GEOCODED

DIGITAL CULTURAL

CONTENT

GEOCODED
DIGITAL CULTURAL
CONTENT

texts by
Franc Zakrajsek
Vlasta Vodeb

Linked Heritage general co-ordinator
Rossella Caffo

design
mt milani,
geo graphic sdf

printed
in the month
of may 2013
by Officine Grafiche
Tiburtine, Tivoli (Roma)

Linked Heritage logo
Susan Hazan

web version
http://www.linkedheritage.org/index.php?en/181/publications

table of contents

Abbreviations

API	Application Programming Interface
CRM	CIDOC Conceptual Reference Model (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 Communication 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
LOD	Level Of Detail
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
LBS	Land Based Services
OWS	OGC Web Services
WSRF	WS-Resource Framework

Preface

The aim of the Linked Heritage project is to support cultural institutions in providing object data for publication in Europeana. A way of supporting them is by providing instruments and tools helping them in enlarging their knowledge on digitisation issues. This new book *Geocoded Digital Cultural Content*, written by Franc Zakrajsek and Vlasta Vodeb, focuses on Geographic location, which is a very important attribute of a cultural heritage item. It can describe provenience, the current institution, as well as the location of the event or other related events.
The added value of the geo-coded cultural content is in the browsing of cultural portals efficiently through space and time, and searching for content in a more user friendly way. This includes searching without necessitating the typing of geographical names, making it possible to discover overlapping cultural content at the same location but originating from different sources and at different times.
Geo-coding maps the cultural content, performing GIS calculations and simulations, overlapping architectural/archaeological heritage with museum objects and intangible heritage, defining the protected areas of monuments, geo-visualisation and historical simulations.
The authors investigate here the possibilities and approaches regarding the use of e-infrastructure in geo-coded digital culture. This work is a result of a strong cooperation between Linked Heritage and the INDICATE (International Network for a Digital Cultural Heritage e-Infrastructure), a concluded European Union FP7 project which aimed to establish a network of common interest made up of experts and researchers in the field of e-infrastructures and digital cultural heritage at Euro Mediterranean level.

Rossella Caffo
Linked Heritage Project Coordinator

Foreword

The geo-coding of cultural content is becoming a very promising technique which is opening new scenarios of exploitation and valorisation of the European heritage. Many applications have been developed and are used in the cultural tourism, teaching, learning research domains. This publication is the result of a wide range of studies, experiments and investigations carried out in the frame of several national and European initiatives in the last year.
One initiative is particularly worth to be mentioned here, the INDICATE[1] project, because it directly contributed to the development of the content of this Handbook.
INDICATE is the acronym of International Network for a Digital Cultural Heritage e-Infrastructure. It has been a coordination action supported by the European Commission in the frame of the Capacities Programme of FP7 and run from September 2010 until November 2012, under the coordination of the Central Institute for the Union Catalogue of Italian Libraries of the Italian Ministry for Cultural Heritage and Activities, with the participation of partners from eight European and Mediterranean countries: France, Greece, Italy, Slovenia, Spain, Egypt, Jordan and Turkey.
The main objective of INDICATE project has been to explore the opportunities opened by e-Infrastructure for the digital cultural heritage and to develop consistent policies and best practices governing the research in this domain. A network of common interest was made up of experts and researchers to share experiences and promote standards and guidelines, constituting a community aiming to be a long-term collaborative group.
The projects was rooted in the reality of research pilots and case studies which acted as exemplars and demonstrators of the issues and processes relevant for the establishment of cultural initiatives on the e-infrastructure based platforms.
The INDICATE pilots studied and experimented the migration of two e-Culture applications to the e-Infrastructure platforms: the semantic search pilot was deployed on the Cloud by the National Technical University of Athens and the e-Collaborative Digital Archive was deployed on the Grid by Consorzio COMETA. They are still accessible at http://indicate-gw.consorzio-cometa.it/home.
In parallel, the INDICATE case studies analysed the potential for e-Infrastructure to be used to address three key challenges in

1 www.indicate-project.eu

e-Culture: long term digital preservation, virtual exhibitions and geo-coding of digital cultural content. Each study produced a report which was discussed in the frame of thematic workshops organised by the partners in Ankara, Amman and Ljubljana.
This document is an extract distilled from the case study on Geo-coded Digital Cultural Content and associated workshop held in Ljubljana, which is published in cooperation with the Linked Heritage Best Practice Network[2], because of the high-interest demonstrated by the cultural institutions about the use and application of the GIS technologies to their digital repositories.

Antonella Fresa
INDICATE and LINKED HERITAGE Technical Coordinator

2 www.linkedheritage.org

Executive summary

Geographic location is one of the most important attribute of any cultural heritage item. It can describe provenience, the current institution, the location of the event or other related locations. The most valuable geographic description is in the form of digital geographic coordinates. Geographic coordinates presented as x, y and possibly z-values define a position in a coordinate system. The added value of the geocoded cultural content is in the browsing of cultural portals efficiently through space and time, searching for content in a more user friendly way, 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, connecting architectural/archaeological heritage with museum objects and intangible heritage, defining the protected areas of monuments, geovisualisation and historical simulations.

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

After the **Introduction** (Chapter 1) into study, the **Concepts and frameworks of GCC** (Chapter 2) are clearly defined. The general concepts and frameworks are applicable for different kinds of individual GCC systems. Content of use, geo feature types, spatial accuracy, geocoding methods, standards, digital object types, linking open data, devices, and e-infrastructure are examined in order to find out what are the dependencies and relation 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 3) 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 of this part is to identify the possibilities and needs for e-infrastructure support. 68 individual use cases are identified in the study, the links to their detailed descriptions are provided as hyperlinks. Use cases are different types of a regularly operating system, prototype, proof of concept, or research.

The experimental part of the study is **Testing of Geoparsing of GCC** (Chapter 4). The testbeds and performing of testing has been done by authors within the Indicate project. The Europeana geoparser v1.0 beta and Athena project data have been used. Geoparsing is the process of assigning geographic coordinates to textual words and phrases. The results of the testing prove that the geoparsing method is a very effective method for assigning geographic coordinates automatically.
The Europena geoparser input could be structured or unstructured attribute data describing a cultural object. It performs natural text mining from textual descriptions of the cultural objects effectively. Geoparsing is quite a good candidate for e-infrastructure grid computing. A huge amount of processing power is required for natural language processing, pattern recognition and web semantics. Distributed gazetteers as local registers, branch registers and other resources require stable distributed data storage. The sample testing proves the hypothesis that the geoparsing is quite useful for upper level of spatial details (as are big towns, regions, countries and up) . On the other side, the output of the Europeana Geoparsing service is not very useful for spatial navigation because the spatial accuracy is not in the range of up to 5 or 10 meters. It is recommended to enhance it with added local databases of geographic names (archaeological and architectural sites, addresses ...).

The present publication originates in the research case study **Geocoded Digital Cultural Content**, which is a part of the INDICATE Deliverable D5.3.

1. Introduction

Geographic location is one of the most important attribute of any cultural heritage item especially when taking into account navigation support by mobile devices. The most valuable geographic description is in the form of digital geographic coordinates. Geographic coordinates are presented as x, y and possibly z-values which define a position in a coordinate system. Examples of coordinated systems are the 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:

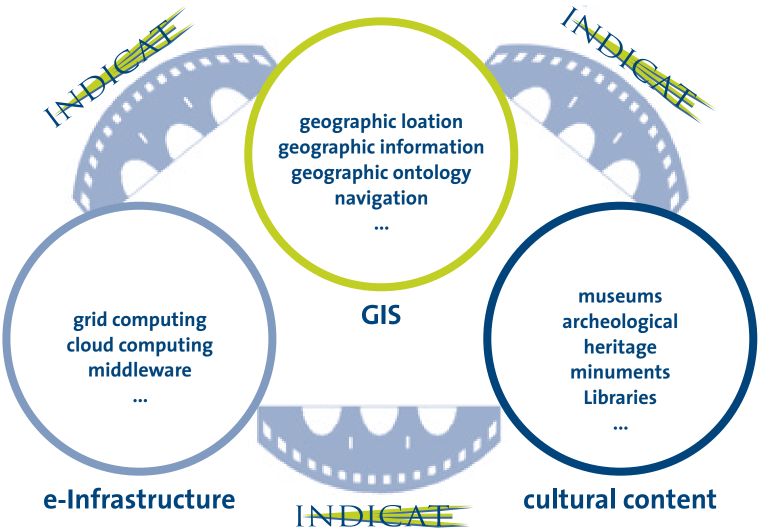
- Browsing cultural portals efficiently through space and time
- Searching for content in a more user friendly way, without the necessity of typing geographical names
- The possibility of discovering overlapping cultural content at the same location but originating from different sources and at different times
- Mapping of the content
- Performing of GIS calculations and simulations
- Overlapping architectural/archaeological heritage with museum objects and intangible heritage
- Defining the protected areas of monuments
- Geovisualisation and historical simulations

The first part of the study Geocoded digital cultural content 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. The main area of the research is dedicated to the 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 last part of the research provides and summarizes the testing of geoparsing and geotagging e-services in digital culture and recommendations for content providers are given.

The meanings of the terms composing the title "Geocoded Digital Cultural Content" are as follows:

- **Geocoded** when digital geographic coordinates are added to description of content , the feature can be associated with or found on the "earth's" surface.
- **Digital** as opposite to analogue; stored as a series of bits in computer like storage, for example: digital camera vis-a-vis analogue camera.
- **Cultural** as in cultural areas retained in memory institutions: libraries, archives, museums, galleries, monuments, audiovisual and live culture as in art, theatre, the media and similar institutions.
- **Content** can be in different forms such as text, images, video, audio, virtual reality and others.

Scheme: The INDICATE project as a bridge between GIS, cultural content and e-infrastructure

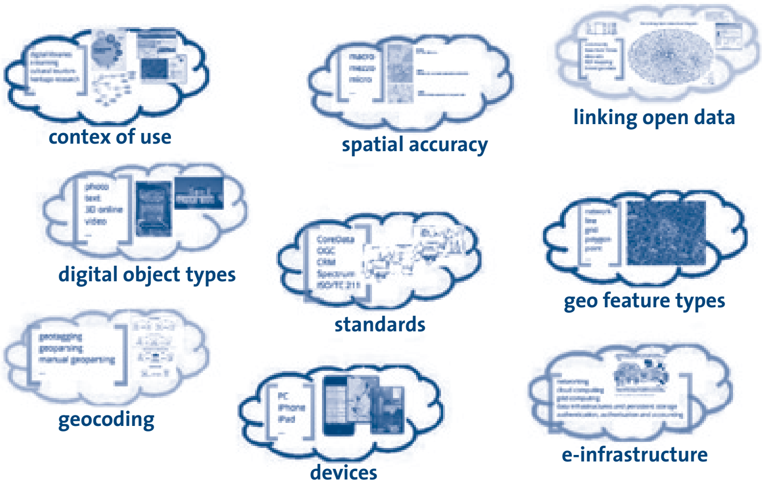


The study is one of the first, as far as we know, that is facing **the three complex areas:** geographic information, e-infrastructure and cultural content. Digital cultural content is very important for the research of cultural heritage, promoting the culture and use / re-use of digital cultural content both in the context of culture and the creative industry. Geographic information and geographic information systems have proven their potential in different areas and platforms. E-infrastructure has gained new content and is thus enabled to support geocoded cultural content.

2. Concepts and framework

The 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 for using **e-infrastructure** in this area. On the schema below the main issues are identified followed by a short description in the subchapters.

Scheme: Framework of digital cultural content



2.1 Context of Use

GCC can be used in many contexts and can be of added value in several areas. For example:

- Researching, discovering cultural heritage objects, studying cultural heritage, investigating archaeological remains, ...
- Restoration, conservation and preservation of cultural heritage
- Promoting cultural heritage on digital library portals, cultural institution portals, integrating heritage into cultural routes
- Extensive use in cultural tourism and the creative industry as well

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

2.2 Geo feature types

Geographical features can be:

- Points on the "earth's surface" as centroid of cultural content
- Segments on the "earth's surface", representing linear objects such as streets and rivers
- Polygons on the "earth's surface", representing detailed boundaries as in archaeological protected areas
- Multipatch which is 3D geometry, representing an area or volume in three-dimensional space as in a three dimensional representation of a building

Geographical features are the components which represent features on the Earth's surface. 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 of the use of e-infrastructure.

2.3 Spatial accuracy

The need for computer processing power and storage depends significantly on how precise the required data is and the geographical area covered. Hence the advantages of using e-infrastructure become clear. For example if a spatial resolution in decimetres is needed instead of 10 meters in the real world, the problem becomes 10,000 times more complex.

Spatial accuracy of the digital geographic coordinates measures the "error" as the distance between the digital geographic coordinates and the actual position on the earth's surface.

Spatial accuracy is very important; for example for routing/ navigation in cultural tourism the accuracy should be within a few meters otherwise the digital coordinates have no value. Spatial accuracy of spatial data might be observed as macro which is appropriate for large scale maps (covering more than 1,000 meters). Mezzo accuracy is between 5 and 1,000 meters and appropriate for orientation. Micro accuracy is less than 5 meters and therefore appropriate for navigation tools.

Therefore:

- A macro location, the level of detail over a larger geographical area (e.g. a country)
- A mezzo location, the location size would cover a smaller geographical area (eg. a city). This is the maximum accuracy of GeoNames
- A micro location, the spatial accuracy of a place (e.g. an address) in meters
- Detail level, the level of detail in cadastral parcels (the accuracy is down to a few centimeters, known as Lidar technology)

2.4 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 its usefulness when geocoding especially:

- When georeferencing and geocoding huge numbers of historical maps
- When geocoding precise locations in a wider geographical area
- When geoparsing and assigning huge numbers of textual digital objects

2.5 Standards

The geocoded cultural content has none or very limited use if the providers do not strictly implement **standards**. Only by upholding standards can the real cultural object be "in the same place" as represented by its digital content. Connection and implementation of standards is therefore an obligation. Widely used standards for digital cultural content are recommended as for example the Dublin Core Metadata Element Set, Europeana Semantic Elements Specifications, SPECTRUM Standard for Collections Information Management, CIDOC Conceptual Reference Model. Especially important are GIS standards. Some GIS standards and related recommendations are shortly described in the paragraphs which follow: ISO/TC 211 Geographic Information/Geomatics, Open Geospatial Consortium Standards and The INSPIRE Directive, and Coordinate systems.

1. ISO/TC 211 Geographic Information/Geomatics

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

The 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 geographic information, 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 had published fiftytwo standards under the direct responsibility of the Technical Committee 211 up to February 2011.

The ISO/TC 211 group of standards provides a fundamental structure of geographic information thus enabling its computational processing. They laid the foundation upon which other developments are possible, for instance the 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 carries out GIS data processing and sharing. They are encouraging the development and implementation of open standards, free and openly available to the market.

The organisation has close connections with other international standards bodies, especially ISO/TC 211 (Geographic Information/Geomatics). The ISO 19100 series under development by the 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 the development of standards and specifications which will establish interoperability in the processing of geographical information. Interoperability is considered as one of the key aspects in designing information systems in the 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 compatibility and usability of the spatial data infrastructures of the Member States in the European Union. To achieve this, the Directive requires that common Implementation Rules are adopted in a number of specific areas: metadata, data specifications, network services, data and service sharing and monitoring and reporting.

These Implementation 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 a great opportunity for use also in the digital cultural heritage field. Firstly, as the implementation rules are set for coordinate reference systems, geographical names and administrative units, they can be used as a methodological background and as technical standards. On the other hand the INSPIRE spatial data as orthoimagery and geographical names can be directly used when representing digital cultural content on web maps.

4. Coordinate Systems

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

For the time being the WGS84 (World Geodetic System 1984) geographic coordinate system is the most widely used. WGS84 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 geocoding cultural content, especially protected architectural and archaeological sites, using the WGS84 is not spatially accurate enough. The use of ETRS89 is demanded by the 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 between the points in the ETRS89 and WGS84 coordinate systems 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/

2.6 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 areas reflect the ways of using e-infrastructure, they could be:

- Still images usually geocoded to points; for cultural areas 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 the right orientation; from simple 3D PDF models to complex 3D City GML models

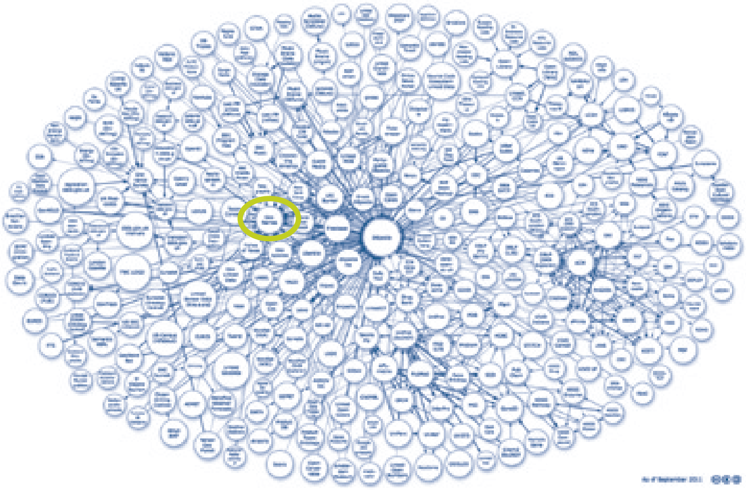
2.7 Linking open data

The real need for using e-infrastructure is demanded when linking open geographical data in order to have a stable ICT environment and then geocoded cultural data is accessible 24 hours a day, 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. The 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 of open data in the geographical content area is not unknown. The example is gazetteer GeoNames.

Scheme: The 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 integrates 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/

The important LinkedGeoData initiative 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 (http://indicate.situla.org/gcc_appendix/gcc.html#2)
- The Linking Open Data: http://lod-cloud.net/
- New York Times Company: http://www.nytco.com/

2.8 Devices

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

Some of the most popular devices are:

- Photo/ video cameras
- Smart phones (built on mobile computing platforms)
- Tablet computers (iPads, ...); especially interesting as a portable dimension of cultural tourism
- Television and other audio visual devices
- Portable computers
- Stationary personal computers
- Display show equipment

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

2.9 E-infrastructure

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

E-Infrastructure[2] is the 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 make 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 through different calculations in parallel for research projects is another yet distinct perspective of the grid computing layer. The DEISA project is an excellent example of what can be achieved when combining the power of the EU supercomputer resources, and the PRACE project is an illustration of the 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 made with regard to standards, since without them, neither connectivity nor interoperability would be possible. Please browse the webpage of the OGF-Europe project and other important European initiatives like ETSI.

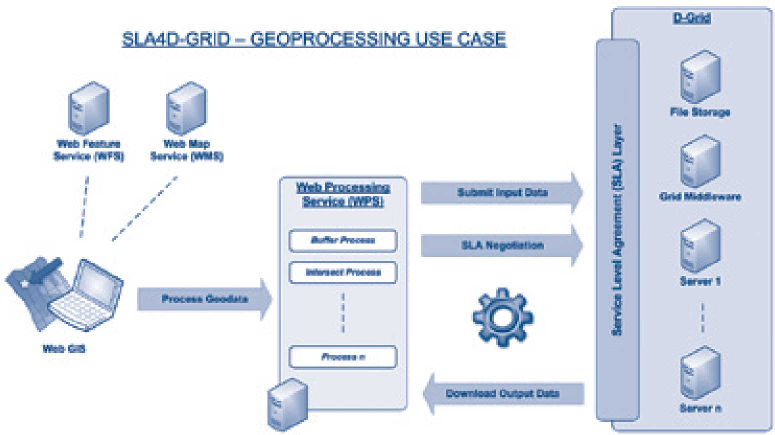
There are many great research or testbed projects that use the e-infrastructures for different research application areas.

1. GRID computing

Grid computing was initially driven by the needs of applications with large volumes of data and complex calculations such as astronomy, biology and physics. As grid computing matures it 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 having different views on what a grid is.

Gridification as a term needs 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 Sla4D grid geoprocessing use case). It is possible to do calculation tasks distributed in the grid and to use grid services for accessing data. In this scenario the existing non-gridified application becomes a bottleneck because it doesn't obtain all the qualities of a general grid service. This kind of gridification is very easy to realize and the implementation is almost independent of the underlying grid middleware. In a more complex approach the existing application or the existing service (in this context an OWS) is fully embedded into the grid middleware (e.g. as a stateful service inside a WS-Resource Framework, WSRF) and obtains all the qualities of a general grid 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 for 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 (http://indicate.situla.org/gcc_appendix/gcc.html#6)
- Grid based 3D animation rendering: research, see Appendix (http://indicate.situla.org/gcc_appendix/gcc.html#9)
- A GEO Grid Implementation for 3D GIS Taiwan: research, see Appendix (http://indicate.situla.org/gcc_appendix/gcc.html#4)
- Grid-Based Digital Libraries: Cheshire3 and Distributed Retrieval: research, see Appendix (http://indicate.situla.org/gcc_appendix/gcc.html#17)
- Using Web Portal for 3D Grid-Based Rendering: research, see Appendix (http://indicate.situla.org/gcc_appendix/gcc.html#15)

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).

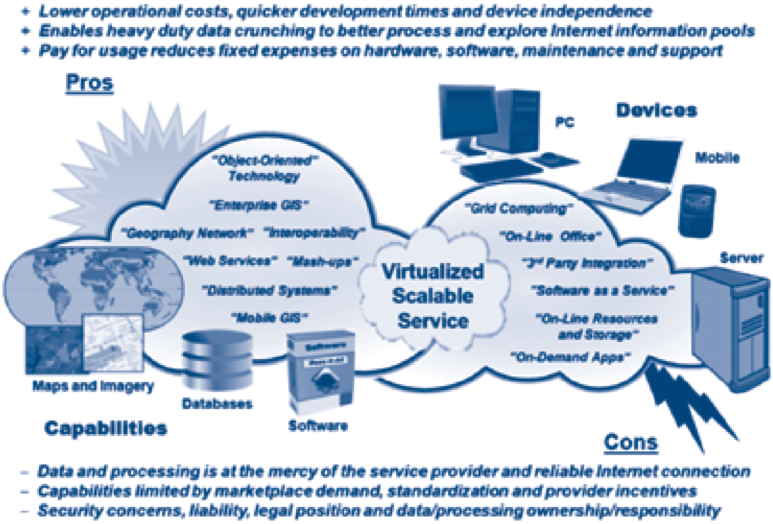
According to the Gartner group (http://www.gartner.com), the attributes of cloud computing are:

- Service-based
- Scalable and elastic
- Shared
- Metered by use
- 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

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



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

Cloud computing offers huge opportunity for ITC support and therefore for GIS areas and cultural sectors such as libraries, museums, archives and other cultural institutions. Cloud computing should be taken into account, especially:

- When introducing new content management systems
- For small and medium size cultural institutions
- When the content is shared as with Europeana for example
- When using and reusing geographical information
- When offered NRN , the National Academic Research Network

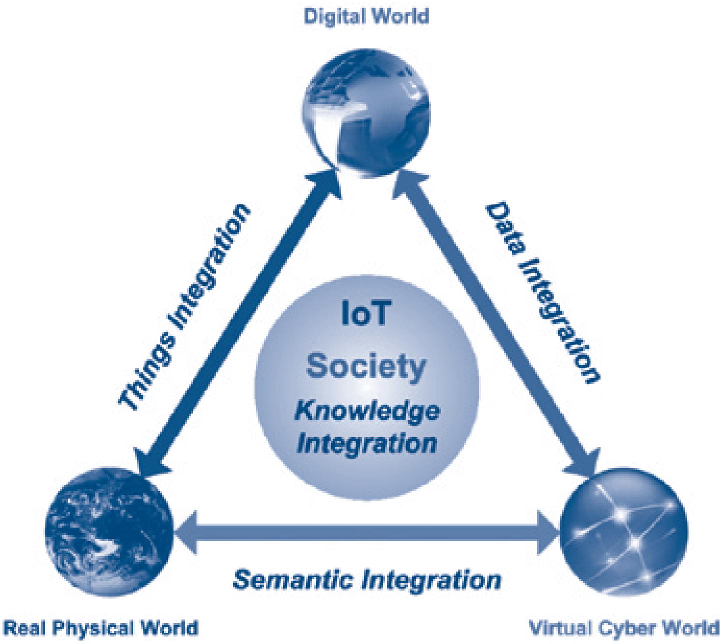
Use cases and links:

- Libraries and the Cloud: research, see Appendix (http://indicate.situla.org/gcc_appendix/gcc.html#20)
- Museums and Cloud Computing: Ready for Primetime, or Just Vapourware? : research, see Appendix (http://indicate.situla.org/gcc_appendix/gcc.html#21)
- Cloud Computing Primer: Steps for using the cloud in Your Museum: research, see Appendix (http://indicate.situla.org/gcc_appendix/gcc.html#19)
- Cloud Computing in the Application of Digital Library: research, see Appendix (http://indicate.situla.org/gcc_appendix/gcc.html#18)

3. Internet of things and other new technological issues

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

The Internet of Things refers to uniquely identifiable objects (things) and their virtual representations in an Internet-like structure. Ashton's original definition of the 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)

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

In the 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 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.

Clearly e-infrastructure and geographic information in the culture heritage field and also the newly established technologies will play a big role in the coming 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 Parliament, the Council, the European Economic and Social Committee 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/

3. Use cases of GCC

The goal of this chapter is to describe the current "state of the art" of the geocoded digital culture. A few hundred use cases were examined and 68 of them are presented here, in the Appendix (available on the internet: http://indicate.situla.org/gcc_appendix/gcc.html).

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 a single use case is that it has to cover at least two topics of: geographic information, e-infrastructure / ICT, cultural content:

- If a 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. a distributed monuments inventory).
- If a use case covers cultural areas 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 a use case covers cultural areas 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 a regularly operating system, prototype, proof of the concept, research, ...

	GRID, CLOUD	GIS	CULTURAL CONTENT
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		■	■
Judaica Europeana Mapsearch		■	■

	GRID, CLOUD	GIS	CULTURAL CONTENT
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			
Flickr geocoded Art's Photostream (Yahoo)			
Kazakhstan project			

3.1. Digital Libraries

A digital library is a collection of digital content from libraries, archives, museums and other cultural institutions. It contains internal collections in e.g. museums or in certain branches 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 libraries objects are becoming more and more important. It would be unimaginable to type the place name when querying more than ten million digital objects instead just to click on the area on the map to retrieve information. For the time being geographic features are simple and are more or less limited to the "point" features.

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

- Digital library as harvester and as one stop portal, could economically benefit with cloud ICT
- Content providers within the cloud environment will provide a more reliable delivery to harvester and
- On line accessibility of digital objects (not temporarily unreached)
- Georeferencing, processing, storage and use of a huge repository of historical maps is a very important subject with regard to grid computing

Use cases and links:

- Europeana culture globe: see Appendix (http://indicate.situla.org/gcc_appendix/gcc.html#57)
- Europeana portal: see Appendix (http://indicate.situla.org/gcc_appendix/gcc.html#58)
- arXiv: see Appendix (http://indicate.situla.org/gcc_appendix/gcc.html#56)
- American Memory: see Appendix (http://indicate.situla.org/gcc_appendix/gcc.html#55)
- Europeana 4D: see Appendix (http://indicate.situla.org/gcc_appendix/gcc.html#59)
- Juidaica Europeana: see Appendix (http://indicate.situla.org/gcc_appendix/gcc.html#31)
- LinkedGeoData.org: see Appendix (http://indicate.situla.org/gcc_appendix/gcc.html#2)

3.2 Architectural / Archaeological 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 the architectural and archaeological heritage field. Firstly they were used just for delineation of the protected areas. Nowadays they are used also for 3D detailed modelling for researching, restoration of monuments and for the 3D presentation of heritage to the general public.

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

Appropriate tasks for grid computing:

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

Potential for the use of grid computing for caching

Steps	Software and hardware	Estimated time
Experiment caching 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	Caching time: 77 minutes
Generalization for the world mainland: 148,429,000 km2, a million times larger area than in the experiment	ArcGIS Server 2x E5450 3GHz (8 threads) 32GB memory	Estimated 77 million minutes or 146 years
Google	• In 2002; upwards of 15,000 servers • A 2005 estimate by Paul Strassmann has 200,000 servers, claiming this number to be upwards of 450,000 in 2006 and 900,000 in 2011	

Use cases and links:

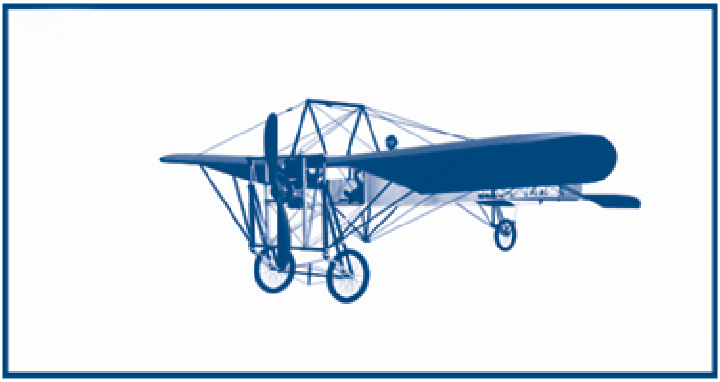
- Register of Immovable cultural heritage of Slovenia: see Appendix (http://indicate.situla.org/gcc_appendix/gcc.html#49)
- National Heritage List for England: see Appendix (http://indicate.situla.org/gcc_appendix/gcc.html#42)
- National Heritage Register Netherlands: see Appendix (http://indicate.situla.org/gcc_appendix/gcc.html#43)
- National Register of Denmark: see Appendix (http://indicate.situla.org/gcc_appendix/gcc.html#44)
- Atlas of Heritage and Architecture, France: see Appendix (http://indicate.situla.org/gcc_appendix/gcc.html#34)

- German Heritage Register Bayern – Nürnberg: see Appendix (http://indicate.situla.org/gcc_appendix/gcc.html#33)
- MIDAS Heritage, the UK Historic Environment Information Standard: http://www.english-heritage.org.uk/
- 4D Cities: http://4d-cities.cc.gatech.edu/atlanta

3.3 Movable and intangible heritage

Movable heritage refers to natural or manufactured objects of heritage significance. Intangible heritage refers to practices, representations, expressions, knowledge, skills that communities recognize as part of their cultural heritage. This study 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.

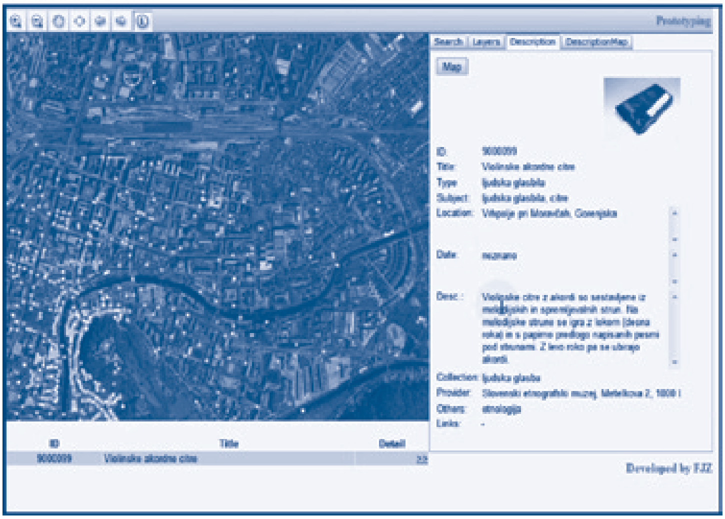
Cloud computing is coming rapidly to cultural institutions such as museums and libraries. It is recommended to conduct a feasibility/business model study when choosing the cloud computing provider. The use of NREN (National Research and Educational Network) as a provider is a preferred choice rather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

Use cases and links:

- 3D Artefact Acquisition: see Appendix (http://indicate.situla.org/gcc_appendix/gcc.html#63)
- Picture War Monuments: research, see Appendix (http://indicate.situla.org/gcc_appendix/gcc.html#46)
- Museums and Cloud Computing: Ready for Primetime, or Just Vapourware?: research, see Appendix (http://indicate.situla.org/gcc_appendix/gcc.html#21)
- 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/

3.4 Cultural tourism

Cultural tourism is concerned with a country's or region's culture. People visit cultural attractions with the intention of seeking out new information and experience 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 theatre. 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)

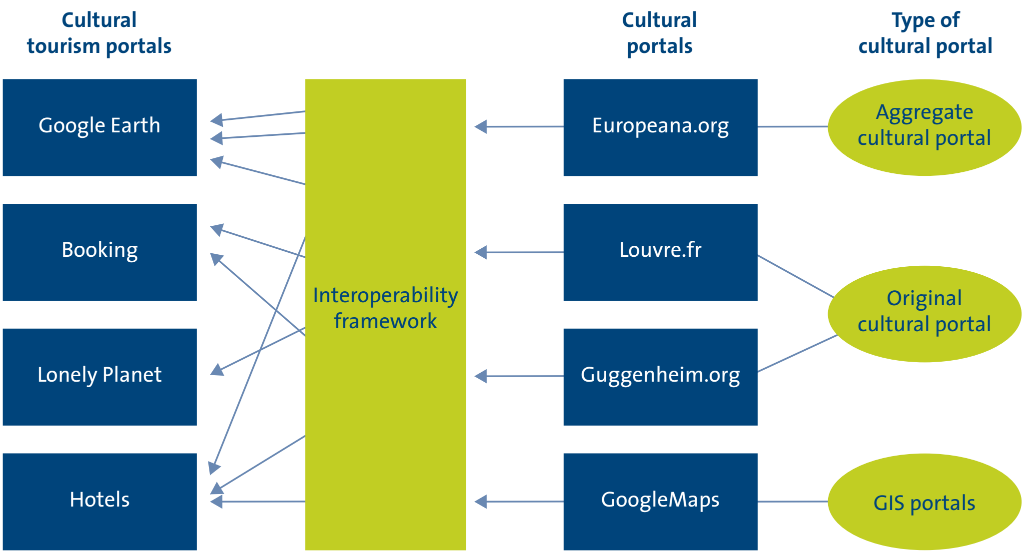
Within this study a short review of some "travelling" portals has been done. The results are presented in the table below.

Review of tourism / booking / routing portals

name	http	framework	Cultural heritage comment	Comment
Michelin	www.viamichelin.com	Web page with searchable information 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 has a link to 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: churches. Cultural heritage can be found under photos, where the user can participate. Historical imaginary: 3D buildings: 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 a few cultural heritage objects and monuments or galleries are displayed.

name	http	framework	Cultural heritage comment	Comment
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. They use Google base maps: street maps and satellite with their embedded symbology for landmarks. The 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. They use Google base maps: street maps 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 the sight 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 objects. 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 in their quest for 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 an important impact on cultural tourism regarding visitor experience of heritage and presenting heritage with new technologies – virtual exhibitions, GIS tools. A further important aspect is the interconnection of tourism web portals with different services, such as navigation, booking, ticketing, etc.

The interoperability framework between travelling services and cultural content services should work more consistantly and reliably therefore the cloud ICT could improve the performance of the framework.

If someone plans to use more computer intensive processing as with the shortest path algorithm or traveller sales algorithm, grid computing should be taken into account.

Use cases and links:

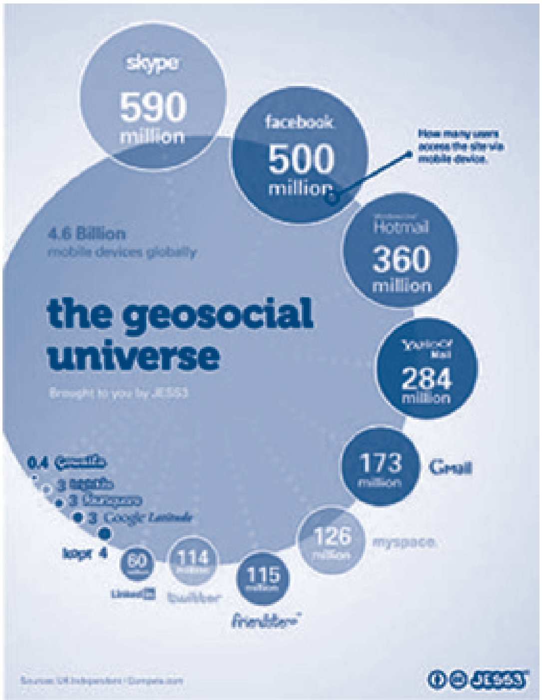
- AskAboutIreland: see Appendix (http://indicate.situla.org/gcc_appendix/gcc.html#62)
- Appia Antica archaeological Park: see Appendix (http://indicate.situla.org/gcc_appendix/gcc.html#65)
- CultureMap London: see Appendix (http://indicate.situla.org/gcc_appendix/gcc.html#24)
- Locating London's Past: see Appendix (http://indicate.situla.org/gcc_appendix/gcc.html#39)
- Lonely planet: www.lonelyplanet.com
- Hotels reservation system: www.hotels.com
- Via Michelin: www.viamichelin.com

3.5 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-centred service whereas online community services are group-centred. Social networking sites allow users to share ideas, activities, events, and interests within their individual networks.

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

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



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.

Cloud computing could improve reliability of the individual provider of cultural services.

Use cases and links:

- Flickr geocoded Art's Photostream: see Appendix (http://indicate.situla.org/gcc_appendix/gcc.html#67)
- 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

4. 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 a certain cultural collection its 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 was 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:

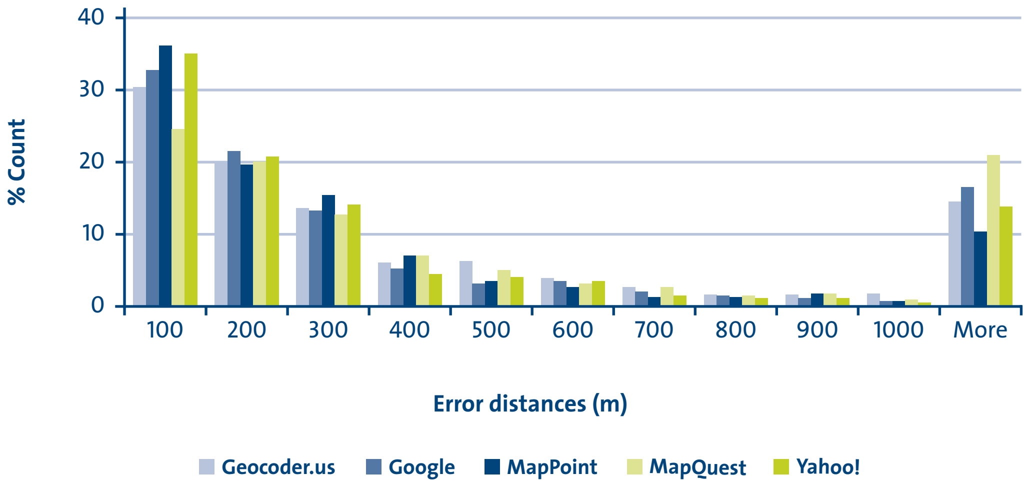
- Could we find geographical coordinates from the textual metadata of certain digital content?
- What strategies and geoparsing services could we use for geoparsing?
- What percentage of the content could be geocoded in this way, at best?
- For what purpose / services could we use the geoparsed geographical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

4.1 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 graph below.

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-lloo



Finally, the Europeana Geoparser v 1.0 Beta has been used for testing data. The Europeana Geoparser has been developed for Europeana content providers and as such is a reasonable tool for testing the data for this report.

4.2 Selection of testing data

The input data for testing has been gathered from the ATHENA project. The 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 to 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 all content items do not have digital geographic coordinates.

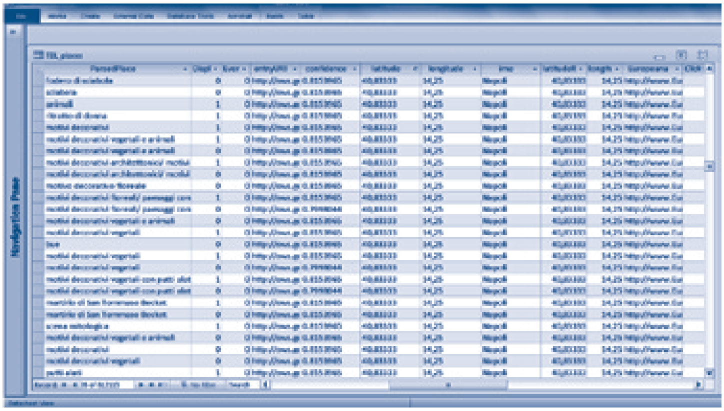
4.3 Customizing parser tool

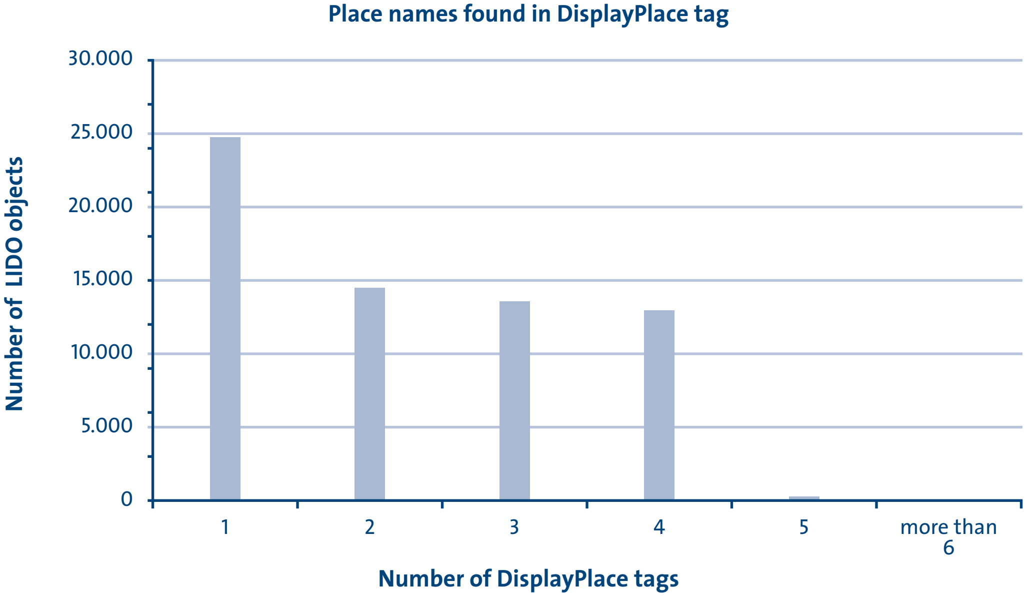
Simple client application has been developed for testing the data. The application uses Europeana Geoparser. Input data are LIDO xml files and the software generates output – data on geoparsed items (coordinates, feature names) in the database.

Input to testing of Europeana geoparser

```
<?xml version="1.0" encoding="UTF-8"?>
- <lido:lidoWrap xmlns:lido="http://www.lido-schema.org" xmlns:xalan="http://xml.apache.org/xalan"
  xmlns:pr2="http://www.opengis.net/gml" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance">
- <lido:lido>
    <lido:lidoRecID lido:type="athena">athena000000000</lido:lidoRecID>
    <lido:category />
  - <lido:descriptiveMetadata xml:lang="ger">
    - <lido:objectClassificationWrap>
      - <lido:objectWorkTypeWrap>
        - <lido:objectWorkType>
            <lido:term lido:addedSearchTerm="no">Plakat</lido:term>
          </lido:objectWorkType>
        </lido:objectWorkTypeWrap>
      - <lido:classificationWrap>
          <lido:classification lido:type="michael_collection" />
        - <lido:classification lido:type="europeanatype">
            <lido:term lido:addedSearchTerm="no">IMAGE</lido:term>
          </lido:classification>
        </lido:classificationWrap>
      </lido:objectClassificationWrap>
    - <lido:objectIdentificationWrap>
      - <lido:titleWrap>
        - <lido:titleSet lido:type="preferred">
            <lido:appellationValue>I. Olympische Winterspiele - 1924, Chamonix</lido:appellationValue>
          </lido:titleSet>
        </lido:titleWrap>
          </lido:legalBodyName>
        </lido:recordSource>
      - <lido:recordInfoSet>
          <lido:recordInfoLink>http://archiv.sportmuseum.info/boxen/boxhandschuhe-muhammad-ali/</li
        </lido:recordInfoSet>
      </lido:recordWrap>
    - <lido:resourceWrap>
      - <lido:resourceSet>
          <lido:linkResource lido:type="image_thumb">http://archiv.sportmuseum.info/wp-content/uploads
          202.jpg</lido:linkResource>
        </lido:resourceSet>
      - <lido:resourceSet>
          <lido:linkResource lido:type="image_master" />
        </lido:resourceSet>
      </lido:resourceWrap>
    </lido:administrativeMetadata>
  </lido:lido>
</lido:lidoWrap>
```


Output to INDICATE testing of Europeana





Scheme: The process of testing the Europeana geoparser

The testing took the following steps:

- Testing of the structure of all (100%) Athena content, especially the occurrence of "place tags"
- Testing of 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 on a map

4.5 Analysis of the results

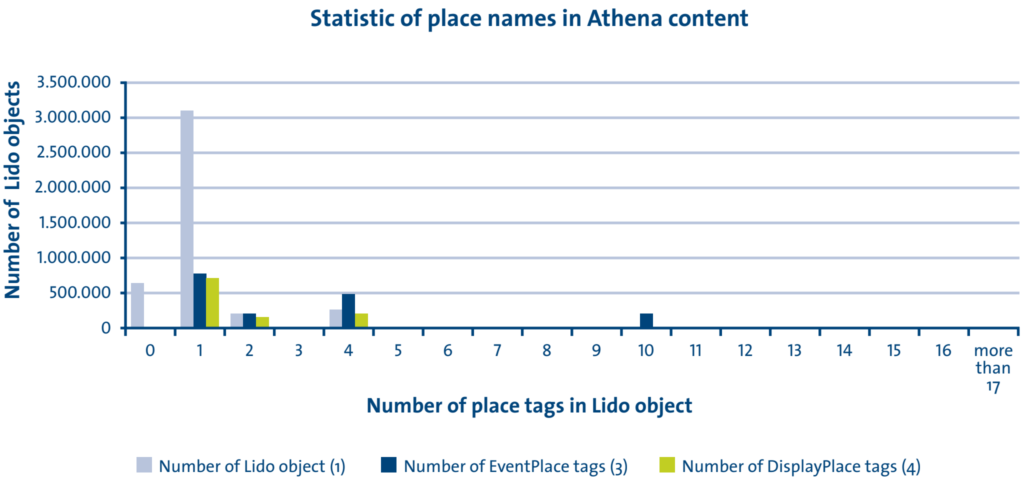
This chapter summarizes the testing of data; some general overview, information about confidence and frequency of found place names and the analysis of found place names in the display and event tag of the LIDO xml.

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. The question as to whether 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 counted how many times Place tags occurred in the 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 objects do not have any Place tag. Again, 3.8% have 2 Place tags, 0.1% have 3 and 5.2% of objects have 4 Place tags. The analysis did not include the analysis of the Place tag itself. The question as to whether the individual place actually exists or is correct was not the subject of this analysis.

Definitions:

»Lido object« is a metadata description of one object delivered to Europeana organized into tags

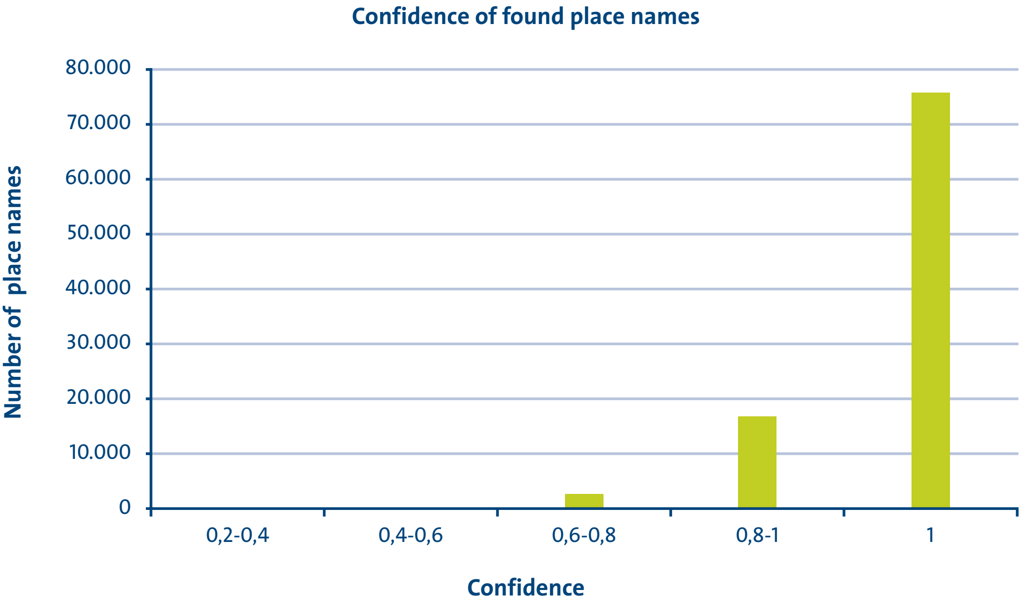
»Place tag« is a lido tag containing the word »place«

»EventPlace tag« is a lido tag containing the word »Eventplace«

»DisplayPlace tag« is a lido tag containing the word »DisplayPlace«

2. Testing of the efficiency of geoparsing

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

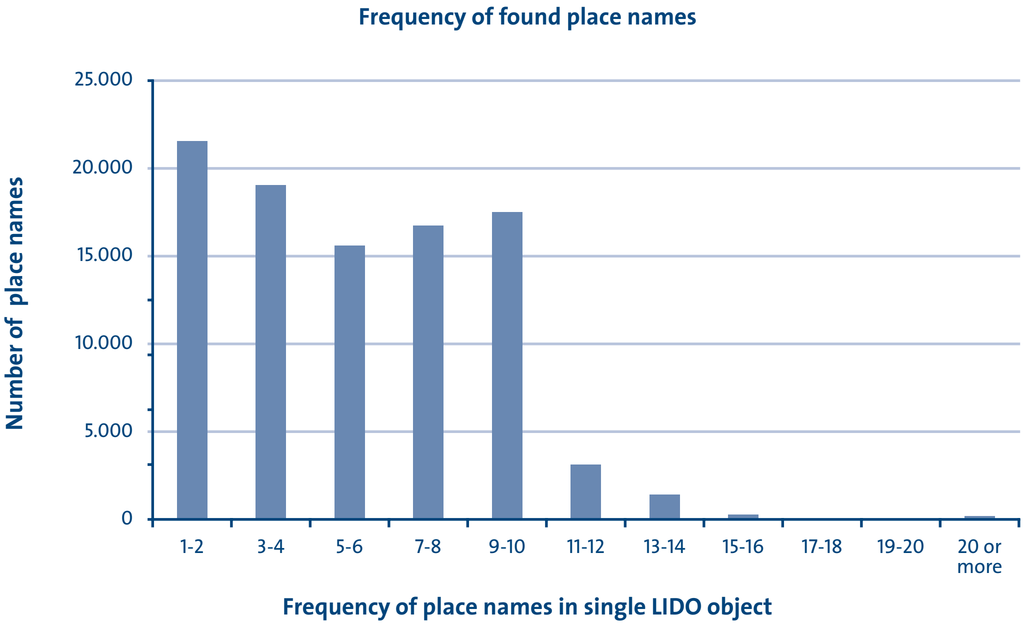


Analysis: The Europeana geoparser found places in 60.37% (94,427) of LIDO objects. The subject of the analysis was not the verification if the found place/geographic coordinates belonging to the place meant by object description, different places could have the same name. The graph above illustrates the confidence of found places, as assigned by the geoparser itself.

3. Testing of the frequency of place names

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

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

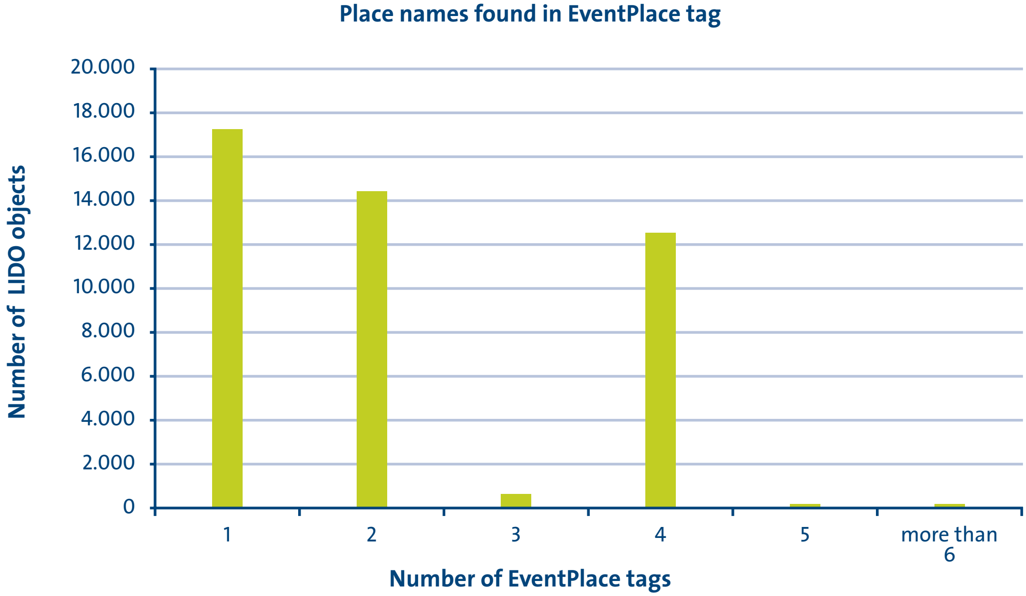


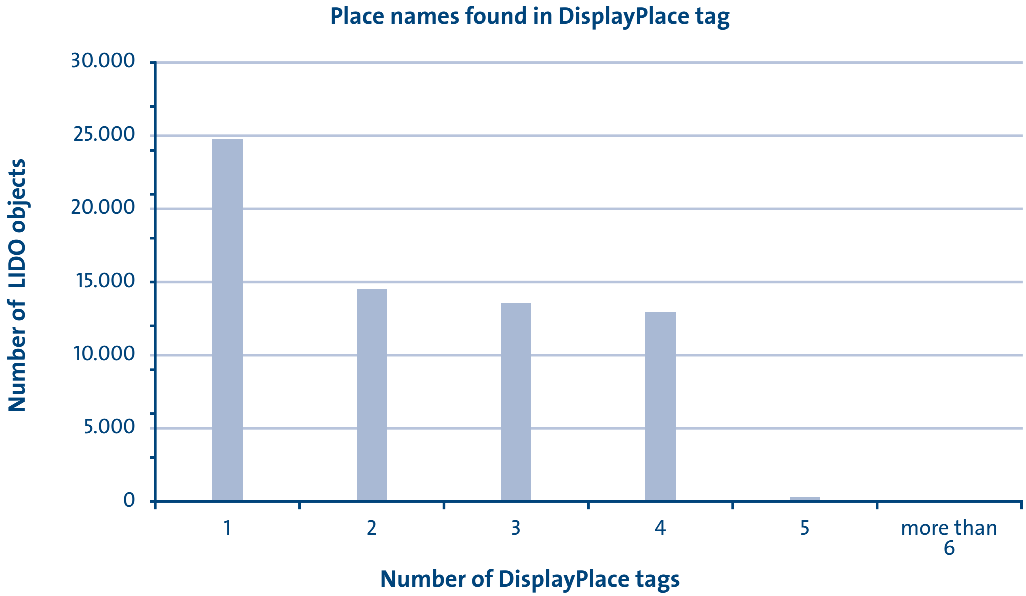
Analysis: The Europeana geoparser found places in 60.37% (94,427) of LIDO objects. The graph above illustrates the frequency of found place names in a single LIDO object. Mostly objects have from 3 to 9 place names in a single LIDO object (around 16,000), but 20 or more place names in an object is rare (the 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 tags and DisplayPlace tags in the LIDO objects.

Sample: 3.84% (156,679) LIDO objects - randomly selected from Athena content





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

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

EventPlace tags were found in 28.53% of the sample LIDO objects delivered to Europeana.

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

5. Analysis of the found place types

The inputs for the analysis are 3.84% (156,679) LIDO objects - randomly selected from Athena content. 641.749 place names are found in the GeoNames Gazetteer (http://www.geonames.org/).

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

Aggregated table of found types

Count	Name
355,141	country, state, region,...
278,692	city, village,...
5,130	spot, building, farm
1,882	park ,area, ...
508	mountain,hill,rock,...
397	stream, lake, ...

Table of found types in detail

Count	Name
277,172	independent political entity
132,978	seat of a first-order administrative division
86,416	capital of a political entity
74,323	first-order administrative division
37,229	populated place
13,722	seat of a second-order administrative division
7,347	seat of a third-order administrative division
3,372	second-order administrative division
2,167	castle
1,649	museum
1,297	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

Count	Name
27	strait
27	peak
26	arch
24	abandoned populated place
23	semi-independent political entity
22	church
20	mountains
19	gate
17	monument
17	theatre
13	political entity
12	dependent political entity
11	airport
10	amphitheatre
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

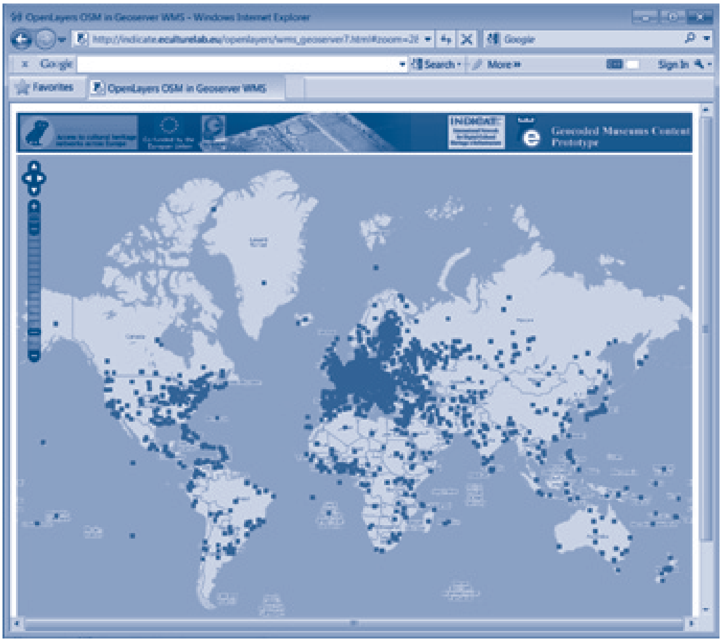
Count	Name
3	wall
3	volcano
2	freely associated state
2	cove(s)
2	reservoir(s)
2	airfield
2	house(s)
2	monastery
2	quay
2	dune(s)
1	section of independent political entity
1	bay
1	anabranch
1	wadi
1	park
1	destroyed populated place
1	administrative facility
1	community centre
1	historical site
1	opera house
1	pyramid
1	valley

4.6 Display results on map

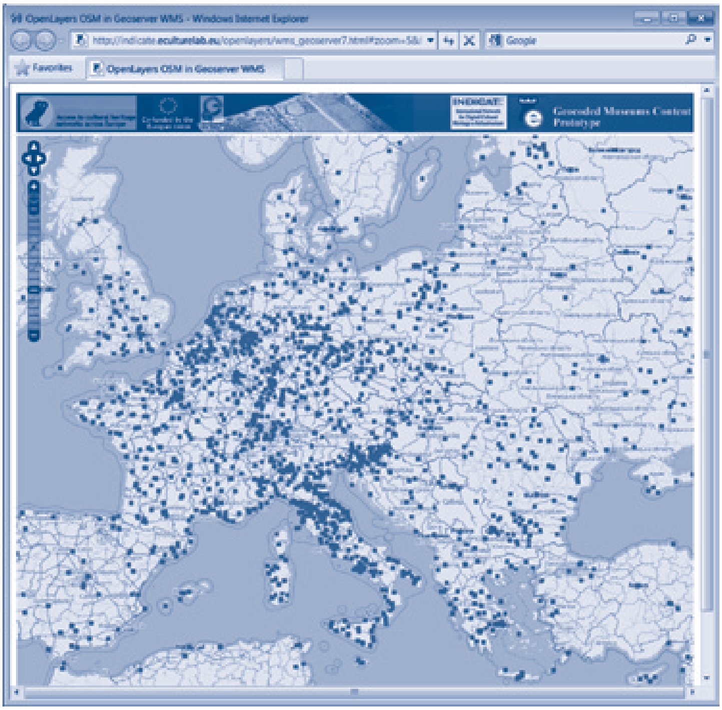
The best way to analyse and estimate the use of the results is when they are displayed on the map. The web application for displaying the geoparsing results has been developed within the INDICATE project. The OpenLayers API, Geoserver, PostGIS server and PostgreSQL data base management system has been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

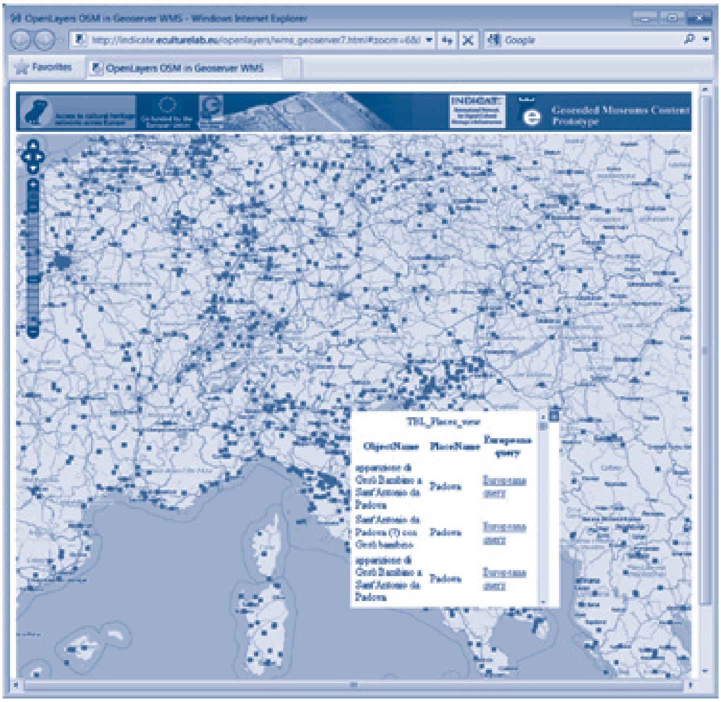
1. **The overall view** extends to the whole world and shows that the Athena content (European museums) relates strongly to the places outside Europe, as in North and South America, Africa, Asia and Australia as well. On the map there are about 800,000 points, a lot of them one on top of the other.



2. The user can simply zoom in, zoom out, pan, change the base map and identify all the LIDO objects on the selected point.



3. When the **user selects a certain point**, the LIDO objects are listed 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**.

4.7 Conclusions

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

2. The Europena geoparser input could be structured or unstructured attribute data describing cultural objects. 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 proves the hypothesis that geoparsing is quite useful for upper level of details (such as big towns, regions, countries and up).

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

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

7. Geoparsing is quite a good candidate for grid computing:

- Huge amounts of processing power for natural language processing, pattern recognition, web semantics
- Distributed gazetteers such as local registers, branch registers and other recources
- Use on line on different systems and various applications

5. 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.: *Georeferencing Locality Descriptions and Computing Associated Uncertainty Using a Probabilistic Approach*, International Journal of Geographical Information Science, 2008, 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.: *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.: *Implementing and Developoing Cloud Computing Applications*. New York: Taylor and Francis group, 2011

- 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.J., 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: *Semantic Web Services Based Process Planning for Earth Science Applications*. International Journal of Geographical Information Science, 2009, vol. 23, no. 9

Michael Culture Association

European Network

providing new experiences of

European digital cultural heritage

www.michael-culture.eu

Parmigianino, *La schiava turca* - Parma, Pinacoteca Nazionale

Education and Culture DG

Culture Programme

Michael Culture Association

A not-for-profit organisation, founded in April 2007 under Belgian law.

• Supports the Minerva network of European professionals working on digital cultural heritage.

It federates professionals and experts through support and development of the transmission and exchange of information, via publications, conferences and events.
It has members from all over Europe, including public agencies, cultural institutions and private organisations.

• Participates in major European initiatives

Michael Culture Association aims at enhancing the digitization and promotion of European cultural heritage.
It is participating in the ongoing construction of Europeana, the European Digital Library.
The association is a member of the executive committee of the EDL foundation and is also a partner in major European projects.

• Promotes and develops the Michael portal

The European catalogue of online digitized cultural collections offers access to over 10.000 collections of high quality cultural content from archives, museums, libraries etc, bringing together national initiatives from across Europe.

JOIN Michael Culture Association

www.michael-culture.eu

contact
info@michael-culture.eu

With the support of the Culture Programme of the European Commission

Culture Programme
Education and Culture DG

minerva Minerva Knowledge Base Digitising Content Together
MEDCULT
Michael Multilingual Inventory of Cultural Heritage in Europe
ATHENA Access to cultural heritage networks across Europe
LINKED HERITAGE
DC NET Digital Cultural Heritage Network
INDICATE
europeana project group member

in the same series

DIGITISATION: STANDARDS LANDSCAPE FOR EUROPEAN MUSEUMS, ARCHIVES, LIBRARIES

PERSISTENT IDENTIFIERS (PIDs): RECOMMENDATIONS FOR INSTITUTIONS

DIGITAL CULTURAL CONTENT: GUIDELINES FOR GEOGRAPHIC INFORMATION

LIGHTWEIGHT INFORMATION DESCRIBING OBJECTS (LIDO): THE INTERNATIONAL HARVESTING STANDARD FOR MUSEUMS

YOUR TERMINOLOGY AS A PART OF THE SEMANTIC WEB RECOMMENDATIONS FOR DESIGN AND MANAGEMENT