Knowledge BasedRecommendation onOptimal Spectral and Spatial RecordingStrategy of Physical Cultural Heritage Objects

Ashish Karmacharya, Stefanie Wefers, Frank Boochs
i3mainz – Institute for Spatial Information and Surveying Technology, Mainz University of Applied Sciences
e-mail: {ashish.karmacharya, stefanie.wefers, frank.boochs}@hs-mainz.de

Abstract—**Ontologies have traditionally been used to represent knowledge of a specific domain. They are also used to provide a base to infer the knowledge present inside them. However, the applications of ontologies within the Cultural Heritage (CH) community have been restricted to providing standard documentation for significant heritage objects. E.g., widely used ontology within CH disciplines, International Committee for Documentation Conceptual Reference Model(CIDOC CRM) is designed to provide standards in documenting archival information of physical CH object. There has been hardly any work relating the objects to their documentation purposes. In this paper, we present the Colour and Space in Cultural Heritage Knowledge Representation (= COSCH^{KR}) ontology – a multi-faceted ontology. With COSCH^{KR}, we present a system that infers inter-woven descriptive semantics of different involved CH disciplines in recording CH objects to recommend optimal spatial and spectral technical solutions to humanities experts and guide through the underlying complexities while recording their objects. It takes numbers of facts into consideration including physical characteristics of the CH objects, the characteristics of their surroundings and even other relevant facts such as budget or staff competence to infer against the characteristics of the technologies for a proper recommendation. In contrast to a typical Recommender System, which does the same for web-based content through stochastic methods, we use descriptive semantics at the concept level.**

Keywords-Ontology development; Description Logics; Humanities;Cultural Heritage; 3D Data;Spectral Data; Descriptive Semantics;Inference

I. INTRODUCTION

In the last decade digital 3D and spectral recording of physical cultural heritage (CH) objects is getting more and more common. The digital representations are not only seen as support for CH expert's tasks (e.g., research studies, monitoring, and documentation) but as useful items especially for dissemination. A wide audience can be addressed, e.g., through websitesand interactive digital applications. For this tasks appropriate data are needed optimally supporting the researcher and/or user.To achieve this, the 3D or spectral data recording, its subsequent data processing, analysis and visualisation has to involve experts from multiple disciplines: (a) CH experts responsible for the knowledge about the constraints given by the CH object itself (e.g., research question, conservation condition, light sensitivity, transportation possibilities); (b) recording experts preparing and executing a digitisation strategy (e.g., recording device needs specific amount of space, limitations of sensors, suitable data accuracy and resolution); (c) IT experts applying proper algorithms on the generated data (e.g., point cloud registration) to allow data analysis; and (d) 3D modellers and communication experts visualising the data for different audiences. All these parameters influence each other. Which digitisation strategy meets the requirements of the CH application (A CH application is connected to CH research questions which can be answered through the generated data – they are a tool illustrating the significance about the existence and significance of CH objects w.r.t. the history of mankind)depends on (1) the parameters of the physical CH object (appearance, size etc.), (2) the limitations and abilities of the devices and methods,and (3) the impact of the data processing tasks. Altogether, the elaboration of a digitisation strategy is a complex collaborative and interdisciplinary task.

The COST Action [34] TD1201: Colour and Space in Cultural Heritage (COSCH) [2] [3] contributes to the conservation and preservation of cultural heritage (CH) by enhancing this mutual understanding among the experts from various disciplines. COSCH is a forum for communication and interdisciplinary networking. Bridging the gap between professionals involved in the recording of physical CH objects through discussions and publications such as guides to good practice, the COSCH community decided to go one step further by developing the knowledge model COSCH^{KR} orthe COSCH Knowledge Representation. COSCH^{KR} is an OWL 2 XML-serialised ontology based model currently under development. It expresses and structures the knowledge of the disciplines involved in CH object recording. The main intention of COSCH^{KR} is to guide CH experts by inferring the optimal spatial and spectral technologies for the recording of their specific physical CH object based on facts on its physical characteristics and the purpose behind the recording.It is therefore comparable to a recommender system in a sense that it identifies and provides recommendations on optimal technologies. A typical recommender system works on available data to develop recommendation algorithms based on stochastic methods [31]. However in our case, we do not have abundant data from successfully completed CH applications for the development of algorithms that filter the recommendations. Instead, we rely on experience and knowledge from experts to create the entire recommendation process and reasoning system within COSCH^{KR}.

The challenge in the development of such an ontology lies in capturing expert's decisions of spatial and spectral recording. It needs to capture the core essence of inter-disciplinary dialogues and activities across multi-disciplinary platforms to achieve common goal in a CH documentation project. The ontology, therefore, needs to describe inter-disciplinary dependencies that echo the real world dependencies across disciplines in such a project. Various disciplines have to work together to answer common research questions of CH applications.

Interpretations and observations on issues and vocabularies vary across the discipline. The ontology needs to fill in these gaps in communication as well. At the core, the ontology needs to address what is required and how to get it. To elaborate, any CH application requires a digital surrogate of the concerned physical CH object providing answers to CH research questions. The nature of digital surrogates depends on the application and its requirements. The requirements for answering the research questions thrown by a CH application dictate what the digital data should contain. Thisrequirements on digitaldata in turn influence theselection technology(ies). These digital data and their nature form a bridge connecting requirements from CH applications and possibilities from recording technologies. We use axis *"Applications – Data – Technologies"*(see Fig. 1) to illustrate this further.

COSCH^{KR} is a knowledge model developed through capturing and structuring knowledge of involved disciplines inside CH recordings. The discipline inherent knowledge-otherwise independent - is interlinked through the description logic (DL) concept constructors, which define descriptive semantics of the concepts inside the ontology [16] (restriction axioms inside the ontology). The descriptive semantics are extensively used to 1) bind different heterogeneous conceptual axioms and theorems inside the ontology and 2) infer the results from the queries. The base axis *"Applications – Data – Technologies"* is supported through other axes that define the underlying semantics of objects and/or other factors that have significant influences on the model and its inference system. Technical process is deterministic to the real world considerations when generating data: technologies applied to a CH object generate data with specific data nature and data content under specific external influences that may be coming from CH objects themselves or other conditions influencing the technologies or CH objects. These considerations have to be logically described and are described through descriptive semantics of the relevant classes inside the ontology. Due to a variety of technologies and their underlying instruments and recording strategies, the importance of expert knowledge on them is further justified while recommending the best suited process. A first attempt to give a structured view on characteristics of spatial recording techniques has been presented in [15].

The descriptive semantics binding technologies parameters to object characteristics should take respective views of involved disciplines into account.CH applications and their conditions on the requirement on data are also described inside the ontology through relevant descriptive semantics. The CH applications that ask for specific data content intercede all these inter-linking descriptive semantics for inferring and navigating through optimal recording techniques.

COSCH^{KR} provides a base for expressing common knowledge on technologies, CH objects and CH applications. The ontology will be exploited through an interactive web based application (COSCH^{KR} platform), which will have interactive Graphical User Interfaces (GUIs) for users to assert their queries through a guided mechanism. The platform will apply those asserted queries to the COSCH^{KR} ontological model to infer recommendationsfor a recording device, strategy, and process, which will support the CH expert to receive spectral and/or spatial data with sufficient content and quality to answer the underlying research questions.

The successful creation of such a platform needs to be based on mutual understanding of experts from the involved fields. It has to start with the consolidation of a common vocabulary with unambiguous terms, continue with the formalisation of domain inherent knowledge, and end with the connection of this formalised knowledge. A special challenge is the content capture and its formalisation. For example, humanities research questions are often directly linked to a specific CH object and domain inherent research question. Moreover, the same physical CH object might be connected to different research questions which ask for differing data requirements. This makes the formalisation of decisive factors in humanities research question a sensible task, which has a strong impact on the identification of the best suitable recording strategy. With this paper, we present our hands-on experiences in developing the ontology COSCH^{KR} and the challenges during data capturing and structuring process. We also present how the descriptive semantics inside the ontology lay the necessary foundation for inferring the recommendations of the optimal technical strategy(ies) in spatial and spectral CH documentation.

The structure of the paper is as follows. In section 2, the state of art, we present the use of semantic and knowledge technologies in CH. We also present the existing stochastic methods based recommendation systems. In section 3, we present our approach, illustrating the purpose and scope and then methodology and principles behind it. We also present an example use case to demonstrate our approach within this section. Lastly, section 4 concludes the paper summarizing the actual state and what is the future outlook.

II. STATE OF ART

Ontologies have evolved as computational artifacts that provide conceptual and computational models of any particular domain of interest. Ontologies populated with concepts are agreed generally to follow the states of uniform knowledge representation and provide a computational model of a particular domain of interest [8] [4] [19]. The main motivation behind ontologies is that they allow for sharing and reuse of knowledge bodies in computational form [20]. Ontologies have become a popular research topic within the communities of the Semantic Web due to the fact that they promise a shared and common understanding of inter-communicable domains, the primary objective of the Semantic Web. The role of ontologies in the Semantic Web and the gradual evolution of Web Ontology Language [35] are discussed in the research paper [21] [23].

A. Description Language Inferences

The Web Ontology Language is a Description Logic based ontology language for the Semantic Web [23]. There are effective reasoning algorithms for Description Logics that can reason with OWL ontologies. Existing DL Reasoners, such as FaCT++ [24], use these algorithms and are quite efficient. A DL comprises of ABoxes and TBoxes where a TBox describes the terminologies expressed through concepts and roles and ABox contains the assertions about the instances of the concepts described through TBoxes. Most OWL ontology based systems apply TBox and ABox inferences for the rightful categorizations and relationships. The application of these inferences for building up possible components of DL ontologies is presented in [25]. The work describes standard and nonstandard inferences.

- A standard inference uses the TBox inference provided through the concept descriptions to subsume the individuals in the ABox. This facilitates computation of hierarchies and internal instance relationships.
- A nonstandard inference uses the semantic descriptions of individuals in ABox to first create, categorize and populate themselves into respected bottom level concepts. Afterwards, with the least common subsumer (lcs) of these bottom level concepts, their super-concepts are defined and created. The practice continues until the final top concepts are created. The concept descriptions of each level are defined in the process with having individuals as building blocks.

With both TBox inference and ABox inference, the inferences are used to build on DL based knowledge representation.

B. Ontologies and their types

Applications of ontologies are required to play a role in analyzing, modeling and implementing domain knowledge and influence problem solving knowledge [20]. Ontologies can generally be classified for intentions of capturing and modeling static and problem solving knowledge.

Static ontologies do not internally reason about the knowledge, but use it for processing natural language [26], achieving interoperability within heterogeneous datasets [12], which facilitate communication, such as in E-commerce. Ontologies within this category fall under Reference Ontologies, whose main inclination is toward realism [27].

Problem solving ontologies areintended for problem solving knowledge and provide views that could be used for reasoning. They are Application ontologies with computational sublogic of full first order logic. The usage of ontologies in the Semantic Web can be found in both categories: the former with core Semantic Web applications like Linked Open Data applications (LOD) and the later with Semantic Web Service Discovery.

Application ontologies combine task/method ontologies (containing terms and reasoning mechanisms of problem solving methods) together with the domain ontologies (containing descriptions of domains of disclosures) to provide overall interpretationof the problem and attempt to provide answers. Such ontologies are preferred in a Recommender System.

C. Ontologies in Recommender Systems

Recommender Systems (RSs) are software tools and techniques providing suggestions for items to be of use to a user [32] [33]. COSCH^{KR}recommends solutions through prior knowledge represented inside its ontology and not through analyzing huge amount of data through statistical methods as a Recommender System does. Recommender Systems are highly influenced by stochastic methods such as machine learning. However, ontologies are routinely used in recommender systems in combination with machine learning and other stochastic methods. Middleton and colleagues presented a recommender system that recommends on-line academic research papers through the profile descriptions of the readers [28]. The system uses classical machine learning algorithms with an ontology based approach to design a recommender algorithm. The inference mechanism is highly influenced by the amount and quality of data for high end results. Similarly, other recommender systems such as [29] extract data from music ontology within LinkedBrainz (A Linked Open Data platform to publish music database) through its SPARQL end points, and then matches results with customized management of user profiles (according to personal preferences). The use of ontologies in Recommender System to use prior knowledge on the resources (academic paper as in [28] and music as in [29]) along with understanding the behavior of the users (through) stochastic methods to provide recommendations to them.

D. Development of ontologies

While developing multi-faceted ontologies any basic metaphors cut across a number of domains [30]. Developments of multi-faceted ontologies are common today both in scientific and commercial communities. Lim and a colleague presented the Multi-faceted product family ontology (MFPFO) that manages the complexity in relationships between physical components with their semantic orientations, such as manufacturing, materials and marketing [17]. The relationships between different facets should be clearly described in such ontology. These knowledge-intensive ontologies need to keep harmony across the people, disciplines and the applications in which they are involved [18] to maintain semantic consistency inside it. Hence, the development issues become much crucial in their development.

Though there is collective experience in designing, developing and using ontologies, there is no common agreed methodology for building ontologies. Different methodologies exist and have been proposed over the years [9] [11] [14] [20] [22]. The commonality among all these propositions is that the ontology should satisfy the purpose of its creation and should not attempt to model the world, should be coherent and extendible and should provide minimum ontological commitment. Another commonality among them is that they prescribe step wise workflow based methodological guidance for ontology engineering. The NeOn methodology presents a different approach through suggesting different pathways for developing ontologies [37]. It presents nine different scenarios covering commonly occurring situations while developing ontologies whereCOSCH^{KR}loosely complies with the first scenario.COSCH^{KR} ontology represents experts' knowledge through inter-linking descriptive knowledge in spatial and spectral CH documentation that could be extended to other technologies and/or other discipline such as architecture as well. The COSCH umbrella provides a base to include experts from different domains to evaluate coherency of ontology and its underlying theorems and axioms through domain specific semantic consistency.

E. Existing domain ontologies

Ontologies for CH disciplines such as CIDOC-CRM [1] are generally designed as standards for stakeholders such as museums who archive CH objects. Though the terminologies used within CIDOC-CRM are of interest for our research and we actually refer to CIDOC-CRM inside our ontology, the intention and application of COSCH^{KR} differs considerably from CIDOC-CRM. Moreover, CIDOC-CRM does not provide a class structure for detailed information about the recording of CH objects. The CARARE 2.0 metadata schema [7] prepared within the frame of the 3D ICONS project provides compatibility to the structure of CIDOC-CRM. The schema is based on CH standards such as MIDAS (English Heritage 2012), an XML based harvesting schema LIDO [6], and EDM-EuropeanaData Model [5]. It harvests meta-, para-, and provenance data of 2D and 3D data of CH objects into Europeana (Europeana Professional). The CARARE 2.0 metadata schema extends the class including technical para- and meta-data of recording strategies. However, it is meant to harvest the content into open knowledge hubs for linking data. The schema does not have provisions to reason itself for choosing optimal para- and metadata from the existing ones when new cases arise. The development of CARARE 2.0 metadata schema thus follows a pattern that is necessary for ontologies managing and harvesting contents. All in all, the core group worked out to develop a new common ontology, not integrating existing domain ontologies since 1) not all involved disciplines have their own well accepted ontology (CH has CIDOC-CRM, but for spatial and spectral technology there is no widely accepted one such as OPPRA cannot be considered as common ontology for spectral domain or no ontology at all as in spatial domain) and 2) ontologies are designed for different purposes and scopes (e.g., CIDOC-CRM is designed for providing standards for museums archiving physical CH objects [1] or OPPRA.owl is designed for 20th century paint conservation [13]) and harmonizing them through inference rules is a long and tedious task.

III. APPROACH

An ontology base system that recommends the optimal spatial/spectral technologies for a CH documentation application requires:

- ontology consisting of descriptive semantics of
  - involved spatial and spectral technologies and data
  - CH object and CH applications
- a recommendation mechanism that infersdescriptive semantics of CH objects, their respective applications against those of technologies

COSCH^{KR} intends to provide recommendations (on spatial and spectral technologies for the CH applications) from the conceptual level and not from the data level through their analysis as no database exists to be used for the purpose. Therefore, there is no possibility of ABox inference inside. This negates any possibilities of implementing any external stochastic mechanisms within inference system. The system hence has to work on pre-defined expert knowledge at schema level to do reasoning. Consequently, it becomes prominent that existing state of art solutions have limited implementations on COSCH^{KR} as

- the inference on knowledge model needs to use the descriptive semantics at concept level for inference and not at data or individual level
- the intention is not to classify the assertions as conventional reasoners in the Semantic Web technologies are meant to but to infer right

relations between technologies and applications through data

- till date, there is no ontology on spatial and spectral technologies that could be used for the case. This limits the assessments of physical objects through the semantics of technologies within an ontology based system
- the existing CH ontology is based for documenting biographical information of the CH objects and has limited scope to define descriptive semantics of concerned object that trigger the technological selection process.

Through COST action we have the leverage of technical and humanities expertise in their respected fields. COSCH^{KR} represents their knowledge and experiences inside within one common framework. They are semantically encoded through DL axioms and theorems.COSCH^{KR} reasoning engine (used for recommendations) reasons these axioms and theorems at TBox level and do not assert any individuals inside. Additionally, the engine distances itself from using any stochastic methods to carry out reasoning.It is solely based on the concepts' descriptive semantics (defined through DL concept constructors), which represent the knowledge and experiences of the domain experts.

A. *Purpose and Scope*

COSCH^{KR} isa system that guides CH experts by inferring the optimal spatial and spectral technologies for the recording of a specific physical CH object based on facts about the physical CH object and the CH application. The purpose is to help the CH end users to answer their competency questions querying for the optimal technical solutions. An example of such questions could be *"What is the right technical solution to record my CH object (a Roman coin) for the CH application determining its origin and time period".* Such a purpose was discussed and agreed on within the COSCH community and through this the scope of involved domains was madeclear: CH domains (archaeology, conservation, art history etc.), spatial technology domains (surveying, computer vision, photogrammetry etc.), spectral technology domains (multi- and hyperspectral imaging etc.), and IT domains (algorithms, data processing etc.). In meantime a core group comprising experts from semantic, spatial and spectral and CH domain responsible for designing the top-level ontology was agreed on, for evaluation of different approaches in knowledge collection, for guidance in knowledge collection, and for regular updating of the entire expert group.

B. *COSCH^{KR}ontology*

COSCH^{KR}ontology represents inter-disciplinary knowledge ofCH recording. It is designed and developed and woven together through rules.

Fig. 1 illustrates the five top level classes of COSCH^{KR}. These top level classes and their specializations are defined through 1) a logical taxonomical structure based on shared concepts 2) the relationships between them and 3) their existence defined through the conditions. The green boxes are classes related to CH domains, the orange box is a class related to spectral and spatial recording domains and the blue box is a class related to data processing domains. The five top-level classes are linked through different properties displayed as arrows.

The intention is to maintain and support the base of the conceptual axis: "*Application – Data – Technologies*" (see grey strip in Fig. 1). We first define the classes under this major axis. Class "*Technologies*" encompasses the technical methods, procedures, tools and their setups to generate or process the generated data. They are presented through specializations of the class where each specialization contains semantic descriptions that describe their best practice, limitations through their characteristics. These specializations generate data which are stored under specializations of class "*Data*". We can illustrate this with a simple example: "*Photography*" is a technology that generates photos. Therefore "*Photography*" will be a specialization of class "*Technologies*". Photos are 2D images i.e., 2D Data – and hence specialization of class "*Data*". The class also includes the capabilities and limitations of instruments that are used to generate data. For instance: "*Photography*" with a mobile camera can have different quality on photo than that with a high end DSLR camera.

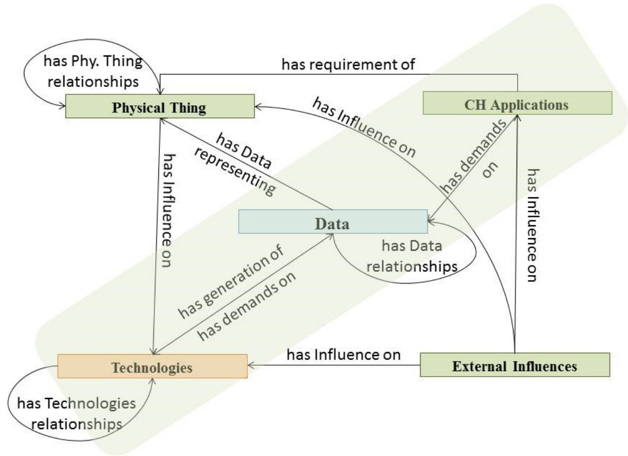

Fig. 1.COSCH^{KR} top-level classes.

The first obvious outlet of class "*Technologies*" is data represented through class "*Data*". This is bridged through the statement *"Technologies generate Data"* which is expressed through the DL concept constructors inside the ontology (see (1)). This bridging statement provides a conceptual crossover between two.

$$\text{Technologies} \equiv \exists \text{hasGenerationOnData.Data} \quad (1)$$

TheseDL concept constructors are ourbase for *inference rules* becausethey will be combined and translated as rules for inferring the content inside COSCH^{KR} system. These DL constructors are defined and formulated in close collaboration with experts in technical fields of spatial and spectral technologies.

Moving on, the purposes and reasons behind the acquiring the data are represented within class "Application". It is a class that determines the nature and quality of data that needs to be acquired and lies on the other side of the axis "*Application – Data – Technologies*". The specializations are again defined through the descriptive semantics through DL concept constructors. They link to data through the expressions of what kind of data is required (see (2)).

$$\text{Applications} \equiv \exists \text{hasRequirementOnData.Data} \quad (2)$$

Let us continue taking example of "Photography" to illustrate our example on applications. We can think of two simple applications of photos here: printingfor i.)a billboardand ii.) a travel album. Even though both require photos as the main product, the qualities required of them differ a lot. The photos needed for the billboard need to have high resolution photos with different degree of sharpness while the same should not holdtrue for the photos for the travel albums. They have impacts on the concerned technology acquiring the photos. The billboard photos should be taken with high end DSLR camera and needs some post- processing while those for travel album can be taken with simple digital cameras or mobile cameras.

The classes outside this axis are equally important in terms that they affect the technologies. Class *"Physical Thing"* represents the main subject to be measured (in our case it is CH Physical Objects). COSCH^{KR} does not define these objects as their real world counterparts. They are defined through the physical characteristics of which they are built-up. For example: churches do not have pseudo representation with class "Church" inside the class "*Physical Thing*". Therefore, they cannot be asserted as "Church". They need to be asserted as composite objects (under sub-class "*CompositeObjects*") built-up with different individual objects (under class "*IndividualObjects*"). These composite/individual objects have certain characteristics like in case of church, they are big in size. COSCH^{KR} therefore does not differentiate between a church or a building. The main reason is: it is not important to know what are being digitized (in terms of how they are called in real world), but important to know whether the physical characteristics of object support or deny its digitization process when certain technology is being used. The core mantra is "the physical characteristics of the objects decide how any technology should be used to digitize them and not objects themselves". Let us roll back to the example of "*Photography*" and its product "Photos". The possibility of "*Close-range Photography*" has high dependency on the size of the object. It cannot photograph big sized objects (see (3)) because it cannot capture entire object in one photo shooting action. So a church cannot be photographed with close range photography.

$$\exists \text{hasSuitabilitiesFor.(Physical-Thing} \sqcap \forall \text{hasSize.}(\neg \text{ Big})) \quad (3)$$

The class "ExternalInfluences" has similar technical implications to that of class "*Physical Thing*". It defines constraining semantics that effect the recommendations of the technologies. They include constraints deriving from the project limitations such as budget, human resource or from the surroundings of the measured objects like available space, lights, access and so on. These factors play major roles in technical solution and need to be defined inside the ontology.

C. Content capture

At the very beginning of development of the ontology, the COSCH community established a core group responsible to collect, manage and structure knowledge from the relevant expert groups. The core group was also responsible to define common vocabulary. It developed theoretical concepts on the basis of the collected unstructured knowledge through: 1) questionnaire; 2) discussion. These theoretical concepts were represented through respective axioms and theorems.

To be able to get an overview of expert's knowledge within the COSCH community, a questionnaire was designed, which has the intention to ask for spectral and spatial recording approaches and technical details applied in various humanities projects. For each group of physical CH objects, which were recorded within a humanities project, one questionnaire requires to be completed, where a group is mainly defined through the purpose of the spectral or spatial recording. The actual version of the questionnaire consists of twelve main questions with subordinate questions asking primarily for technical details [36]. The completed questionnaires are supporting the analysis to structure the content, to define work areas through the determination of relevant terms and vocabularies, and to identify contact persons having a specific expertise and being available for discussions and feedback. All in all, it should be highlighted that this tool cannot have the intention to collect knowledge, which is already structured and ready for the integration into the ontology (see below). In contrast, the specific content related to one physical CH object and application, which is described within the completed questionnaires, gives evidence for structuring theoretical concepts included in the ontology.

D. Case Study example

The analysis of the completed questionnaires led to the following strategy: the ontology will be developed using case studies as framework since it provides concrete facts for a discussion with experts from different domains. These facts on the one hand are the basis for a common understanding and on the other hand are helpful to stay focused. Furthermore, a case study provides added advantage important for the development of theoretical

 ISBN: 978-1-61208-507-4

concepts: the case study discussion can be expanded easily by fact modifications even after it has been included through theoretical concepts inside the ontology. For example, if the original case study was related to small physical CH objects a fact modification could be to imagine the physical CH object being very large. Depending on the facts under discussion this approach helps to extend branches of the ontology.

The first selected case study dealt with waterlogged wood as through unavoidable conservation treatment the shape and volume of these objects is modified. To be able to measure the influence of various conservation treatments a high number of samples of waterlogged wood were recorded in 3D before and after conservation. And through a comparison of the two 3D models representing one sample changes could be evaluated giving information on the influence of the conservation on the shape and volume of the objects [10].

The physical CH objects of interest are samples from different time periods having a minimum size of 100 x 60 x 60 cubic mm (Fig.2 CH Object Size “small”). The material condition of the samples before conservation treatment was an important issue as the archaeological waterlogged wood samples had a dark brown to black appearance and were partly shiny. The translucent and reflective surface of the untreated samples had impact on the data quality. However, this impact was reduced to a minimum through careful toweling of the samples before recording (Fig.2 CH Object Reflectivity “low”). After conservation the appearance of the samples sometimes changed immensely as the water inside the wood is gone and conservation materials stabilized the object causing sometimes a colour change to light brown. However, all sample surfaces were dry after treatment which means they were not reflective anymore (Fig.2 CH Object Reflectivity “low”). A crucial factor was the high number of samples: All in all 777 objects were recorded before and after treatment (Fig.2 CH Object Number “large”) why an industrial recording device – a structured light scanner – was chosen as selected processing steps could be automated and controlled through scripts of associated software (Fig. 2 Workflow Method “Automated, semi-automated”). The workflow control was applied for quality management of the required data and accuracy (Fig. 2 “3D” and “high”). Due to this the operating staff of the structured light scanner could be changed without major impact on the workflow and data quality as the number of possible error sources was reduced to a minimum (Fig.2 technical competence “low”). However, the varying operating staff needed supervision by a 3D recording expert. Especially the above mentioned workflow control possibilities determined the choice of the 3D recording technique.

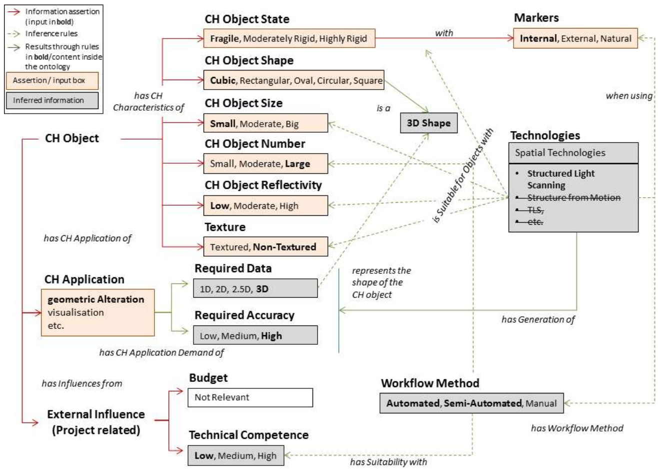

Fig.2.Simulation of a GUI for the case study “conservation of waterlogged wood”. The red boxes represent the user input and the grey boxes represent the inferred information.

$$\text{Applications} \sqcap \geq 2\ \text{hasRequirementOnData.(Data} \sqcap (\leq 1\ \text{hasRepresentationOf.PhysicalObjects)} \qquad (4)$$

The CH application in this case study is to compare the deformation of geometry (Fig.2 CH Application “geometric Alteration”). This CH application requires high level of accuracy to detect any changes, which also means it requires high resolution datasets for the comparison. This needs to be explicit while describing semantics of the class representing the CH Application Geometric Alteration (class *“Deformation Analysis”* – a specialization to top level class *CH Application*). Through these descriptive semantics (see below), we relate and compare different classes to derive suitable answer.

Going back to our *“Applications – Data – Technology”* (see Fig. 1) conceptual axis, the CH application *“Deformation Analysis”* demands for data representing the objects with high accuracy e.g. depending on their size. *“Deformation Analysis”* is an immediate sub-class of *ChangeDetection* (again a specialization of class *CH Applications*), defined through the descriptive semantics stating that at least two dataset of the same object is required (see (4)).

Besides inheriting these descriptive semantics encoded in DL class constructors of parent classes, *“Deformation Analysis”* describes the semantics of required data such as nature (3D) and accuracy (high) (see (5)).

$$\text{ChangeDetection} \sqcap \exists\,\text{hasRequirementOn-Data.(Data} \sqcap \exists\,\text{hasRepresentationOf.(PhysicalObjects} \sqcap \exists\,\text{has-ObjectShape.Shapes)} \sqcap \exists\,\text{hasSpatialAccuracy.High)} \qquad (5)$$

Once the requirement on data is known in the axis Applications – Data – Technologies, COSCH^{KR}uses these semantic descriptors to infer right technology(ies). In this case, the descriptive semantics of measurement method *StructuredLightScanning*, which is a specialized class under *MeasurementMethods*(specialization of class *Technology)*is inferred as the recommended because

1) nature of data and accuracy it can generate with

$$\exists \text{hasGenerationOnData}.(\text{3D\_Data} \sqcap \exists \text{hasSpatialAccuracy.High}) \tag{6}$$

2) evaluating the technology against the object characteristics

a) Size/volume

$$\exists \text{hasSuitabilitiesFor}.(\text{PhysicalObjects} \sqcap (\exists \text{hasObjectSize.2DSize\_Small} \sqcup \exists \text{hasObjectVolume.3DVolume\_Small})) \tag{7}$$

b) textured: non textured

$$\exists \text{hasSuitabilitiesFor}.(\text{PhysicalObjects} \sqcap (\exists \text{hasObjectTexture.NonTextured}) \tag{8}$$

c) number of objects: large with 777 samples

In order to determine the effectiveness of the technology to manage large number of objects, the technology should provide automated or semi-automated work flow. Therefore, semantic description of the technology (Structured Light Scanning) is semantically described through the workflow and number of objects with a single DL statement (see (9)).

$$(\exists \text{hasSuitabilitiesFor}.(\text{PhysicalObjects} \sqcap (\exists \text{hasObjectQuantity.LargeNumber})) \sqcap (\exists \text{hasWorkflowMethod}.(\text{SemiAutomatedWorkflow} \sqcup \text{AutomatedWorkflow}) \tag{9}$$

d) reflectivity: low

$$(\exists \text{hasGenerationOnData}.(\text{3D\_Data} \sqcap \exists \text{hasSpatialAccuracy.High})) \sqcap (\exists \text{hasSuitabilitiesFor}.(\text{PhysicalObjects} \sqcap \exists \text{hasObjectReflectance.Low\_Reflectivity})) \tag{10}$$

The initial reflectivity of the waterlogged wooden samples was high because of the higher reflectance of water. COSCH^{KR}in such a case does not provide any technology to generate high accuracy data with the highly reflected objects. Therefore, the ontology does not provide any technical solutions for highly reflected wooden samples and checks with the user if the reflectance could be lowered. In our case, the sample could be wiped to lower the reflectance. This again infers *StructuredLightScanning*as right technology.

e) fragile: no possibilities to put markers

This again gives no results. The standard setup of the technology (Structured Light Scanning) represented through class *MeasurementSetups*(a specialization of class Technology) will check whether one can stick any markers into the object. If the answer is yes then the rule of high accurate 3D data will be possible with the structured light scanning. The standard setup of the method is described through class StandardStrcturedLightScanning-Setups (specialization of class *MeasurementSetups*with relation to class *MeasurementMethods*) .

$$(\exists \text{hasGenerationOnData}.(\text{3D\_Data} \sqcap (\exists \text{hasSpatialAccuracy.High})) \sqcap (\geq 1 \text{hasImplementingInstruments}.(\text{InternalMarkers} \sqcup \text{NaturalInternalMarker}) \tag{11}$$

3) evaluating technology against external influencing characteristics

Technical competence is the characteristics of the project influences. In this case the technical competence among operating staff is low. The case resembles the case of ii.c. Therefore, the technical competence depends on the kind of workflow the technology provides. If it provides automated workflow like in this case, the technology will only require operating staff with low competence. We defined this inside the class of *StructuredLightScanning*

$$(\exists \text{hasOperatingProject}.(\text{ProjectInfluences} \sqcap (\exists \text{hasOperatingStaffCompetence}.(\text{Competence\_Medium} \sqcup \text{Competence\_Low}))) \sqcap (\exists \text{hasWorkflowMethod}.(\text{SemiAutomatedWorkflow} \sqcup \text{AutomatedWorkflow}) \tag{12}$$

Here class *ProjectInfluences*is specialization of top-level class *ExternalInfluences*.

This summarises that the Structured Light Scanning is the optimal recommended technology for scanning wooden samples in order to estimate deformation. The technologies are sorted out while the ontology processes knowledge inside to infer at different level. For example, at the very beginning when the requirement was 3D data, the ontology suggested all technologies that generate 3D data including Terrestrial Laser Scanning (TLS), Structure from Motion (SFM) and so on. As more semantic constraints were applied, technologies were filtered out. E.g., when the requirement was highly accurate data SFM was ignored, and when the object size was asserted “small” TLS was ignored. All technologies are semantically defined to support or deny the conditions they will be inferred against. At the end, semantically defined rules of Structured Light Scanning supported all the asserted conditions so the system recommended the technology.

 ISBN: 978-1-61208-507-4

Different technology might be recommended if and when the situation changes or other different constraints are added into. The knowledge model will alter the parameters of this case study to simulate other situations, e.g., instead of a high number of physical CH objects a low number is assumed, to identify why and how the recording strategy would have changed.

We are working with two other case studies. They are still under development and not yet integrated in COSCH^{KR}.They concentrate on a CH application related to the spectral recording and visualization domains (www.cosch.info/case-studies). Through these two case studies the better part of the technical classes could be developed. One of the most important reasons choosing the case study related to the spatial recording was the fact that the spatial recording expert was personally available for face-to-face. The development of a common understanding might be a longer iterating exchange of views and content, the number of iterations increases with the distance between the science fields, why for matter of convergence it is proposed to use face-to-face discussions. All in all, it is recommended to center the discussions around a case study in a process-related manner to stay focused and to create a common understanding between the different experts. The aim of the discussion is to develop theoretical concepts, which could be integrated into the ontology as formal axioms presenting the descriptive semantics and which finally display the case study as theoretical concepts addressing and linking all top-classes within the ontology. After the integration of both case studies as theoretical concepts in the ontology further identified work areas will be approached through discussions with other experts creating theoretical concepts related to other case studies.

IV. CONCLUSIONS

CH is arguably one of the most multi-disciplinary areas of research where disciplines from highly diverse disciplines (incl. human science, technologies and even pure science like chemistry) are actively involved. Developing ontology that not only smoothen the communication problems but also provides inter-disciplinary understandings to support recommendation on the best possible technical approach for a CH application requires a platform where experts from individual respective domain are open to exchange inter-disciplinary discussions. COSCH – the COST Action TD1201 provides such a platform where experts from spatial and spectral technologies discuss on specific CH research questions with humanities experts to suggest on best usages of the technologies for answering them. The underlying knowledge from those discussions are captured and encapsulated within ontology COSCH^{KR}.

In this paper, we have presented the experiences we gained in developing COSCH^{KR}. With COSCH^{KR} we intend to address issues relevant in the area of CH, spatial and spectral technologies and the Semantic Web technology itself. The usage of semantics within CH communities is mostly limited to knowledge management and rarely to knowledge processing. They are mostly used to capture, document and re-use information on CH objects through knowledge management technologies.We see huge potential in using semantics to go beyond knowledge management; they can be used for knowledge processing with their in-built reasoning capabilities.

COSCH^{KR} exploits Description Logics reasoning capabilities by encoding knowledge already at concept level. This has an added benefit against conventional recommender systems. We use experts with prior knowledge and experience in CH documentation that can be already encoded inside the knowledge model and not rely on stochastic on huge amount of data at data level. These encoded knowledge sets can then be exploited by any interpreting systems to infer the right recommendations. In addition, the existing databases and knowledge hubs with para-, meta- and provenance information could benefit from COSCH^{KR} for evaluating their own data.

COSCH^{KR} is developed within the conceptual axis of requirement on data by CH application – generation of data by technologies. Other concepts are woven around this axis. Though we use the concept with the application field of CH, it could be applied in other domains as well. We are currently working on a mechanism that interprets descriptive semantics encoded through DL concept constructors into inferencing rules. These descriptive semantics will be parsed into rule based statements that could be reasoned by existing reasoning engines to provide the recommendations.

ACKNOWLEDGMENT

This work was partly supported by COST under Action TD1201: Colour and Space in Cultural Heritage (COSCH). Furthermore, we would like to thank Dipl.-Ing. (FH) Guido Heinz M.Eng., Uwe Herz (both RGZM), Marcello Picollo, Tatiana Vitorino (both IFAC-CNR), Dr. Julio del Hoyo Melendez (The National Museum in Krakow) and Christian Degrigny (HE-Arc) for discussions and support in developing the classes and rules.

REFERENCES

[1] P.L. Boeuf, M. Doerr, C.E. Ore, and S. Stead, Definition of the CIDOC Conceptual Reference Model. ICOM/CIDOC CRM Special Interest Group, 2013.

[2] F. Boochs, “COSCH - Colour and Space in Cultural Heritage, A New COST Action starts”. The 4th International Conference of EuroMed 2012. Ioannides M., Fritsch D., Leissner J., Davies R., Remondino F. and Caffo R. eds., pp. 865-873, 2012. Limassol, Cyprus: LNCS.

[3] F. Boochs et al., “Towards Optimal Spectral and Spatial Documentation of Cultural Heritage. COSCH – An Interdisciplinary Action in the COST Framework”.XXIV International CIPA Symposium. Strasbourg: International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, pp. 109-113, 2013.

[4] C. Brewster and K. O'Hara, “Knowledge representation with ontologies: Present challenges - Future possibilities”.International Journal of Human-Computer Studiesm vol. 65, no. 7. pp. 563-568, 2007.

[5] V. Charles, “Introduction to the Europeana data model”. 17th International Conference on Theory and Practice of Digital Libraries. Malta: International Conference on Theory and Practice of Digital Libraries, 2013.

[6] E. Coburn and R. Stein, *LIDO*. [Online] Avaliable from:http://network.icom.museum/cidoc/working-groups/lido/. [retrieved: November, 2015]

[7] K. Fernie, D. Gaverilis, and S. Angeli, TheCARARE meta data schema v2.0. Europeana Carare project, 2013.

[8] G. Jakus, V. Milutinovic, S. Omerovic, and S. Tomazic,“Concepts, Ontologies and Knowledge Representation”. SpringerBriefs in Computer Science, 2013.

[9] D. B. Lenat and R. V. Guha,Building large knowledge-based systems: Representation and inference in the cyc project. Boston: Addison-Wesley Longman Publishing Co., Inc, 1989.

[10] C. Mazzola, “What to do with “Large Quantity Finds in Archaeological Collections” - A KUR-project”. News in Conservation 6 (December 2009).

[11] T.R. Gruber,“A Translation Approach to Portable Ontology Specifications”. Knowledge Acqusitionvol. 5 no. 2,pp. 199-220, 1993.

[12] A. Młka, B.Kryza, and J. Kitowski,“Integration of heterogeneous data sources in an ontological knowledge base”.Computing and Informaticsvol. 31, no. 1.pp. 189-223, 2014

[13] S.A. Odat, A Semantic e-Science Platform for 20th Century Paint Conservation, PhD dthesis. University of Queensland, Australia, 2014.

[14] M. Uschold andM. Gruninger,“Ontologies: Principles, methods and applications”. Knowledge Engineering Review, pp. 93-136, 1996.

[15] A.-K. Wieman, S. Wefers, A. Karmacharya, and F. Boochs, “Characterisation of Spatial Techniques for Optimised Use in Cultural Heritage Documentation”.M. Ioannides, N. Mageanat-Thalmann, E. Fink, R. Zarnic, A. Yen, E. Quak, eds. Digital Heritage,Lecture Notes in Computer Science, pp. 374-386, 2014.

[16] I. Horrocks, P. F. Patel-Schneider, D. L. McGuinness, and C.A. Welty, OWL: a Description Logic Based Ontology Language for the Semantic Web, 2004.

[17] S. Lim, Y. Liu, and W. B. Lee, “A methodology for building a semantically annotated multi-faceted ontology for product family modelling”. Advanced Engineering Informatics, vo. 12, no. 2, pp. 147-161, 2011.

[18] Y. Sure, J. Angele, and S. Staab, “OntoEdit: Multifaceted Inferencing for Ontology Engineering”. Journal on Data Semantics, 2003, pp. 128-152, 2003.

[19] S. Grimm, P. Hitzler, and A. Abecker, “Knowledge Representation and Ontologies: Logic, Ontologies and Semantic Web Languages”. R. Studer, S. Grimm and A. eds.,Semantic Web Services. Springer Link, pp 51-105, 2007.

[20] R. Studer, V. R. Benjamins,and D. Fensel, Knowledge Engineering: Principles and Methods, Karlsruhe, 1998.

[21] S. Grimm, A. Abecker, J. Völker,and R. Studer, “Ontologies and the Semantic Web”. J. Domingue, D. Fensel,and J. A. Hendler, eds., Handbook of Semantic Web Technologies, Springer, pp. 507-579, 2011.

[22] X. Wang, J. Almeida,and A. Oliviera, “Ontology Design Principles and Normalization Techniques in the Web”.Data Integration in the Life Sciences, 5th International Workshop, DILS 2008, Evry, France, Springer Berlin Heidelberg, pp 28-43, 2008.

[23] I. Horrocks, D. Patel-Schneider, D. L. McGuinness, and C. A. Welty, “OWL: a Description Logic Based Ontology Language for the Semantic Web”. F. Baader, D. Calvanese, D. McGuinness, D. Nardi,and P. F. Schneider, eds, The Description Logic Handbook: Theory, Implementation, and Applications (2nd Edition), Cambridge University Press, 2007.

[24] D. Tsarkov and I. Horrocks, “FaCT++ Description Logic Reasoner: System Description”.Automated Reasoning, Bd. 4130, Nr. LNCS, pp. 292-297, 2006.

[25] F. Baader and R. Küsters, “Nonstandard Inferences in Description Logics: The Story So Far”.Mathematical Problems from Applied LogicI, Bd. 4, Springer New York, pp. 1-75, 2006.

[26] D. Estival, C. Nowak,and A. Zschorn, “Towards Ontology-Based Natural Language Processing”.The Workshop on NLP and XML (NLPXML-2004): RDF/RDFS and OWL in Language Technology, Strousburg, Association for Computational Linguistics, pp 59-66, 2004.

[27] C. Menzel, “Reference Ontologies - Application Ontologies: Either/Or or Both/And?”The KI2003 Workshop on Reference Ontologies and Application Ontologies, 2003.

[28] S. E. Middleton, D. D. Roure,and N. R. Shadbolt, “Ontology-Based Recommender Systems”.Handbook on Ontologies, Springer, pp. 779-796, 2009.

[29] M. Á. Rodrígues-García, L. O. Mendoza, R. Valencia-García, A. Lopez-Lorca,and G. Beydoun, “Ontology-Based Music Recommender System, Distributed Computing and Artificial Intelligence”. 12th International Conference of the series Advances in Intelligent Systems and Computing, Vol. 373, Springer International Publishing, pp. 39-46, 2015.

[30] G. Lakoff and M. Johnson, Metaphors We Live By. The University of Chicago Press, 1980.

[31] F. Ricci, L. Rokrach,and B. Shapira, “Introduction to Recommender Systems Handbook”.Recommender Systems Handbook, Springer, pp. 1-35, 2010.

[32] R. Burke, “Hybrid web recommender systems”.The Adaptive Web, Springer Berlin / Heidelberg, pp. 377-408, 2007.

[33] P. Resnick and H. R. Varian, “Recommender systems”. Communications of the ACM vol. 40, no. 3, pp. 56-58, 1997.

[34] COST, COST Action, *COST European Cooperation in Science and Technology.* [Online] Available fromhttp://www.cost.eu/COST_Actions, 7 July 2015. [retrieved: July, 2016]

[35] S. Bechhofer, F. v. Harmelen, J. Hendler, I. Horrocks, D. L. McGuinness, P. F. Patel-Schneider, and L. A. Stein, “OWL Web Ontology Language”. *W3C Recommendation.* [Online]Available from http://www.w3.org/TR/owl-ref/ [retrieved: September, 2016]

[36] COSCH, COSCH^{KR}, *Colour and Space in Cultural Heritage.*[Online] Available from www.cosch.info/coschkr [retrieved: July 2016]

[37] M. C. Suárez-Figueroa, A. Gómez-Pérez, and M. Fernandez-Lopez. "The NeOn Methodology framework: A scenario-based methodology for ontology development." *Applied Ontology* 10, no. 2,pp. 107-145, 2015.