Project Acronym:	INDICATE
Project Title:	International Network for a Digital Cultural Heritage e-Infrastructure
Contract Number:	261324
Starting date:	1 September 2010
Ending date:	31 October 2012

Deliverable Number:	D 5.3
Title of the Deliverable:	Case study report – Geocoded digital cultural content
Task/WP related to the Deliverable:	Task 5.3 Geocoded digital cultural content
Type (Internal or Restricted or Public):	Public
Author(s):	Franc J. Zakrajsek Vlasta Vodeb Jurij Stare Darija Modrijan
Partner(s) Contributing:	All

Contractual Date of Delivery to the CEC:	**month 18**
Actual Date of Delivery to the CEC:	**August 2012**

Project Co-ordinator

Company name :	Istituto Centrale per il Catalogo Unico (ICCU)
Name of representative :	Rosa Caffo
Address :	Viale Castro Pretorio 105, I-00185 Roma
Phone number :	+39.06.49210427
Fax number :	+39.06. 06 4959302
E-mail :	rcaffo@beniculturali.it
Project WEB site address :	http://www.dc-net.org

Context

WP 5	Research case studies
WPLeader	I2CAT
Task 5.1	Task 5.3 Geocoded digital cultural content
Task Leader	AAS
Dependencies	All WPs
Starting date	September 2010
Release date	August 2012

Author(s)	Franc J. Zakrajsek, Vlasta Vodeb, Jurij Stare, Darija Modrijan
Contributor(s)	All
Reviewers	Matteo Lorenzini (ICCU), Bahadir Aydinonat (SGB)

HISTORY

Version	Date	Author	Comments
1.0	31.8.2012	Franc J. Zakrajsek, Vlasta Vodeb	Completed version, including the executive summary
06	1.5.2012	Franc J. Zakrajsek, Vlasta Vodeb	Draft
05	1.4.2012	Franc J. Zakrajsek, Vlasta Vodeb, Darija Modrijan	Collecting and describing use cases
04	1.3.2012	Franc J. Zakrajsek,	Added content according to Workshop: Geocoded cultural content in Ljubljana
03	31.1.2012	Franc J. Zakrajsek, Vlasta Vodeb, Jurij Stare	First Draft
02	25.11.2011	Franc J. Zakrajsek	Added index, chapters
01	19.9.2011	Franc J. Zakrajsek	First version, framework

Table of Contents

Abbreviations

API:	Aplication programming interface
CRM :	CIDOC Conceptual Reference Model[24] (CRM)
DCMES:	The Dublin Core metadata element set
ESE:	Europeana Semantic Elements Specifications
ETRS89:	European Terrestrial Reference System 1989
GCC:	Geocoded Cultural Content
GIS:	Geographic Information System
GPS:	Global Positioning System
HTML5:	5th Revision of the HTML Standard
ICT:	Information Comunication Technology
INSPIRE:	Infrastructure for Spatial Information in the European Community
IoT:	Internet of Things
ISO/TC211:	Geographic Information/Geomatics: Technical Committee 211 of International Organization for Standardization Standards
NREN:	National Research and Educational Network
SPECTRUM standard:	SPECTRUM Standard for Collections Information Management
RDF:	Rich Data Format
WGS84:	World Geodetic System Dating from 1984

1 Executive summary

The research case study **Geocoded Digital Cultural Content** (GCC) investigates the possibilities and approaches regarding the use of e-infrastructure in geocoded digital culture.

Geographic location is one of the most important attribute of every cultural heritage item. It can describe provenience, current institution, location of event or other events. The most valuable geographic description is in the form of digital geographic coordinates. Geographic coordinate presented as x, y and possibly z-values defines a position in a coordinate system. The added value of the geocoded cultural content is in the browsing cultural portals efficiently through space and time, searching for content in a more user friendly way, without need to type geographical names, making it possible to discover overlapping cultural content at the same location but originating from different sources and at different times, mapping the cultural content, performing GIS calculations and simulations, overlapping architectural/archeological heritage with museum objects and intangable heritage, defining the protected areas of monuments, geovisualisation and historical simulations.

After the **Introduction** into study (chapter 2) the **Concepts and frameworks of GCC** (chapter 3) are clearly defined. The general concepts and frameworks are applicable for different kinds of individual GCC systems. Content of use, geo features types, spatial accuracy, geocoding methods, standards, digital objects types, linking open data, devices, and e-infrastructure are examined in order to find out what are the dependencies and relations with the need for processing power, computer storage and distributed processing, and therefore the need for e-infrastructure.

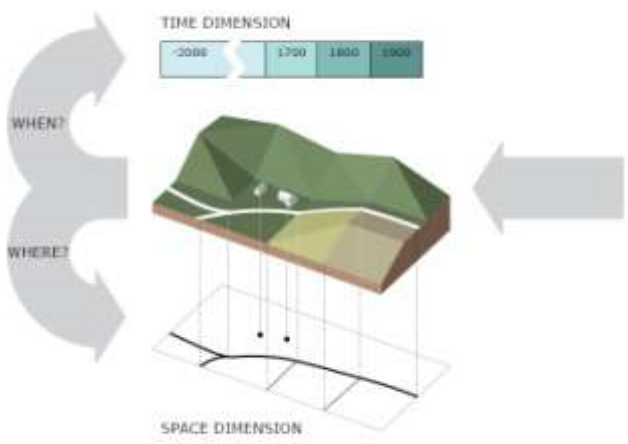
The next part **Use cases of GCC** (chapter 4) summarizes the “state of the art” of the geocoded digital culture in digital libraries, architectural/archaeological heritage, movable and intangible heritage, cultural tourism, and social networking. The focus is the identification of the possibilities and needs for e-infrastructure support. The detail descriptions of 68 individual use cases are enclosed in Appendix 2. Selected use cases are different types and could be regularly operating system, prototype, proof of the concept, or research.

The experimental part of the study is presented in **Testing of geoparsing of GCC** (chapter 5). The development of test beds and performing the testing has been done within Indicate project. The Europeana geoparser v1.0 beta and Athena project data have been used (European museums have delivered 4.082.619 objects to Europeana). The results of testing prove that geoparsing is very effective method for assigning geographic coordinates automatically. The Europeana geoparser input could be structured or unstructured attribute data describing cultural object. It performs natural text mining from textual descriptions of the cultural objects effectively. The sample testing prove the hypothesis that the geoparsing is quite useful for upper level of details (as are big towns, regions, countries and up) but its output is not very useful for spatial navigation. Geoparsing is quite good candidate for grid computing because of huge amount of processing power for natural language processing, pattern recognition, web semantics, the use of distributed local gazetteers, and the on line use on different systems and in various applications.

2 Introduction

2.1. The scope of the case study

The first part of the research **Geocoded digital cultural content** (GCC) reviews the current approaches and new R&D on geocoding of cultural content in digital libraries, cultural tourism, heritage, e-learning, and other cultural areas. Main area of the research is dedicated to identification of the possibilities and benefits of using e-infrastructure. The focus is primarily on cloud and grid computing and data infrastructures when dealing with geocoded digital cultural content. The second part of the research provides and summarizes the testing of geoparsing and geotagging e-services in digital culture and the recommendations for content providers are given.



cultural objects and events

architectural buildings
archeological sites
museum objects
books
paintings
monuments
maps
theater shows
exibitions
concerts
festivals

The following **meanings** of the terms are used in this study:

- *Geocoded* when digital geographic coordinates are added that the feature could be associated or found on the "earth" surface.
- *Digital* as opposite to analogue; stored as series of bits in computer like storage, digital camera vis-a-vis analogue camera.
- *Cultural* is cultural areas as are memory institutions: libraries, archives, museums, galleries, monuments, audiovisual and other institutions and live culture as is art, theatres, medias and other institutions.
- *Content* could be in different forms as text, image, video, audio, virtual reality and others.

2.2. Opportunities of GCC

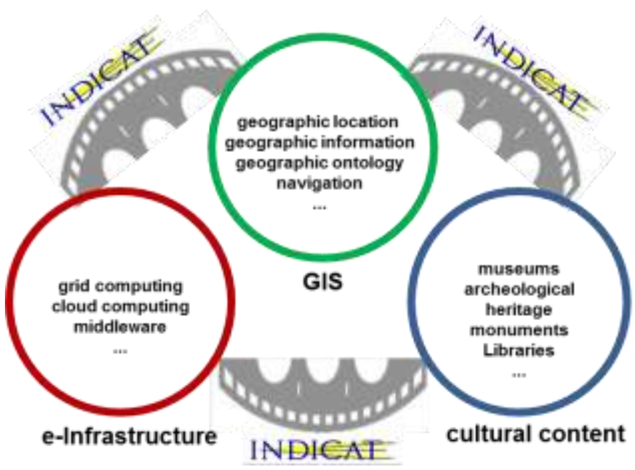
Geographic location is one of the most important attribute of every cultural heritage item. It can describe provenience, current institution, location of event or other events. The most valuable geographic description is in the form of digital geographic coordinates. Geographic coordinate (also coordinate) is presented as x, y and possibly z-values and define a position in a coordinate system. Examples of coordinate systems are system of latitude and longitude, used on the Earth's surface, and the Cartesian system.

The **added value** of the geocoded cultural content is in the:

- browsing cultural portals efficiently through space and time
- searching for content in a more user friendly way, without need to type geographical names
- making it possible to discover overlapping cultural content at the same location but originating from different sources and at different times
- mapping of the content
- performing of GIS calculations and simulations
- overlapping architectural/archeological heritage with museum objects and intangable heritage
- defining the protected areas of monuments
- geovisualisation and historical simulations

2.3. Conclusions

Scheme: INDICATE project as bridge among GIS, cultural content and e-infrastructure



The study is one of the **first**, as known for us, **facing three complex areas** as are geographic information, e-infrastructure and cultural content. Digital cultural content is very important for research of cultural heritage, promoting the culture and use / re-use of digital cultural content in cultural and creative industry as well. Geographic information and geographic information systems has been proven its potential on different areas and platforms. E-infrastructure has got new content and thus is enabled to support geocoded cultural content.

3 Concepts and framework of GCC

Concepts and framework of the topics have crucial meaning when performing this kind of study. Geocoded cultural content refers to different issues and topics which are very important when researching the state of the art or investigating the possibilities to use **e-infrastructure** in this area. On the schema bellow the main issues are identified and afterwards shortly described in subchapters.

Scheme: Framework of digital cultural content

3.1. Context of use

The context of use of GCC is wide and could be value added in the several areas as are:

- Researching, discovering cultural heritage objects, studying cultural heritage, investigating archaeological remainings, ….
- Restoration, conservation and preservation of cultural heritage
- Promoting cultural heritage on digital libraries portals, cultural institutions portal, integrating heritage into cultural routes
- Extensive use in cultural tourism and creative industry as well

E-infrastructure could be efficiently used on all the areas above because of the need for processing and/or storage power where using e-infrastructure is economically more cost effective. There are certain areas where cloud computing are moving very fast in, but grid computing is still in the research and feasibility phase.

3.2. Geo feature types

Geographical features could be:

- Point on the "earth surface" as centroid of cultural content
- Segment on the "earth surface", representing linear object as are streets and rivers
- Polygons on the "earth surface", representing detail boundaries as of archaeological protected areas, …
- Multipatch is 3D geometry, representing area or volume in three-dimensional space as 3D building

Geographical features are the components which are representing features on the Earth. There are two types of geographical features, namely natural geographical features and artificial geographical features. Natural geographical features include but are not limited to landforms and ecosystems. For example, terrain types, bodies of water, natural units (consisting of all plants, animals and micro-organisms in an area functioning together with all of the non-living physical factors of the environment). Meanwhile, human settlements, engineered constructs, etc. are types of artificial geographic features.

The need for e-infrastructure is highly correlated to the complexity of the spatial feature. 3D spatial and 3D on line rendering are very good examples for the use of e-infrastructure.

3.3. Spatial accuracy

The need for computer power processing and storage highly depends on how precise data are and what geographical area is covered and then follows the recognition of need for e-infrastructure. For example if the spatial resolution of decimetre is needed instead of 10 meters in real word, the problem is 10.000 times more complex.

Spatial accuracy of the digital geographic coordinates measures the "error" as the distance of the digital geographic coordinates and the actual position on the earth surface. The spatial accuracy is very important; for example for routing/navigation in cultural tourism the accuracy should be few meters otherwise the digital coordinates have no value.

Spatial accuracy of spatial data might be observed as macro that is appropriate for large scale maps (more than 1000 meters). Mezzo accuracy is between 5 and 1000 meters and appropriate for orientation and micro is less than 5 meters and therefore appropriate for navigation tools.

We could distinguish:

- Macro location, the level of detail is larger geographical area (e.g. country)
- Mezzo location, the location size would be for example city (this accuracy is the maximum accuracy of geoNames)

- Micro level, the spatial accuracy of place is for example address in meters
- Detail level, the level of detail are cadastral parcels (the accuracy is few centimeters, known as Lidar technology)

3.4. Geocoding methods

Geocoding is the process of assigning geographic coordinates to the location of real world entities such as houses, streets, parcels. The geocoded location can then be used in GIS. Geocoding is the process of finding associated geographic coordinates (often expressed as latitude and longitude) from other geographic data, such as street addresses, or zip codes (postal codes). With geographic coordinates the features can be mapped and entered into GIS, or the coordinates can be embedded into media such as digital photographs via geotagging.

Reverse geocoding is the opposite: finding an associated textual location such as a street address, from geographic coordinates. A geocoder is a piece of software or a (web) service that helps in this process. Three main methods of geocoding are available: by street address, by postal code; and by boundary. Geocoding is performed using a reference layer.

E- infrastructure proves the use when geocoding especially:

- When georeferencing and geocodining huge amount of historical maps
- When geocode precise location on the wider geographical area
- When geoparsing and assigning huge amount of textual digital objects

3.5. Standards

The geocoded cultural content has none or very limited use if the providers do not strongly implement **standards**. Only using standards the real cultural object is "on the same place" as is represented by its digital content. Connection and implementation of standards is therefore an obligation. Widely used standards are recommended as Dublin Core metadata element set, Europeana Semantic Elements specifications, SPECTRUM Standard for Collections Information Management, CIDOC Conceptual Reference Model. Selected GIS standards are shortly described: ISO/TC 211 Geographic Information/Geomatics, Open Geospatial Consortium Standards and The INSPIRE Directive.

1. ISO/TC 211 Geographic Information/Geomatics

Standardization in the field of digital geographic information is in the domain of Technical Committee 211 of International Organization for Standardization (ISO/TC 211 Geographic information/Geomatics).

Technical Committee 211 is working towards establishing "a structured set of standards for information concerning objects or phenomena that are directly or indirectly associated with a location relative to the Earth".

Standards address the infrastructure for geospatial standardization, data models for geographicinformation, geographic information management, geographic information services,encoding of geographic information and specific thematic areas. More specifically they include methods, tools and services for data management and also its definition and description.

They concern acquiring, processing, analyzing, accessing, presenting and transferring geographical data in digital form between different users, systems and locations. ISO has published fiftytwo standards under the direct responsibility of Technical Committee 211 by February 2011.

The ISO/TC 211 group of standards provides for a fundamental structure of geographic information thus enabling its computational processing. They lay the foundation on which other developments are possible, like INSPIRE Directive.

2. Open Geospatial Consortium Standards

The Open Geospatial Consortium (OGC) is a voluntary consensus organization that is leading the development of standards for geospatial content and location based services and also for GIS data processing and sharing. They are encouraging the development and implementation of open standards, free and openly available to the market.

Organisation has close connections with other international standards bodies, especially ISO/TC 211 (Geographic Information/Geomatics). The ISO 19100 series under development by Technical Committee 211 will progressively replace the OGC abstract specification. The OGC standards Web Map Service, Geography Markup Language (GML), and Simple feature access have become ISO standards.

The main concern of OGC is development of standards and specifications which would establish interoperability in the processing of geographical information. Interoperability is considered as one of the key aspects in designing information systems in cultural heritage field.

The OpenGIS standards have formed the basis for the development of open source software which is frequently used in the cultural heritage field, for example OpenLayers and Geoserver.

3. The INSPIRE Directive

The INSPIRE Directive aims to establish spatial information infrastructure in Europe in order "to support Community environmental policies, and policies or activities which may have an impact on the environment". It came into force on 15 May 2007 and its implementation will follow various stages until full implementation in 2019.

The INSPIRE Directive ensures compatibilty and usability of the spatial data infrastructures of the Member States in European Union. To achieve that the Directive requires that common Implementing Rules are adopted in a number of specific areas: metadata, data specifications, network services, data and service sharing and monitoring and reporting.

These Implementing Rules are adopted as Commission Decisions or Regulations, and are binding in their entirety.

The Directive is addressing 34 spatial data themes organized in three annexes. INSPIRE spatial infrastructure provides great opportunity to be used also in the digital cultural heritage field. Firstly the implementation rules as they are set for coordinate reference systems, geographical names and administrative units can be used as a methodological background and as technical standards. On the other side the INSPIRE spatial data as orthoimagery and geographical names could be directly used when representing digital cultural content on web maps.

4. Coordinate Systems

Coordinate system in which cultural content is geocoded is one of the main issue. The geographical coordinate system describes coordinates on the sphere. The geographical coordinate system is used on globe (Earth) or on continental level. The projected coordinate system uses projected coordinates to plane. Projected coordinate systems are used more on the national or regional level.

At the time being WGS84 geographic coordinate system is the widely used. WGS84 (also World Geodetic System 1984) consists of a three-dimensional Cartesian coordinate system and an associated ellipsoid, which enables the description of positions as either XYZ Cartesian coordinates or latitude, longitude and ellipsoid height coordinates. WGS84 (dating from 1984 and last revised in 2004) is the reference coordinate system used by the Global positioning system.

When gecoding cultural content, especially protected architectural and archeological sites, using the WGS84 is not spatially accurate enough. The use of ETRS89 is demanded by INSPIRE directive for such cases. ETRS89 (European Terrestrial Reference System 1989) is used as the standard precise GPS coordinate system throughout Europe. It is tied to the European continent, and hence it is steadily moving away from the WGS84 coordinate system. In 2000, the difference of a point coordinates is about 25cm, and increasing by about 2.5 cm per year.

Use cases and links:

- ISO/TC 211 Geographic Information/Geomatics: http://www.isotc211.org/
- Open Geospatial Consortium Standards: http://www.opengeospatial.org/
- The INSPIRE Directive: http://inspire.jrc.ec.europa.eu/
- Dublin Core metadata element set: http://dublincore.org/documents/dces/
- Europeana Semantic Elements specifications: http://pro.europeana.eu/documents/900548/dc80802e-6efb-4127-a98e-c27c95396d57
- SPECTRUM Standard for Collections Information Management: http://www.dcc.ac.uk/resources/standards/diffuse/show?standard_id=160
- CIDOC Conceptual Reference Model: http://www.cidoc-crm.org/

3.6. Digital object types

Digital objects can be conceived as a compound artefact that wraps digital material in terms of four elements: its content, its metadata, its relationships with other objects and its behaviour. It is evident here that all objects in a single group share near-identical structure and behaviour.
The types of digital object used in cultural area reflect the ways of using **e-infrastructure**, they could be:

- still images usually geocoded to points; for cultural area high definition pictures are interesting,
- video geocoded also to mid points
- text annotated with geotags
- 3D models exactly allocated in 3-dimensional space and with right orientation; from simple 3D PDF model to complex 3D City GML models

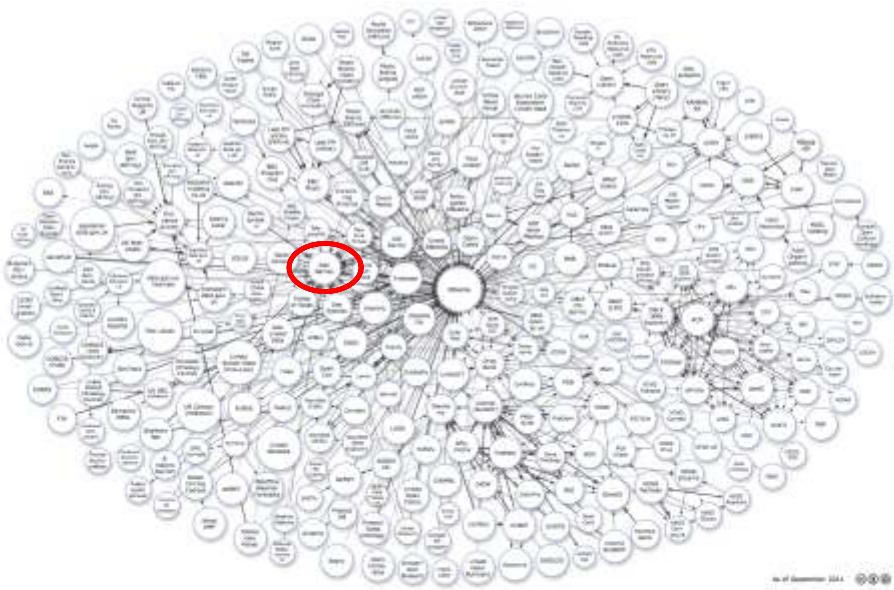
3.7. Linking open data

The real need for using e-infrastructure is demanded when linking open geographical data in order to have stable ICT environment and then geocoded cultural data are accessible 24 hours a day and 7 days a week.

In computing, linked data describes a method of publishing structured data so that it can be interlinked and become more useful. Method builds upon standard Web technologies such as HTTP and URIs, but rather than using them to serve web pages for human readers, it extends them to share information in a way that can be read automatically by computers. This enables data from different sources to be connected and queried.

The linking open data in the geographical content area is not unknown. The example is gazetteer GeoNames.

Scheme: The Linking Open data cloud diagram



Linking Open Data cloud diagram by Richard Cyganiak and Anja Jentzsch, (source: http://lod-cloud.net/)

GeoNames is a geographical database that contains over 10 million geographical names and consists of over 8 million unique features. GeoNames is integrating geographical data such as names of places in various languages, elevation, population and others from various sources. All lat/long coordinates are in WGS84. The data is accessible free of charge through a number of web services and a daily database export. GeoNames is already serving up to over 30 million web service requests per day.[1]

[1] http://www.geonames.org/

Important iniciative LinkedGeoData is also an effort to add a spatial dimension to the Web of Data / Semantic Web. LinkedGeoData uses the information collected by the OpenStreetMap project and makes it available as an RDF knowledge base according to the Linked Data principles. It interlinks this data with other knowledge bases in the Linking Open Data initiative.

Use cases and links:

- LinkedGeoData: see Appendix 2
- The Linking Open Data: http://lod-cloud.net/
- New York Times Company: http://www.nytco.com/

3.8. Devices

The subchapter introduces the simple question about devices that enable viewing and using geocoded cultural heritage. This devices are not produced just for reading the content but also for updating the content and retrieving or identifying geographic location, e.g. when accompanied with GPS.

Some most popular devices are:
Photo / video cameras

- Smart phones (built on mobile computing platform)
- Tablet computers (iPad, ...); especially interesting for portable dimension of cultural tourism
- Television and other audio visual devices
- Portable computers
- Stationary personal computers
- Display show equipment

The application and production of ICT environment of services is usually developed to be used on different devices.

3.9. E-infrastructure

E-Infrastructure[2] is a term used for the technology and organizations that support research undertaken in this way. It embraces networks; grids, data centers and collaborative environments, and can include supporting operations centers, service registries, single sign-on, certificate authorities, training and help-desk services. Most importantly, it is the integration of these that defines e-infrastructure.

E-Infrastructure comprises 6 layers or 'perspectives':

Connectivity: The first is high speed connectivity. GEANT2 provides continuous top-of-the range connectivity with much higher levels of performance to researchers, educators and students in order to lower access barriers to distributed resources and instrumentation. Other projects like SEEGRID2 and EELA-2 help extend its reach to regions worldwide and ensure the real creation of virtual research communities.

Grid computing: Another essential layer is "Grid computing" which allows researchers to do huge calculations using many computers simultaneously. For example, EGEE-III which is a key component of storing and elaborating the data from the LHC, and many other collaborative scientific research projects. See GridTalk for more information on grid computing.

Supercomputing: Harnessing the power of supercomputers to run in different calculations on parallel for a research projects is another yet distinct perspective of the grid computing layer. The DEISA project is a great example of what can be achieved when combining the power of the EU supercomputer resources, and the PRACE project is illustrative as pioneering work done in this area.

Scientific Data: The third layer is a coherent and managed eco-system of repositories of scientific data that projects can share within and between different communities of research practice. Europe is defining consistent policies to enhance access to this scientific information, and ensure its sustained use and value in the long term.

Global Virtual Research Communities: With the maturing of the different 'layers' of e-Infrastructure, a new paradigm of research is developing, where communities of researchers in Europe and globally will work together sharing best practices, software and data virtually. These global virtual research communities will ensure societies reap the high innovation potential of multidisciplinary e-Infrastructure enabled research.

Standards: Finally, when speaking of virtual global research communities an important reference needs to be done to the standards, since without them neither connectivity neither

[2] Cited from: http://www.beliefproject.org/

interoperability is possible. Please browse the webpage of the OGF-Europe project and other important European initiatives like ETSI.

There are many other great research or testbed projects that use the e-infrastructures for different research application areas.

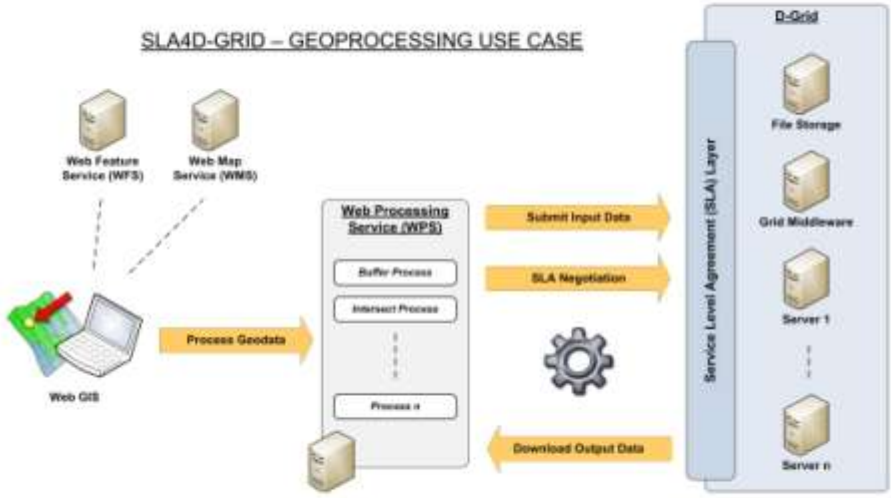
1. GRID computing

Grid computing was initially driven by the needs of applications with large volumes of data and complex calculations as astronomy, biology and physics. As grid computing matures is being extended to other sciences and for other uses.

The term 'grid computing' is a diffuse phrase and there are many definitions available. This lack of a sole definition leads to many people working with grid technology and having different views on what a grid is.

Gridification as a term need further explanation. »It is possible to distinguish between two categories of gridification of an OWS. In a simple approach the existing application or the existing service (in this context an OWS) stay primarily unchanged and the grid is used as a computation- or a data-resource (a low level gridification; see Figure 1a). It is possible to do calculation tasks distributed in the grid and to use grid services for accessing data. In this scenario the existing not-gridified application becomes the bottleneck because it doesn't obtain all qualities of a general grid service. This kind of gridification is very easy to realize and the implementation is nearly independent of the underlying grid middleware. In a more complex approach the existing application or the existing service (in this context an OWS) is full embedded into the grid middleware (e.g. as a stateful service inside WS-Resource Framework, WSRF) and obtains all qualities of a generalgrid service (a high level gridification). It is possible to do calculation tasks distributed in the grid and to use grid services for accessing data. This solution is not easy to implement and there is a need of a proxy to stay OWS compliant and to handle the communication (see Figure 1b).«[3]

[3] Baranski, B., 2008. Grid Computing Enabled Web Processing Service. GI Davs 2008: Interoperability and spatial processing in GI applications. Munster, Germany. (http://www.gi-days.de/archive/2008/downloads/acceptedPapers/Papers/Baranski.pdf)



Sla4D grid geoprocessing use case (Source: http://www.sla4d-grid.de/solutions/live_demo)

Use Cases and links:

- Distributed Geo-rectification of Satellite Images using Grid Computing: research, see Appendix 2
- Grid based 3D animation rendering: research, see Appendix 2
- A GEO Grid Implementation for 3D GIS Taiwan: research, see Appendix 2
- Grid-Based Digital Libraries: Cheshire3 and Distributed Retrieval: research, see Appendix 2
- Using Web Portal for 3D Grid-Based Rendering: research, see Appendix 2

2. CLOUD computing

Cloud computing[4] is the dynamic delivery of information technology resources and capabilities as a service over the internet. Cloud computing is a style of computing in which dynamically scalable and often virtualized resources are provided as a service over the internet. It generally incorporates infrastructure as a service (IaaS), and software as a service (SaaS).

According to Gartner group (http://www.gartner.com), the attributes of cloud computing are:

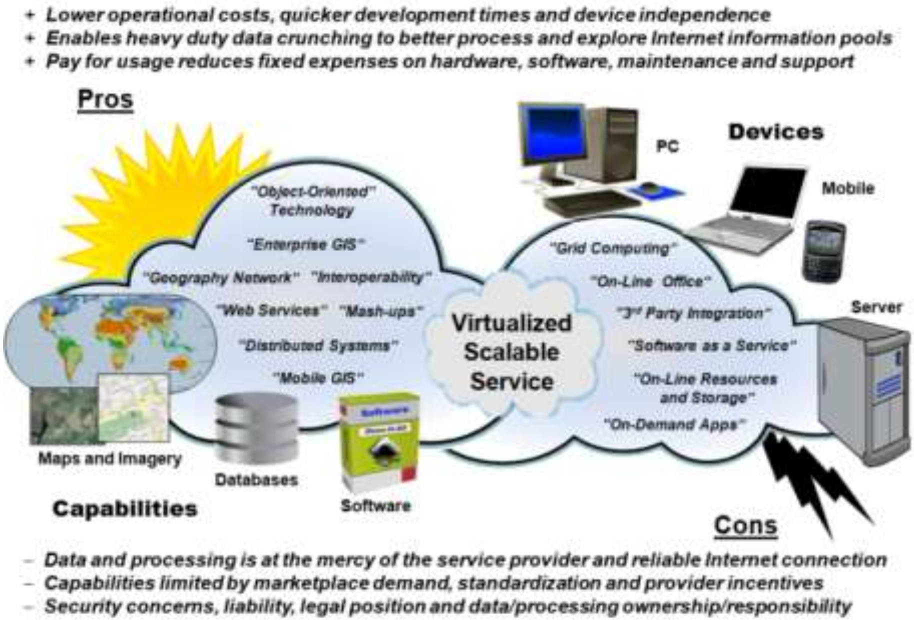
- service-based,
- scalable and elastic,
- shared,
- metered by use,

[4] Sarna, D.E.Y. (2011). Implementing and developoing cloud computing applications. New York: Taylor and Francis group

- use of internet technologies.

The most frequently cited benefits of cloud computing are:

- it is agile, with ease and speed of deployment,
- its cost is use-based, and will likely be reduced,
- in-house IT costs are reduced,
- capital investment is reduced,
- the latest technology is always delivered,
- the use of standard technology is encouraged and facilitated.



Source: GIS Evolution and Future Trends. In: MAP Analyses, GeoTec Media 2007

Cloud computing offers great opportunity for ICT support and therefore for GIS areas and cultural sectors as are libraries, museums, archives and other cultural institutions. The cloud computing should be taken into account, especially:

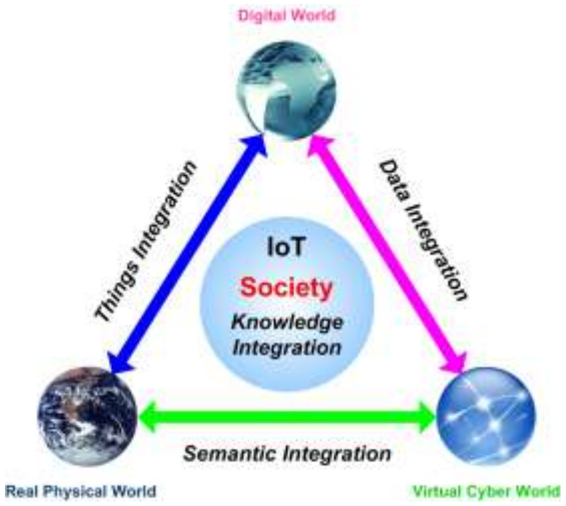
- when introducing new content management systems,
- for small and medium size cultural institutions,
- when share the content as is Europeana for example,
- to use and reuse the geographical information,
- when offered NRN, National Academic Research Network.

Use cases and links:

- Libraries and the Cloud: research, see Appendix 2
- Museums and Cloud Computing: Ready for Primetime, or Just Vapourware? : research, see Appendix 2
- Cloud Computing Primer: Steps for using the cloud in Your Museum: research, see Appendix 2
- Cloud Computing in the Application of Digital Library: research, see Appendix 2

3. Internet of things and other new technological issues

The Internet of Things refers to uniquely identifiable objects (things) and their virtual representations in an Internet-like structure. Ashton's original definition of Internet of things began with: "Today computers—and, therefore, the Internet—are almost wholly dependent on human beings for information."[5]



Internet of Things (source: http://www.iot-a.eu)

In an Internet of Things[6], the precise geographic location of a thing—and also the precise geographic dimensions of a thing—will be critical. Currently, the Internet has been primarily used to manage information processed by people. Therefore, facts about a thing, such as its location in time and space, has been less critical to track because the person processing the information can decide whether or not that information was important to the action being taken, and if so, add the missing information (or decide to not take the action). Note that some things in the Internet of Things will be sensors, and sensor location is usually important. The GeoWeb and Digital Earth are promising applications that become possible when things can become organized and connected by

[5] Ashton, K: That 'Internet of Things' Thing. In: *RFID Journal*, 22 July 2009. Abgerufen am 8 April 2011.

[6] Wikipedia, Internet of Things, http://en.wikipedia.org/wiki/Internet_of_Things

location. However, challenges that remain include the constraints of variable spatial scales, the need to handle massive amounts of data, and an indexing for fast search and neighbour operations. If in the Internet of Things, things are able to take actions on their own initiative, this human-centric mediation role is eliminated, and the time-space context that we as humans take for granted must be given a central role in this information ecosystem. Just as standards play a key role in the Internet and the Web, geospatial standards will play a key role in the Internet of Things.

Concerning **e-infrastructure** and geographic information in culture heritage field also the other technological newness will play big role in next years. First of all HTML5 and Geolocation API specification, x3D, CityGMLand and other new standards should be considered.

Use cases and links:

- Internet of Things - An action plan for Europe, Communications from the Commission to the European Parlamient, the Souncil, the European Economic and Social Commitee and the Committee of the Regions, Commission of the European Communities, Brussels, 18.6.2009, COM(2009) 278 Final
- HTML 5 Reference, source: http://dev.w3.org/html5/html-author/

4 Use cases of GCC

The goal of this chapter is to describe current "state of the art" of the geocoded digital culture. Few hundreds of use cases were examined and here 68 of them are presented in the Appendix 2.

Each use case is presented by:

- Name
- Type and geographical area
- Short description
- Main links and/or sources
- Graphical display
- Comments

The criterion for the selection of particular use case is that it has to cover at least two topics among: geographic information, e-infrastructure / ICT, cultural content:

- If use case covers **grid or cloud computing and GIS** (example: geogrid) it is assumed that these experiences could be transferred also to cover certain geocoded cultural content (e.g. distributed monuments inventory).
- If use case covers **cultural area and GIS** (example: Museum content on GIS layer) it could be transformed to cloud technology or by gridification to grid computing (e.g. distributed on line geoparsing).
- If use case covers **cultural area and grid or cloud computing** (e.g. grid storage of high resolution cultural images) the new services for geographic location could be added (e.g. by spatial retrieval of cultural images).

Use cases may be regularly operating system, prototype, proof of the concept, research...

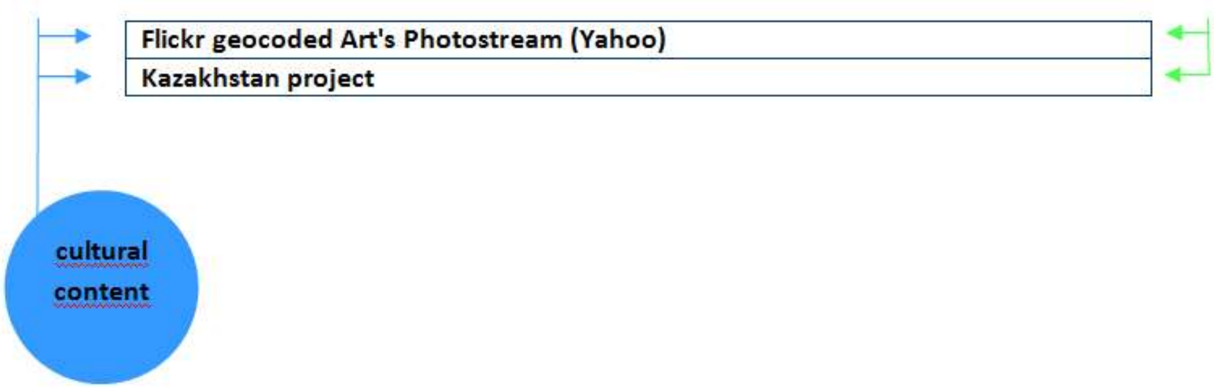
GEOCODED CULTURAL CONTENT

grid, cloud

GIS

A Geo Grid Implementation for 3D GIS Taiwan
Design of a Grid-based Geo-service Architecture
Distributed Geo-rectification of Satellite Images using Grid Computing
Geographic Information and Grid Computing : An introduction
GeoMiddleware to Support Interoperability for Grid Computing
Grid based 3D animation rendering
Grid Computing Enabled Web Processing Service
Research of the application of grid computing on geographical information system
Using a Computational Grid for Geographic Information Analysis: A Reconnaissance
Use of grid computing for modelling virtual geospatial products
Using Grid Computing for Rendering to Support 3D Animation Training Courses
Using Web Portal for 3D Grid-Based Rendering
Digital library grid: A roadmap to next generation digital libraries using grid technologies
Grid-Based Digital Libraries: Cheshire3 and Distributed Retrieval
Cloud Computing in the Application of Digital Library
Cloud Computing Primer: Steps for using the cloud in Your Museum
Libraries and the Cloud
Museums and Cloud Computing: Ready for Primetime, or Just Vaporware?
Geographical Linked Data: a Spanish Use Case
LinkedGeoData
A GIS in cultural heritage based upon multiformat databases and hyper medial personalized queries
Accessing Heritage Documents according to Space Criteria within Digital Libraries
Advanced GIS technologies to support georeferencing of the Cultural Heritage
Geographic Information Contribution and Retrieval - An Agenda for the Next Generation Gazetteer
AskAboutIreland, culture on the interactive map
Atlas of Heritage and Architecture
Connecting Historical Authorities with Links, Contexts and Entities
CultureMap London
Cultnat' experience in Geo-coding culture heritage content

Development of a GIS Based Information and Management System for Cultural Heritage Site, Case Study of Safranbolu
Developing a Spatial Data Infrastructure for Cultural Heritage
Digital Atlas on the History of Europe since 1500
Embedding GeoCrossWalk Final Report
Explorative user interfaces for browsing historical maps on the Web
Europeana Mapsearch
Geocode your Twitter network with NodeXL
German Heritage Register Bayern – Nürnberg
Gis & Social Media Integration
GIS system for the Catalan Cultural Heritage
GIS technologies for the study of the Roman agricultural landscape
Locating London's Past
MEGA-J Middle East Geographical and Archaeological database
Mobile cultural heritage guide: location-aware semantic search
NAC Locator - A Universal Geocoding Solution for the Entire World
National Heritage List for England
National Heritage Register Netherlands
National Register of Sites and Monuments Denmark - Fund og Fortidsminder
Novel approach to 3D archeology, 3D semantics, open sources and open standards, experiences of geoparsing CulturaItalia
Odysseus, www server of the Hellenic Ministry of Culture
Past places - place names: www.hgis-germany.de?
Picture War Monuments: Creating an Open Source Location Based Mobile Platform
Putting Museum Collections on the Map: Application of Geographic Information Systems
Reorganizing the Topographic Databases of the Institut Cartogràfic de Catalunya applying generalization
Register of cultural heritage of Slovenia (RCHS)
Use of the Edinburgh Geoparser in the GeoDigRef and Embedding GeoCrossWalk Projects
Virtual Museum via Flaminia Antica
3D Artefact Acquisition (3D COFORM Tools & Expertise for 3D Collection Formation)
American Memory
ArXiv
Europeana Culture Globe
Europeana portal
Europeana4D
3D historical maps
A guide to the magnificent Awqaf Mosques of Al Darb Al Ahmar Area
Appia Antica archaeological Park
EuropeanaConnect



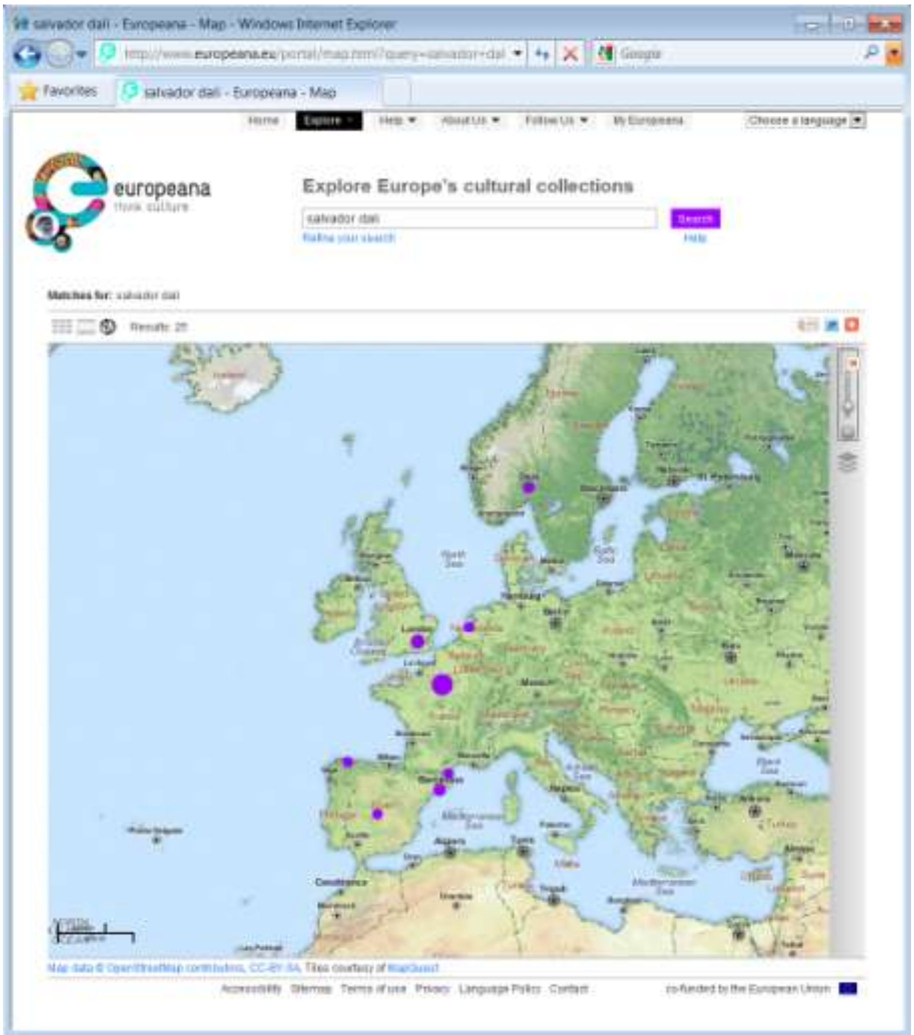
The remaining of the chapter summarizes the examined use cases. Use cases are grouped into five topics:

- Digital libraries
- Architectural / archeological heritage
- Movable and intangible heritage
- Cultural tourism
- Social networking

The focus of subchapters is to identify the possibilities and needs for e-infrastructure support. The detail descriptions of 68 individual use cases are references here and enclosed in Appendix 2. Use cases are different types of regularly operating system, prototype, proof of the concept, or research.

4.1. Digital libraries

Digital library is a collection of digital content from libraries, archives, museums and other cultural institutions. It contains internal collections in e.g. museum or in certain branch e.g. movable heritage and resides at national level, european level and world level.



Europeana digital library (source: http://www.europeana.eu/portal/)

Geographic coordinates for description of the geographic coverage or the location of digital objects are becoming more and more important. It could not be imagined to type the place name when querying among ten millions of digital objects instead just to click on the area on the map and retrieve the data. At the time being the geographic features are simple and are more or less limited to the “point” features.

As far as could be estimated now, the following topics of digital libraries will significantly benefit from **e- infrastructure:**

- digital library as a harvester and as one stop portal could economically benefit with cloud ICT
- content providers will delivery to harvester with cloud environment more reliable
- on line accessibility of digital object (not temporary unreached)
- georeferencing, processing, storage and use of huge repository of historical maps is very good subject of grid computing

Use cases and links:

- Europeana culture globe: see Appendix 2
- Europeana portal: see Appendix 2
- arXiv: see Appendix 2
- American Memory: see Appendix 2
- Europeana 4D: see Appendix 2
- Juidaica Europeana: see Appendix 2
- LinkedGeoData.org: see Appendix 2

4.2. Architectural / archeological heritage

Architectural and archaeological heritage refers to a place, locality, natural landscape, settlement area, architectural complex, archaeological site, or standing structure from inventories, management, restoration,...

Ljubljana 3D-City (source: https://urbanizem.ljubljana.si/3durbanizem)

Geographic information systems have a long tradition of use in architectural and archaeological heritage field. At the beginning were used just for delineating of the protected areas. Nowadays are used also for 3D detail modelling for research, restoration of monuments and for the 3D presentation of heritage to general public.

Geographic information in archaeological / architectural sectors is used when capturing data, for management the repositories, and processing and displaying data on the maps. The level of detail goes to individual site or object.

Appropriate tasks for **grid computing**:

- Risk scenarious simulations
- Risk management for cultural heritage
- Climate changes simulations
- 3D visualisation
- Spatial statistics
- Spatial analyses
- Geoprocessing services (WPS)

Potentials of use grid computing for caching

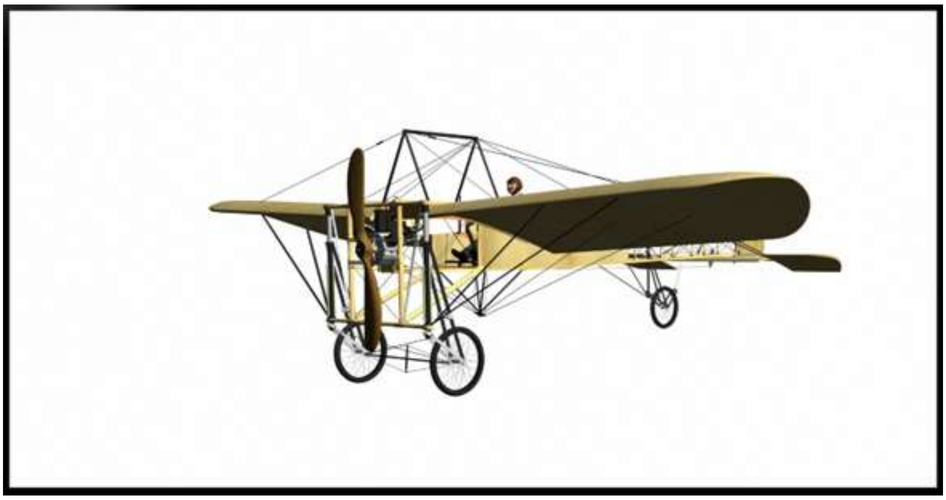
Steps	Software and hardware	Estimated time
Experiment Cashing area: 152 km2 scale 1:76 (approx. 2D 1:1000) Tiles: 512x512 pixels (finally 104.000 tiles (3/4 tiles))	ArcGIS Server 2x E5450 3GHz (8 threads) 32GB memory	Cashing time: 77 minut.
Generalization for the world mainland: 148.429.000 km2, million times larger area than in experiment	ArcGIS Server 2x E5450 3GHz (8 threads) 32GB memory	Estimated 77 milions minutes or 146 years
Google	• In 2002; upwards of 15,000 servers • A 2005 estimate by Paul Strassmann has 200,000 servers claimed this number to be upwards of 450,000 in 2006 900.000 (2011)	

Use cases and links:

- Register of Immovable cultural heritage of Slovenia: see Appendix 2
- National Heritage List for England: see Appendix 2
- National Heritage Register Netherlands: see Appendix 2
- National Register of Denmark: see Appendix 2
- Atlas of Heritage and Architecture, France: see Appendix 2
- German Heritage Register Bayern – Nürnberg: see Appendix 2
- MIDAS Heritage, the UK Historic Environment Information Standard: http://www.english-heritage.org.uk/
- 4D Cities: http://4d-cities.cc.gatech.edu/atlanta

4.3. Movable and intangible heritage

Movable heritage refers to natural or manufactured object of heritage significance. Intangible heritage refers to practices, representations, expressions, knowledge, skills that communities recognize as part of their cultural heritage. This report covers inventories of movable and intangible heritage, research on movable and intangible heritage, their management and other uses.



3D model of Bleriot XI ready for wind tunnel simulation (http://www.bleriot.arts-et-metiers.net/)

Geographic coordinates and geographical data are relatively new in the area of museums, libraries and archives. The concepts what could be geocodes of intangible heritage are still unclear. Geographic information can be revealed from provenience, place of origin, place of creation, geographical area where the practice continues; spatial data can be used for processing and displaying data on the maps.

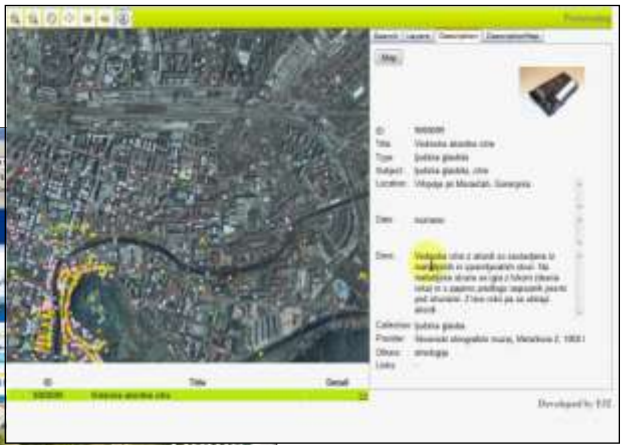
Cloud computing is coming rapidly to cultural institutions as are museums and libraries. It is recommended to conduct feasibility/business model study when choosing the cloud computing provider. The use of NREN (National Research and Educational Network) as provider is preferred choice than established (E2c) or "from the street" provider.

Use cases and links:

- 3D Artefact Acquisition: see Appendix 2
- Picture War Monuments: research, see Appendix 2
- Museums and Cloud Computing: Ready for Primetime, or Just Vapourware?, research, see Appendix 2
- Zakrajšek, F., Vodeb, V.: Digital Cultural Content: Guidelines for Geographic Information, Athena project, 2011
- Musee des Artes Metiers: http://www.bleriot.arts-et-metiers.net/

4.4. Cultural tourism

Cultural tourism is concerned with a country or region's culture. Persons visit cultural attractions with the intention to gather new information and experiences to satisfy their cultural needs. Cultural tourism includes all products associated with promotion, visits to attractions and sites, museums, and the indigenous products, festivals or theater. It includes heritage documentation, inventories of tourism related cultural sites and data management systems to manipulate information on tourism.



Museum Gecoded Content: Guidelines for Geographic Location Description of Digital Cultural Content (http://www.athenaeurope.org/)

Hotel reservation system (http://www.booking.com)

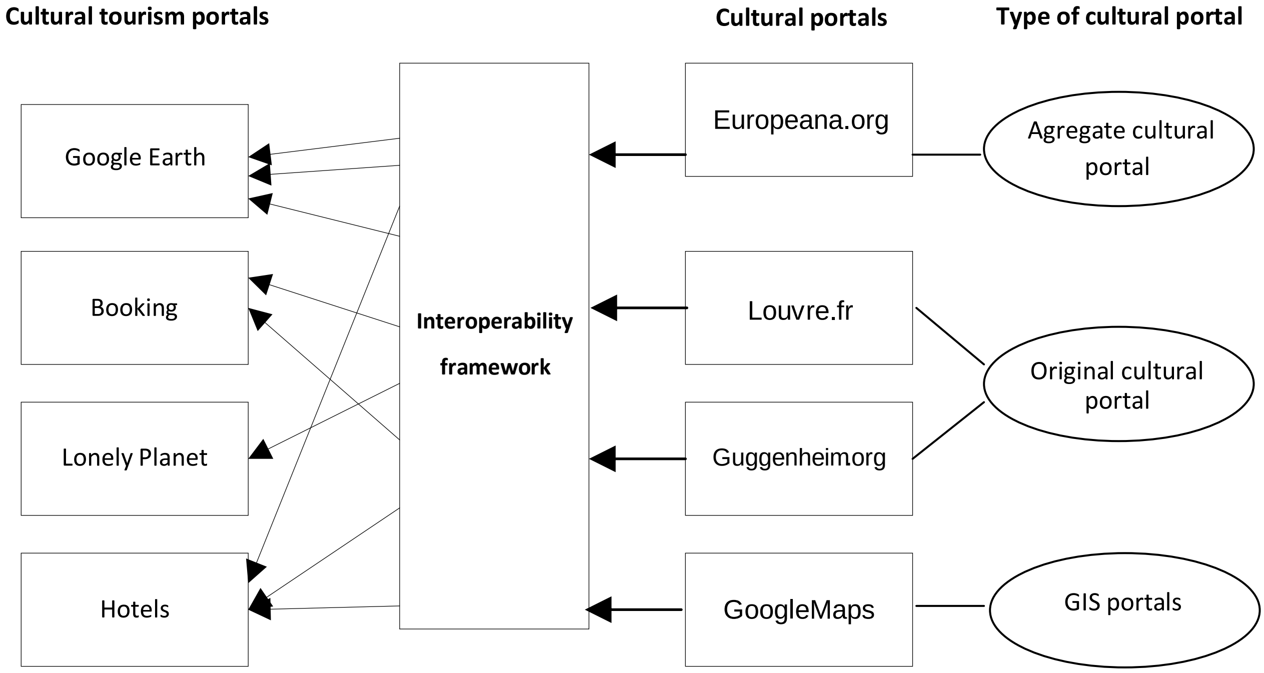
Within this study the short review of some “travelling” portals has been done. The results are presented in the table below.

Review of tourism / booking /routing portals

name	http	framework	Cultural heritage comment	Comment
Michelin	www.viamichelin.com	Web page with searchable information on driving directions, maps, weather forecast, hotel booking, restaurants and travel guides.	Cultural heritage is found on a map under Tourism > Destinations and then each destination had a link where tourist attractions are listed, described with photos and detailed map.	Ljubljana is the only town with listed and described cultural heritage among tourist attractions.
Google earth	www.google.co.uk/intl/en_uk/earth	Free version: maps with different cartographic layers: satellite imagery, maps, terrain, 3D buildings, galaxies in outer space and the depths of the ocean.	Cultural heritage can be identified under different layers. Primary database lists heritage under landmarks and City marks. Cultural heritage can be found under photos, where user can participate. Historical imaginary contain heritage. 3D buildings layer contains 3D models of monuments, fountains, bridges, towers, museums and more. User can participate 3D models on the map.	Only few cultural heritage objects and monuments or galleries are displayed.
Booking	www.booking.com	Online hotel reservations website is available in 41 languages and offers over 188,467 accommodations in 163 countries.	Cultural heritage is marked on a map among other landmarks.	Only few cultural heritage objects and monuments or galleries are displayed.

Hotels	www.hotels.com	Online hotel reservations website offers over 140,000 accommodations worldwide.	Cultural heritage Map is searchable by landmarks, among them are listed also some immovable cultural heritage sites. Google basemaps are used; street map and satellite with their embedded symbology for landmarks. User can also check a landmark to see it on a map.	Only few Cultural heritage objects and monuments or galleries are displayed.
Lonely Planet	www.lonelyplanet.com	Lonely Planet displays information regarding traveling for world destinations. They employ around 450 employees and over 200 authors.	Information on cities and countries is mapped. Google basemaps are used; street map and satellite with their embedded symbology for landmarks. Cultural heritage is marked on a map as a sight. They provide basic information on a cultural heritage with address, short description and with the possibility to zoom on a sights and with the link to read more about the heritage.	Only few Cultural heritage objects and monuments or galleries are displayed.

The “travelling” portals usually include the interactive **geographic map** with locations of hotels, streets, and other interesting object. For cultural tourism it is a pity that just a few more or less randomly cultural objects are presented with very poor linked information. Therefore it is strongly advised to merge the travelling services with cultural content services.



Cultural tourism exists to satisfy cultural travellers to get cultural experience. According to statistics their travel is longer in time and distance. Beside this economic aspect cultural tourism raises the potential of creative tourism. Cultural tourists are no longer satisfied observing cultural heritage and events – they want to participate in the creation or development of local cultures.

Technology has important impact on cultural tourism regarding visitor experience of heritage and presenting heritage with new technologies – virtual exhibitions, GIS tools. Further important aspect is interconnection of tourism web portals with different services, as navigation, booking, ticketing, etc.

The interoperability framework between travelling services with cultural content services should work very stable and reliable therefore the **cloud ICT** could improve the performance of framework.

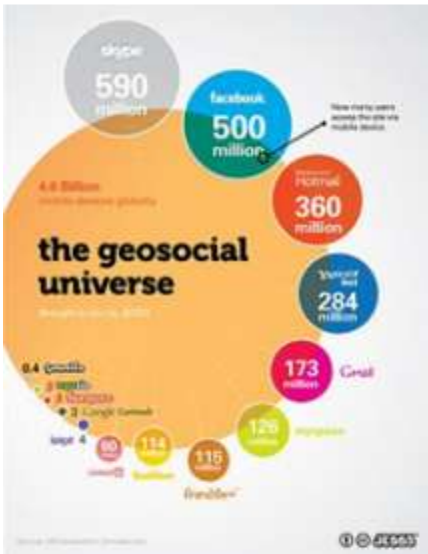
If someone plans to use more computer intensive processing as is shortest path algorithm or traveler sales algorithm also **grid computing** should be taken into account.

Use cases and links:

- AskAboutIreland: see Appendix 2
- Appia Antica archaeological Park: see Appendix 2
- CultureMap London: see Appendix 2
- Locating London's Past: see Appendix 2
- Lonely planet: www.lonelyplanet.com
- Hotels reservation system: www.hotels.com
- Via Michelin: www.viamichelin.com

4.5. Social networking

A social networking service[7] is an online service, platform, or site that focuses on facilitating the building of social networks or social relations among people who, for example, share interests, activities, backgrounds, or real-life connections. A social network service consists of a representation of each user (often a profile), his/her social links, and a variety of additional services. Most social network services are web-based and provide means for users to interact over the Internet, such as e-mail and instant messaging. Online community services are sometimes considered as a social network service, though in a broader sense, social network service usually means an individual-centered service whereas online community services are group-centered. Social networking sites allow users to share ideas, activities, events, and interests within their individual networks.



JESS3 Labs: The Geosocial Universe for year 2010 (source: http://jess3.com/geosocial-universe/)

[7] http://en.wikipedia.org/wiki/Social_networking_service

Geosocial networking is a type of social networking in which geographic services and capabilities such as geocoding and geotagging are used to enable additional social dynamics. User-submitted location data or geolocation techniques can allow social networks to connect and coordinate users with local people or events that match their interests. Geolocation on web-based social network services can be IP-based or use hotspot trilateration. For mobile social networks, texted location information or mobile phone tracking can enable location-based services to enrich social networking.

The cloud computing could improve reliability of the individual provider of cultural services.

Use cases and links:

- Flickr geocoded Art's Photostream: see Appendix 2
- Geosocial networking, Interpol: https://www.europol.europa.eu/sites/default/files/publications/geosocialnetworking.pdf
- List of social networking websites: http://en.wikipedia.org/wiki/List_of_social_networking_websites
- Twitter: http://www.twitter.com
- Facebook: http://www.facebook.com
- Skype: http://www.skype.com
- Gmail: http://www.google.com

5 Testing of geoparsing of GCC

Geographic information in the form of digital geographic coordinates makes cultural content more effective and usable. If the geographic coordinates do not exist in the description of certain cultural collection their automatic retrieval from non-structural or structural text information is one possibility.

Geoparsing is the process of assigning geographic coordinates to textual words and phrases (e.g. “The author has been born in Rome”). Geoparsing is capable of handling ambiguous references in unstructured content. Geoparsed features can then be mapped and entered into a geographic information system. A geoparser is a piece of software or a (web) service that helps in this process.

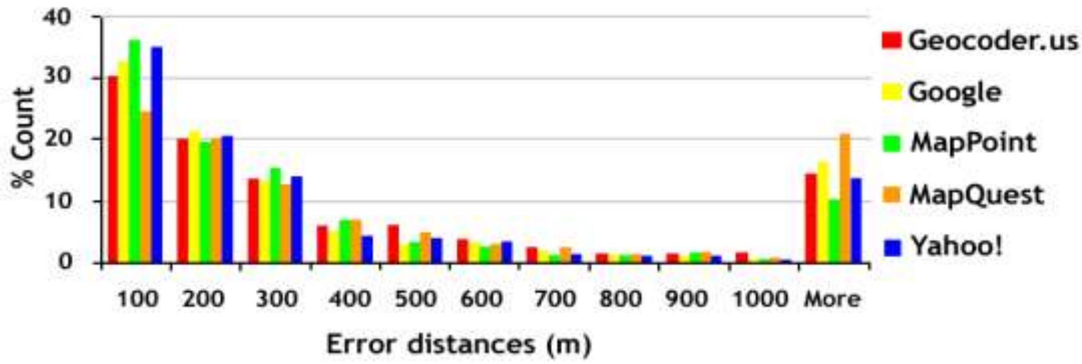
Purpose of testing:

- Could we retrieve the geographical coordinates from the textual metadata of the certain digital content?
- What strategies and geoparsing services could we use for geoparsing?
- What percentage of the content could be geocoded in this way, at the best?
- For what purpose / services could we use the geoparsed geographical coordinates (spatial accuracy)?
- To plan the real production of geoparsing.

5.1. Review and selecting the geoparser

Several geoparsers are available for use: GeoCrossWalk GeoParser, Edinburgh geoparser, Klokan geoparser, Yahoo!'s placemaker, MapQuest's geocoding service, Geocoder.us, Google Maps geocoding service, MapPoint Web Services' FindAddress.

Usually the distribution of error distances computed between each geocoded point and its corresponding baseline for each service is reviewed as in the bellow graph.



Source: Roongpiboonsopit D., Kmimi H.A. (2010). Comparative evaluation and analysis of online geocoding services. In: International Journal of Geographical Information Science, Vol. 24, No. 7-8, July-August 2010, p.1081-ll00

Finally, the **Europeana Geoparser v1.0 Beta** has been used for testing data. The Europeana Geoparser has been developed for Europeana content providers and as such reasonable choice for testing the data for this report.

5.2. Selection of testing data

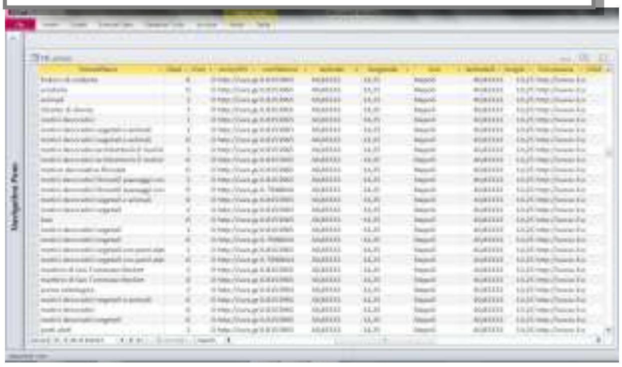
The **input data for testing** has been gathered from ATHENA project. ATHENA project might be presented as a Network of Best Practice within the eContentplus Programme. The project brought together relevant stakeholders and content owners from museums and other cultural institutions all over Europe and evaluate and integrate specific tools, based on a common agreed set of standards and guidelines to create harmonised access to their content. ATHENA has contributed 4.082.619 LIDO objects to Europeana. Almost any content items do not have digital geographic coordinates.

5.3. Customizing parser tool

Simple **client application** has been developed for testing the data. The application uses Europeana Geoparser. Input data are LIDO xml files and the software generates output – data on geoparsed items (coordinates, feature names) in database.

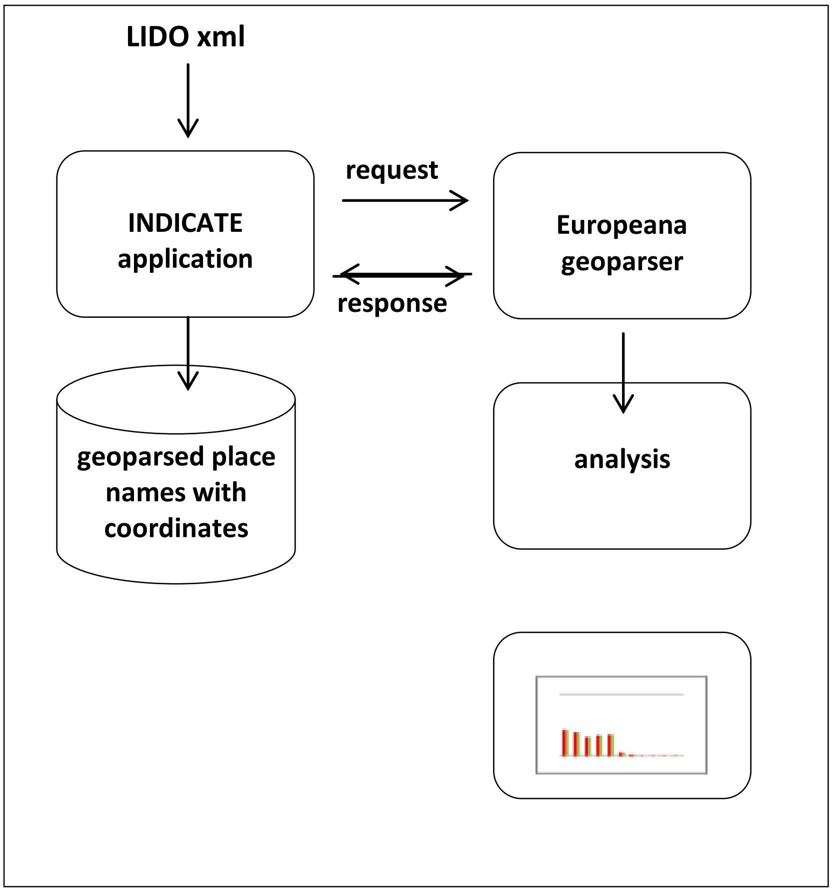
Input to testing of Europeana geoparser

Output to INDICATE testing of Europeana



5.4. Performing the testing

Scheme: The process of testing theEuropeana geoparser



The testing had the following steps:

- Testing of the structure of all (100%) Athena content, especially the occurrence of "place tags"
- Testing the efficiency of geoparsing; how many place names are found, and what is the confidence measured for them
- Testing of the frequency of place names
- Testing of the frequency of the EventPlace tag and DisplayPlace tag
- Analysis of the found place types
- Preparation of display results

5.5. Analysis of the results

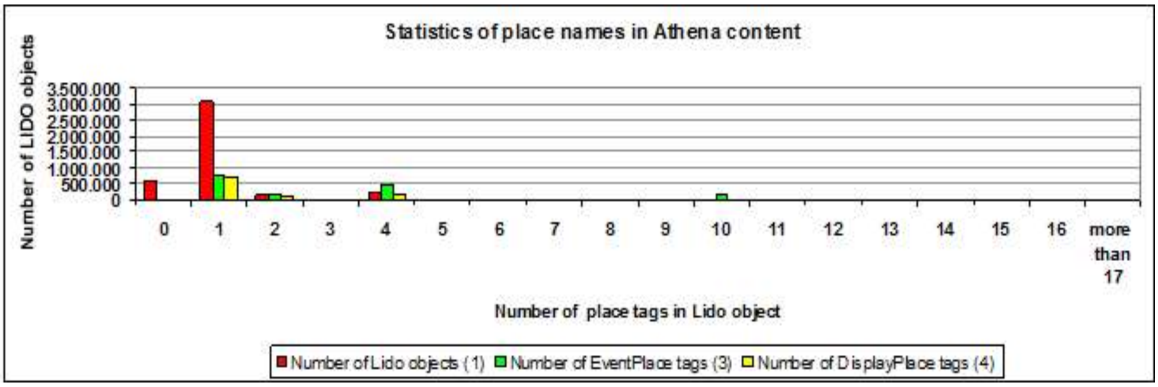
This chapter **summarizes the testing of data**; some general overview, information about confidence and frequency of found place names and the analysis of founding place names in the display and event tag of the LIDO xml.

The majority of Lido objects has at least one »Place« tag: 75,53% and 14,83% of the objects do not have »Place« tag. The analysis did not include the analysis of the »Place« tag itself. If the places exist or their syntax was not the subject of the analysis**.**

1. Testing of »place tags« in LIDO objects

Purpose: The testing checked the LIDO objects and count how many times Place tag occurred in LIDO object.

Sample: 100% (4.082.619) of Athena content LIDO xml objects**.**



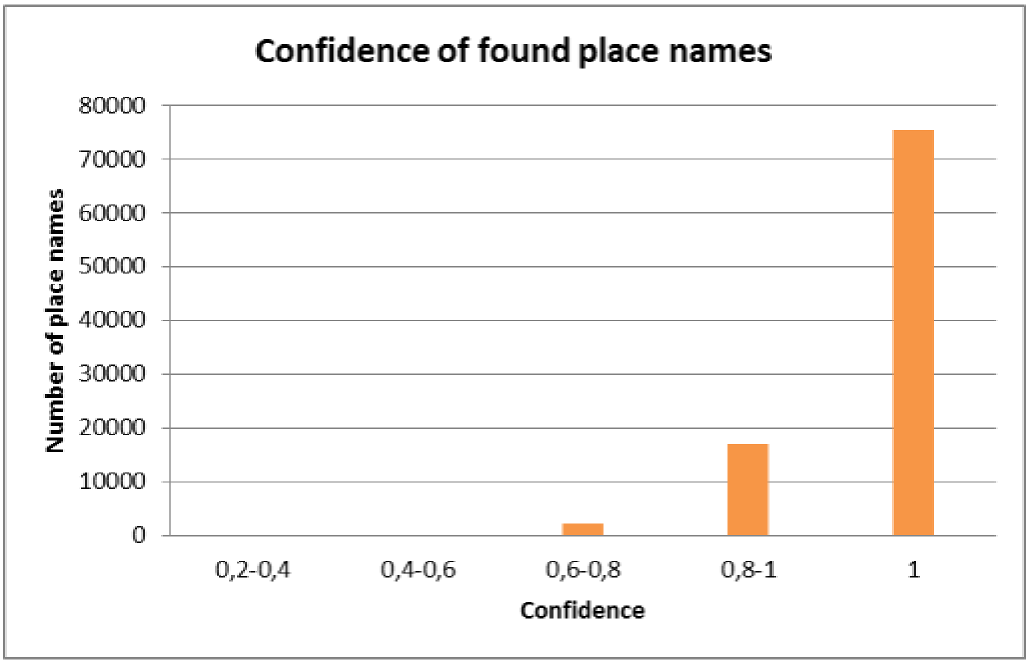
Analysis: The majority of Lido objects (85,17%) has at least one Place tag and 14,83% of the object do not have any Place tag, further numbers are 3,8% has 2 Place tags, 0,1% has 3 and 5,2% of objects have 4 Place tags. The analysis did not include the analysis of the Place tag itself. If the individual place actually exists or is right was not the subject of this analysis.

Definitions:

»Lido object« is a metadata description of one object delivered to Europeana organized into tags.
»Place tag« is a lido tag containing word »place«.
»EventPlace tag« is a lido tag containing word »Eventplace«.
»DisplayPlace tag« is a lido tag containing word »DisplayPlace«.

2. Testing of the efficiency of geoparsing

Purpose: This testing tried to found out how many place names could be found in this way and the confidence of the found places names. The whole LIDO object has been geoparsed.
Sample: 3.84% (156.679) LIDO objects - randomly selected from Athena content

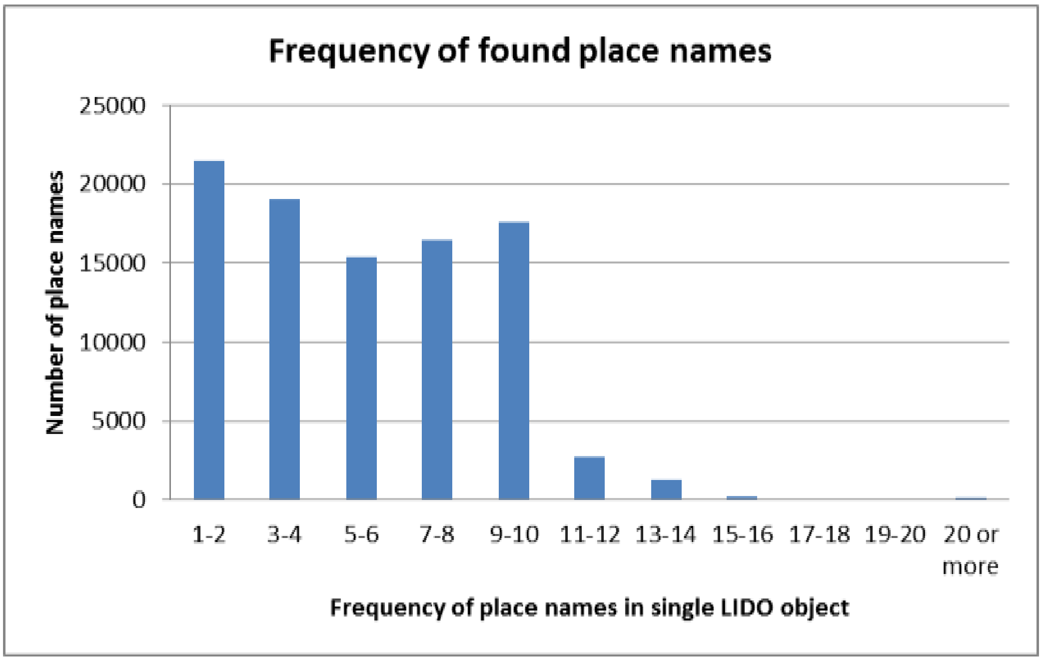


Analysis: The Europeana geoparser found places in 60,37% (94.427) of LIDO objects. The subject of the analysis was not the verification if the found place/geographic coordinates belong to the place meant by object description, different places can have the same name. The graph above illustrates the confidence of found places, as assigned by geoparser itself.

3. Testing of the frequency of place names

Purpose: The testing considered the frequency of the place names in single LIDO object found by geoparser. The whole LIDO object has been geoparsed.

Sample: 3.84% (156.679) LIDO objects - randomly selected from Athena content

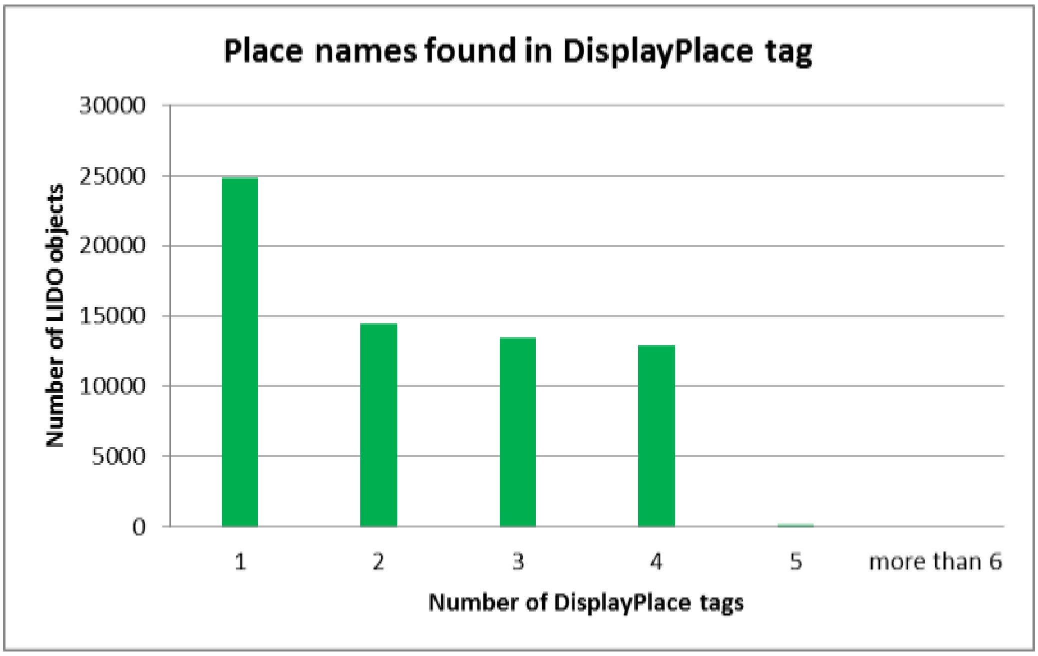


Analysis: The Europeana geoparser found places in 60,37% (94.427) of LIDO objects. The graph above illustrates the frequency of found place names in single LIDO object. Mostly objects have from 3-9 place names in single LIDO object (around 16.000), but 20 or more place names in object is rare (geoparser found just 102 objects with 20 or more place names).

4. Testing of the frequency of the EventPlace tag and DisplayPlace tag

Purpose: This analysis considered the frequency of EventPlace tag and DisplayPlace tag in the LIDO objects.

Sample: 3.84% (156.679) LIDO objects - randomly selected from Athena content.



Analysis: The forth analysis purpose has been to reveal the use of EventPlace tags and DisplayPlace tags in the the collections metadata.

DisplayPlace tags are found in 42% of the sample LIDO objects delivered to Europeana.

EventPlace tags are found in 28,53% of the sample LIDO objects delivered to Europeana.

The result clearly shows geographic data are not integral part of collections metadata for most of the cases.

5. Analysis of the found place types

The input data were 156.679 LIDO xml files. About 800.000 places and 2773 different geonames locations were identified.

The analysis revealed what type of place names are found. The geoparser found the identified administrative units as country, state, region and villages or cities. Also other features were found. Clearly the lack of smaller territorial units, culture areas and objects are revealed and though the accuracy of the geoparsing is evident.

Agreggated table of type

Count	Name
355141	country, state, region,...
278692	city, village,...
5130	spot, building, farm
1882	parks,area, ...
508	mountain,hill,rock,...
397	stream, lake, ...

Table of types in detail

Count	Name
277172	independent political entity
132978	seat of a first-order administrative division
86416	capital of a political entity
74323	first-order administrative division
37229	populated place
13722	seat of a second-order administrative division
7347	seat of a third-order administrative division
3372	second-order administrative division
2167	castle

1649	museum
1297	continent
845	building(s)
833	seat of a fourth-order administrative division
412	region
215	ruin(s)
201	cape
186	lake
161	stream
137	area
135	section of populated place
124	island
113	administrative division
104	third-order administrative division
40	spur(s)
38	mountain
35	locality
34	islands
31	ancient site
31	square
28	railroad station
27	strait
27	peak
26	arch
24	abandoned populated place
23	semi-independent political entity
22	church
20	mountains
19	gate
17	monument
17	theater
13	political entity
12	dependent political entity
11	airport
10	amphitheater
8	plateau
6	fourth-order administrative division
6	school
6	rock
5	gulf
5	farm
5	hotel
4	sound
4	sea

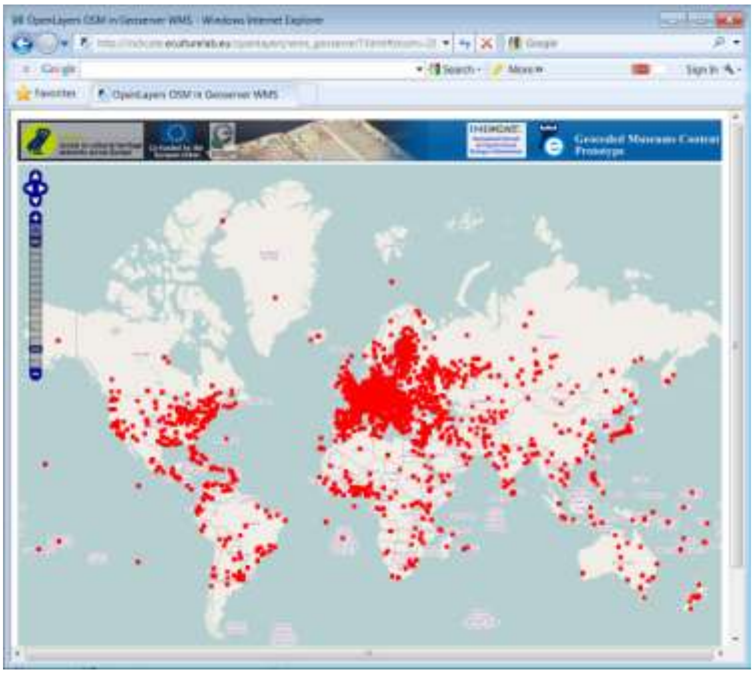
4	seat of government of a political entity
4	university
4	hill
3	canal
3	populated locality
3	cemetery
3	palace
3	wall
3	volcano
2	freely associated state
2	cove(s)
2	reservoir(s)
2	airfield
2	house(s)
2	monastery
2	quay
2	dune(s)
1	section of independent political entity
1	bay
1	anabranch
1	wadi
1	park
1	destroyed populated place
1	administrative facility
1	community center
1	historical site
1	opera house
1	pyramid
1	valley

5.6. Display results on map

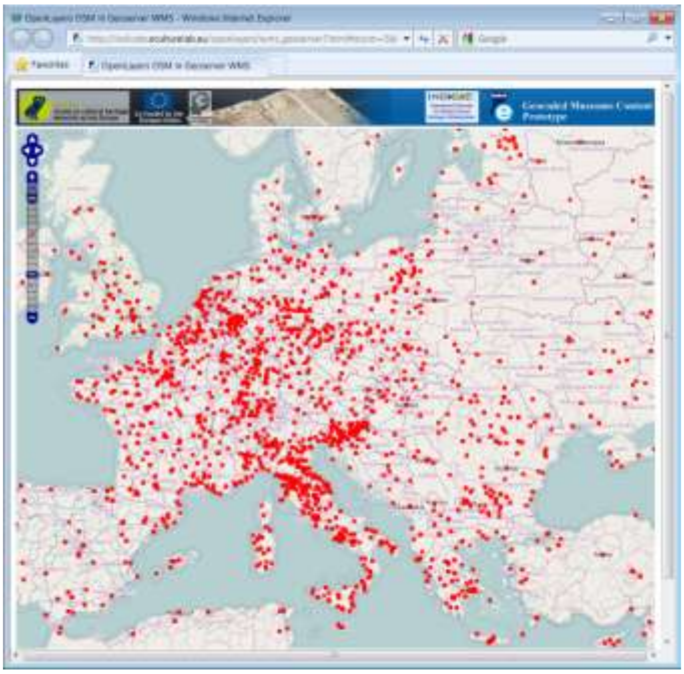
The best way to analyse and estimate the use of the results is when they are displayed on the map. The web application for displaying the geoparsing results has been developed within INDICATE project. The OpenLayers API, Geoserver, PostGIS server and PostgreSQL data base management system has been used. All components are open source. The OpenStreets maps and Googlemaps and Google orthophoto are used as basemaps.

The purpose of the developed interactive map is to represent the geoparsed testing data and its connection to Europeana. Application also generated hyperlink to a Europeana collection with an image and full description as given by content provider. The maps clearly visualise the enrichment of cultural heritage metadata with coordinates. Clearly the tested items have worldwide range, either as provenience, location, or other spatial attributes.

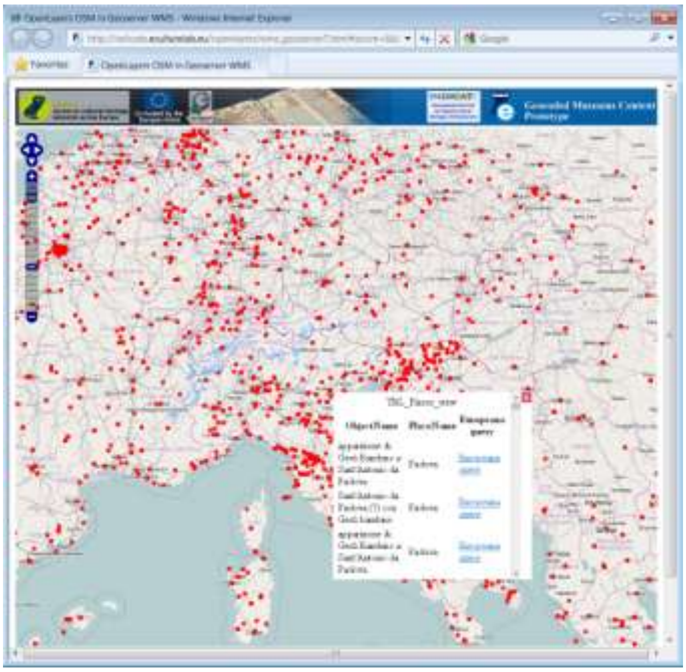
1. **The overall view** extends to the whole world and shows that Athena content (European museums) relates highly to the places outside Europe, as is North and South America, Africa, Asia and Australia as well. About 800.000 points, many of them one on the other are displayed on the map.



2. **The user** could simply zoom in, zoom out, pan, change the basemap and identify all the LIDO objects on the selected point.



3. When user selects certain point, the LIDO objects georeferenced to this point are listed in pop-up window with the name of the object, place name and link to Europeana query.



4. In this way the user is directly **navigated** to the selected object in the **Europeana portal**.

5.7. Conclusions

1. The geoparsing method is very effective method for assigning geographic coordinates automatically, in the cases where there is no available coordinates for the cultural object or the additional coordinates are needed.

2. The Europena geoparser input could be structured or unstructured attribute data describing cultural object. It performs natural text mining from textual descriptions of the cultural objects effectively.

3. The Europeana geoparser is relatively simple to use, and for testing could be used also by the end user directly.

4. The sample testing prove the hypothesis that the geoparsing is quite useful for upper level of details (as are big towns, regions, countries and up).

5. Europeana Geoparsing service could also appear as an useful tool for validation of geographical coordinates. It performs validity check if proper geographical coordinates are assigned to certain cultural heritage object after all projections transformations. In cases where this is not the case it is informative for the content provider.

6. The output of Europeana Geoparsing service is not very useful for spatial navigation because of spatial accuracy is not in the range up to 5 or 10 meters. It is recommended to enhance the service with added local databases of geographic names (archaeological and architectural sites, addresses …).

7. Geoparsing is quite good candidate for grid computing:
    - Huge amount of processing power for natural language processing, pattern recognition, web semantics
    - Distributed gazetteers as local registers, branch register and other recources
    - Use on line on different systems and various applications

6 Literature and sources

- Bubestinger, P: Free Software and Open Formats: virtual immortality and independence for digital archives, presentation at EuropeanaTech Conference, Vienna, Austria, October 4-5 2011
- Clough P.: Extracting Metadata for Spatially-Aware Information Retrieval on the Internet, University of Sheffield, 2010
- Coar, K.: Open Software:For the Public Good, presentation at EuropeanaTech Conference, Vienna, Austria, October 4-5 2011
- Correndo G. et al.: Geographical Service: a compass for the Web of Data, Conference Linked data on the Web (LDOW2010), April 27th, 2010, Raleigh, North Carolina
- Cousins, J.: Europeana: an Open Philosophy, , presentation at EuropeanaTech Conference, Vienna, Austria, October 4-5 2011
- Engels, R.H.P.: Where is the music in Semantics?, presentation at EuropeanaTech Conference, Vienna, Austria, October 4-5 2011
- Freire, N.: M5.5.5 The Europeana Geoparser – Second Prototype, version 0.1, Europeanna*Connect*, 24.6.2010
- Goldberg, D. W., Wilson, J. P., Knoblock, C. A. (2007). From Text to Geographic Coordinates: The Current State of Geocoding. In: URISA Journal 19(1)
- Guldbæk Rasmussen, K., et al.: D 3.2.3 – Recommendations for Conducting User Tests, EuropeanaConnect, 15.02.2011
- Guo, Q., Liu, Y. and Wieczorek, J.(2008). Georeferencing locality descriptions and computing associated uncertainty using a probabilistic approach, International Journal of Geographical Information Science, vol. 22, no. 10
- Haskiya, D.: Europeana Remix: an interactive experience around the story of an unlikely friendship during the First World War, presentation at EuropeanaTech Conference, Vienna, Austria, October 4-5 2011
- Heinen,D., Hesselmann, T.: D3.4.3 – Rich mobile client for accessing Europeana, EuropeanaConnect, 31.7.2010
- Hesselmann, T.: Developing innovative, usable and stable user interfaces using open source software, , presentation at EuropeanaTech Conference, Vienna, Austria, October 4-5 2011
- Isaac, A.: data.europeana.eu – the Europeana Linked Open Data Pilot, , presentation at EuropeanaTech Conference, Vienna, Austria, October 4-5 2011
- Jentzsch, A.: Linked Data Evolving the Web into a Global Data Space, , presentation at EuropeanaTech Conference, Vienna, Austria, October 4-5 2011
- Marco, M.: Europeana, Innovation and Cultural Content online Where do we stand, challenges and opportunities, presentation at EuropeanaTech Conference, Vienna, Austria, October 4-5 2011
- Martins, B., et al.: A geo-temporal information extraction service for processing descriptive metadata in digital libraries, In: e-Perimetron, Vol. 4, No. 1, 2009

- McKeague, P., Corns, A., Shaw, R.: Developing a Spatial Data Infrastructure for Cultural Heritage, Article under Review for the International Journal of Spatial Data Infrastructures, Research, submitted 31.5.2011
- Miksch, S.: Visual Analytics – Detect the Expected and Discover the Unexpected, , presentation at EuropeanaTech Conference, Vienna, Austria, October 4-5 2011
- Molendijk, J.: Europeana and Open Source, presentation at EuropeanaTech Conference, Vienna, Austria, October 4-5 2011
- Nicholas, D., Clark, D., Rowlands, I.: Culture on the go, presentation at EuropeanaTech Conference, Vienna, Austria, October 4-5 2011
- Pedrosa, G., Georg, P., Concordia C., Aloia, N.: Europeana OAI-PMH Infrastructure – Documentation and final prototype, EuropeanaConnect, 11.10.2010
- Roongpiboonsopit D., Kmimi H.A. (2010). Comparative evaluation and analysis of online geocoding services. In: International Journal of Geographical Information Science, Vol. 24, No. 7-8, July-August 2010
- Sarna, D.E.Y. (2011). Implementing and developoing cloud computing applications. New York: Taylor and Francis group
- Schade S., Lutz M.: Opportunities and Challenges for using Linked Data in INSPIRE, paper presented at Workshop on linked spatiotemporal Data 2010 in conjunction with *the* 6th International Conference on Geographic Information Science *(GIScience 2010), Zurich*, 14-17th September, 2010
- Smart, P.D. et al. A Framework for Combining Rules and Geo-ontologies. In: M. Marchiori, J.Z. Pan, and C. de Sainte Marie (eds.): RR 2007, LNCS 4524, pp. 133–147, 2007. Springer-Verlag Berlin Heidelberg, 2007
- Thompson, B.: The Finest Digital Drawing Room in Europe, presentation at EuropeanaTech Conference, Vienna, Austria, October 4-5 2011
- Vilches-Blázquez L.M., et al.: An Approach to Publish Spatial Data on the Web: The GeoLinked Data Case, paper presented at Workshop on linked spatiotemporal Data 2010 in conjunction with the 6th International Conference on Geographic Information Science (GIScience 2010), Zurich, 14-17th September, 2010
- Zakrajšek, F., Vodeb, V.: Digital Cultural Content: Guidelines for Geographic Information, Athena project, 2011
- Yue, Peng, Di, Liping, Yang, Wenli, Yu, Genong, Zhao, Peisheng and Gong, Jianya. (2009). Semantic Web Services based process planning for earth science applications. International Journal of Geographical Information Science, vol. 23, no. 9