

Article

# On the Use of Geographic Information in Humanities Research Infrastructure: A Case Study on Cultural Heritage

Albina Mościcka <sup>1,\*</sup>  and Agnieszka Zwirowicz-Rutkowska <sup>2</sup>

<sup>1</sup> Faculty of Civil Engineering and Geodesy, Military University of Technology, 00-908 Warsaw, Poland

<sup>2</sup> Faculty of Geodesy, Geospatial and Civil Engineering, University of Warmia and Mazury, 10-719 Olsztyn, Poland; agnieszka.zwirowicz@uwm.edu.pl

\* Correspondence: albina.moscicka@wat.edu.pl; Tel.: +48-503-510-470

Received: 14 January 2018; Accepted: 12 March 2018; Published: 14 March 2018

**Abstract:** As an invaluable source of knowledge about the past, cultural heritage may be an important element of the humanities research infrastructure, along with other elements, such as spatial references. Therefore, this paper attempts to provide an answer to the questions concerning the ways in which spatial information can contribute to the development of this infrastructure and the aspects of storytelling based on cultural resources that can be supported by such infrastructure. The objective of the methodology that was used was to combine the aspects that refer to spatial information and cultural items into a single, common issue, and to describe them in a formalized way with use of Unified Modeling Language (UML). As a result, the study presents a proposal of the Humanities Infrastructure Architecture based on spatially-oriented movable cultural items, taking into account their use in the context of interoperability, along with the concept of creating spatial databases that would include movable monuments. The authors also demonstrate that the ISO 19100 series of geographical information standards may be a source of interesting conceptual solutions that may be used in the process of the standardization of geographical information that was recorded in the descriptions of cultural heritage items in form of metadata and data structure descriptions.

**Keywords:** information infrastructure; e-Science; spatial database; ISO 19100 series standards; metadata; interoperability

---

## 1. Introduction

### 1.1. Big Data on Cultural Heritage

Informing people with use of geo-visualized stories and supporting space-time inference requires a combination of different kinds of data, such as text, photos, audio and video materials, maps, etc. Their common processing, analysis and visualization is difficult but valuable, as it can lead to discovering new knowledge. This process is all the more effective and valuable, the more heterogeneous and comprehensive the data used. Therefore, the biggest challenges in using datasets lie in exploring huge amounts of various data, called big data [1]. Data is called “big” not only because of the size of the data volume, which is just one of several dimensions of such large data sets [2]. Big data management poses at least three major challenges, the so called Three Vs, suggested as volume, variety, and velocity [3], which have emerged as a common framework to describe big data [4,5]. Effective exploration of these Three Vs—high-volume, high-velocity, and high-variety datasets—requires modern forms and advanced technologies of data storage, dissemination, management, analysis and visualization [6]. Some further powerful challenges lie in the integration of big data originating from different sources [7], which can result in invaluable knowledge discovery.

Much of the big data remains in the hands of its creators, but the institutions that are involved in the recording of cultural heritage are becoming increasingly more often involved in capturing, storing, analyzing, and visualization of big cultural datasets [8]. National libraries, museums, and archives digitize physical collections on a large scale. Europeana alone [9] collects about 51 million works of art, books, films, and audio files from all over Europe, while the London Metropolitan Archives put online over 20 million parish records [10] and the US Library of Congress has 25 million digitized items, which represent less than 16% of the Library's total 160-million item collection [11]. The importance of cultural big data and the need to provide access to cultural resources to the widest possible audience resulted in defining the guidelines for Member States of the European Union on the policy on big data for culture [12].

Big data on cultural heritage is a rich source of information about the past, as it describes people, places, events, and at the same time providing material evidence of past times. We do not always realize that all of it is related to the places where people used to live and events happened. It refers to places that existed in the past, etc., which is why it is a great source for storytelling with the use of geographic space.

### *1.2. Humanities Research Infrastructure*

Effective use of big cultural data in geographic environment, that is, geo-visualization of storytelling, demands proper infrastructure, which is the cornerstone of big data architecture. Having the right tools for data storage, management, analysis and visualization, as well as integration, is crucial in any big data project [13]. This infrastructure is a mix of hardware, software, digital content, data and archives, as well as human resources, knowledge, and expertise. The core of the infrastructure is the idea of collaboration and sharing between and across communities in order to enable new forms of enquiry, as well as to generate and understand new research questions. It does not depend on whether it is the research data, computing power, or other resources that are shared [14]. In the discourse, infrastructure is also called cyberinfrastructure, which is defined as a layer of information, expertise, standards, policies, tools, and services, but also the less tangible layer of expertise understood as the best practices, standards, tools, collections and collaborative environments. All of them should be shared widely across communities of inquiry [15]. It should be noted that the above definitions are focused on infrastructure provided with the primary goal to support research resulting in the concept of research infrastructure. The European Commission stated that the term "research infrastructure" refers to facilities, resources and related services used by scientific community members to conduct top-level research in their respective fields. Research infrastructure may be "single-sited" (a single resource at a single location), "distributed" (a network of distributed resources), or "virtual" (if the service is provided electronically) [16].

Resources that are included in the infrastructure—also in the research infrastructure—include such humanity's resources as e.g., cultural heritage items. Andrew Prescott claims that "scientists want to map the Universe, humanities scholars want to map the universe in a single poem" [17]. Therefore, one can ask if the humanities really need an analogous infrastructure considering that "the single poem" does not require "big data" methods. It seems that we rather need to develop a different narrative about what infrastructure is for those working in the humanities and to define what "humanities research infrastructure" is. Anderson & Blanke claim that "we are at the very beginning of understanding what a humanities research infrastructure is and could be, and what technologies are the best suited for supporting and enhancing our research practices and processes" [14]. They argue that research infrastructures help disciplines to redefine themselves around a shared set of devices that support their research. According to Anderson & Blake, humanities research infrastructures have been theorized as digital ecosystems without a center and constituted through heavily interconnected online platforms. Thus, thinking of research infrastructures as digital ecosystems entails the commitment to identifying them as services that are built around communities. Infrastructures are then the sum and integration of

these services that are shared through a platform. Communities as crowds work together on a common goal and become the most important resource for the sustainability of the infrastructure [18].

In order to develop the humanities research infrastructure, one has to ask a lot of questions concerning the relationship between the humanities' communities and the digital space including the big structures within it, as well as about the potential benefits and threats that result from working with these new research infrastructure types. Probably the most important question refers to the potential solutions that can be proposed to develop humanities infrastructure in order to help discover new knowledge and reap the benefits. These benefits can be much bigger if infrastructure development and its use result from researchers' own practices. In such cases, they fill gaps in the existing solutions and solve the identified problems and difficulties. Infrastructure will be used and developed if it is an integral part of the research goals [19]. Moreover, the most successful infrastructures appear to be those that take the existing practices as their starting point, and then add something new [20]. This is especially important in infrastructures of archives in the library sector. They are focused on infrastructure for access, using digital technologies to enhance searches, and utilizing the linked data approaches to connect content across collections, and, in some cases, even across different institutions. These developments are not starting from scratch but are based on the existing metadata that are already created to match archive and library standards [19].

### *1.3. Aim of the Study*

Taking the above into consideration, the aim of the paper is to analyze the role of geographical information used in the humanities research infrastructure. The research is based on cultural heritage resources that contain an enormous amount of information about the past and fast-growing big humanities data, as well as their potential as a source of information that is used for storytelling in geographical space. Therefore, the research questions are as follows: how can spatial information contribute to the development of humanities infrastructure and what aspects of storytelling using cultural resources can be supported by humanities infrastructure based on spatial reference? The authors' experience in the area shows that humanities research infrastructure plays an important role in spatially-related research [21].

In this paper, the authors demonstrate that geographical information can be an important element supporting the integration and standardization in the area of cultural heritage, significantly improving the interoperability of big data on cultural heritage, and consequently, the visualization of stories in geographical space. Moreover, integrating dispersed and fragmented archives of relevance into a specialized area of research has the potential to stimulate new topics of enquiry and give rise to new research questions [22]. Answers to these questions can be found using spatial information, the contribution of which on almost every stage of work with cultural data means that it can be one of the essential elements of the humanities research infrastructure.

## **2. Materials and Methods**

The research methodology is based on previous research achievements in the field of use of geographic reference (information about location) as the most powerful integrator of data scattered across a wide spectrum of disciplines, formats, and languages [23], and the link between cultural heritage goods and geographical space in different semantic relations [24]. However, previous research in this area treated the above aspects separately, independently and it failed to provide any formal definitions. The aim of the methodology that is used in this article is to combine the aspects related to spatial information as an integrator of different kind of data with spatially-oriented cultural goods into a single, common issue and to describe all the necessary aspects in a formalized way with the use of the Unified Modeling Language (UML) (standard OMG—Object Management Group), which will enable the implementation of the proposed solutions in any tool environment. The research focuses on a specific type of cultural heritage—movable monuments. These monuments can be easily moved

from one place to another, so their reference to geographical space is not unambiguous, as it can be changed over time.

The above objective is realized by developing the Humanities Infrastructure Architecture proposal based on spatially-oriented cultural items. The formal description of selected technological aspects of the Humanities Infrastructure Architecture was made using the Unified Modeling Language [25], including:

- use case diagram,
- activity diagram, and
- class diagram.

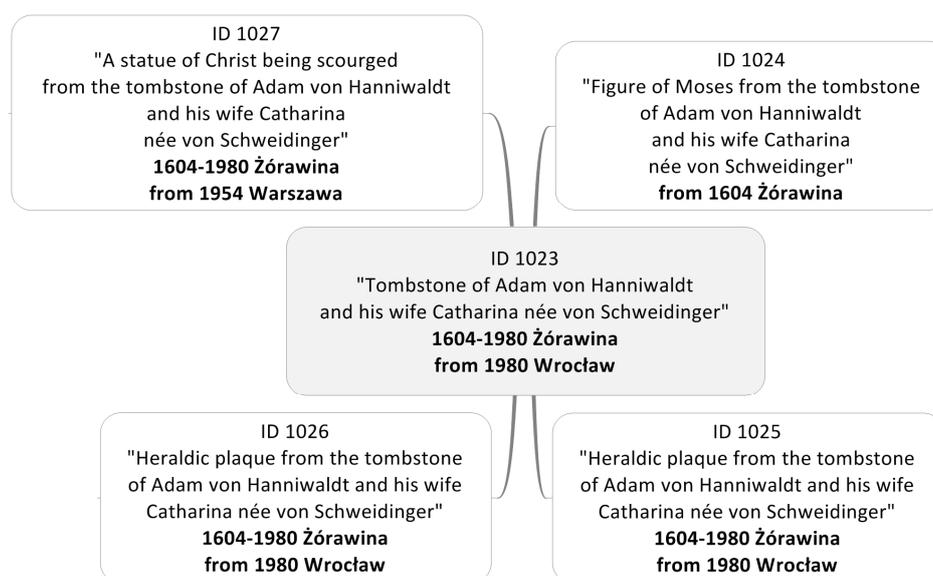
Considerations on the use of geographic information are based on selected ISO 19100 standards. Standard 19109 [26] describes the rules of application schemas, 19103 [27] describes the UML profile in the field of geographic information, while 19107 [28] describes aspects of geometry and topology.

The Object ID [29] and EAD (Encoded Archival Description) [30] standards and ontology of the CIDOC Conceptual Reference Model (CRM) [31] were used to develop the concept of structure of spatial database for monuments. An exemplary spatial database for movable objects was developed based on the UML application scheme in accordance with the MDA (Model Driven Architecture) methodology [32] and support of the Enterprise Architect, and it was implemented in Microsoft Access and the GIS GeoMedia Professional 6.1 environment. In addition, the Erlangen CRM/OWL, the ontology for the Semantic Web—an OWL-DL 1.0 implementation of the CIDOC CRM [33] was used.

### 3. Case Study

The study was supported by 2 datasets of digital copies of cultural heritage items.

The first dataset is the collection of digital copies of works of art documented in Object ID metadata standard. This dataset included ninety-eight works of art from the Holy Trinity Church in Żórawina (near Wrocław, Poland). They represent almost all artistic periods—from Gothic to Modernism. They were originally created as part of the church's interior design, but now most of the objects are dispersed and held in various museums and churches—mainly in the Silesia area, but also throughout the territory of Poland. Additionally, nine immovable monuments that were related to these works of art were taken into consideration. They are mainly places of former or current storage of works of art from Żórawina. An example of a monument from Żórawina and its scattered parts is shown in the Figures 1 and 2.

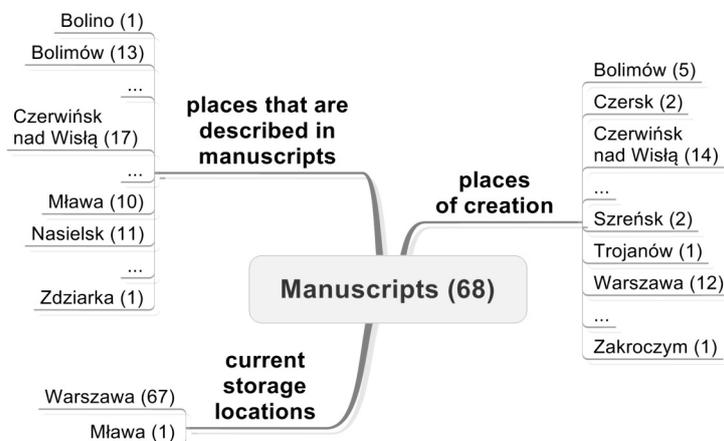


**Figure 1.** An example of a monument from Żórawina and its scattered parts (source: own work).



**Figure 2.** A statue of Christ being scourged from the tombstone of Adam von Hanniwaldt and his wife Catharina neé von Schweidinger: as the part of the tombstone in Holy Trinity Church in Żórawina in the past (a) and in the National Museum in Warsaw at present (b) (source: [34]).

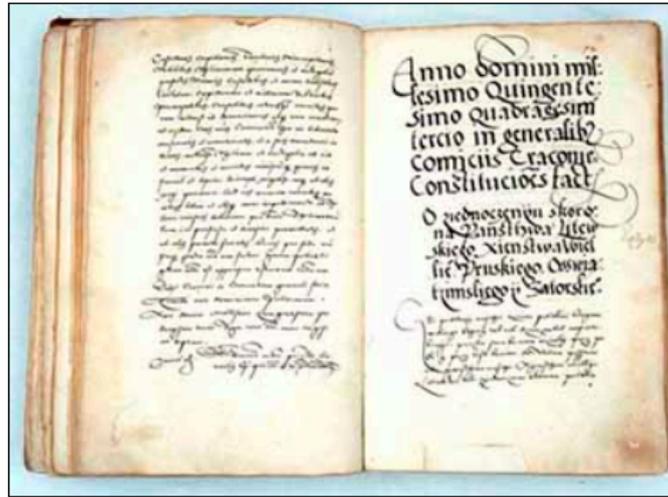
The second dataset is the collection of archival documents prepared in EAD metadata standard. This dataset included one hundred-thirty-one digital copies of archival documents, whose originals are stored in seven different Polish cultural institutions located all over the country, including five archives, one library, and one museum. Archival documents represent the digital collection “Cities in archival documents” of the Central Archive of Historical Records in Warsaw, which include documents referring to cities that used to play important roles in the past, but have lost their leading role since then. The test dataset concerns 6 towns: Bolimów, Czerwińsk, Latowicz, Mława, Szreńsk, Nasielsk (located in the Masovia region in central Poland) and contains various types of documents: manuscripts, old prints, drawings, maps, etc. created in period from prehistory to mid-20th century. An example of spatial references of the manuscripts is presented in the Figure 3. An example of a description of the manuscript is presented in the Figure 4.



**Figure 3.** Spatial references of the manuscripts (source: own work based on the digital collection “Cities in archival documents”).

### Parliamentary constitution of 1543

From the collection of the Central Archives of Historical Records in Warsaw (AGAD)



1543, Krakow

The oldest parliamentary constitution written in Polish during the session of Sejm that took place in Krakow in 1543. It concerns unifying Lithuania, the Royal Prussia, Oświęcim and Zator with the Crown.

**Description:** Original, manuscript, in Polish, 21 sheets of paper sized 215x325 mm, the part of book bound in gilt, embossed leather.

**Collection:** The Central Archives of Historical Records, Metryka Koronna 65 (Register of the Crown 65), leaves 52 – 72.

[Find out more in Polish](#)

*The Central Archives of Historical Records in Warsaw*

**Figure 4.** Example of a description of the manuscript (source: The Central Archive of Historical Records in Warsaw, formerly presented on [www.poland.pl](http://www.poland.pl)).

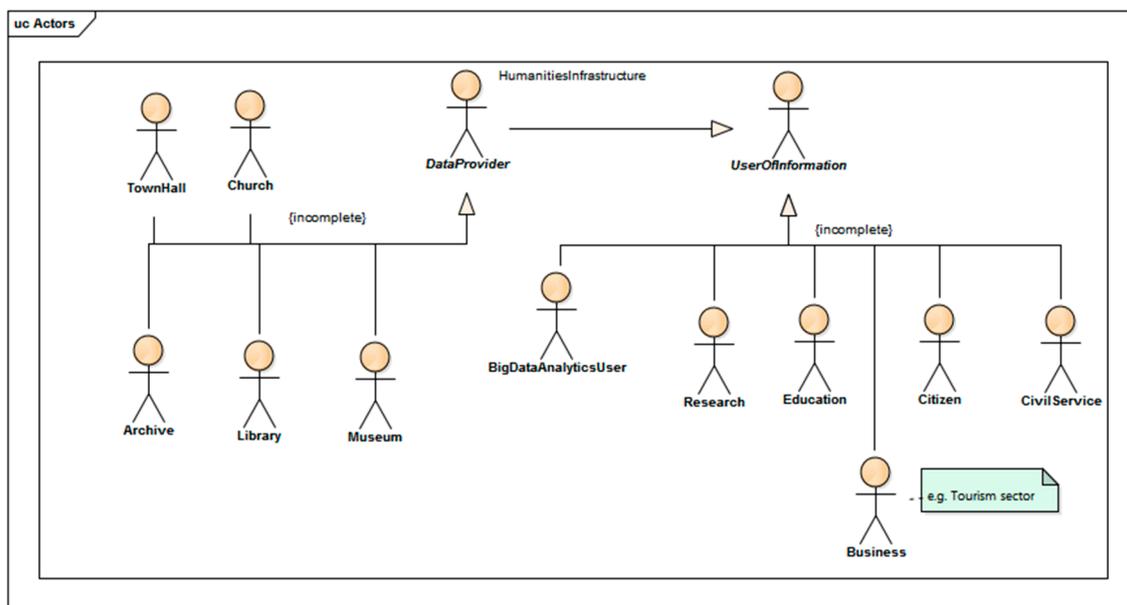
## 4. Results

The information needed for interdisciplinary research is typically distributed over a vast range of domains, formats, and languages, reflecting the many different perspectives to be considered [35]. The same applies to movable heritage, collected in thousands of cultural institutions and databases. Information and communication technology influences the development of means of cultural heritage data presentation, access and exchange, as well as technological solutions.

Humanities infrastructure is a complex structure consisting of the following components: descriptions of collections (metadata), databases, access points to resources (including web portals), procedures, standards, and (Figure 5) users, operators, and data providers.

The spatially-oriented Humanities Infrastructure Architecture is very diverse in terms of technology. Collections that have not been digitized can be distinguished. The translation of analogue data into digital formats provides not only the possibility for better protection and the management of huge amount of data, but also focuses on interdisciplinary challenges in using this information effectively. Some of these challenges lie in the effective use of cultural heritage resources, both as a value in itself and as a data source for a number of new research projects. Resources collected within the infrastructure are part of the Big Data trend—they are extensive datasets, they are characterized by volume, variety, velocity, and/or variability, and require a scalable technology for efficient storage, manipulation, management, and analysis. Digitization can take place at the level of individual

institutions collecting resources: e.g., libraries, archives, and museums. IT systems for resource management at the level of local institutions can also be distinguished. More advanced solutions take the form of federated systems in which local files are aggregated as single access points. Referring to the Semantic Web idea, one can indicate the semantic integration of heterogeneous sources in the form of a knowledge base (KB).



**Figure 5.** Data providers and users of Infrastructure (source: own work).

The issue of humanities infrastructure is considered in the following subsections in two dimensions:

- the vision of humanities infrastructure using spatial information as the integrator of different data in the area of cultural heritage, and
- the issue of ensuring interoperability between individual infrastructure elements.

#### 4.1. Use of the Spatially-Oriented Humanities Infrastructure

Integration and aggregation of information systems and databases that co-create the Humanities Infrastructure with geographic information can be considered in several perspectives.

From the point of view of institutions that are collecting information on monuments, it is an important issue (Figure 6) to prepare databases that would become a part of the developed geographical information system (as thematic data). Monuments descriptions (metadata) could be essential for creating database structures. A geographic information system that integrates thematic data with such reference data as: boundaries of administrative units or geographical names, provides an opportunity to prepare various forms of visualization of data on monuments, such as maps. Access to data can also be obtained via a web portal that can be built in a variety of technologies.

From the perspective of users of historical monument databases, the need to present monuments may refer to (Figure 7) single artifacts, groups of objects, or entire collections. Single artifact is understood as a single piece of movable monument, which is treated as a monument itself. It is an independent work of art. Several artifacts can be associated with each other, e.g., being parts of the equipment of a church, created at the same time and constituting a coherent artistic whole. In such a case, they constitute a thematic collection. In such collection, a single artifact can be treated as a part of the collection. There can be also a group of objects, which are not a coherent artistic whole, but are grouped according to any rules e.g., the same image topic or place of storage/origin. The spatial

reference for monuments may refer to the place of its creation, the place of storage and places of geographic references. According to the selected location, one can indicate artifacts that are belonging to different collections, collections, as well as various types of monuments.

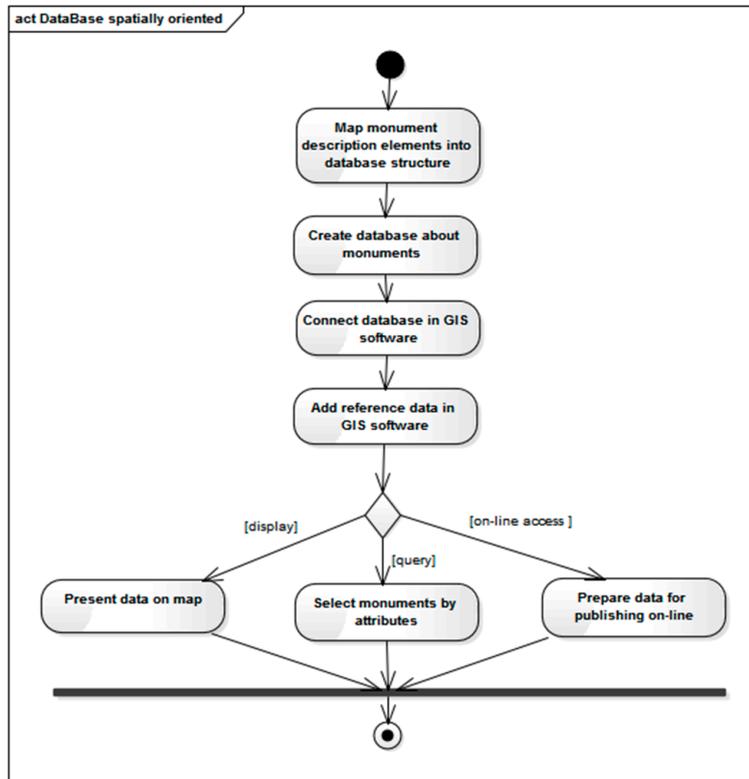


Figure 6. Development of spatial database about monuments (source: own work).

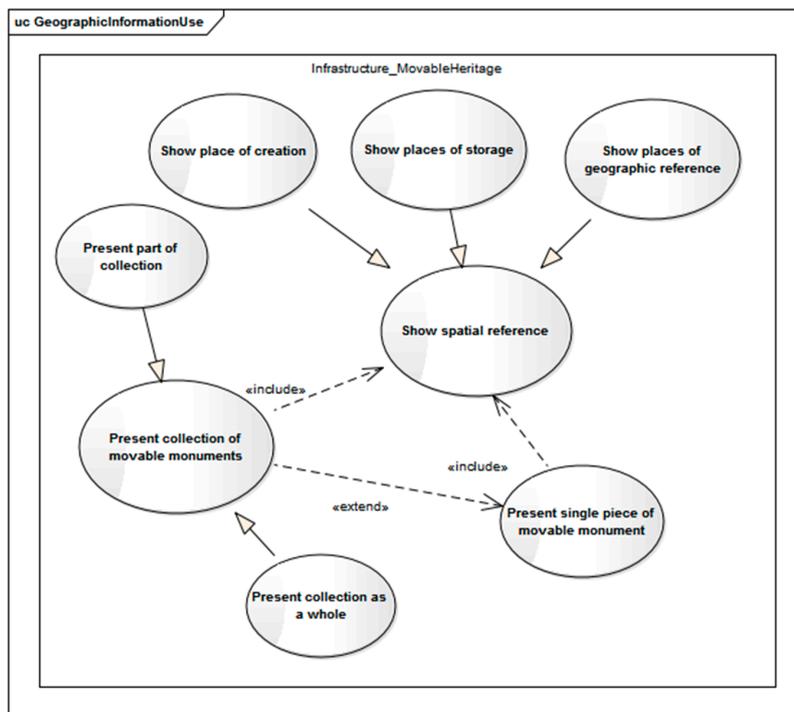


Figure 7. Presentation of different spatial relations of monuments (source: own work).

Figure 8 shows the presentation of monuments on the map, as well as the results of the attribute queries concerning the places of creation and of actual storage of the data sets discussed in the Case Study section. The concept of access to monuments through the geoportal is shown in Figure 9.

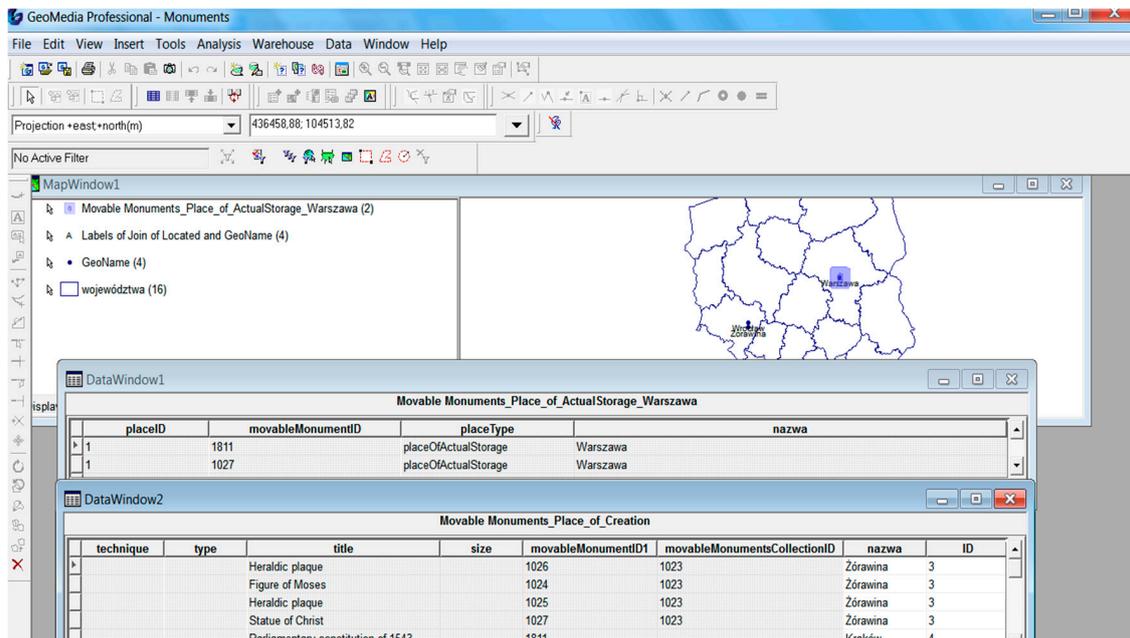


Figure 8. Presentation of results of monument selection in the table and on the map (source: own work).

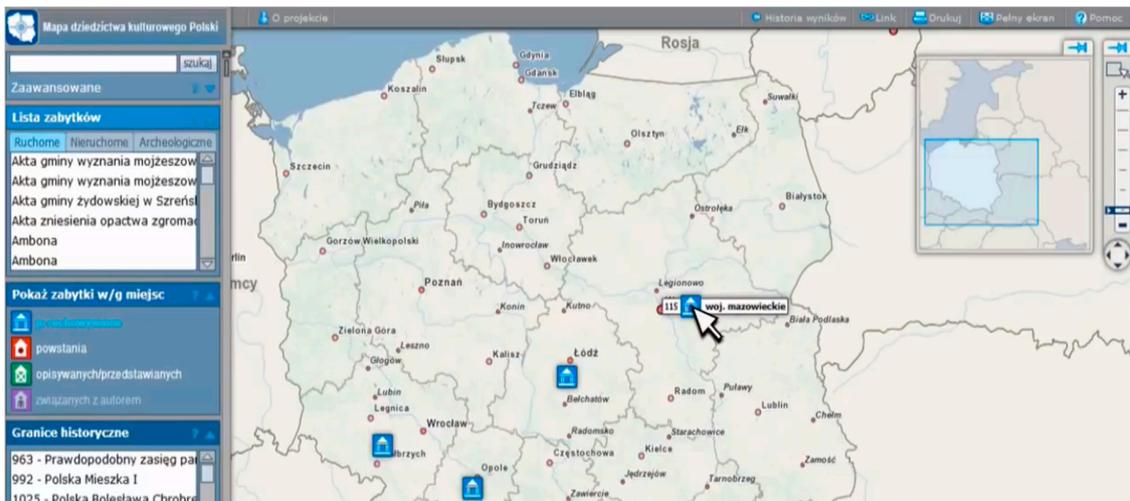


Figure 9. On-line access (geoportal) to the database about monuments with spatial reference (source: own work, formerly presented on [www.geoheritage.polska.pl](http://www.geoheritage.polska.pl)).

#### 4.2. Interoperability Assumption in the Humanities Infrastructure

Digitization of movable heritage data is common, but many different types of database management systems (DBMSs) and information systems that are used by various institutions can be distinguished. In the analyzed testing datasets, works of art described in Object ID standard are an example of a category in the architecture of the infrastructure modeled as business system with metadata, while documents that are described in EAD standard were collected in database with metadata available via a web portal. One of the challenges is to support data integration and exchange in this area. The solution is interoperability.

Interoperability [36] is the ability of a system or system component to provide information sharing and inter-application co-operative process control. Interoperability refers to the ability to find information and processing tools when they are needed, independently of their physical location. Moreover, interoperability refers to the ability to understand and employ the discovered information and tools, no matter what platform supports them, whether local or remote.

There are three levels of interoperability [37]: systems, data, and institutional.

Interoperability between systems has several aspects. First of all, network protocol interoperability is required for basic communication between systems. This communication occurs on two levels. On the higher level, there is the communication between applications. The lower level refers to the transmission of signals. File system interoperability requires that a file can be opened and displayed in its native format in another system. This includes interoperability for transferring and accessing files, as well as naming conventions, access control, access methods, and file management. Search and access databases provide the ability to query and manipulate data in a common database that is distributed over different platforms. Interoperability challenges include the location and access to the stored data. There is also application interoperability that refers to the ability for different applications to use and represent data in the same manner. In the analyzed case study, metadata were originally saved in DTD (document type definition) or in text files (works of art). The need for interoperability resulted in developing them into xml files.

At the data level interoperability means a common (interoperable) interpretation of the semantics of the data. Semantic interoperability refers to applications interpreting data consistently in the same manner in order to provide the intended representation of the data. In currently used standards for the description of cultural goods, the same information is recorded in different metadata elements. Moreover, information that is saved in one metadata element in one standard, can be dispersed in several metadata elements in another standard. For example, information about materials and techniques in the Object ID standard is saved in Materials & Techniques. The same information can be found in at least two EAD metadata elements: in Genre/Physical Characteristic and in Physical Description. Furthermore, in CIDOC this information is saved in: E57 Material, general techniques as instances of E55 Type, and specific techniques as instances of E29 Design or Procedure.

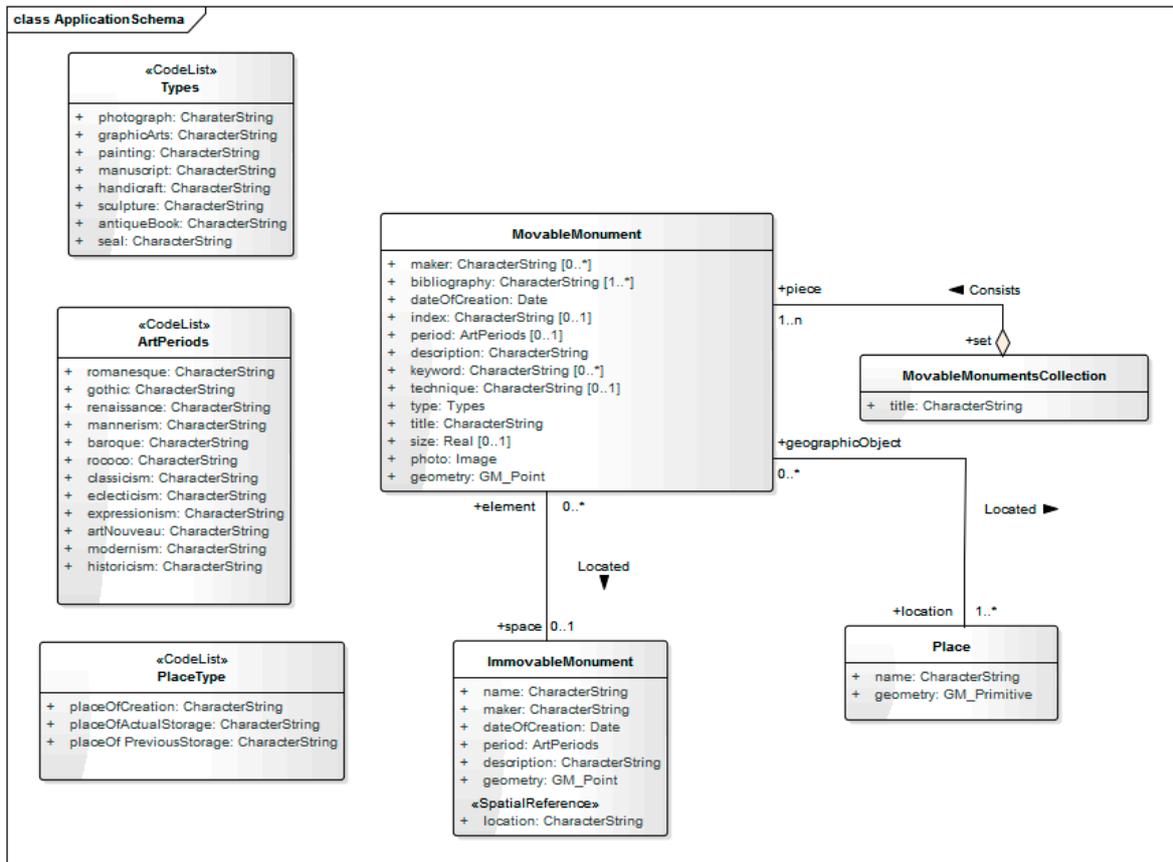
The institutional interoperability relates to laws, regulations, agreements, people, etc.

Interoperability of spatially oriented movable monuments systems concerns the following issues: dispersion of objects belonging to the same collection (Figure 1), different structures of descriptions of the objects from the same collection (data and metadata structures) and the spatial and non-spatial integration of objects for the same geographic reference (Figures 1 and 3, see Warsaw, Poland). One of the problematic issues is the description of objects of the same type using different standards of description, depending on the classification of a given object (i.e., as an archival object, a work of art or library resource) resulting from the current location (institution) of its storage. For example, an archival document can be described by archives in the EAD standard, and through a digital library in the Dublin Core standard. Integration of monuments using spatial information requires the unification of the way spatial information is stored in metadata. Currently, each institution describing monuments applies its own rules or does not apply any rules at all. This also applies to the test dataset. Due to the lack of common rules, the Central Archive of Historical Records in Warsaw (an institution managing archives at the national level) describes documents in a slightly different way than its regional branches. This results in situations where the same place can be saved in metadata by using contemporary names, historical names or names from different languages. This, in turn, leads to difficulties in obtaining full and unified information about specific sites associated with monuments, and also significantly hinders the transfer or even makes it impossible to transfer the information between systems or institutions.

The following subsections focus on data level interoperability assumptions, taking into consideration selected technological context and foundations.

### 4.2.1. Spatial Database Structure

Creating spatial databases about monuments (Figure 6) requires the development of database structures. The structure can be expressed using the language of the UML conceptual scheme (ISO 19109 and 19103), as application schemas (Figure 10). This approach allows for an unambiguous tool-independent description of semantics, thus enabling the implementation of semantic interoperability assumptions.



**Figure 10.** An Unified Modeling Language (UML) class diagram describing the concept of spatial database structure for movable monuments (source: own work).

Referring to the MDA methodology (MDA OMG), the UML application scheme can be mapped to various platform specific models (PSM), and then transformed into a variety of physical models. In the case at hand, an essential PSM is an SQL script defining tables in a relational database and implementations, such as e.g., Microsoft Access (Figure 11) and GeoMedia Professional 6.1, or PostgreSQL/PostGIS and QuantumGIS.

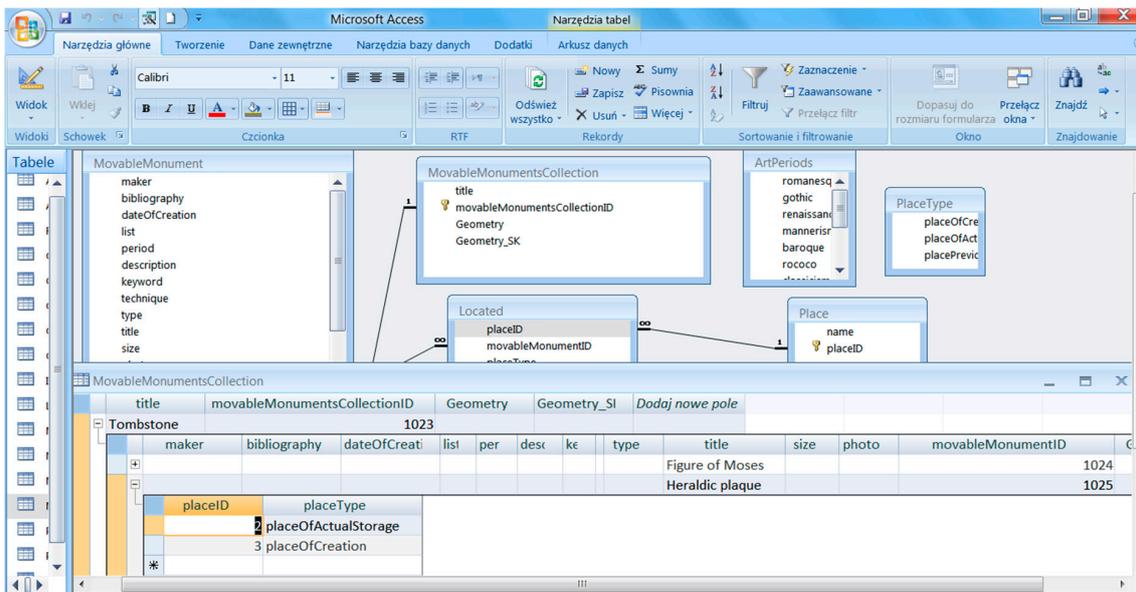


Figure 11. Database structure implementation in Microsoft Access (source: own work).

#### 4.2.2. Ontologies for Semantic Web

At present, the Internet is of particular importance as it provides a variety of information and communication facilities. But, in addition to the classic “Web of documents”, the paradigm of a complex global knowledge base, known as the Semantic Web [38], is also starting to make inroads into the area of cultural heritage. Standards, such as Resource Description Framework (RDF) and Web Ontology Language (OWL), are aimed at information integration.

The Erlagen CRM/OWL, the ontology for the Semantic Web, is an OWL-DL 1.0 implementation of the CIDOC CRM. The whole process of knowledge modeling (Figure 12) for the two databases that are presented in case study section was supported by the Protege tool. In addition, the Geonames ontology was used.

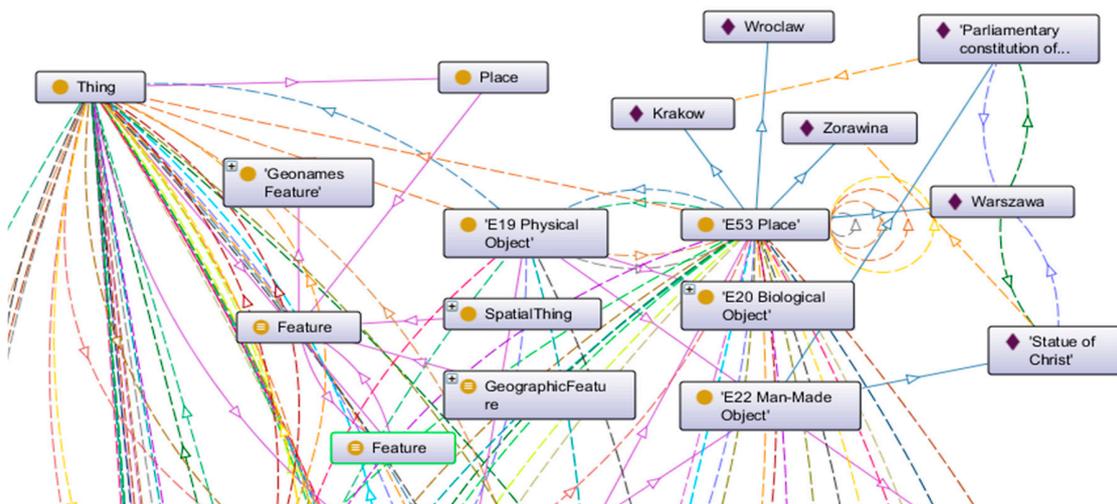


Figure 12. A sample of the ontology’s graph representation of the example datasets (source: own work).

### 5. Discussion

Sections 4.1 and 4.2.1 provide certain proposals that may be applied to databases that are maintained by individual institutions, information systems for resource management,

and interoperability between heterogeneous data sources and the creation of federated systems by aggregating local collections in the form of single access points. Section 4.2.2 addresses the issue of data source integration based on semantic technologies and the creation of knowledge bases.

Inclusion of data collected by cultural institutions in the humanities infrastructure, as well as its enrichment with spatial reference extends the use of movable monuments as a source of knowledge about the past, thus providing the answer to the first of the research questions posed. The use of spatial information often allows for a new way of interpreting data, which, when enriched with geographical location, allows for the extraction of new knowledge that used to be inaccessible, e.g., due to the lack or incompleteness of data. Geographical locations can fill these gaps, making it possible to acquire knowledge that was not directly accessible, which can only be obtained by interpreting existing data in the geographical space. What is more, the integration of heritage resources from various sources enables their comprehensive exploration, reaching all possible data on the topic that interests us. Spatially-oriented humanities infrastructure could derive from the area of geographical information systems, standards and spatial data infrastructure considering such issues as metadata, spatial data structures and visualization, interoperability, as well as web services. Hence, this study develops the issue of movable heritage spatial reference in the context of database structure design and implementation, as well as spatial data presentation on the map. The concept of spatial database structure for movable monuments that are proposed in this study has been developed based on the Object ID and EAD standards and it is intended to be a proposal addressed to various institutions dealing with management of data referring to monuments. The description in UML language and the reference to the principles of schema construction according to the rules given in the ISO 19100 standards enable the use of the MDA methodology and preparation of many different implementations depending on the solutions that are used in various institutions. The concept can be developed and expanded, depending on the needs or requirements of individual institutions or projects. In the UML application scheme, the reference to the geographic information is expressed by the attribute describing the geometry of the object presented on the map (MovableMonument class, ImmoveableMonument class, Place class), as well as the association with the Place class (MovableMonument). Both in the spatial database and on the map, individual places are identified by coordinates. Descriptions of movable monuments and the location of the objects are referred to the map layer "Located" table. This table is created automatically during the implementation of the UML application scheme, which links a given cultural heritage artifact with a given place and at the same time indicates the meaning of this place (PlaceType), classifying it as either the place of creation, place of actual storage, or place of previous storage. This is due to the fact that, for any given object, the mobile heritage object may be associated with many places, and a given place may be associated with different objects.

The utility and performance of such technologies as RDF and OWL have recently been investigated in the context of large-scale projects, such as the Europeana Digital Library. The process of building the knowledge base following the ontology life-cycle consists of four phases [39]: (a) knowledge specification (determination of the scope, aims, taxonomies, main and sub-concepts, and contents); (b) KB conceptualization (establishing relations among collected concepts); (c) KB formalization (transferring the created collection into a computer program); and, (d) knowledge reuse and exploitation (representation of the previously obtained knowledge for further applications, including accident investigation). In this study the first three phases are under investigation in terms of types of the movable heritage presented in the case study. The knowledge specification has already been defined for the test dataset in Object ID and EAD metadata standard and the assumed target taxonomy is CIDOC CRM. Description of test examples in the Erlagen CRM/OWL ontology required the mapping of information categories (metadata elements) in the Object ID and EAD standards, on the classes and properties used in the CIDOC CRM model. The description of particular information categories that exist in Object ID is much more complex in CIDOC CRM, and it requires providing direct reference to the object (instance) that is being described. This is due, among others, to the distinction of many sub-categories of object types in CIDOC CRM. In the phase of the KB conceptualization the role

of spatial relations is particularly emphasized by the authors. They can be supported by ontology, like Geonames and WGS84 Geo Positioning. Geographic information helps to integrate sources by using geographic location, which has remained unchanged for centuries. The location of the place in the geographic space is constant, independent of borders and political affiliation. Therefore, an information system based on spatial information (GIS) plays a very important role in the study of the past, as well as in the integration of cultural heritage.

The research was carried out on two metadata standards—Object ID and EAD—but the proposed solutions can also be used for other standards, including, among others, Dublin Core, which is used by digital libraries to describe many cultural works. Both the principles of database development and the presentation of various spatial references of monuments are very similar in this standard and it should be included in further research.

The effective maintenance and use of cultural heritage metadata requires a number of technical solutions related to the way of writing values of their metadata, as well as searching for datasets, including data related to geographical information. These solutions include applications both for the cultural heritage users—metadata catalogues which can be linked with a map viewer to present the results of datasets search, and for institutions that are responsible for metadata—metadata editors and validators. From the point of view of metadata management the important issue is to adopt rules of unifying recorded data values, which may then be implemented in the metadata applications. First of all, code lists and thesauruses should be developed. A code list is a list of allowed values that can be entered as the value of a given metadata element. Such lists can be created for e.g., object types. The lists limit the number of the allowed values for each of the metadata elements, and make it impossible to enter incorrect data. Furthermore, it would be necessary to develop and implement a thesaurus, including thesauruses for geographic information such as gazetteers. These should be gazetteers using geographical names that existed in different historical periods, together with the coordinates of locations. Such historical names may no longer be in use or can exist in different language now but they are often used in heritage resources metadata as used in the time when the goods were created. The best solution would be to use temporal gazetteers, which would provide the former counterparts and its form in other languages (e.g., in the language of the country where the town was located before the border was changed) for any given name (e.g., of towns). The way of recording coordinates should be based on ISO standard rules.

Regardless of the direction of change, it is imperative to work out and implement specific recommendations (and even requirements) for people preparing metadata of cultural resources. Implementation of such solutions into practice requires the preparation of detailed step-by-step instructions, as well as conducting specialized training to help people understand the principles of description and the methods of their implementation. This would create the solutions that would allow for different institutions to describe cultural resources, and consequently to exchange them in the same ways.

The humanities infrastructure analysis can also be carried out from the perspective of the big data phenomenon. For people that are involved in finding lost works of art, art collectors seekers of data sources about the past of a specific place, infrastructure is a huge source of data and knowledge in various forms and structures, which must be properly processed and prepared for the users' own needs. This data also requires the use of appropriate methods of analysis and visualization. There are also geoinformation methods of analysis, e.g., attribute and spatial queries performed in relational databases using SQL, as well as visualization, e.g., in the form of maps (static, animated, interactive).

## 6. Conclusions

The goal of the paper was to analyze the role of geographical information use in the humanities research infrastructure. The idea of using spatial information as an integrator of spatially related cultural resources has been presented. It was supported by a case study based on 2 collections of cultural heritage items: works of art documented in Object ID metadata standard

and archival documents documented in EAD metadata standard. An attempt was made to formulate an architectural outline of the humanities infrastructure, indicating its diversity in terms of technology. Methods of creating and the scope of spatial databases on movable monuments, as well as the possibilities of using spatial information in connection with cultural heritage objects were presented in a methodological approach using UML and from a practical point of view. Attention was also paid to the need to ensure interoperability within the infrastructure in relation to the selected technological possibilities.

A methodology for creating spatial databases on movable monuments was presented along with certain aspects of the use of spatial information in cultural heritage. The results add to examples of ISO 19100 series (e.g., [40]) use in the area of cultural heritage demonstrating that they are a source of the interesting conceptual solutions that can be used in the process of standardization of movable cultural goods' data structures.

The study contributes to current achievements by attempting to describe what humanities infrastructure is from a technological perspective and identifying user groups and data providers on cultural heritage. The discussion on how to use the infrastructure has been extended to cases that are related to the linking of monuments with spatial information and the use of spatial databases.

As it was presented in this study, a few methodological solutions derived from the geographical information science could be essential for the humanities infrastructure. However, taking into account the thematic scope of documentation, as well as the fact the standards or specifications are written by experts for experts [41], it is essential both to popularize them among developers of the humanities infrastructure, when considering different forms of knowledge transfer (e.g., workshops, e-learning courses, seminars) and the presentation of topics (e.g., tutorials, step-by-step instructions, "best practices", or case studies descriptions), and also to promote the elements of the infrastructure (e.g., metadata catalogues, (geo)portals) among users by e.g., social media or websites of institutions providing data on cultural heritage (short presentations, news, user's tutorials, forum).

The proposed vision of the humanities infrastructure is based on existing solutions both in humanities research and in geoscience. The contribution of the presented research to the development of the humanities infrastructure is the proposal of the way to connect these two areas of science. This process concerns two aspects—the method of development of spatial databases about monuments as well as the integration of spatially referenced heritage. It is based on using semantic interoperability rules that are used for spatial information integration in cultural heritage resources. It results in the possibility of formulating new research questions based on spatial information and providing new knowledge that would be impossible to be discovered without the presentation and analysis of cultural resources with this additional parameter—spatial reference. Research questions can be typical attribute or spatial questions but related to the historical data as well as their resource (e.g., which documents describe Warsaw and where they are stored now, which villages on the cross-border area belonged to our pre-war voivodship before the war). The most interesting aspect seems to be the exploration of knowledge based on the combination of historical information from archival documents and spatial locations. New knowledge can be provided in the event of lack of the data about some area. Missing information can be "interpolated" on the basis of data related to the neighboring area. For example, if we have archives confirming that almost all of the villages on the analyzed area have been visited by the tax collector, and we do not have documents about collecting taxes in a village nearby, it can mean either that taxes were not collected or maybe only that the documents with the information about this specific village are missing. Using other archival documents, containing data about flood at the time of taxes collection, and the knowledge that the omitted village lies on the other side of the river, we can suppose that tax collector could have problems with crossing the river and collecting the taxes. The other example of knowledge exploration can be related to the prediction of random events based on long-term data analysis (several hundred years old instead of decades). There are some areas where the cycle of repeating the flood is longer than the life span of one generation. This results in the belief that there are no floods in a given area. People build houses there, and then the flood comes,

because it happens in an 80-year cycle. It can be predicted on the basis of spatially-related historical data. Similar application is related to the spatial conditioning in conducting warfare, conditions of spatial development of cities, as well as capturing various types of anthropogenic processes, such as changing settlement areas or communication routes. Moreover, the sources of historical data exist in different semantic references, which broaden the scope of use and application of the possible analysis. All of these aspects confirm that spatial information contribute to the development of humanities infrastructure.

The humanities' infrastructure based on spatial reference supports two main aspects of storytelling: data access and visualization. The proposed solutions integrate various types of data (text, photos, maps, audio) with data from different sources. This integration gives the possibility to access all such data from one common platform. As a result, users can find resources distributed all over the world. Geo-visualization based on such data offers the possibility of telling more complex stories, joining various—spatial and non-spatial—aspects of the knowledge about the past. Moreover, it makes it possible to present data that was originally non-spatial as well as distributed around the world in a single common space—geographical space. An additional advantage of such visualization is the possibility to present cultural heritage resources in different semantic relations to the space, which is only dependent on the users' needs or expectations. These solutions are only possible with the access to interoperable data; therefore, data standardization and integration was the key point in the research.

The considerations focus on the spatial aspect and data structures of cultural heritage. An important matter for further research is to take into account the issue of time and changeability of information both in time and in space. This issue concerns the fact that movable monuments have moved for years, the place of their storage has been changed over time and the elements of the collection were separated from one another at the same time. It would be interesting to undertake research into the visualization of such issues in geographical space and to develop a comprehensive way of presenting the time-space history of movable monuments. These issues require close cooperation of GIS specialists, cartographers, and historians. Also, the issue of the ISO standards uses in the area of spatially—oriented humanities infrastructure is worth exploring further, e.g., the topic of metadata [42] or web services [43].

**Acknowledgments:** The research was carried out as part of the statutory work "Spatio-temporal characteristics of data and geographic objects in the context of their effective use" realized in years 2015–2018 at the Military University of Technology in Warszawa with the help of University of Warmia and Mazury in Olsztyn, both in Poland.

**Author Contributions:** Albina Mościcka and Agnieszka Zwirowicz-Rutkowska designed the study; Albina Mościcka performed the experiments; Agnieszka Zwirowicz-Rutkowska analyzed the data; Albina Mościcka and Agnieszka Zwirowicz-Rutkowska wrote the paper.

**Conflicts of Interest:** The authors declare no conflict of interest. The founding sponsors had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, and in the decision to publish the results.

## References

1. Ward, J.S.; Barker, A. Undefined By Data: A Survey of Big Data Definitions. The Computing Research Repository 2013. Available online: <https://arxiv.org/pdf/1309.5821> (accessed on 13 January 2018).
2. Gandomi, A.; Haider, M. Beyond the hype: Big data concepts, methods, and analytics. *Int. J. Inf. Manag.* **2015**, *35*, 137–144. [CrossRef]
3. Laney, D. 3-D Data Management: Controlling Data Volume, Velocity and Variety. Application Delivery Strategies by META Group Inc. 2001. Available online: <http://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf> (accessed on 13 January 2018).
4. Chen, H.; Chiang, R.H.L.; Storey, V.C. Business intelligence and analytics: From big data to big impact. *MIS Q.* **2012**, *36*, 1165–1188.

5. Kwon, O.; Lee, N.; Shin, B. Data quality management, data usage experience and acquisition intention of big data analytics. *Int. J. Inf. Manag.* **2014**, *34*, 387–394. [CrossRef]
6. TechAmerica Foundation's Federal Big Data Commission. *Demystifying Big Data: A Practical Guide to Transforming the Business of Government*; TechAmerica Foundation's: Washington, DC, USA, 2012; 40p. Available online: <http://www.gartner.com/it-glossary/big-data/> (accessed on 13 January 2018).
7. Lee, I. Big data: Dimensions, evolution, impacts, and challenges. *Bus. Horiz.* **2017**, *60*, 293–303. [CrossRef]
8. Pymm, R.; Carroll, M.; McCausland, S.; Kennan, M.A. Big data: Opportunities and barriers across cultural heritage sectors. In *Annual Review of Cultural Heritage Informatics*, 1st ed.; Arns, J.W., Ed.; Rowan & Littlefields: Lanham, MD, USA, 2016; pp. 119–128.
9. Europeana Collections. Available online: <https://www.europeana.eu> (accessed on 13 January 2018).
10. City of London, Ancestry Digitisation Project. Available online: <http://www.cityoflondon.gov.uk/things-to-do/visiting-the-city/archives-and-city-history/london-metropolitan-archives/family-history/Pages/Digitisation.aspx> (accessed on 13 January 2018).
11. Nield, D. The US Library of Congress Just Put 25 Million Records Online, Free of Charge. Science Alert. Available online: <https://www.sciencealert.com/the-us-library-of-congress-just-put-25-million-records-online-free-of-charge> (accessed on 13 January 2018).
12. Council of Europe. Recommendation CM/Rec (2017) 8 of the Committee of Ministers to Member States on Big Data for Culture, Literacy and Democracy. Available online: [https://search.coe.int/cm/Pages/result\\_details.aspx?ObjectId=0900001680750d68](https://search.coe.int/cm/Pages/result_details.aspx?ObjectId=0900001680750d68) (accessed on 13 January 2018).
13. McNulty, E. Understanding Big Data: Infrastructure. Available online: <http://dataconomy.com/2014/06/understanding-big-data-infrastructure/> (accessed on 13 January 2018).
14. Anderson, S.; Blanke, T. Taking the Long View: From e-Science Humanities to Humanities Digital Ecosystems. *Hist. Soc. Res.* **2012**, *37*, 147–164. [CrossRef]
15. Atkins, D.E.; Droegemeier, K.K.; Feldman, S.I.; Garcia-Molina, H.; Klein, M.L.; Messerschmitt, D.G.; Messina, P.; Ostriker, J.P.; Wright, M.H. Revolutionizing Science and Engineering through Cyberinfrastructure. Available online: [https://www.immagic.com/eLibrary/ARCHIVES/GENERAL/US\\_NSF/N030131A.pdf](https://www.immagic.com/eLibrary/ARCHIVES/GENERAL/US_NSF/N030131A.pdf) (accessed on 13 January 2018).
16. European Commission. ESFRI: Research Infrastructures for Europe. Available online: [http://europa.eu/rapid/press-release\\_MEMO-12-772\\_en.htm?locale=en](http://europa.eu/rapid/press-release_MEMO-12-772_en.htm?locale=en) (accessed on 13 January 2018).
17. Digital Riffs. Available online: <http://digitalriffs.blogspot.co.uk/> (accessed on 13 January 2018).
18. Blanke, T.; Kristel, C.; Romary, L. Crowds for Clouds: Recent Trends in Humanities Research Infrastructures. In *Cultural Heritage Infrastructures in Digital Humanities*; Benardou, A., Champion, E., Dallas, C., Hughes, L., Eds.; Routledge: Oxford, UK, 2017; pp. 48–62, ISBN 978-1-4724-4712-8.
19. Anderson, S. What are research infrastructures? *Int. J. Humanit. Arts Comput.* **2013**, *7*, 4–23. [CrossRef]
20. Schraefel, M.C. What is an analogue for the semantic web? In Proceedings of the Eighteenth Conference on Hypertext and Hypermedia, Manchester, UK, 10–12 September 2007; pp. 123–132.
21. Mościcka, A. The CENDARI infrastructure in GIS-based historical research. *Program* **2017**, *51*, 132–151. [CrossRef]
22. Anderson, S.; Blanke, T. Infrastructure as intermeditation—From archives to research infrastructures. *J. Doc.* **2015**, *71*, 1183–1202. [CrossRef]
23. Janowicz, K. The Role of Space and Time for Knowledge Organization on the Semantic Web. *Semant. Web—Interoperability Usability* **2010**, *1*, 25–32.
24. Mościcka, A. Europeana Data Model in GIS for movable heritage. *Geografie* **2015**, *120*, 527–541.
25. Object Management Group. *OMG Unified Modeling Language TM (OMG UML), Version 2.5*; Object Management Group: Needham, MA, USA, 2015.
26. International Organization for Standardization. *ISO 19109:2015 Geographic Information—Rules for Application Schema*; International Organization for Standardization: Geneva, Switzerland, 2015.
27. International Organization for Standardization. *ISO 19103:2015 Geographic Information—Conceptual Schema Language*; International Organization for Standardization: Geneva, Switzerland, 2015.
28. International Organization for Standardization. *ISO 19107:2003 Geographic Information—Spatial Schema*; International Organization for Standardization: Geneva, Switzerland, 2013.
29. Object ID. Available online: <http://archives.icom.museum/objectid/> (accessed on 13 January 2018).
30. Encoded Archival Description (EAD). Available online: <http://www.loc.gov/ead/> (accessed on 13 January 2018).

31. ICOM/CIDOC CRM Special Interest Group. *Definition of the CIDOC Conceptual Reference Model, Version 6.2*; ICOM/CIDOC CRM Special Interest Group: Crete, Greece, 2015.
32. Miller, J.; Mukerji, J. (Eds.) *MDA Guide Version 1.0.1*; OMG (Object Management Group): Needham, MA, USA, 2003.
33. Erlangen CRM/OWL. Available online: <http://erlangen-crm.org/documentation> (accessed on 13 January 2018).
34. The Pearl of Mannerism. Church of the Holy Trinity in Żórawina [In Polish: Perła manieryzmu. Kościół Trójcy Świętej w Żórawinie]. Available online: <https://perlamanieryzmu.wixsite.com/zorawina/zabytki-poza-kosciolem> (accessed on 9 February 2018).
35. Kuhn, W. Core concepts of spatial information for transdisciplinary research. *Int. J. Geogr. Inf. Sci.* **2012**, *26*, 2267–2276. [CrossRef]
36. International Organization for Standardization. *ISO 19101:2005 Geographic Information—Reference Model*; International Organization for Standardization: Geneva, Switzerland, 2005.
37. Frank, A.U. (Ed.) *Panel—GI Pan European Link for GI and GIS*; GeoInfo Series 21; Technical University Vienna: Wien, Vienna, 2000.
38. Semantic Web. Available online: <http://www.w3.org/standards/semanticweb/> (accessed on 13 January 2018).
39. Mohammadfama, I.; Kalatpour, O.; Golmohammadi, R.; Khotanlou, H. Developing a process equipment failure knowledge base using ontology approach for process equipment related incident investigations. *J. Loss Prev. Process Ind.* **2013**, *26*, 1300–1307. [CrossRef]
40. Dimopoulou, E.; Gogolou, C. LADM as a Basis for the Hellenic Archaeological Cadastre. In Proceedings of the 5th Land Administration Domain Model Workshop, Kuala Lumpur, Malaysia, 24–25 September 2013.
41. Paasch, J.M.; Paulsson, J. Standards for Land Use and Planning—Only for Specialists? Available online: <http://oes.tuwien.ac.at/sektor/periodical/titleinfo/1345577> (accessed on 8 February 2018).
42. International Organization for Standardization. *ISO 19115-1:2014 Geographic Information—Metadata—Part 1: Fundamentals*; International Organization for Standardization: Geneva, Switzerland, 2014.
43. International Organization for Standardization. *ISO 19128:2005 Geographic Information—Web Map Server Interface*; International Organization for Standardization: Geneva, Switzerland, 2005.



© 2018 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).