.
• ECCLES = All church related geographies
• MILITARY = All military related geographies
• JUDICIAL = Only court related geographies
• STATS = Geographies which only existed for reporting statistics and not in reality
• HEALTH = Geographies which relate solely to health administration
• EDUCATION = Geographies which relate solely to education administration
• DRAINS = Geographies which relate solely to sewerage and drainage administration

Note that the current implementation allows each unit type to have only one function. We discussed allowing each type to have multiple functions, but this would have complicated the system, required much additional research but would still have involved much historical simplification. In particular, many types of unit changed their functions over time; and many of the finer points of unit status are precisely about variations in exact function between units of the same type; for example, an English County Borough controlled its schools while a Municipal Borough did not; but both have the type LG_DIST.

3.7.8 Multiple Languages
To support multiple-languages the structure has specific modifications in certain areas. The only support for multiple languages in the data tables is the recording of the languages of place-names. Most of the information held in the data tables other than place-names is not language specific, and no one had time to translate everything in the notes columns. The unit types and unit status tables include columns for labels and an explanatory note in both a national language, identified by the value in the g_language column, and in English. This is as agreed: the only language that explanations of Estonian unit types should be translated into is English. The unit types table includes additional columns to hold the plural of the type label in both the national language and in English; for example, in English the plural of district is districts, but the plural of parish is parishes, so you cannot just add an S. The tables defining the available types of unit levels, relationships, name statuses and authorities all have child tables able to hold meanings in any number of languages. The child table for relationships can also hold text giving the relationship in reverse: has as a part instead of is part of. The child table for unit levels can also hold a short explanation of the level in multiple languages.

3.7.9 Authorities
The system is very careful in how the sources of information are recorded, storing two separate “authorities” for each row in each of the five main data tables (excluding the “places table”, which is essentially derived). Each of the five data tables in the ontology for -- Units, Names, Statuses, Relationships and Footprints -- contains four columns designed to hold authority information. Although these columns are not currently defined as an object, they could be. The four columns are:
• \textit{im\_auth} = Immediate authority: This column is constrained to hold only values defined in the \texttt{g\_authority} column of the \texttt{g\_authority} table, and is used to identify where the information was computerized from. For now, the two national archives are defined as being themselves authorities, just as we accepted the Scottish National Archives as being an authority for Vision of Britain.

• \textit{im\_auth\_note} = Note on immediate authority: This can be left empty, but if the immediate authority were a book then this would contain a page number.

• \textit{ul\_auth} = Ultimate authority: Mostly, the immediate authorities in use are compilations which took their information from somewhere else, and identify their source by footnotes; these sources are recorded as ultimate authorities. They are likely to be legal documents. Here again, the only strings that can appear are those defined in the \texttt{g\_authority} table.

• \textit{ul\_auth\_note} = Note on ultimate authority: These provide more information on the ultimate authority. E.g. many of the boundary changes in VoB have been computerized from census reports, but the census often gave the source as some kind of government order. All that goes into the \textit{ul\_auth} column is \textit{BC\_LIST}, and the details are given in the \textit{ul\_auth\_note} column; for example, “M. of H. and Local Government Order No. 28164. The County of Chester (Borough of Ellesmere Port) Confirmation Order, 1967”.
4. Data Section

4.1 Introduction to the data

The history of administrative units is, by definition, messy. European countries have experienced both subtle and sometimes drastic changes to their internal administrative landscape over several hundred years, since the concept of an administrative unit came into being. Of course in addition to the internal alterations there are territorial disputes, which have arisen between neighbouring states and the fairly continual story of invasion, occupation and reclaimed independence, which defines much of the changing peace-time map of Europe. Further there are the added complications of international wars and administrative changes made by governments not officially recognized by other contemporary states. The intricacies of war-time have not been investigated by the QVIZ project team as this was deemed beyond the scope and scale of the QVIZ prototype interface. As the three countries represented by the partners supplying the administrative data are geographically dispersed and have no common boundaries there was no requirement to develop a methodology for handling disputed border areas across national frontiers or for ensuring the vector boundaries supplied fitted together. If further countries were to be added to the QVIZ system in the future, which did have a common border with one of the existing data providers this might require re-examination. Significant work has however gone into creating a running commentary of the changing boundaries under peacetime conditions since 1815. Prior to this date the scale of the fragmentation of Europe following the end of the expansion of Napoleonic France and the collapse of the Holy Roman Empire would require a very large project, which was not possible within the timescale of QVIZ. Disparate sources have been trawled and data compiled to produce the most factually accurate listing of boundary changes possible.

A strategy was developed to cope with the disappearance and reappearance over time of particular units with the same name. Instead of having individual units with multiple dates of existence, which was considered confusing and would create data inconsistencies within the database, each new start date was taken as the creation date of a new unit. For example the Republic of Estonia which existed between 1918 and 1940 is one unit within the system whilst the Republic of Estonia which has existed since 1991 is a completely separate unit. This strategy is employed throughout the database content at all unit levels. It can sometimes lead to the proliferation of units with the same name and same type, but the difference in dates of existence is always present and serves to differentiate between those units. Units can also have ‘SucceededBy’ relationships. These are generally used to link an earlier unit to a later one, often with the same name.

4.2 The World data

A single unit, “World”, has been created as the root unit for the whole system and is the only administrative unit which is permitted to have no ‘Is Part Of” relationship. Below this units were created for the Continents and Sub-Continents that could be identified. The Encyclopaedia Britannica was used as the source for the existence and relationships associated with these units and the textual entries explain the relationships where they may not be immediately obvious. No states beyond Europe or satellites of European states located outside Europe have been created as units within the ontology as part of this project, but the structure of the system is such that they could be inserted at a later date should funding become
available. International organizations can also be inserted into the structure at the same level as sub-continents or continents depending on whether they are continental or world level organizations.

4.3 European states data
Created from scratch, the European data consists of a list of all the recognized states of Europe since 1815 based on a list compiled from Rossiter, S and Flower, J., 1986, *The Stamp Atlas*, London: Macdonald; operating a postal service is a clear indication that a functioning state existed, not just a political front of some kind. In addition to this the existence of two periods of independence for Estonia, the control of the Polish-Lithuanian Commonwealth over Estonia in the sixteenth and seventeenth centuries and the existence of the League of Nations in 1920 meant further additional state values had to be added to allow for linkage within the relationships table. The boundary changes for the European states have been compiled from a combination of sources. Sir Edward Hertslet’s book *The map of Europe by Treaty: showing the various political and territorial changes which have taken place since the general peace of 1814* published in 1875 lists the states of Europe and their various changes during the mid nineteenth century. For later change details a combination of various internet resources and books have been employed including A. Crispin Jewitt’s book, *Maps For Empire; the first 2000 numbered War Office Maps* (1992) and the Delimitation treaties infobase on the United Nations website.

Additional preferred names for each state from other languages (French, German, Swedish and Estonian) were added using various dictionaries detailing translations into English where they could be identified. The names for the state in the national language(s) were also obtained from the EUROSTAT website, where the national language could be easily identified. Alternate names were also added where the names of the states varied from the preferred name on the 1920 *Times Atlas of the World* map.

The European vector boundaries have been compiled from various sources. Firstly the digital boundaries of the states of the world in 2007 were downloaded from the EUROSTAT website on 13th August 2007. The European state boundaries were then selected and copied to another layer, which was imported into the database as the modern coverage of Europe. The European state boundaries were then copied to a new layer and the internal boundaries removed to leave just the external boundaries following the coastline of Europe. These external boundaries were then used for the creation of new layers for different years to allow exact matching on the coastline boundaries whilst the internal boundaries changed according to the political landscape. The vectors for the states were assigned dates according to the date of the maps rather than periods of existence as the dates for individual states would be different in every case. They were assigned longitude and latitude co-ordinates for compatibility with the scanned map images.

4.4 The Estonian data
The Estonian data required a significant amount of work by both UoP and The National Archives of Estonia (NAE) to work it into a compatible format. The Estonians supplied a database of manors, a separate list of information for parishes along with various shape-files containing vector data and an administrative unit typology. This typology assigned various unit types to levels, recorded relationships between unit types and listed the higher level units and their dates of
existence. The individual unit data incorporated in the typology was transferred to the temporary data table.

4.4.1 Dates

Many of the database units had text strings indicating their start or end dates. These were translated into English from Estonian, although the date objects can only hold a single text string, so where multiple language date strings exist the Estonian version of the string has been inserted into the date object for that unit. The QVIZ interface requires relationship dates in order to function properly, but does not work well with text string dates so a best-year date had to be inserted for the relationships. An automated process was developed, but it did not work well, and significant manual editing was also required. For example: where a string appeared “Rajatud 17. sajandi 1. poolel” (founded in the first half of the 17th Century) the automated process interpreted this to mean 1700. At very least it should have been interpreted as 1600, but in terms of historical accuracy it should be 1600-1649. Compromise was reached by employing a mean date value implemented as follows:

<table>
<thead>
<tr>
<th>Start dates</th>
<th>End dates</th>
</tr>
</thead>
<tbody>
<tr>
<td>xx13 = First quarter ...</td>
<td>xx25 = xx20s</td>
</tr>
<tr>
<td>xx38 = Second quarter ...</td>
<td>xx29 = xx25/xx29</td>
</tr>
<tr>
<td>xx63 = Third quarter ...</td>
<td>xx89 = xx89</td>
</tr>
<tr>
<td>xx88 = Fourth quarter ...</td>
<td>xx50 = xxth century</td>
</tr>
<tr>
<td>xx25 = First half of ...</td>
<td>1930 = between 1922 and 1938</td>
</tr>
<tr>
<td>xx75 = Second half of ...</td>
<td>1944 = between 1938 and 1950</td>
</tr>
<tr>
<td>xx00 = Start of ...</td>
<td>1969 = between 1959 and 1979</td>
</tr>
<tr>
<td>xx50 = Middle of ...</td>
<td>1984 = between 1979 and 1989</td>
</tr>
<tr>
<td>xx99 = end of ...</td>
<td></td>
</tr>
<tr>
<td>xx50 = Created in the ...</td>
<td></td>
</tr>
</tbody>
</table>

4.4.2 Units from Vectors

Unit information, which was not already within the database was also extracted from the shape-files and used to create new units for insertion into the temporary table. Where units were added from the vector files and dates for individual units were unknown approximate dates were assigned based on the dates of existence for the unit type or the assistance of archival partners and the appearance of the units in the series of vector files for different dates. For example if a külanõukogu (Village Council) appeared in the vector files for 1959 and 1979 but had gone by 1989 the start date would be indicated to be 1950 and the end date between 1979 and 1989, i.e. starting between the start date for the unit type (1945) and the first set of vectors it appeared in (1959) and ending after the last set of boundaries it appeared in (1979) and the next set it did not (1989).

4.4.3 Multiple units with the same name and type

The existence of multiple unit types with the same type and relationship just different dates meant statuses had to be used and hint numbers employed. This in turn meant the relationships within Estonia are extremely complex and much discussion was conducted between NAE and UoP in an attempt to correctly represent all the administrative diversity within the system. An example would be at the regional unit type level (6) the kubermang units. Each had a different status type, different dates of existence and relationships and a different hint value. For
example between 1583 and 1713 there was a kubermang called Livonia, in 1713 to 1722 there was a different one of the same name, 1722-1783 another one, and 1796-1917 a fourth. When considering that each level in Estonia had similar examples the necessity for a structured order of content becomes obvious.

4.4.4 A single root
The only simplification made to the ontology was to create a single unit with a unit type of NATION to allow all lower level units to link to a single high level unit and to show the changing relationships with other states over time, including periods of Estonian independence. NAE did not supply all Estonian administrative data. Instead they supplied data about which they have information, dates and sources and omitted any units, which could not be easily manipulated to fit within the data set.

4.5 The Swedish data
Although the GB Historical GIS database framework was modelled around the pre-existing Swedish National Topographical Database, a component of the National Archival Database (NAD), the Swedish data was not immediately compatible with the QVIZ framework. Significant work went into creating a hierarchical diagram of the relationships and functions of the many different types of Swedish administrative unit. This was done to enable each unit type to be assigned a unit level. Once the diagram and levels had been assigned by the Portsmouth team to the satisfaction of the Swedish National Archives (SNA) the actual insertion of the data was achieved by the insertion of all the administrative unit data into temporary tables before loading the unit names, types, relationships, dates and vectors into the permanent tables, which proved relatively straightforward. Various errors were encountered with the original data supplied by SNA and through discussion with SNA manual editing corrected these issues.

4.6 The British data
The British administrative unit system was used as the basis from which the QVIZ framework was developed. It therefore fitted neatly into the table structure with little need for editing. Of course some individual manual edits were necessary as the database is not entirely error-free across all its 50,000 units. Another problem was that the best version of the British data did not have any polygons connected to it. Therefore when the administrative unit data was imported from the current database some manual work was required to match up the polygons from a slightly older and less complete database.

4.7 Relationship to Archival Resources
To build in linkage between the administrative unit ontology and the cultural heritage institution resources it became clear that some kind of identifier that was common to both had to be implemented. Although the Swedish units already had a suitable identifier the Estonian units did not. In order for the implementation to be uniform across the system unique identifiers as arbitrary numbers continued to be assigned to individual units as had previously been done in the British system. To cope with the linkage to the recognized statistical reporting unit identification coding system employed across Sweden the existing identifier was inserted in the g_name table using the coding C_KALLAID as an extra name for each unit to enable look-up and cross-referencing. Estonia did not have an equivalent existing system. Instead a dumb identifying number was added to the Estonian
administrative unit data to be used to match up the archival documents to the administrative units, this was inserted into the framework as an extra name for the individual Estonian units with the g_name_status of C_NAEID. Any further matching would just require the insertion of further dumb numbers as names with the same g_name_status for units existing within the system, which do not currently have one.

As there are no partners currently involved in the QVIZ project that hold British archival digital content there was no need to construct similar linkages for Great Britain. However, this procedure would be simple to replicate if more unit data were to be added to the system.
5. Toolkits for validation

5.1 Methodologies employed
All of the methodology described here was implemented using SQL scripts with insert and update statements thus allowing repeated redefinition of all tables and data. However once the final load of metadata and data into the temporary tables had been done, some manual work was required to ensure as full a match as possible on the units and their relationships to allow maximum success in insertion into the permanent tables. Although the database skills of the creators of the administrative unit ontology are mostly Oracle-based the other QVIZ partners are most familiar with Postgres and this was how the deliverable was expected. Once all the data had been inserted into the relevant permanent tables within the framework in ORACLE the data was then copied across from Oracle to Postgres using command scripts for the second and third prototypes.

5.1.1 Prototype 1
The first prototype involved the supply of Isle of Wight data to Regio for a first mapping interface prototype. Also Estonian data was inserted for a single date to highlight the progress being made on Estonian data and allow investigation into the issues surrounding complex poly-hierarchical data structures. The administrative unit data was viewable in a text-only pl/sql interface overlying ORACLE. At this time UoP did not have any staff capable of supplying the data in Postgres format so it was supplied in ORACLE format together with shape files and edited by partners at Umeå University for their needs.

5.1.2 Prototype 2a
The second prototype involved all Swedish data, British data from the older database, which straightforwardly contained the polygons and a subsection of the Estonian data – for a single date. Polygons for Great Britain, Sweden and Estonia (for one date) were included as were the initial listing of international level units and unit types and a single set of international boundaries for modern European states. This data was supplied in PostgreSQL and PostGIS format.

5.1.3 Prototype 2b
The final prototype contained a full listing of the continents, sub-continents, European states including various language preferred names, all the relationships. Also included was the current British data, the Swedish data with significant cleaning done, all the Estonian data and polygons where they existed for all units, including state level in 1920. Further, before the launch of the QVIZ interface at the symposium further international boundaries changed will be added to display the changing boundaries across Europe during the twentieth century.

5.1.4 Text-only Interface
In addition to the prototypes there is also a text-only interface to the data in Oracle. This was initiated during the development for the first prototype and has been used in the development of the content ever since. It displays all individual units including their names, languages, statuses, unit types, unit statuses, relationships to other administrative units, authorities and any dates associated with this information. It also shows who is responsible for controlling the information displayed on the page. This tool was helpful for identifying errors in
the data loading and for the data contributors to assess the accuracy of methods of
data compilation employed. This interface is web accessible but not designed for a
broad audience, the address is available on request.

5.2 Administrative unit creation

The tables for the database were all created initially as empty tables with
constraints attached within Oracle. All relevant data was then loaded directly into
the metadata tables. Then a complete import of the British administrative unit data
was implemented from the most up-to-date version of the British database. Finally
temporary tables were constructed for each of the other datasets; international,
Swedish and Estonian. These data were checked for the legality of their
relationships and arbitrary unit numbers assigned to individual units. For the
international and Estonian units this was done as a staged sequence. Every unit
must have an 'Is Part Of' relationship with another higher level unit which means
the higher level unit must exist before it can be matched. Therefore the highest
level units were inserted into the permanent tables first once they had unique
identifiers (g_unit). Once they were inserted the g_unit number of the container
unit could be identified and inserted into the temporary table for the next lower
level of units, a new g_unit number was ascribed to each unit in this lower level
and these units then inserted into the database and so on.

5.3 Vector creation

The vectors were created in ArcGIS using on-screen digitizing and assigned a
longitude-latitude WGS84 co-ordinate system. They were then translated into
shape files and loaded into the Oracle database using MapBuilder, which
converted the shape files into oracle tables containing the vectors as geometric
objects. These individual tables were then compiled into a single table inside the
database and the geometric objects were copied into a new column utilizing a
Lambert Conformal Conic projection conversion command. Date durations were
then set to correspond to the date of the map. For example the vectors originating
from the Times Atlas of the World was set from 1920 to 1920 to reflect the view
of the states as they existed precisely in 1920. The modern boundaries were set
from 2007 to 5000. This approach was employed after discussions with partners
following delivery of the final prototype data, instead of giving each map a start
and end date dependent upon the existence of the next temporal vector set.

5.4 Scanned mapping

The default basis for the background mapping will be the “Digital Chart of the
World” which is freely accessible and downloadable over the web. However in
keeping with the historical content of the QVIZ interface a whole GSGS 4072 map
series at 1:500,000 scale has been scanned for use. They were created during the
1940’s from flights made by the RAF and therefore are out of copyright. The
majority of these maps were loaned to UoP for scanning as part of this project and
have been geo-rectified and mosaicked together using the Lambert Conformal
Conic projection as agreed with partners at the Consortium meeting in Budapest in
October 2007. Unfortunately the European coverage is incomplete. Areas that
were irrelevant to wartime activities because of their neutrality are omitted from
the series. There are also a couple of gaps in the series we hold, the sheets for
Cartagena, Bolzano and Graz, that despite repeated requests to relevant parties has
failed to produce copies. Where possible the neutrality gaps have been plugged
using maps of a similar scope and scale. The historical map scans are made
available via hosting on MapServer using an interface, which has been developed using OpenLayers, an open source mapping software. This layer of historical mapping is available to be called by the QVIZ interface and is viewable beneath the QVIZ portal to give historical context to the mapping environment.

5.5 Quality Assurance
Due to the loading of units occurring in a sequence it is obvious when units are not being inserted. All administrative units require their container unit to exist before they can be inserted into the system. Failure to be inserted is recorded in an error message and a list of unloaded units is output to be corrected. This could be because the meta-data or the data is incomplete or does match the existing information within the system. Once the units are inserted into the system, the text-only interface constructed from pl/sql acts as a reference tool for understanding what is happening in the database tables. It identifies the errors in the unit details.

The vectors are loaded into the system and are visible within ORACLE using the Geo-Raptor extension. This tool allows the operator to visually check that the vectors look approximately correct. However, it does not indicate if vectors are misaligned with one another where they have been inserted from separate layers, or if the co-ordinate systems are different. Errors of this nature were discovered by a quick visual check of the data once inserted into Postgres comparing the co-ordinate values. The only way to check was to extract the vector data from the database and insert in ArcGIS where it is possible to compare different layer values, although even then the package will show the user the best representation of the data, rather than the actual values it encounters. Errors in the assignment of co-ordinate reference systems and the inclusion of some rogue values skewed the vectors and manual editing had to be used to correct this.
6. Conclusion

Deliverable 3.2 in the application submitted to the European Commission identified the task output to be “To design an ontology of European administrative units defining temporal and spatial relationships”. During the QVIZ project partners have collaborated to create an ontology, which achieves this goal. It is designed to record in detail the sources, variant names and poly-hierarchical relationships for all administrative units within Sweden, Estonia and Great Britain with an overarching collection of European state level data, boundary polygons where available and date values of varying precision. It has been developed to function with multi-lingual interfaces and typologies of units, to display the complexity of administrative unit history whilst supporting visualization and querying interfaces. Covering three countries in great detail, it permits access to units down to village level and allows the comparison of units at similar theoretical levels in the hierarchical structure. By holding the unit names in a duplicate column in upper case devoid of all punctuation and special characterization there has been the opportunity to develop a name searching over and above the faceted browser querying. Significant manual manipulation was needed to make new Estonian administrative data fit the schema, whilst the Swedish data was inserted fairly easily once the general structure had been identified. This process also helped identify inconsistencies in all three countries as the cascading input format meant any undefined relationships could be easily identified.

The QVIZ interface provides a unique portal to access similar kinds of digital data from a variety of disparate national archives. As European cultural heritage institutions gather ever increasing amounts of digital content the QVIZ user interface offers an opportunity to facilitate access to that content in a single user environment incorporating administrative history, a faceted browser and a mapping interface with the additional benefit of support and discussion between users of the same content through Communities of Practice. The interaction of users and institutions holding resources should allow more exchange of information across Europe, but none of this would be possible without the flexibility of the database structure of the administrative ontology beneath it.
Appendix 1

Administrative Unit Levels

The following levels are used in the QVIZ ontology:

1. **ROOT UNIT = WORLD**: following on from problems with the British structure the highest level had to be definitely the highest level. Note that modern France includes *departments*, which are not in Europe, and there is scope to include colonial possessions. Note that by definition there can be only one root unit, so there is only one type defined at this level and no status values.

2. **CONTINENTS**: i.e. Europe etc. International organizations can also exist at this level.

3. **SUB-CONTINENTS**: this includes, for example, *Scandinavia and Iberia*, as well as *British Isles*, which is the root unit for the British system.

4. **STATE**: (e.g. Estonia, Sweden, and United Kingdom: a unit that is politically independent, and can belong directly to the European Union and the United Nations.

5. **NATION**: e.g. England, Wales, Scotland. There are no units at this level within modern Sweden, but they might well be needed for Spain or Russia, not to mention various historical empires. Nations are not completely independent politically, but are usually defined in terms of ethnicity and/or language, not just geography. They usually either have some kind of government led by a parliament, or a political campaign to have one. Estonia has been created as a nation to enable linkage to various other states when it was under occupation as well as having a single stable unit to which all lower level units could be linked.

6. **REGION**: For example within England, East Anglia and the South West. England is usually divided into 8 to 10 of these. They are bigger than counties, but are NOT nations. Please be clear that Regions and Nations will often have similar populations and geographical areas, but they differ in status and maybe in aspirations.

7. **COUNTY**: In England, this means a unit focused on a fairly big town, probably with a cathedral, which is the centre for administering, for example, education. The county will usually contain 5 or 6 market towns as well as the county town. In Britain, if you come from a village you would not normally expect other people to know the village so if they asked where you came from you would usually say the name of the county. Names of villages are often duplicated within Britain, so a full statement of where you came from would be something like “Cradley, Herefordshire” where “Herefordshire” is the name of a county. County towns are likely to be big enough to usually have several churches and parishes within them.

8. **High-level DISTRICT/MUNICIPALITY**: We have to have three levels of district because several systems of administration in Britain had two levels of unit between the counties and the parishes, and the system of “Ancient Districts” had three. High-level districts will be large subdivisions of counties, and may exist only in the largest counties: for example, Yorkshire was divided into three “Ridings” and Lincolnshire into three “Parts”.

30
9. **Middle-level DISTRICT/MUNICIPALITY**: If you have only one level of “District” between counties and parishes they should usually be classed as “Middle-level districts”.

10. **Low-level DISTRICT/MUNICIPALITY**: Lower-level districts will often be based on EITHER a market town OR the surrounding villages; districts which combine the two will usually be classed as “Middle-level districts”.

11. **PARISH**: These exist in all parts of Britain, in Estonia and in Sweden. Legally parishes at least begin as units of religious administration, but in practice they are based on either a single large village or a cluster of small ones. In some countries, such as Ireland, more than one religion existed at the same time so you have more than one kind of parish at the same time; but these can be handled as separate types.

12. **Sub-PARISH**: This covers anything more detailed than parishes, which exists as legally defined administrative units. These will often be based around large land-holdings, and settlements that consist of a big house for the landowner, the buildings of a farm, and the houses of the people who work on the estate’s land.

13. **Sub sub-PARISH**: This covers the smallest units of administration

The key assumptions are that in every country:

- We can identify units corresponding to “Counties”
- We can identify units corresponding to “Parishes”
- There are never more than three administrative levels between counties and parishes
- There will be only a very limited number of levels below parish, as the system aims to describe government structures, not land holdings
Appendix 2

A2.1 Representation of some administrative units levels 1-5 in the ontology.
A2.2 Example of Estonian unit type levels 4-12 showing relationships of Tartu and Kursi.
A2.3 Example of British unit type levels 5-12 showing relationships of Hampshire and Newport.
A2.4 Example of Swedish unit type levels 4-12 showing relationships of Uppsala län and Uppsala.
Appendix 3

A3.1 Statistics relating to Administrative unit content.

A3.1.1 Count of units and polygons per unit type where some units exist.

<table>
<thead>
<tr>
<th>UNIT LEVEL</th>
<th>UNIT TYPE</th>
<th>UNITS PER TYPE</th>
<th>POLYGONS PER TYPE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>ROOT</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>CONTINENT</td>
<td>7</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>WORLD_ORG</td>
<td>6</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>SUBCONTINENT</td>
<td>6</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>CONTINENTAL_ORG</td>
<td>10</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>STATE</td>
<td>90</td>
<td>2185</td>
</tr>
<tr>
<td>5</td>
<td>NATION</td>
<td>8</td>
<td>15</td>
</tr>
<tr>
<td>6</td>
<td>EST_ASEH</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>EST_KORP</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>EST_KUB</td>
<td>10</td>
<td>2</td>
</tr>
<tr>
<td>6</td>
<td>EST_OBL</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>7</td>
<td>EST_LILN</td>
<td>7</td>
<td>0</td>
</tr>
<tr>
<td>7</td>
<td>EST_MAAK</td>
<td>83</td>
<td>114</td>
</tr>
<tr>
<td>7</td>
<td>EST_VOJ</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>9</td>
<td>EST_LINN</td>
<td>7</td>
<td>2</td>
</tr>
<tr>
<td>11</td>
<td>EST_KIHEL</td>
<td>114</td>
<td>106</td>
</tr>
<tr>
<td>12</td>
<td>EST_KULAN</td>
<td>428</td>
<td>709</td>
</tr>
<tr>
<td>12</td>
<td>EST_MOIS</td>
<td>2425</td>
<td>1469</td>
</tr>
<tr>
<td>12</td>
<td>EST_VALD</td>
<td>3291</td>
<td>2361</td>
</tr>
<tr>
<td>13</td>
<td>EST_ASULA</td>
<td>4673</td>
<td>4673</td>
</tr>
<tr>
<td>6</td>
<td>ECC_PROV</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>IRL_PROV</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>MOD_REG</td>
<td>9</td>
<td>18</td>
</tr>
<tr>
<td>7</td>
<td>ADM_CNTY</td>
<td>65</td>
<td>231</td>
</tr>
<tr>
<td>7</td>
<td>ANC_CNTY</td>
<td>53</td>
<td>276</td>
</tr>
<tr>
<td>7</td>
<td>DIOCESE</td>
<td>46</td>
<td>0</td>
</tr>
<tr>
<td>7</td>
<td>IRL_CNTY</td>
<td>32</td>
<td>67</td>
</tr>
<tr>
<td>7</td>
<td>MOD_CNTY</td>
<td>42</td>
<td>84</td>
</tr>
<tr>
<td>7</td>
<td>PR_CNTY</td>
<td>55</td>
<td>217</td>
</tr>
<tr>
<td>7</td>
<td>SCO_CNTY</td>
<td>40</td>
<td>71</td>
</tr>
<tr>
<td>8</td>
<td>ANC_CDIV</td>
<td>42</td>
<td>23</td>
</tr>
<tr>
<td>8</td>
<td>ARCHDEACON</td>
<td>122</td>
<td>0</td>
</tr>
<tr>
<td>8</td>
<td>MOD_DIST</td>
<td>408</td>
<td>816</td>
</tr>
<tr>
<td>8</td>
<td>PR_DIST</td>
<td>723</td>
<td>1369</td>
</tr>
<tr>
<td>UNIT LEVEL</td>
<td>UNIT TYPE</td>
<td>UNITS PER TYPE</td>
<td>POLYGONS PER TYPE</td>
</tr>
<tr>
<td>------------</td>
<td>--------------</td>
<td>----------------</td>
<td>-------------------</td>
</tr>
<tr>
<td>9</td>
<td>ANC_DIST</td>
<td>1683</td>
<td>11</td>
</tr>
<tr>
<td>9</td>
<td>LG_DIST</td>
<td>2180</td>
<td>4220</td>
</tr>
<tr>
<td>9</td>
<td>R_DEANERY</td>
<td>1521</td>
<td>0</td>
</tr>
<tr>
<td>9</td>
<td>SAN_DIST</td>
<td>1567</td>
<td>1</td>
</tr>
<tr>
<td>9</td>
<td>SCO_BURGH</td>
<td>532</td>
<td>225</td>
</tr>
<tr>
<td>9</td>
<td>SCO_COMB</td>
<td>47</td>
<td>0</td>
</tr>
<tr>
<td>9</td>
<td>SUB_DIST</td>
<td>2610</td>
<td>0</td>
</tr>
<tr>
<td>11</td>
<td>ECC_PAR</td>
<td>17211</td>
<td>0</td>
</tr>
<tr>
<td>11</td>
<td>PAR_UNIT</td>
<td>20255</td>
<td>45165</td>
</tr>
<tr>
<td>11</td>
<td>SCO_PAR</td>
<td>901</td>
<td>1679</td>
</tr>
<tr>
<td>12</td>
<td>MANOR</td>
<td>2258</td>
<td>0</td>
</tr>
<tr>
<td>7</td>
<td>SWE_LAN</td>
<td>30</td>
<td>80</td>
</tr>
<tr>
<td>8</td>
<td>SWE_HOV</td>
<td>6</td>
<td>20</td>
</tr>
<tr>
<td>8</td>
<td>SWE_KON</td>
<td>269</td>
<td>539</td>
</tr>
<tr>
<td>8</td>
<td>SWE_MILIN</td>
<td>107</td>
<td>0</td>
</tr>
<tr>
<td>8</td>
<td>SWE_STIFT</td>
<td>13</td>
<td>37</td>
</tr>
<tr>
<td>10</td>
<td>SWE_FOG</td>
<td>442</td>
<td>959</td>
</tr>
<tr>
<td>10</td>
<td>SWE_HARAD</td>
<td>297</td>
<td>423</td>
</tr>
<tr>
<td>10</td>
<td>SWE_KOM</td>
<td>2934</td>
<td>3543</td>
</tr>
<tr>
<td>10</td>
<td>SWE_RAD</td>
<td>371</td>
<td>521</td>
</tr>
<tr>
<td>10</td>
<td>SWE_TINGT</td>
<td>110</td>
<td>143</td>
</tr>
<tr>
<td>11</td>
<td>SWE_KYRK</td>
<td>2867</td>
<td>3081</td>
</tr>
</tbody>
</table>
A3.1.2 Histogram showing the numbers of polygons over time

![Histogram showing the numbers of polygons over time](image)


A3.1.3 Count of number of names per language

<table>
<thead>
<tr>
<th>LANGUAGE</th>
<th>NUMBER OF NAMES</th>
</tr>
</thead>
<tbody>
<tr>
<td>ENGLISH</td>
<td>62603</td>
</tr>
<tr>
<td>SWEDISH</td>
<td>7507</td>
</tr>
<tr>
<td>ESTONIAN</td>
<td>11973</td>
</tr>
<tr>
<td>GERMAN</td>
<td>5032</td>
</tr>
<tr>
<td>WELSH</td>
<td>1069</td>
</tr>
<tr>
<td>FRENCH</td>
<td>61</td>
</tr>
<tr>
<td>GREEK</td>
<td>3</td>
</tr>
<tr>
<td>ITALIAN</td>
<td>3</td>
</tr>
<tr>
<td>RUSSIAN</td>
<td>2</td>
</tr>
<tr>
<td>TURKISH</td>
<td>2</td>
</tr>
<tr>
<td>OTHER LANGUAGES WITH 1 NAME EACH</td>
<td>26</td>
</tr>
</tbody>
</table>
### A3.1.4 Count of units and names per jurisdiction

<table>
<thead>
<tr>
<th>JURISDICTION</th>
<th>TOTAL UNITS</th>
<th>TOTAL NAMES</th>
<th>% UNITS WITH MULTIPLE NAMES</th>
</tr>
</thead>
<tbody>
<tr>
<td>GBHGIS</td>
<td>52535</td>
<td>63950</td>
<td>17.5</td>
</tr>
<tr>
<td>NAE</td>
<td>11049</td>
<td>16885</td>
<td>40.7</td>
</tr>
<tr>
<td>SNA</td>
<td>7446</td>
<td>7446</td>
<td>0</td>
</tr>
</tbody>
</table>

### A3.1.5 Count of units with names in two or more languages by authority

<table>
<thead>
<tr>
<th>NAME AUTHORITY</th>
<th>COUNT OF UNITS WITH MULTIPLE NAMES</th>
</tr>
</thead>
<tbody>
<tr>
<td>UK CENSUS OF POPULATION</td>
<td>22</td>
</tr>
<tr>
<td>EUROPA WEBSITE</td>
<td>40</td>
</tr>
<tr>
<td>GREAT BRITAIN HISTORICAL GIS PROJECT (GBHGIS)</td>
<td>1</td>
</tr>
<tr>
<td>THE MAP OF EUROPE BY TREATY (E. HERTSLET)</td>
<td>67</td>
</tr>
<tr>
<td>WELSH ADMINISTRATIVE AND TERRITORIAL UNITS (M.RICHARDS)</td>
<td>810</td>
</tr>
<tr>
<td>NATIONAL ARCHIVES OF ESTONIA (NAE)</td>
<td>4703</td>
</tr>
<tr>
<td>GBHGIS PROJECT SOURCE DOCUMENTATION SYSTEM</td>
<td>1</td>
</tr>
<tr>
<td>GUIDE TO THE LOCAL ADMINISTRATIVE UNITS OF ENGLAND: NORTHERN VOLUME (F. YOUNGS)</td>
<td>23</td>
</tr>
<tr>
<td>GUIDE TO THE LOCAL ADMINISTRATIVE UNITS OF ENGLAND: SOUTHERN VOLUME (F. YOUNGS)</td>
<td>5</td>
</tr>
<tr>
<td>COLLINS GERMAN DICTIONARY</td>
<td>57</td>
</tr>
<tr>
<td>COLLINS ROBERTS FRENCH DICTIONARY</td>
<td>59</td>
</tr>
<tr>
<td>ENCYCLOPAEDIA BRITANNICA</td>
<td>9</td>
</tr>
<tr>
<td>ESTONIAN - ENGLISH - ESTONIAN DICTIONARY</td>
<td>65</td>
</tr>
<tr>
<td>OXFORD ENGLISH DICTIONARY</td>
<td>1</td>
</tr>
<tr>
<td>SPRÅKRÅDET SWEDISH-ENGLISH DICTIONARY</td>
<td>9</td>
</tr>
<tr>
<td>STANDARD SWEDISH DICTIONARY</td>
<td>52</td>
</tr>
</tbody>
</table>
A3.1.6 Count of relationship types.

<table>
<thead>
<tr>
<th>RELATIONSHIP TYPE</th>
<th>COUNT</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>AbolishedToCreate</td>
<td>806</td>
<td>0,373</td>
</tr>
<tr>
<td>AbolishedToEnlarge</td>
<td>207</td>
<td>0,096</td>
</tr>
<tr>
<td>AdministeredBy</td>
<td>20</td>
<td>0,009</td>
</tr>
<tr>
<td>IsPartOf</td>
<td>201727</td>
<td>93,267</td>
</tr>
<tr>
<td>PrecededBy</td>
<td>3</td>
<td>0,001</td>
</tr>
<tr>
<td>ReducedToCreate</td>
<td>1918</td>
<td>0,887</td>
</tr>
<tr>
<td>ReducedToEnlarge</td>
<td>12</td>
<td>0,006</td>
</tr>
<tr>
<td>SucceededBy</td>
<td>11596</td>
<td>5,361</td>
</tr>
</tbody>
</table>