Database Design for a Cultural Artifact Repository

Todor Todorov
St. Cyril and St. Methodius University of Veliko Turnovo, and Department of Mathematical Foundations of Informatics, Institute of Mathematics and Informatics, Bulgarian Academy of Sciences, Veliko Tarnovo, Bulgaria

Galina Bogdanova
Nikolay Noev
Department of Mathematical Foundations of Informatics, Institute of Mathematics and Informatics, Bulgarian Academy of Sciences, Veliko Tarnovo, Bulgaria

Abstract
The purpose of this investigation is to present challenges and solutions in the process of designing database from a special data. This includes data gathered in the process of digitization and organization of cultural artifacts. We present some basic steps and algorithms for development of digital archives of information gathered from artifacts, analysis, optimization, and addition of metadata for indexing of digital data, compression and data protection, prevention of data loss, and design. We investigate the methods of protection with watermarking which can be used against illegal use of data. We illustrate our methods with two such concrete archives.

INTRODUCTION
Protection and preservation of cultural heritage is particularly an actual problem today. Digitization of analog materials and the creation of digital resources in the field of cultural heritage is a major contributor to e-Europe.

Main tasks in the research are:
- Development of audio, photo, and video archive with information collected by artifacts;
- Analysis and indexing the digital data;
- Design and maintenance of digital archive;
- Compression and optimization of the archive;
- Preventing data loss;
- Development of software for adding watermark against illegal use of data;
- Development of functions for creating, formatting, and protection of samples for additional applications and web sites.

We demonstrate our methods with two projects that we have managed to successfully complete all these tasks. The project “BellKnow” was to develop an archive containing detailed description of church bells, as well as to develop a digital archive (using advanced technologies) for analysis, reservation, and data protection. To accomplish this we have to document the main bells’ characteristics: design, form, type, geometric size, decorative and artistic scheme, weight, material, state, characteristics of chime, data about the producer and owner of the bell, and estimation of its historical value. So, in case unexpected circumstances destroy a bell, the archive will store the specific details to be investigated by different specialists. Objective of the second project “Balkan wars” was the digitizing of various documentary collections and artifacts related to the Balkan wars and the creation of electronic records and various public events and publications to promote in Bulgaria and Europe, the events during this historical period for the country. Project is implemented through digitization of historical artifacts and development of 3D digital model of some items, creating a common electronic archive of documentary heritage through digitization and processing of specimens found, construction of bibliographic and full-context databases, and presenting them on the Internet.

In the second section we consider modern methods of digitization and data collection. In the third section we provide an analysis of the objects in order to determine the metadata. The fourth section contains research of advanced technologies and methods for semantic data organization.

DIGITIZATION AND DATA COLLECTION
Modern technologies have changed the way one presented the information in the archives and made possible new services that were unthinkable a few years ago. Digitization is creation of an object, image, audio, document, or a signal (usually an analog signal) from a discrete set of its points or samples. The result is called “digital image,” or more specifically “digital images” of the object and “digital form of the signal.”
The tasks of digitalization can be synthesized in certain key areas:

- Retention of funds and records—many of digitalized objects are fragile; brittle structure is influenced by weather conditions and over time their digitization is the only hope for preservation;
- Simultaneous access to materials—most objects are subject to the digitization of rare and unique items of historical past and have a priceless value; the process of digitization will allow more users to touch them;
- Conservation funds in digital formats—archives, websites, and digital libraries;
- Strengthening international exchange and promotion;
- Providing access via computer networks—easy access to digital archives, and access to records of persons with disabilities;
- Provide new opportunities to work with digitalized materials funds—all the functionality is available to users of web space to be copied, multiplied, forwarded, etc., without jeopardizing its integrity and strength;
- Full text search—digital archives’ organization contributes to easier detection of the searched object among all the crowd, advanced search, as well as unification of search results;
- Classification of digital funds via metadata—entire photo metadata wealth funds may be accompanied by important information about copyright, creation date, identification number, etc.

Development of digital technology hit the storage and processing of information. Today, almost every unit of information is created digitally: digital photography, digital sound recording, digital communications, text files, videos, movies, multimedia presentations stored on digital media, etc. Storage of such digital multimedia data in digital archives was subject to the same challenges, such as archives, we know before the invention of digital computing devices.

DIGITALIZATION PROCESS

Digitizing or digitization is creating an object, image, sound, document, or a signal (usually an analog signal) by a discrete set of its points or samples. The result is called “digital presentation” of the object. Some methods of digitalization used in the project are:[2]

- TEXT, PHOTO (photographing documents)—Photographing documents and images is preferred because of the speed of digitization, accessibility, and relatively good quality of the received digital copies;
- TEXT, PHOTO (scanning with flat scanner)—The system actually represents a scanning device which is connected to a personal computer with appropriate software components for performing the scanning process;
- VIDEO, AUDIO (digitizing analog video and stream)—The process of digital video and audio recording is done with the system that operates in real time. This is analog playback device according to media digital recording. To perform digitization adequate equipment and appropriate software process management are needed;
- 3-DIMENSION IMAGING(3D scanning)—Process of 3D scan is performed with a suitable system; The object is scanned by all the visible sides and then specific software builds a full three-dimensional representation.

We use specialized hardware tools such as:

- Digital camera—Sony Alpha DSLR-A100;
- Video-camera—Sony HDR-SR8E HD AVCHD Camcorder;
- Audio-system—PULSE11 from Bruel and Kjer;
- 3D scanner—DAVID SLS system.

3D Scanning of Historical Objects

3D scanning is a new way for protection and promotion of cultural heritage material.[3–5] 3D digital recordings have more information about the volume, shape, and structure of the surface of the object than the classic photo shooting.

3D Scanning Technology

There are a variety of technologies to digitally recreate the shape of the 3D object. Classification of 3D scanners starts with a division into two types: contact and noncontact.

Contact 3D scanners use probe to touch the surface of the object to determine the distance and hence the shape of the object.

Noncontact scanners are divided into two main types: active and passive. Active 3D scanners use radiation of some kind of light, sound, ultrasound etc., and offset the effects of this radiation, which can determine the distance to certain points of the object or its form. Passive scanners capture multiple images of an object and based on them, determine distances and spatial location of individual points on the surface of the object.

We use “structured light” technology for 3D scanning.[3]
Structured light 3D scanning

The technology for 3D scanning using a structured light uses light with certain patterns and according to captured impact of this model it is possible to define its shape and spatial positioning of the various items of its surfaces. The positioning of the light source (projector), the camera, and the object is schematic in the form of a triangle defined by positional parameters. This example can be seen in Fig. 1. By projecting a series of known patterns onto an object and measuring the deformation of those patterns with a camera, the position in space of each pixel can be determined through triangulation calculations. Triangular calculations are performed by specialized software, taking into account the position of the light source (projector), camera trapping effects of lighting schemes, and reflection from the surface of the scanned object. These calculations are performed in parallel for each pixel within the camera producing a depth map or “point-cloud.” Fig. 1 shows vertical black and white lines designed by the projector and captured by the camera angle.

3D scanning system using structured light consists of a projector, a digital camera, and a computer system.

Basic methodology of 3D shooting is as follows:

1. Positioning System—is important and is a key point in the calculation of spatial location of points on the surface of the scanned object;
2. Calibration of the system—is also important for constructing a 3D digital model;
3. 3D scanning of the side/part of the side—the visible part of the object is scanned. To obtain a complete

3D digital model of the object, all visible parts should be scanned separately;
4. Digital processing—by processing the received 3D images and combining them into a comprehensive 3D model, one can obtain the appropriate digital formats.

3D scanning of historical objects

For 3D scanning of cultural objects, we use “structured light” technology with DAVID SLS-1 system.[3,6] Fig. 2 shows the system working. The “structured light” technology requires certain conditions of controlled light and background environment.

For an example of 3D scanning and digital processing, we will show part of a 3D image of the gun “Schwarz-loze.” The example shows six different 3D scans (Fig. 3) (a total of thirty four scans are required for complete coverage of the subject). Also is provided their consistent “assembly” (Fig. 4), a complete 3D image of the assembled six views with color separation for each individual scan (Fig. 5), and views from different perspectives (Fig. 6).

To build a complete 3D digital model of the example of Fig. 3, we captured 34 individual appearances of the artifact. After scan and construction of the 3D model, we add indexing textual data and the model is archived. We index the text content of the object: name, type, digitizer, date, author, archive/library number, innkeeper, brief description, copyright, etc., (necessary and explanatory data). This metadata is collected by different specialists from partner cultural and historical organizations: historians, librarians, etc., and processed by mathematicians and computer scientists. This is explained in details in.[3] Because of the specifics (file format—OBJ) of digital model we transform it to PDF or HTML format with embedded flash 3D images.
ANALYSIS, OPTIMIZATION AND INDEXING THE DIGITAL DATA

Essence of Metadata

Extensible Markup Language (XML) is a markup language that defines a set of rules for encoding documents in a format which is both human-readable and machine-readable.

The Extensible Metadata Platform (XMP) is an ISO standard for the creation, processing, and interchange of standardized and custom metadata for digital documents and data sets.

Metadata are text fields, built-in media files or additional text files (XML, XMP) for recording information on the nature of the digital resource. In other words, metadata is “data about data” describing any electronic or nonelectronic source of information on pre-established standard.

Metadata contribute to finding and sharing information. In other words, this is the last stage of information management. Once digitized (implemented in electronic form) and structured (arranged in a specific sequence and line), information is visualized in appropriate form. To fully justify the efforts of both phases scheduled for managing the information, it needs to be found and used by as many users. This is possible thanks to the metadata.

The presence of metadata with correctly placed points of connection ensures speed and accuracy of the application, and interactive user interaction.

Analysis of the objects in order to determine the metadata of selected artifacts collections and problem areas

In case of “BellKnow” archive we have the following data that should be indexed:\[1,2,7\]

- The main bells’ characteristics: design, form, type, geometric size, decorative and artistic scheme, weight, material, state, characteristics of chime, data about the producer and owner, estimation of its historical value;
- Digital photos and video recordings of the bells while being tolled;
- The frequency spectrum of the bells during a stroke;
- The bells’ frequency spectrum after transitive process;
- Charts representing the sound fade by time, sound stream, sound pressure, and other acoustic characteristics.

Modern Techniques for Intellectual Property Protection

With the development of digital technologies increasingly part of the audio, video, and any other information is available for fast, easy, and high quality copying. This fact entails the problem of protecting information from unauthorized distribution. Research in this area is considered in several aspects. One of the most important of these is steganography\[^{8,9}\]. Steganography is subdealing with the concealment of information; it hides the message which should remain hidden.

Like steganography, watermark protection carries hidden information. However, there are significant differences between the two techniques.

Digital watermark is visible or preferably invisible to the identification code that is permanently embedded into digital data and maintains a presence in them after extracting it.\[^{8}\]

Methods for image watermarking in the spatial region

In these methods, data are incorporated directly into the original image. The main advantage is that the key is not necessary to do any precondition transformations. The watermark is embedded by changing the
illumination or color components. The main disadvantage is the low resistance. An example of this method is the method of Kutter.\[^9\]

To derive the value of the embedded bit we calculate the difference between predicted and actual value of the pixels. The sign of the difference determines the value of the embedded bit. Extracting bits is done without the knowledge of the original message. The method is robust to filtering, JPEG compression, and geometric transformations.

**Methods for audio watermarking using low-bit coding**

By replacing the least significant bit of each sampling point by a coded binary string, we can encode a large amount of data in an audio signal.

The major disadvantage of this method is its poor immunity to manipulation. Encoded information can be destroyed by channel noise, resembling, etc. We improve robustness using by error-correcting codes.

**Visual watermarking**

Visible mark added in digital picture and video records (Fig. 7).

**Digital Warehouse**

Considering that there is a digital archive for unique Bulgarian bells, and there is lot of interesting information hidden in digital resources, we make an intelligent annotation of knowledge. A digital archive “BellKnow” is developed by using advanced technologies for analysis, reservation, and data protection, and it contains:

- The main bells’ characteristics: design, form, type, geometric size, decorative and artistic scheme, weight, material, state, characteristics of chime, data about the producer and owner, and estimation of its historical value;
- Digital photos and video recordings of the bells while being tolled;
- The frequency spectrum of the bells during a stroke;
- The bells’ frequency spectrum after transitive process;
- Charts representing the sound fade by time, sound stream, sound pressure, and other acoustic characteristics.

Organization of the “BellKnow” archive:\[^2\]

- Digital data:
  - More than 3000 digital records with added digital steganographic sign (invisible watermark);
Including photo pictures, video clips, and audio records;
- Technical data, historical references, passports, diagrams, etc.
- Tree file structure:
- Digital files format, parameters, coding;
- Specific signature for file name;
- Additional META textual data for indexing of media files:
  - Title (name of subject);
  - Creator (name of digitalizer);
  - Description (additional data);
  - Date (date of creation);
  - Type (type of media);
  - Format (file format, codec, and parameters);
  - Identifier (geographic coordinates);
  - Rights (owner of property rights).

Organization of the archive can be separated in the following steps:
- Create archive;
- Creating own functions—actions, macros;
- Adding objects to a digital archive;
- Converting file format, codec, and size;
- Adding metadata to objects in a digital media collection;
- Adding watermark—visible and invisible.

Documentation and Passportization

Based on these and other data we prepare a passport model for each test bells.\cite{1,2,10,11} Passports are summaries of all the information gathered about an object, in this case several sets of bells in one place. This includes photographs, historical reference, technical data, charts, and research done in the electronic version of the passport-embedded multimedia files with recorded audio and video clips (Fig. 8).

This passportization is a publication of all data collected from various studies in a document. The document includes sections in which information is divided into:
- Home page of the document and location of the bells;
- Page with a presentation of the draft study and identification of church bells in Bulgaria;
- History of the location of the bells (monastery, church, museum, etc.);
- Scheme of the bells of a given place.
- For each bell following data is included:
  - Picture materials;
  - Technical data shown in the next scheme (dimensions, location, type, material, year and place of creation, creator, etc.);
  - Art Design of the bell (historical data, captions, pictures, decorations, ornaments, etc.).
Data from sound analysis, supported by diagrams;[10] Digital recording—video clips and sound recordings.

This model of passport could be used to document other bells in the future (Fig. 9).

### SEMANTIC DATA ORGANIZATION

#### Semantic Web

The most widespread standards for the description of resources are Standard Generalized Markup Language (SGML), XML, Resource Description Framework (RDF), Web Ontology Language(OWL), etc. RDF is a framework for describing and exchanging data. It offers a modeland syntax for metadata, so that they can be used by independent components. At its core, RDF contains nodes and attached there to pairs of attributes and values. Nodes can be any Web resources (pages, servers, virtually all of which can be given Universal Resource Identifier [URI], or even other types of metadata). Attributes are properties of nodes and their values are either atomistic (text strings, numbers, etc.) or other resources or metadata. This mechanism allows to build boxes (labeled directed graphs), which could later be converted into XML. OWL is a Semantic Web language designed to represent rich and complex knowledge about things, groups of things, and relation between things. OWL is a computational logic-based language such that knowledge expressed in OWL can be exploited by computer programs, e.g., to verify the consistency of that knowledge or to make implicit knowledge explicit.

We make an experimental semantic annotation, based on the current W3C Semantic Web initiative standards[12,13] of the resources in digital archive of unique bells. We use the RDF data model, because it provides a model for describing resources of bells. Digital resources have properties (attributes or characteristics of bells). RDF defines a digital resource as any object that is uniquely identifiable by an URI. OWL is a Semantic Web language designed to represent rich and complex knowledge about things, groups of things, and relation between things. OWL is a computational logic-based language such that knowledge expressed in OWL can be exploited by computer programs, e.g., to verify the consistency of that knowledge or to make implicit knowledge explicit.

#### Linguistic Approaches of Semantic Research

The linguistic explanation is a very essential task in our days. Basically the semantic systems are built of different types of thematic dictionaries, which explain nature (give a meaning) of the information.[7,14]

Numerous ontology dictionaries with different functions are a basis of our experimental bell ontology system. These dictionaries are made using XML/RDF/
OWL technology to provide lexical explain and description of bell knowledge. Combination of descriptions, text explanation, and lexical rules provides basics of experimental bell ontology system.

The main tools to represent semantics are dictionaries, in which concepts are arranged in a subtype–supertype hierarchy, thus forming taxonomy.

**Frequency dictionary for bell science**

The main purpose of this project is to build a text-research system to create frequency dictionary from texts of bell knowledge descriptions. To construct this system are built a hierarchical datasets and WEB interface. The texts are divided into thematic headings (bell science domains).

The dictionary aims to cover big corpuses of texts that contain terms and conceptions in the field of Bulgarian bell’s knowledge. Frequency dictionary is designed for large arrays of such texts. Separate phrases and words are representatives of different domains. The dictionary aims to give information on how often a particular word or phrase is used in a particular corpus of texts.

Full text search could be made with the system. After the indexing of the data, searching of the appropriate index is saved in a catalogtable. The result shows information such as: in how many files in one particular area’s domain the number of words and phrases are founded. A hierarchical structure of data (tree) is used for organization of the information. The hierarchical structure of data has tables included for administration of categories and growing of the tree structure is allowed in volume and depth.

The system offers an easy and fast search, due to the hierarchy of the data. It enables introduction of many different domains and they do not influence the speed of searching. The individual tables contain only the names of domains, as well as their keys for organization the hierarchy. The help table contains all texts of all domains, organized with the help of indexes, which enables a fast access to the relevant texts and domains. There is an option for construction of a dynamic growing of the tree of tables in depth.

**Term dictionaries for bell science**

For the bell science domain, more suitable dictionaries are term dictionaries. The term dictionaries present the explanation of bell knowledge in a definite corpus of texts. It is considered that the facts in the term dictionary are enough. The term dictionaries give versatile information: text context of different terms, presence/absence of define terms, relations with other terms, lexical rules, and groups of different bell domains.

Depicted is a part of XML dictionary data (Fig. 10):

```
<!DOCTYPE bdict[
<!ELEMENT item (idbt, term, description, type, classOf, ref, transl, bnote) +>
<!ELEMENT classOf (#PCDATA)>
<!ELEMENT ref (href) +>
<!ELEMENT transl (bword) +>
<!ELEMENT idbt (#PCDATA)>
<!ELEMENT term (#PCDATA)>
<!ELEMENT description (#PCDATA)>
<!ELEMENT type (#PCDATA)>
<!ELEMENT bnote (#PCDATA)>
<!ELEMENT c1Of (#PCDATA)>
<!ELEMENT href (#PCDATA)>
<!ELEMENT bword (#PCDATA)>
]>
<bdict>
<item>
<idbt>1</idbt>
<term lang = "en">bell</term>
<description lang = "en">music device</description>
```
OWL ontologies

In recent years the development of ontologies has become very popular in different scientific areas. Ontologies have become common especially on the World Wide Web.

Ontology defines a common vocabulary for research- ers who need to share information in a domain. It includes machine-interpretable definitions of basic concepts in the domain and relations among them.

Sharing common understanding of the structure of information among people or software agents is one of the more common goals in developing ontologies.

There are a lot of definitions of ontology in the literature. Ontology is a formal explicit description of concepts in a domain of discourse called classes, properties of each concept describing various features and attributes of the concept, and restrictions on properties. If we consider ontology with a set of individual instances of classes, this results in a knowledge base.

Classes are one of the main parts of most ontologies. Classes describe the concepts in the domain.

A class can have subclasses that represent concepts that are more specific than the superclass. Properties describe features and attributes of classes and instances. In practical terms, developing ontology includes basic methodology of 3D shooting which is as follows:

1. Defining classes in the ontology;
2. Arranging the classes in a subclass–superclass hierarchy;
3. Defining properties and describing allowed values for them;
4. Filling in the values for properties for instances.

We can then create a knowledge base by defining individual instances of these classes filling property value information and additional restrictions.

The protégé platform

Protégé is a free, open-source platform that contains powerful tools to create ontologies. Protégé implements a rich set of knowledge-modeling structures and actions that support the creation, visualization, and manipulation of ontologies. The environment can be extended to work as a plug-in or as a Java-based Application Programming Interface (API) for building knowledge-based tools and applications.

The Protégé platform supports two main ways of modeling ontologies:

The Protégé-Frames editor: these are tools to create ontologies that are frame-based, in accordance with the Open Knowledge Base Connectivity protocol (OKBC). In this model, an ontology consists of a set of classes organized in a hierarchy, a set of slots associated to classes to describe their properties and relationships, and a set of instances of those classes.

The Protégé-OWL editor: these are tools to create ontologies for the Semantic Web, in particular in the W3C’s Web Ontology Language (OWL). “An OWL ontology may include descriptions of classes, properties, and their instances. Given such an ontology, the OWL formal semantics specifies how to derive its logical consequences, i.e., facts not literally present in the ontology, but entailed by the semantics. These entailments may be based on a single document or multiple distributed documents that have been combined using defined OWL mechanisms” (see the OWL Web Ontology Language Guide).

Protégé’s model is comparable to object-oriented and frame-based systems. It basically can represent ontologies consisting of classes, properties (slots), property characteristics (facets and constraints), and instances. Protégé provides an open Java API to query and manipulate models.

Using Protégé’s user interface one can create classes, assign properties to the classes, and then restrict the properties’ facets for certain classes. Finally Protégé is able to automatically generate user interfaces that support the creation of instances. For each class in the ontology, the system creates one form with editing components (widgets) for each property of the class.
BellKnow Ontology

Using the information of metadata annotation we define an ontological describing the Bulgarian bells. The basic schema of these digital resources is shown in Fig. 11.

This scheme shows the basic BellKnow ontology classes (bell, history information, technical data, location, media files, and audio characteristics) and their properties. Next we will consider basic OWL ontology components and how these components are realized in the BellKnow ontology in Protégé.

An OWL ontology consists of individuals, properties, and classes, which roughly correspond to Protégé’s framework objects:

- **Individuals**—represent objects in the domain that we are interested in. For example an individual is each different bell.
- **Properties**—Properties are binary relations between individuals. For example, the property hasLocation might link the individual Bell1AN to the individual StAleksanderNevskiCatedral, or the property hasBell might link the individual StAleksanderNevskiCatedral to the individual Bell1AN. Properties can have inverses. For example, the inverse of hasLocation is isLocationOf. Properties can be limited to having a single value—i.e., to being functional. They can also be either transitive or symmetric.
- **Classes**—OWL classes are sets that contain individuals. For example, the class Bell would contain all the individuals that are bells in our domain of interest. Classes may be organized into a superclass–subclass hierarchy. OWL classes are built up of descriptions that specify the conditions that must be satisfied by an individual for it to be a member of the class.

We start creating ontology using ‘Create New OWL Ontology’ from Protégé’s main menu. Next we could start building a class hierarchy using ‘Classes Tab’. The empty ontology contains one class called ‘Thing’. The class Thing is the class that represents the set containing all individuals. Because of this, all classes are subclasses of Thing. First we create BellKnow’s main class Bell. After that we could continue building the BellKnow
ontology hierarchy by adding classes and subclasses as shown in Fig. 11.

The next step is to create OWL Properties. There are three main types of properties: Object properties, Datatype properties, and Annotation properties (Fig. 12). Object properties are relationships between two individuals. Annotation properties can be used to add metadata to classes, individuals, and properties. Using the Datatype property, we can add a restriction to the class states that all individuals of the class must meet specific restrictions. Properties may be created using the Object Properties tab.

Example properties:

- An Object property linking the individual Bell1AN to the individual Bell1ANBeat (Fig. 13).
- A Datatype property linking the individual Bell1AN to the data literal 140, which has a type of an xsd:float (Fig. 14).
- An Annotation property, linking the class Bell to the data literal (string) Geogri Ivanov (Fig. 15).

We can also add additional Datatype and Range restrictions to the Properties. Finally we can use one of the key features of OWL-DL language reasoner. It helps to test whether or not one class is a subclass of another class and whether or not it is possible for the class to have any instances.

**Singing RDF Graph**

An RDF statement involves a name if it has that name as subject or object. An RDF graph involves a name, if any of its statements involves that name. Given an RDF statements, the Minimum Self-contained Graph (MSG)

contains that statement, written MSG(s), is the set of RDF statements comprised of the following:

- The statement in question;
- Recursively, for all the blank nodes involved by statements included in the description so far, the MSG of all the statements involving such blank nodes.

This definition recursively build the MSG from a particular starting statement; Theorems proved in\cite{16,17} show that the choice of the starting statement is arbitrary and this leads to an unique decomposition of the an RDF graph into MSGs.

In\cite{17,18} is proven that an RDF model has a unique decomposition in MSGs. The MSG definition and properties say that it is possible to sign a MSG attaching the signature information to a single, arbitrary triple composing it. Along with the signature, an indication of the public key to use for verification might be provided. By “attach” we mean by using a verification procedure. Using the same procedure more signatures can be attached to the same MSG either independently or “layered” thus providing a mechanism for countersigning.

**CONCLUSION**

We present some challenges related to the process of designing database from a special data.

In the survey are presented some algorithms for development of digital archives of information gathered from cultural artifacts. We investigate the methods of protection with watermarking which can be used against illegal use of data. All presented approaches are illustrated in two such concrete archives.

**REFERENCES**


