Chapter 2:

REGATTA – Regional Aggregator of Heterogeneous Cultural Artefacts

Emil Hadjikolev, George Vragov, George Totkov, Elena Somova

1 Introduction

Digital libraries offer modern technological solution for presenting cultural heritage artefacts and providing semantic access to them. The main prerequisite for their effectiveness is the structuring of content through standardized collections of metadata. The European digital library Europeana plays a major role in bringing together cultural heritage content from various countries.[90] One of the issues it faces is the uneven distribution of materials it currently presents from different countries and on different subjects. While Europeana has already developed its strategy to include new digital objects through a network of aggregators, dealing with objects of specific types, the relatively low presence of objects from some countries could be explained by the lack of digitization strategies and respectively, a critical mass of digitized resources. As already emphasized in Chapter 1, currently the European Commission drives digitisation towards setting quantitative goals which possibly will address also existing gaps.

Currently, the technology allows creating "digitized images" of cultural artefacts and placing them into our cultural space through the Web. As suggested in [Chen et al, 2005] "research on significant cultural and

---

[90] At the end of 2010 Europeana had 15 million objects from over 2,000 museums, libraries, archives and audiovisual collections across the 27 countries of the European Union (see [Cousins, 2011], p.69)

historical materials is important not only for preserving them but for preserving an interest in and respect for them". The geographical disposition, relief and climate of the Bulgarian lands, especially the Upper Thracian Valley, make it an attractive living place over the centuries. From Chalcolithic ages till now, many cultural layers have piled up. The creation of a common space for presenting different periods and different types of cultural marks gives the opportunity to receive a complex notion.

This chapter presents the results of work done in 2010-11 on the development of a repository to be used within an aggregator of digitized collections of cultural objects. The specific task is to create an environment for the maintenance of various collections through the aggregator in order to integrate them into the European digital library Europeana and to establish a shared environment for representing the rich regional cultural heritage in this part of Bulgaria. The chosen approach supports the idea of preserving the valuable national monuments in the European cultural environment while highlighting their identity and specificity.

Information technologies are widely applied in the collection (including recovery), storage, processing and dissemination of information concerning cultural and historical objects (CHO). Information on cultural and historical objects is stored in digital documents in the form of data, located on various media (with local or global access), organized in static Web pages, databases or digital storages.

The collection of data on cultural objects includes their digitalisation and input of relevant metadata. After the process of digitalisation is completed, the artefact is presented in a digital format, while the metadata (structured information about the digitalised object) is used in order to enable resource discovery, access, management, etc.

Nowadays, the extraction of relevant metadata and the so-called *interoperability* between different descriptions (metadata schemas) of digital artefacts are amongst the major challenges in the digital cultural heritage domain. Interoperability between metadata systems allows procedures and items associated with the processing of digital objects to be managed in the same manner, providing opportunities for the exchange of data between different systems. During the design and creation of software systems for digital object processing it is necessary to ensure that not only the *technical compatibility* (related to the use of common technical standards – file types, metadata, etc.) but also the *semantic interoperability* using common thesauri (glossaries for the used terminology) will be supported.

The access to digital artefacts can be supported by various means: promotional sites, virtual catalogues, virtual tours, virtual museums, libraries and archives, cultural-historical portals, business applications (in the fields of tourism, auctions, genealogy, history of art, criminology), etc.

The work presented in this chapter aims to create a digital repository for digitalised objects and a technological environment which will be used for the maintenance of a range of distributed museum collections in an aggregator; the long-term goal is to prepare the ingested materials for integration in the European digital library.

The content of that library is expected to be maintained by selected collections from the museums and galleries fund from the Plovdiv region. This task is synchronised with the idea of storing the valuable historical monuments in the European cultural environment and the complete preservation of their identity and particulars.

Certain experience has been accumulated with the application of appropriate metadata. When the structure of the collections from the museum fund is specified, positive results are observed in the process of retrieving data from the description of the object available in them.

2 Aggregators of Digital Content for Cultural Artefacts in EU

Some of the conclusions reached by the study made in early 2010 in the EU-countries [Piccininno, 2009] for aggregators of metadata for cultural objects and used technologies to deliver content to Europeana, are:

- All aggregators share the crucial goal to provide with integrated access to digital cultural resources via the Internet;
- Initiatives to create aggregators have been shown by creating and working on projects related to Europeana and the involvement of cultural institutions that support digital content. In both cases, similar approaches and technical solutions are available. The period of their creation and development is 2002-2010, and the appearing of the new aggregators has escalated in the last two years.
- Sixty percent of the aggregators are related to national portals. EU-funded initiatives, supporting about twenty percent of them, and the lowest (about seven percent) are regional aggregators. There is a trend of a growing interest for developers and institutions to develop regional aggregators.

According to their application area there are two basic types of aggregators:

- Aggregators with a global purpose – they are the result of larger initiative to improve online accessibility and usability of digital resources of libraries, archives and museums, to promote research for development search functions and the retrieval of integrated information to accelerate the digitization and to improve the training process for application of modern technologies;

– Aggregators focused on specific areas – they provide the technological tools for documenting and searching of specific subjects and topics (special cultural objects, musical instruments, educational problems, biodiversity, etc.).

The main characteristic of both types of aggregators is related to their searching capabilities. Services provided on both aggregators share the same features: they are portals for semantic search and navigation for various types of digital objects (text, images, video and audio files) and have options for storing and sharing content.

Most of the aggregators (over sixty percent) are designed as a resource for preparation of data digitalized by cultural institutions that lack the capacity to develop and maintain their own digital repository. Only one third of the aggregators that provide access to metadata and digital content for heritage contain four types of digital objects (audio, video, text and image).

The main conclusions that can be outlined are about the importance of the contents of the collections as a substantial criterion for assessing, the development of new features and services, and the need for its continuous enrichment.

All aggregators are ready to deliver content to Europeana. While the survey was taking place, in early 2010, only 20% of aggregators were ready to deliver content and 60% planned to implement this in the period of 2010-2011. 15% of the content has arrived, or will reach Europeana through the development of projects (TEL, EuropeanaLocal, ATHENA).

An aggregator, in the context of Europeana, is an organization that collects metadata from a group of content providers and then transmits them to Europeana. Aggregators collect data from individual organizations and standardize file formats and the corresponding metadata in accordance with the procedures for Europeana. The administrators of the aggregator are committed to support the efforts of the content providers through technological assistance, consulting and training.

3 The Prototype REGATTA–Plovdiv

The task was to create a regional aggregator of digital artefacts based on the standard used by Europeana. It is presumed that the aggregator is accessible to each regional cultural and historical institution (for storing digitalized artefacts), as well as end users (for resource discovery). The creation of a regional aggregator is the first step towards presenting and promoting the rich heritage of Plovdiv and its region in the European digital space.

The open structure of the aggregator enables the creation of data models for various types of digitized cultural objects. This allows different types of collections to be presented, including museum collections, archaeological sites, and immovable heritage from Ancient, Mediaeval and the National Enlightenment periods in Bulgaria. The chosen approach supports the idea of preserving the valuable national monuments in the European area of culture, keeping their identity and uniqueness. The experimental "**REG**ional **A**ggregator of he**T**erogeneous cul**T**ural **A**rtefacts" (REGATTA) is the basic practical outcome. Currently the first application of REGATTA is applied for the Plovdiv region and can be accessed on http://www.plovdiv-eu.com. It was designed following the standards of Europeana [EAH, 2010] and characteristics specified in the so-called "*passport of cultural values*" [Reg.6, 2009]. Its purpose is to bring together objects from the collections of museums and other cultural and historical institutions.

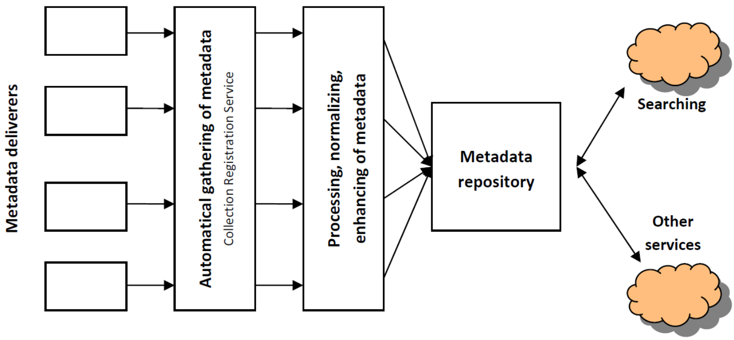

Figure 4. Aggregation of metadata

The foundation of the model for digital library is the structure of the metadata collections for specific items. In historical context, the process of creating such structures recaps the efforts and resources, well-known from the systems of cataloguing, search and retrieval of information in library systems.

The structure of the present project [Hadjikolev et al, 2010/UBS] includes modules for the aggregating of metadata from corresponding resources, their storing in a repository and providing services through their processing. An important component for the functionality and performance of this architecture is the metadata aggregator. Figure 4 shows the technology of metadata aggregation.

3.1 The Functional Scheme of REGATTA

The selection and preparation of structured metadata is the basis for designing the digital library. The metadata is well-known and is used as a tool for information library services, particularly for searching and finding information. The modern technologies for its use are applied by the Digital Scientific Library NSDL[91].

Modern museums tend to keep and serve two kinds of collections: physical and digital. The textual description of the subject of a digital image is regarded as metadata associated with the image. As previously stated, prerequisites for the project were the effective structuring of object collections and maintenance of metadata according to relevant standards – in this case the standards used by Europeana.

The functional structure of REGATTA follows the framework suggested in the Open Archival Information System (OAIS) [OAIS, 2002] which has also been adopted as the international standard ISO 14721:2003. This model finds its successful implementations as a common framework with concretizations in application areas for the so-called GLAM (Galleries, Libraries, Archives, Museums). Figure 5 shows the functional schema of REGATTA conforming to OAIS.

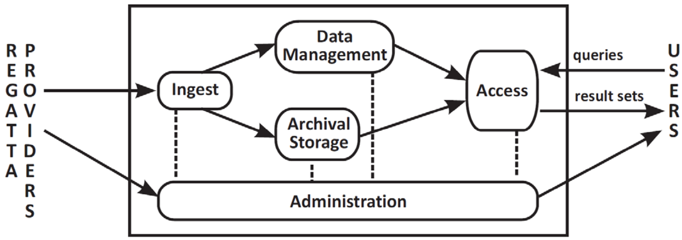

Figure 5. REGATTA Functional Entities

The aggregator of digital collections is a web-based technology, enabling the numerous different users not only to publish objects on the Internet but also to create their own models of data, related to these objects [EAH, 2010]. All the objects with the same data model are combined in one collection.

[91] http://nsdl.org

The main features of the technology are:

- Multiusers – users with different roles are organized in a hierarchy with defined access rights to services. Each user can log in and change the basic details of his/her account;
- Multilingual – objects can be introduced in one or more languages. For the realization of this mechanism, localized files with basic terms are used as well as localized database;
- Catalogue of objects – the standard options for entering, editing and deleting objects and for different types of searches (keywords, categories, dates) are established;
- Categories of objects – the standard options for input, edit and delete categories are provided;
- Collections of objects –modelling of objects data is enabled.

Treating collections as part of the model is very natural for GLAM institutions and in the case of REGATTA it differs from Europeana's approach. Browsing within a specific collection is natural to visitors of cultural institutions and has a number of benefits in the digital space because it contextualizes the objects and allows easy discovering of multiple objects which are thematically or chronologically related.

A good example of that is Ireland's gateway to Irish digital collections and resources DHO:Discovery [92]. It supports the interdisciplinary and inter-institutional sharing of knowledge throughout the Humanities Serving Irish Society (HSIS) consortium and across digital research collections of Irish interest [Gourley and Viterbo, 2010]. A key requirement of DHO's website development was the need to support both the development of thematic research collection project websites, led by the partners and focusing on a single collection of resources, and the need to provide a generic cross collection interface to discover and re-use resources across the whole repository. The goals behind REGATTA's creation were similar – to keep the specificity of each collection in order to assure most colourful representation of each object in its natural environment and to build a common frame which allows different kinds of objects having some semantic coherence on different levels.

3.2 Data Model in REGATTA

In the case of uniform collections, the different kinds of producers usually apply the same metadata model. Here, the collections present different objects. For instance: texts and images for movable artefacts; 3D representation for immovable sites; music or video for representing

[92] http://discovery.dho.ie

folklore and customs, etc. and each object of these different types is supplied with corresponding metadata description. Because of this, REGATTA allows producers to add their own object characteristics. The incorporated technology that reflects the specifics of data input includes [Hadjikolev et al, 2010/UBS]:

- Creation of a basic scheme for object description;
- Hiding some of the fields from the forms during data input;
- Adding additional specific fields/characteristics from the producers;
- Defining names of the models.

The technology adopts the objectively arisen lack of correspondence between the basic characteristics described in the standards and the concrete available data of the institutions. For example, the standard fields in the passport [Reg.6, 2009] are 26 well defined fields. In the passports of the Ethnographic Museum in Plovdiv the objects are described in 37 fields. They also contain fields that are not relevant to the main features. Moreover, not all objects cover the full set of features required.

The creation of data models ensures additional classification of the objects and facilitates data input.

Collection in the sense of the created technology is a set of objects with the same data model. Using the data modelling mechanism, the different institutions can create a collection of object descriptions based on the already existing metadata schemas or can create their own.

The main requirements of the system are:

- Compatibility with the Europeana standards, i.e. the objects provided in the portal can be easily exported to the Europeana portal;
- Compatibility with the characteristics required for the movable cultural properties' passport in Bulgaria; the comparison of the national standards with the Europeana standards, where in result many common features are found. The correlation between the names of the characteristics of the two standards is described in the help information provided for each entry field. For example, "Name of the object" of the movable cultural property's passport has a corresponding label in Europeana "dc:title". Different characteristics of the two standards are divided into subpages in order to describe the shape of an object – "Metadata" and "Passport". The "Metadata" are specific features of the Europeana's standards, which are rarely used. The "Passport" contains specific characteristics of Bulgarian museum exhibits. This possibility to introduce specific passport data enables the institutions to use the system as a data repository as well.

3.2.1 Functional Elements

Ingest: The process of incorporating the digital objects in REGATTA takes place in three phases: Preliminary Phase, Transfer Phase, and Validation Phase (Figure 6). The Preliminary phase includes identification of information about objects that will be presented in the aggregator. The content provider creates a model of the collection or selects any of the already defined models in the REGATA collections. Here, the assessment of the resources (time, people, financial) is done as well. In the Transfer phase, the data input is carried out. Each assistant can manage only objects stored by them. The content providers can monitor the work of their assistants. In the presence of well-structured digital information, the provider may prepare a tool for automatic transfer by defining the appropriate mappings. The Validation phase includes verification of data entered and the deletion of errors and omissions. Only after this phase, digital objects become part of the REGATTA's public record and can be accessed by users. Most of the functions are already established, however some of the functions are in the process of planning and development.

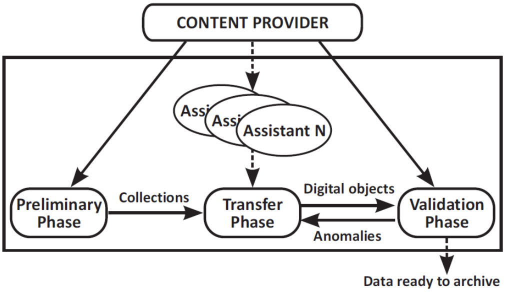

Figure 6. REGATTA Ingest

Archival Storage: This module provides services and functions for storage, maintenance and restoration of digital objects. This includes the receiving of digital objects from the Ingest and adding them to the backup repository, the performance of routine and special checks for faults that periodically backup (duplicate) data for recovery after a system failure.

Data management: The module provides services and functions for implementation, maintenance and access to both descriptive information, which identifies the owner of the archive, and to the administrative data used to manage the archive. This includes functions for managing database records, updating the archive, performing operations on data retrieval efficient sets and generating relevant reports and more.

Administration: The administration module provides services and functions used for the overall functioning of the aggregators: registration and maintenance of accounts, defining collections, data entry in aggregators (digital objects and descriptive information), management of providers' standard sites, search, retrieval and formatting the data resulting from user requests and more.

Access: This module is concerned with the access to the aggregator by the two main types of users which can be discerned (Figure 7): humans and robots. The humans use the REGATTA-content through the base portal. Each content provider has a standard website, containing their objects and additional information for them. The second type of users is web applications, which extract data from the aggregator and put it under additional processing. The simpler types are different kinds of search engines that use data directly from the basic model. The more sophisticated web applications use data and links given by the aggregator for incorporating into static or dynamic websites, 3D-tours, virtual excursions, etc. For these purposes REGATTA provides the standard services for generating content. There is an option to define user styles, used by REGATTA to return organized content, which can be incorporated directly into the external web-application without additional processing.

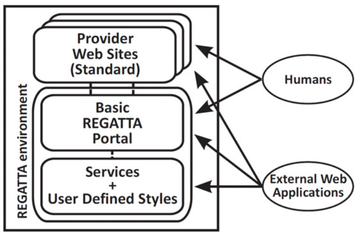

Figure 7. REGATTA Access

3.3 Technological Aspects

The fundamental technological aspects of the current aggregators are in accordance with the protocol OAI-PMH (*Open Archives Initiative Protocol for Metadata Harvesting*) established and developed within the *Open Archives Initiative* community[93]. It is used for extraction and collection of metadata from descriptions in the information providers. The developments, based on this protocol, prevailingly maintain the metadata standard Dublin Core[94].

---

[93] http://www.openarchives.org
[94] http://dublincore.org

The design of the systems associated with the Europeana initiative complies with the established rules for preparation of digital libraries, described in [DL, 2007].

The tasks performed by these projects are:

- Preparatory work with digital objects – the expert services of museum professionals are widely used;
- Introduction and management of digital objects collections – images, text, video, audio and more;
- Retrieval and management of metadata;
- Analysis of objects and their descriptions aimed to support their structure and clustering;
- Development of modules for access and (integral, semantic, context-based) search of digital objects;
- User requirements: a graphical interface, review services and interactive presentation of digital content and objects, multilingual support;
- Administration, supporting roles for the various categories.

3.3.1 User Aspects

The system is maintained in two languages – English and Bulgarian. Every object can be introduced simultaneously in both of these languages and it is displayed in the language version of the portal. Unfortunately most of the descriptions of museum exhibits in Bulgaria do not have English translation.

The roles – in the completed system there are several types of users. Each of them has a role that is usually chosen when the user is registering and that specifies what rights they have to use the services:

- *Unregistered user* – only deals with content – static pages, pages of sites and institutions, carries out different types of searches;
- *Registered user/Institution* – there are two types of registered users – generic users and Institutions. They can create, edit and delete categories, collections, sites and websites (as a sub-domain). The difference is that institutions can activate their objects, while the viewers can activate their objects only when they are authorized by the system administrator. This restriction is necessary for content security reasons;
- *Consumers to institutions* – each institution may set up helpers for the entering of objects. After a review is made over the additions, the institution activates the objects entered by the helpers. After this operation, the objects become inaccessible for the helpers who

entered them. In some cases, such helpers could be students or employees temporarily engaged with the institution.

– *Administrator* – controls major categories and collections of the portal as well as objects created by ordinary users.

The work on the development of the online cultural and historical websites continues through the creation of additional opportunities for both rounds of consumers – the museum workers and other visitors with diverse interests.

4 Virtual Tours in REGATTA

The virtual tours provide a realistic way to create a full exhibition of architectural sites, museums and galleries on the Internet. Virtual tours are recommended as a good alternative way to visit, especially for people with special educational needs. They should include the necessary information, in order for consumers to receive the same knowledge as in an actual site visit.

According to the method of presentation, there are two kinds of virtual tours: *three-dimensional* and "*presentation*" type. In a three-dimensional virtual tour (created with the help of photogrammetry) images are projected onto surfaces like walls, buildings floors, etc. In the second type of virtual tours (so-called "presentations"), panoramic images are used.

In terms of ways and means for the implementation of virtual tours, the following types can be distinguished: *text-based*, *photo-based*, *video-based*, *panoramic* and *virtual reality* in real time [Sobota et al, 2009].

Text virtual tours are a narrative description of the space and content of the display site. When using a program to convert text to speech the tour is converted to a *virtual audio tour*.

In the *photo-based virtual tours,* objects are displayed through a series of images and their textual description. Depending on the software, photo-based virtual tours can be interactive (e.g. clicking an artefact can zoom, or trigger audio or text description). For example, front presentation in the site of Van Gogh Museum [95] is constructed as a photo-based virtual tour.

In the *video-based virtual tours,* the video view of a typical visitor walking supplemented with some audio information (speech guide, music and/or special effects) and/or text information are proposed. One example is Video and Audio Tour in the British Museum London[96], which can be seen on YouTube.

[95] http://www.vangoghmuseum.nl/

[96] http://www.youtube.com/watch?v=b71Oi677irI

The building of *panoramic virtual tours* is based on a series of consecutive overlapping images that are "sewn" together in order to create a continuous 360° view of the object, or it is done by using a set of special panoramic shots. These tours are interactive. An excellent example is shown on the website of the Vatican Museums[97].

Virtual reality in real time (3D-virtual tours) are built on the basis of software modelling of three-dimensional objects that are used to make the user feel like being among the exhibits. Users can manage their way through virtual reality as if walking among real objects.

4.1 Panoramic Virtual Tours

Panoramic virtual tours provide the opportunity to view the panoramic images to be viewed in an interactive way. These technologies enable users to examine (crawl, spin, zoom and see additional information about a specific artefact) panorama of the situation as though users are inside the real object.

Since 1996 the team of Panoguide[98] has been aiming to provide a free central resource of information and discussion about panoramic photography.

The virtual tour is prepared by using multiple panoramic images that are linked through the so-called hot-spots. A hot-spot is a part of the panoramic images which allows interaction with and this can provoke an action – moving to another panoramic image or displaying further information. The most common example is a hot-spot on the door, which transfers the user to the panorama of the room behind the door. For a more realistic representation, sometimes the virtual tours are accompanied by sounds.

Panorama is created in various shapes and sizes depending on the chosen projection, showing how the prospect of panoramic images is changed by software to provide a full or partial 3D-scene or a realistic 2D-scene on the computer screen.

There are several types of projections used in the creation of panoramas:

- Full ball formats – displayed on all of the surrounding space, visible 360° by horizontal, 90° up and 90° down. Two types of techniques are used here:
  - Ball – a panoramic image is projected onto the sphere inside, and the panorama ratio height:width is 1:2;

[97] http://mv.vatican.va/3_EN/pages/MV_Visite.html
[98] www.Panoguide.com

    - Cubic – a spherical view is based on the cube, i.e. using six photo-walls in the ratio height:width equal to 1:1;
- Partial formats – a partial view is shown as a horizontal visible maximum of 360°, but the vertical maximum is 120°. The used techniques are:
    - Cylindrical – a display area inside the surrounding wall of the cylinder or part of it (used for landscape panoramas);
    - Straightforward – the horizontal and vertical visibility is restricted to 120° (used for architectural objects);
    - Partial ball – implemented as a full ball, by cutting the highest or lowest point of the panorama.

For recording digital 360° panoramic images the following technologies are used [Maas and Schneider, 2004]:

- "Stitching" of images through an ordinary camera;
- A simple camera with wide angle lens (180°);
- A simple camera with hyperbolic mirror;
- Camera with rotating sensor linear typesetting;
- Camera with multisensory system (four or more sensors, equipped with wide angle lenses).

Also, there is a wide variety of software for making panoramas and panoramic virtual tours (PTViewer[99], Spi-V[100], QuickTime[101], 0-360 UnWrapper[102], Panoweaver[103], etc.), as well as file formats for storage (QTVR, JPG, IVR, etc.).

4.2 3D-Virtual Tours

Virtual tours can offer a way to travel back in time by producing objects that are currently not-existing or an earlier view of the existing ones. Also these tours can be used for online visits to existing sites, offering a simulation of an actual visit. For this purpose, 3D computer graphics are used to create mathematical 3D-models of the objects in the scope of the virtual tour. The resulting 3D-model is "decorated" to mimic the real object (texturing elements of the model; adding windows, doors, curtains, furniture, artefacts, etc.).

The specialized software grasps the opportunity to crawl and view the model in its individual parts. The resulting virtual tour allows movement

[99] http://www.fsoft.it/panorama/ptviewer.htm
[100] http://fieldofview.com/projects/spv
[101] http://www.apple.com/quicktime/
[102] http://www.0-360.com/software.asp
[103] http://www.easypano.com/

and exploration of objects in real time without any "jumps" in space as hot-spots of panoramic virtual tours.

There are four popular ways to create a 3D-model:

- Polygonal modelling – a form of the model that allows to be drawn on with the use of the polygon tool, which is then divided and screened until a desired 3D-shape is acquired and finally, smoothing to make the object look realistic;
- NURBS (Non Uniform Rational BSpline) modelling – mathematical curves are painted by a set of equations that have control points that can change the shape of the curve;
- Modelling with splines;
- Modelling with primitives – use of geometric primitives such as spheres, cylinders, cones and cubes, which help to build a more complex model.

There are many software programs for 3D modelling: 3DS Max and 3DS Max Design[104], Maya[105], Blender[106], DAZ Studio[107], Cinema 4D[108], Houndini[109], Poser[110], ZBrush[111], Google SketchUp[112], etc.

5 Presentation of Plovdiv Ethnographic Museum in REGATTA

The pilot implementation of presenting movable and immovable artefacts in REGATTA was made in collaboration with the Plovdiv Ethnographic Museum.

5.1 Movable Artefacts

Concerning technical interoperability [IDABC, 2004], the processes of migration between other kinds of presentation of the artefacts have to be decided for each case separately. But the first question is "Is there compatibility between the fields of Europeana and the fields in the existing passports of the objects?".

The exhibits of the Regional Ethnographic Museum – Plovdiv are allocated to funds/departments and collections defined in [Reg.6, 2009]. The departments in the museum are "Agriculture", "Crafts", "Woven

[104] http://usa.autodesk.com/3ds-max/
[105] http://usa.autodesk.com/maya/
[106] http://www.blender.org/
[107] http://www.daz3d.com/
[108] http://www.maxon.net/
[109] http://www.sidefx.com/
[110] http://poser.smithmicro.com/
[111] http://www.pixologic.com/
[112] http://sketchup.google.com/

Fabrics and Apparel", "Furniture and Interior", "Ritual Musical Instruments and Props", "Photo Library and Artworks". The Crafts department contains collections such as "Jewels", "Wrought Iron", "Cold Steel" and others that the user himself can create.

The Ethnographic Museum in Plovdiv maintains two types of documents in electronic format: (1) passports of the objects in Word format and (2) inventory books in Excel format. They contain the so-called "Scientific passports", made under state requirements [SR, 2009] and reflect the metadata that are mandatory for Bulgarian museum institutions and in the meantime are conformable to MARC standard.

Table 2 shows the compatibility between existing data for digital object of the Plovdiv Ethnographic Museum and metadata for Europeana. The sign 'Y' indicates that the metadata are available (in one of the two documents); sign 'N' means that metadata are not available; sign 'A' points that metadata could be created automatically during the process of storing objects into the corresponding digital repository.

Table 2. Relation between metadata (Europeana) and data concerning digital objects of the Plovdiv Ethnographic museum

Strongly recommended		**Additional**		**Europeana**	
dc:title	Y	dc:format	A	europeana:country	A
dcterms:alternative	Y	dcterms:extent	Y	europeana:hasObject	A
dc:creator	Y	dcterms:medium	Y	europeana:isShownAt	A
dc:contributor	Y	dc:identifier	Y	europeana:isShownBy	A
dc:date	Y	dc:rights	A	europeana:language	A
dcterms:created	Y	dcterms:provenance	Y	europeana:object	A
dcterms:issued	A	dc:relation	Y	europeana:provider	A
Recommended		dcterms:conformsTo	A	europeana:type	A
dc:coverage	Y	dcterms:hasFormat	A	europeana:unstored	Y
dcterms:spatial	Y	dcterms:isFormatOf	N	europeana:uri	A
dcterms:temporal	N	dcterms:hasVersion	N	europeana:usertag	A
dc:description	Y	dcterms:isVersionOf	N	europeana:year	A
dcterms:isPartOf	Y	dcterms:hasPart	Y		
dc:language	A	dcterms:isReferencedBy	N		
dc:publisher	Y	dcterms:references	N		
dc:source	Y	dcterms:isReplacedBy	N		
dc:subject	Y	dcterms:replaces	N		
dc:type	Y	dcterms:isRequiredBy	N		
		dcterms:requires	N		
		dcterms:tableOfContents	Y		

The work with these archives requires analysis of the readiness for creation of metadata in accordance with *Europeana* standards [EAH, 2010]. The research of the "Crafts" fund of the Ethnographic Museum helped us to specify the basic classes and subclasses of elements in the corresponding groups required by the *Europeana* metadata scheme.

"Identification signs" – contains descriptive criteria (typical for the traditional metadata description as well) such as title, archive number, period, place of residence, type, annotation, etc. of the specific object.

"Technical information" – includes information related to the digitalization of the object.

As an illustration we present part of the xml description in accordance with the requirements of Europeana:

```
<!-- Strongly recommended elements -->
<dc:title> Pitcher </dc:title> <! Title of the object -->
<dcterms:alternative>Earthen Jar</dcterms:alternative> <!-- Alternative
„folk" title -->
<dc:creator>Unknown master from Troyan</dc:creator> <!-- Author of the
original -->
<dc:contributor>Darin Kambov</dc:contributor> <!-- Grantor -->
<dc:date>1860–1870 </dc:date> <!-- Date of the creation of the original -->
<dcterms:created>Revival</dcterms:created>
<dcterms:issued>2007</dcterms:issued><!-- Specifying element of <dc:date>
                Date of publishing of the digital object or the original -->
```

Then some "additional elements" follow – the examination of the scientific passports of the exhibits reveals a satisfactory amount of information for their retrieval – a description of the original object, physical characteristics, data related to conservation, digitalisation, etc.

Several questions are examined concerning the automatic transfer of data objects from the collections of the Ethnographic Museum:

- is there a correspondence between the fields of Europeana and the scientific fields in the passport of the object;
- can data be automatically extracted from the inventory book (Excel-file) and transferred into the database;
- can data be automatically extracted from the passport (Word-file) and transferred into the database;
- can we analyze the information from the inventory book and the passport, concerning one particular object, in order to optimize performance.

Despite the aggregator was created quite recently, the automatic transfer of data concerning the objects from the collection of the Ethnographic museum was successful [Hadjikolev et al, 2010/MathTech].

The Excel-files with simple transformations (e.g. word processing functions or MS Access) were transferred to a database table (in this case – MySQL). After the initial transfer in a secondary table, the data were distributed into the original tables of the object.

The transformation of Word-documents into a format, suitable for automated processing, has proved a more difficult task. There are technological solutions, but their use for treatment of a particular file format is meaningful if a joint decision can be accomplished, related to other similar tasks as well.

Table 3 shows the bijection between fields of the scientific passport and REGATTA-metadata on the example of one concrete exhibit (shown in Figure 8) of the museum.

Currently, the catalogue includes more than 4500 objects from the "Crafts" fund at the Ethnographic Museum.

Table 3. Scientific passport N4501 of the Plovdiv Regional Ethnographic Museum and correspondence to aggregator metadata

Passport fields	Values	Aggregator metadata
Section	Crafts	Created by the administrator
Collection	Wrought Iron	Created by the administrator
Name	Candlestick, wall, Double-arm	dc:title
Folk name		dcterms:alternative
Inventory number	4501	dc:identifier
Dating	2005	dc:date
Number of exemplars	1	
Material	Wrought iron	dcterms:medium
Sizes	H=78cm; W=38cm	dcterms:extent
Weight		dcterms:extent
Carats		dcterms:extent
Object is one with inv.numbers		dc:relation
Producing place	Plovdiv	
Keeping place	Craft Fund of REM-Plovdiv	
Author of the original	Georgi Manolov	dc:creator
Technique	Craft mastering	
Technology		
History of the object		dcterms:provenance
The object is from the group of	Professional	
Condition	Very good	
Conservation and restorations		

Object description	Candlestick, wrought iron, wall, double-arm; examination work of Georgi Manolov to obtain a master title from the Association of Masters of Folk Crafts	dc:description
Literature		dcterms:references
Surrogates	no	
Object moving		
Remarks		
Passport maker	Sonya Semerdjieva	
Appointment	Chief organizer	

Automatically for this object the system filled in:

europeana:isShownAt	http://www.plovdiv-eu.com/object.php?id=3
europeana:isShownBy	http://www.plovdiv-eu.com/images/user_objects/3/1374422587129031399.jpg
europeana:type	IMAGE
europeana:provider	In progress

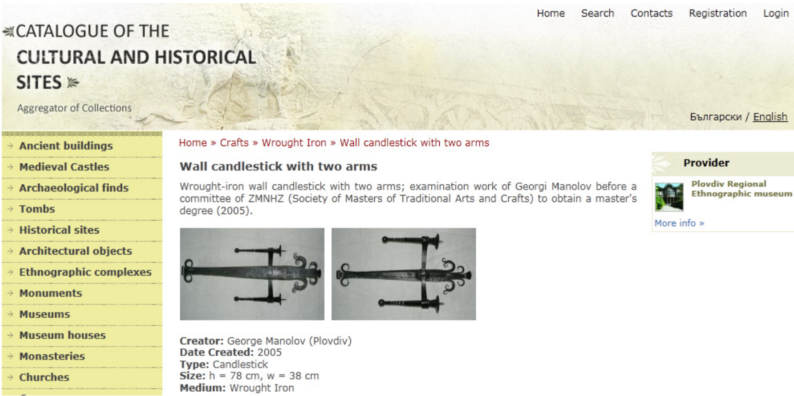

Figure 8. Exhibit N 4501 of the Plovdiv Regional Ethnographic Museum

5.2 Virtual Tours of the Plovdiv Ethnographic Museum

The information for immovable sites in REGATTA is still in the process of unification and enrichment. Currently the process of building proper presentation of such sites, including gathering historical materials, building design and scenarios go in parallel with such processes of assuring rights for representing immovable sites in digital form concerning different institutions, such as Ministry of culture, the Holy Synod, municipalities, etc.

The successful examples of integrating a learning process with collaboration opportunity to develop a shared syllabus and associated teaching and learning resources for humanities visualization, such as educational and research project Second Life, presented in [Denard et al, 2010], gives us assurance to start programs in informatics specialties in Plovdiv Universities for using 3D and VR-representation of these sites as practical works.

The first steps for presenting immovable sites are already made [Stoyanov et al, 2011]. 3D and VR representations of the Plovdiv Regional Ethnographic Museum have been created. The connections have been established between a VR representation and already digitized craft collection. Now, the process of incorporating these resources in REGATTA is ongoing. In parallel, the construction of the collections that represent immovable sites are being expanded with the information in correspondence with [CARARE, 2010] in order to establish easy incorporation into Europeana.

Below, two kinds of virtual tours of the Plovdiv Ethnographic Museum are presented.

5.2.1 Panoramic Virtual Tour

An experiment was carried out to create a spherical panoramic virtual tour of the Plovdiv Ethnographic Museum. A simple camera with a hyperbolic mirror is used to capture the panoramic view.

Figure 9. 360° panoramic photo image of the museum

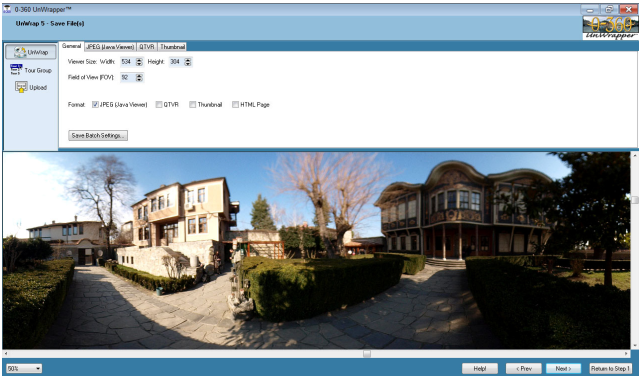

Figure 10. "Extended" 360° image of the museum

The resulting picture is shown in Figure 9. The software 0-360 UnWrapper[113] is used for generating the spherical panorama (Figure 10).

The construction of the panoramic virtual tour itself is realized through the Tourweaver[114] software. The resulting file format ".Swf" does not need a special software to be displayed and the virtual tour can be seen using the popular Adobe Flash Player.

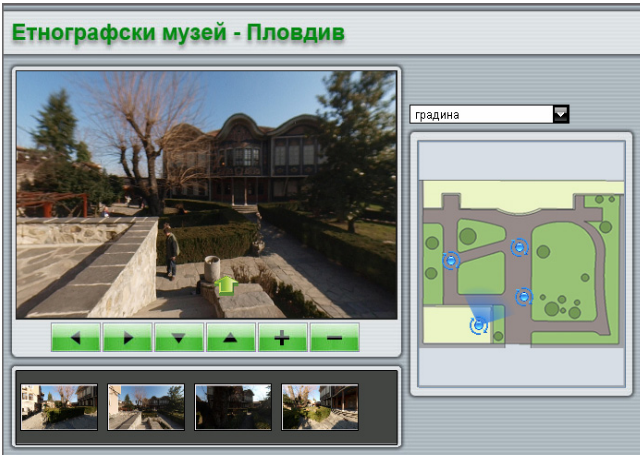

Figure 11. View from the panoramic virtual tour of the museum

---

[113] www.0-360.com
[114] http://tourweaver.en.softonic.com/

For easier orientation of the user, additional maps are inserted as an accompanying element of the virtual tour in order to allocate possible entry points views of the active view point of the museum space (Figure 11). Entry points to the input views are hot-spots. The point of view is an active radar, which is a hot-spot with very specific action – to show the position and direction of the current view displayed in the map. Maps of the garden and every floor of the museum were created with the Google SketchUp[115] Software.

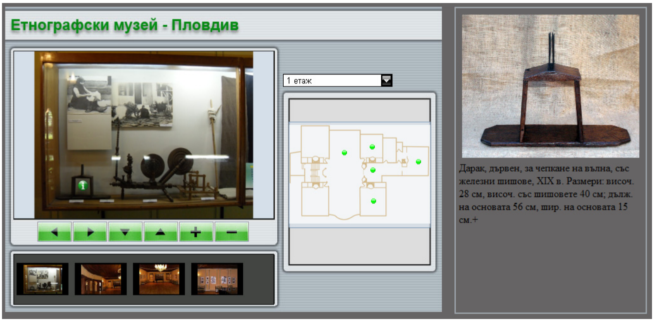

Figure 12. Presentation of an artefact in the frame of a panoramic tour

Besides the standard navigation elements (direction, zooming in and out, skip a particular place) a lot of hot-spots are placed for quick access to other views (rooms) or for more important artefacts of the museum. The artefacts are presented by the image and text information (metadata) such as name, origin, date and time of creation, author, etc. (Figure 12). Performed acts of hot-spots are generated by JavaScript.

5.2.2 3D-Virtual Tour

Here is presented the development of 3D-model of the Plovdiv Ethnographic Museum and its use for the realization of 3D-virtual tours.

Modelling primitives are used for the building of the museum model. The complex objects are split into built primitive forms. The construction is done piece by piece, starting with the walls and ending with the roof and the environment. Special attention is paid to the modeling of some artistic details of the building. A 3D-model of the museum has been built (Figure 13), both from the outside and the inside. The modelling is based on the original scaling dimensions of the museum.

---

115 http://sketchup.google.com/

The software program Google SketchUp was used for the 3D modelling of the museum. The program is specifically designed for architects and civil engineers. Google SketchUp is easy to learn and use.

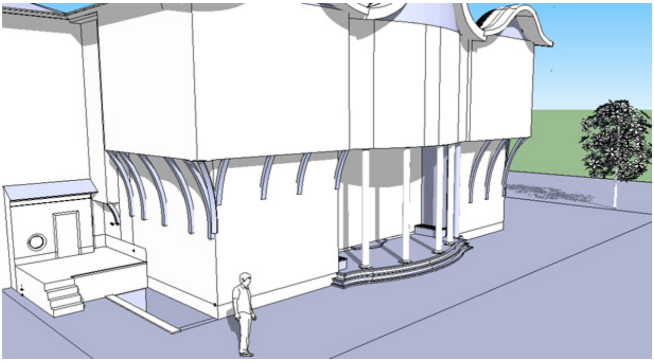

Figure 13. Monochromatic 3D model of the museum

Figure 14. Figure 6. Textured 3D model of the museum

The monochromatic 3D-model is textured with real images from the museum in order to obtain a photorealistic model of the building (Figure 14). Photographs of the museum, which is used for the textures, need to be edited for the establishing the correct perspective of the model (to change the angle of shooting), removing unnecessary objects from the picture, like trees and people, and adjusting the image resolution. For such purposes, the images are pre-processed with Adobe Photoshop[116].

The textured 3D-models are imported into the Unity3D[117] software to build a virtual tour. This software allows composing of the scene to make animations, adjusting the lighting and the movement inside the museum. Such movement is done through the figure of a man crawling out of the building, guided by the user.

The hot-spots for pointing the more significant artefacts of the museum can also be placed here. The artefact will be also enhanced through image and text information (metadata) (Figure 15). Performed acts of hot-spots are implemented through JavaScript.

Figure 15. Show metadata artefacts in 3D-virtual tour

6 The Next Step – Enforcing the Data Management with Data Mining Tools

In the frame of the project which supports the realization of the REGATTA aggregator, several tools were invented that use different kinds of data mining techniques for automated metadata extracting and categorization.

An approach for indirect spatial data extraction by learning restricted finite state automata is presented in [Blagoev et al, 2009]. It uses heuristics to generalize initial finite-state automata that recognizes the positive examples without extracting any non-positive examples from the training data set. The created system InDES was tested over extraction of spatial metadata from websites and shows promising results. It gives us assurance that such an approach can be used for metadata extraction

116 http://www.adobe.com/
117 http://unity3d.com/

from objects descriptions and this way can be applied in the process of migration from older representations of the objects in cases when the descriptions are in non-structured form.

Association rule mining (ARM) is a popular and well researched method for discovering interesting rules from large collections of data. It has a wide range of applicability, such as market basket analysis, gene expression data analysis, building statistical thesaurus from text databases, finding web access patterns from web log files, discovering associated images from huge sized image databases, etc. One approach for association rule mining, which uses the possibilities of the multidimensional numbered information spaces as a storage structures is presented in [Mitov et al, 2011]. The ArmSquare algorithm is implemented in data mining environment system PaGaNe [Mitov, 2011]. The possibilities of extracting frequent item-sets can be used for enforcing connections between metadata elements within created ontology.

Based on similar techniques, but in the field of categorization, are class-association rules (CAR) algorithms. Compared to other classification methods, associative classifiers hold some interesting advantages [Zaiane and Antonie, 2005]. Firstly, high dimensional training sets can be handled with ease and no assumptions are made about the dependencies between attributes. Secondly, the classification is often very fast. Thirdly, the classification model is a set of rules which can be edited and interpreted by human beings. The created algorithm PGN, which is a kind of CAR-algorithm is also implemented in PaGaNe. It was implemented in the field of analyzing semantic attributes, extracted from art images using content-based image retrieval. The rules, extracted by PGN, form the semantic profiles of the examined movements [Ivanova, 2011]. Similarly within the frame of the data management and the access of the aggregator the classifier PGN can be used for enforcing information discovery.

The integration of data in the virtual space has to be examined also in the context of global Semantic Web. As it is mentioned in [Berners-Lee, 2009] "the Semantic Web isn't just about putting data on the web. It is about making links, so that a person or machine can explore the Web of Data". This Web of Data enables new types of applications. RDF links enable you to navigate from a data item within one data source to related data items within other sources using a Semantic Web browser. Such links can also be followed by the crawlers of Semantic Web search engines, which may provide sophisticated search and query capabilities over crawled data [Bizer et al, 2009]. In order to supply such integration

W3C created Library Linked Data incubator group[118], whose mission is to help increase global interoperability of library data on the Web, by bringing together people involved in Semantic Web activities – focusing on Linked Data – in the library community and beyond, building on existing initiatives, and identifying collaboration tracks for the future.

7 Conclusion

Work on the development of online cultural and historical objects continues through the creation of additional new opportunities for both rounds of consumers – the museum curators and other visitors with diverse interests.

In the modern museology the conception that the value of the artefact depends not only on information contained herein, but also on the facilitated distributed access to it is already emerging. Digital libraries are the appropriate solution to make museums information centres for a wide range of information services for museum professionals as well as for visitors and users of museum information.

The presented technology supports dynamic creation of categories and standardized collections of objects. The opportunities for modifying allow associating additional collections in the system. It can be used in the construction of catalogues with various objects – cultural, historical, natural, personal data, etc.

Bibliography

[Berners-Lee, 2009] Berners-Lee, T.: Linked Data. http://www.w3.org/DesignIssues/LinkedData.html

[Bizer et al, 2009] Bizer, C., Heath, T., Berners-Lee, T.: Linked data – the story so far. Int. Journal SWIS, 5(3), 2009, pp. 1-22.

[Blagoev et al, 2009] Blagoev D., Totkov, G., Staneva, M., Ivanova, K., Markov, K., Stanchev, P.: Indirect spatial data extraction from web documents. IBS ISC 14: New Trends in Intelligent Technologies, Sofia, 2009, pp. 89-100.

[CARARE, 2010] White Paper on CARARE Technical Approach. http://www.carare.eu/fre/Media/Files/White-paper-on-CARARE-technical-approach

[Chen et al, 2005] Chen, C.-C., Wactlar, H., Wang, J. Z., Kiernan, K.: Digital imagery for significant cultural and historical materials – An emerging research field bridging people, culture, and technologies. Int. Journal Digital Libraries, 5(4), 2005, pp. 275-286

[Cousins, 2011] Cousins, J.: The cultural heritage of Europe: building the value proposition for Europeana. In: The Journal for the Serials Community, 24 (1), 2011, pp. 69-78.

[118] http://www.w3.org/2005/Incubator/lld/

[Denard et al, 2010] Denard, H., Salvatori, E., Simi, M.: An integrated approach to digital cultural heritage. 3rd Int. Conf. dedicated on Digital Heritage EuroMed2010, Limassol, Cyprus, 2010: Short Papers, Archaeolingua, Budapest, pp.73-78.

[DL, 2007] Digital Libraries: Research and Development. LNCS, Vol. 4877, 2007, pp. 22-35

[EAH, 2010] Europeana Aggregators' Handbook. http://version1.europeana.eu/c/document_library/get_file?uuid=94bcddbf-3625-4e6d-8135-c7375d6bbc62&groupId=10602

[EM-TR, 2007] Bulgaria – Collective Memory and National Identity. Tech. Report, Ethnographical Museum with Institute – BAS, 2007.

[Gourley and Viterbo, 2010] Gourley, D., Viterbo, P.: A sustainable repository infrastructure for Digital Humanities: the DHO experience. 3rd Int. Conf. dedicated on Digital Heritage EuroMed2010, Limassol, Cyprus, 2010: Project Papers, LNCS 6436, pp. 473-481.

[Hadjikolev et al, 2010/MathTech] Hadjikolev, E., Vragov, G., Totkov, G.: Aggregator for standardized collection of metadata for cultural and historical objects. National Conference with International Participation MathTech, Shumen, Bulgaria, 2010, UI "Ep. K. Preslavski", 2011, vol.1 pp. 141-149 (in Bulgarian)

[Hadjikolev et al, 2010/UBS] Hadjikolev, E., Totkov, G., Vragov, G.: Digital technologies for presenting museum collections. Proc. of Annual Scientific Conference of Union of the Bulgarian Scientists, Plovdiv, Bulgaria, 2010, pp. 192-198 (in Bulgarian)

[IDABC, 2004] European Interoperability Framework for pan-European eGovernment Services. http://ec.europa.eu/idabc/en/document/3761/5845.html

[Ivanova, 2011] Ivanova, K.: A Novel Method for Content-Based Image Retrieval in Art Image Collections Utilizing Color Semantics. PhD Thesis, Hasselt University, Belgium, 2011.

[Krasteva, 2007] Krasteva, St.: The Museology – Meeting between Alpha and Omega of Self-knowledge. Sofia, 2007 (in Bulgarian).

[Maas and Schneider, 2004] Maas, H.-G., Schneider, D.: Photogrammetric processing 360 panoramic images. GIM International 7(4), 2004, pp. 68-71.

[Mitov et al, 2011] Mitov, I., Ivanova, K., Depaire, B., Vanhoof, K.: ArmSquare: an association rule miner based on multidimensional numbered information spaces. Proc. of First Int. Conf. IMMM, Barcelona, Spain, 2011, pp.143-148.

[Mitov, 2011] Mitov, I.: Class Association Rule Mining Using Multi-Dimensional Numbered Information Spaces. PhD Thesis, Hasselt University, Belgium, 2011.

[Nedkov, 1998] Nedkov, S.: Museums and Museology. Sofia, 1998.

[OAIS, 2002] Reference Model for an Open Archival Information System (OAIS): Blue book. Consultative Committee for Space Data Systems, January 2002, http://public.ccsds.org/publications/archive/650x0b1.PDF

[Piccininno, 2009] Piccininno, M.: Analysis of the Europeana and Athena survey for the aggregators. www.athenaeurope.org/getFile.php?id=609

[Reg.6, 2009] Regulation Number 6 from 11.12.2009 for Creating and Managing of Museum Funds. Ministry of Culture of the Republic of Bulgaria – State Gazette 2/2010, (in Bulgarian).

[Sobota et al, 2009] Sobota, B., Korecko, S., Perhac, J.: 3D modeling and visualization of historic buildings as cultural heritage preservation. In 10th International Conference on Informatics, Herl'any, Slovakia, 2009.

[Somova et al, 2010] Somova, E., Vragov, G., Totkov, G.: Toward regional aggregator of digitalized cultural-historical objects. Proc. of National Conf. Education in Information Society, Plovdiv, Bulgaria, 2010, pp. 154-161 (in Bulgarian)

[Stoyanov et al, 2011] Stoyanov, K., Yordanova, B., Somova E., Totkov, G.: Two experimental virtual tours in the Plovdiv Ethnographic Museum. Proc. of National Conf. Education in Information Society, Plovdiv, Bulgaria, 2011, pp. 35-43 (in Bulgarian)

[Zaiane and Antonie, 2005] Zaiane, O., Antonie, M.-L.: On pruning and tuning rules for associative classifiers. In Knowledge-Based Intelligent Information and Engineering Systems, LNCS Vol.3683, 2005, pp. 966-973.