Digitization of Cultural Heritage – a Service-Oriented Approach

Boris Shishkov

Institute of Mathematics and Informatics, Bulgarian Academy of Sciences / IICREST
Sofia, Bulgaria
b.b.shishkov@math.bas.bg

Abstract. The digitization of Cultural Heritage assumes an interdisciplinary approach, collaborative knowledge co-creation, and distant communication. This can only be fulfilled by ICT means requiring in turn alignment between those who deliver the technical (IT) support and those who are in the role users. A Service-Oriented approach is considered adequate in this regard because of users' composing services at high level (with underlying technical complexity remaining hidden) and developers' considering the software components whose functionality is delivered through the services. All this requires however complex regulations to conform with. Semiotic norms in combination with workflows are considered in this regard. In this paper, we propose an approach based on Service-Oriented Computing and Organizational Semiotics, directed to Cultural Heritage digitization. The approach is partially illustrated by means of an example.

Keywords: Cultural heritage digitization, Service-oriented computing, Organizational semiotics, Workflow management

1 Introduction

The digital recording of Cultural Heritage (CH) is a multidimensional process depending on the nature of the subject of recording and the purpose of its recording. The process involves the 3-dimensional digitization, digital data processing and storage, archival and management, representation and reproduction as well as the knowledge (co)-creation and/or dissemination concerning the content. The current paper focuses on the business processes concerning the management of artefacts (digitized items reflecting corresponding Cultural Heritage assets) and also the learning process related to this.

Aiming at increasing effectiveness in artefact management and related learning, current concepts and strategies stress upon: (i) collaborativeness; (ii) on-line access. The CH Digitization is of interdisciplinary essence because domain knowledge is to be brought together with multimedia and data management, such that a widely accessible digital archive is achieved that is adequately related to corresponding real CH assets. This assumes collaborative knowledge co-creation and intensive distant communication. With regard to COLLABORATIVENESS: Both individual knowledge

and organizational processes should be considered in relation to create a synergetic effect when introducing new artefacts and distributing corresponding digital data, and also when it is necessary to train people in (non-core) competences needed for the digitization process. Regarding this, the possibility to collaborate with third parties is advantageous, especially in co-creating and/or using learning content. With respect to ON-LINE ACCESS: Inspired by the fact that many educational institutions have acknowledged the importance of providing on-line access to student laboratories, we claim that the web portals could be useful in making virtual and remote knowledge available on the Internet, in the form of digital data reflecting CH assets. This could help in the classification of existing digitized items, in the recognition and processing of images, and in the specification and maintenance (accordingly) of metadata.

We take a knowledge perspective on this, presenting a problem statement through the three main sub-perspectives, as introduced in [1]: Knowledge core, as a global perspective, Knowledge additions, as a local utilization perspective, and Knowledge representation, as concerning the personal approach. In particular: (i) There is a 'global' drive first of all for disseminating a knowledge core (for example, a digitized item reflecting a CH asset), pushed either by Society or by legislation or by the global management of a (distributed) institution, to just mention several possibilities. (ii) There is a 'local' utilization (on top of the global drive) that concerns the way the knowledge is introduced in a particular societal segment or organization. (iii) There is finally a personal issue that concerns the way the knowledge is actually delivered to a person having his/her own personal preferences taken into account.

The knowledge core is (often) of global concern, the knowledge additions are driven usually by the need for a local (knowledge) utilization, while the knowledge presentation is about the way knowledge is actually delivered and this should be driven by individualism – making the user in control of how (s)he would be utilizing the information – with more/less explanations, at high/medium quality, for instance. Elaborating further on those perspectives and their relevance to CH Digitization:

- The CH "raw" information (digitized items) is available on a global scale – one could spot within seconds images from Egypt, Rhodes, Sofia.
- Still, this information has as well a particular local "meaning" that concerns historical, cultural, traditional, and other aspects.
- Finally, there is a "personalized" use of this information in the end – the archeologist would look at one thing, the historian would look at another thing, and so on.

Hence, each of those perspectives would need to be architecturally reflected in a corresponding IT environment, to guarantee that the global, local, and individual demands will be satisfied.

We claim that this can be possible with service-oriented solutions that allow stakeholders to compose "own" services at high level, counting however on underlying technical consistency. For this reason, we get inspiration from the Service-Oriented Computing (SOC), in general and the Web-Service concept – in particular. As already mentioned, composing web services at high-level assumes "hiding" their underlying technical complexity – this complexity is with the software components who are im-

plementing the corresponding service(s). *Composability, traceability*, and *interoperability* are thus of crucial importance in web service provisioning [2].

This, in combination with Web Portal technologies [3] could also provide advanced CH data brokerage, for the benefit of effectively managing and globally distributing digital data [4], counting on effective match-making and advanced access control mechanisms (for example: through embedding JavaScr. code fragments to uploaded items).

Finally, such service-portal Internet-driven solutions could only be implemented if there are rigorous rules and regulations "underneath", that govern all processes, aligning in this way (existing) CH digitized items and related processes [to be served] and the technical support [serving those processes]. For this, we find Organizational Semiotics useful, especially in combination with Workflow Management, as studied in [5]. In particular, we consider one Semiotics method, namely the Norm Analysis Method that allows for specifying rules and regulations by means of semiotic norms (rules) that can be usefully combined with workflow patterns, for better visualization and possibility for simulations.

Hence, the contribution of the current paper is two-fold: we firstly propose solution directions with regard to identified challenges that concern the digitization of CH (done in the current Section), and on this basis, we discuss each of the proposed paradigms / methods (as it will be done in the further sections of the paper).

The remaining of this paper is organized as follows: In Section 2, we consider SOC and e-Brokerage, especially sticking to its relevance to CH digitization. In Section 3, we outline and discuss the Semiotics and Workflow paradigms and their possible role for solving problems in the mentioned context. In Section 4, we discuss the combined application of SOC and semiotic norms. Section 5 contains the conclusions.

2 Service-oriented Solutions for Cultural Heritage Digitization

The emergence of SOC is considered as a move towards combining real-life business concerns and technological concerns [2], envisioning a service (of a component/entity) as defining the goal, capabilities and/or behavior (of the component/entity) as observed by and relevant to the users (of the component/entity) [6].

A Web Service is considered as a self-contained, Internet-enabled service component capable not only of performing business activities on its own but also possessing the ability to engage other web services to form higher-order business transactions [7].

We distinguish between composite and constituent web services – a composite web service consists of (is provided by an orchestration of) multiple constituent web services, and a constituent web service is an 'elementary' web service, i.e. a web service which can be used on its own or in a composite web service [8][9].

In order to be usable on a large scale, web services (which are based on specific sets of standards) should be somehow reflectable in certain abstractions, as an instrument for their application in any platform through which the Internet user accesses them. Moreover, web services usually should not require design 'from scratch' be-

cause this would make them expensive. They should instead be re-usable, using one web service as a basis for developing another, by making use of its core functionality [10].

We thus consider it innovative that multiple users are able to access web services, personalize them and finally use them. Our first conclusion is that this usage of web services implies advanced infrastructures and application platforms that utilize and coordinate such (globally) re-usable services. Furthermore, employing such generic web services for work in domain-specific business environments means that the service use has to be driven by appropriate underlying business models. Prior to their use, web services should have been discovered (by matching requirements to advertised names) and subjected to negotiation (since the user must accept using a particular web service).

All those considerations have contributed to the emergence of the Service-Oriented Architecture – SOA which goes beyond the sole consideration of web services [11], being a useful paradigm that can support engineers in their designing, building and using distributed software systems. SOA facilitates the establishment of ICT support for business processes, which is readily available, flexible and easily maintainable across multiple organizations and platforms. The concept of service / web service adopted by SOA, has evolved from modular object/component middleware approaches, such as CORBA, DCOM and J2EE [8]. However, web services have become the technology of choice for implementing service-oriented software systems, primarily because they are based on ubiquitous Internet standards, such as HTTP and XML, and because they support 'loose coupling'. Whereas the uptake of web-services-based SOA is impressive, there are still important fundamental challenges not addressed by this technology, as recognized in [12]: Firstly, the 'plug and play' interoperability of web services to enable ad hoc cooperation of new partners is limited [13]. For on-demand composition of services in an open service-oriented world, interoperability has to be ensured at different levels (syntactic and semantic) and in different dimensions (information and behavior). Current research in this direction is using, for example, Semantic Web and ontology technologies [7]. Secondly, the property of 'loose coupling' is not appropriate for many applications that involve stateful components. Hence, the benefits of web services and SOA would be limited for developers of such applications if they themselves have to solve the issues of stateful interaction, notification of state changes, support for sharing and coordination [14]. It should thus be aimed that those concerns are placed at the service infrastructure level or that another solution is enforced. Thus, our second conclusion is that enhancement needs to be achieved in the way applications which are by nature not loosely coupled, are supported by SOA-related technology. *Finally*, we argue that web services alone are insufficiently capable of guaranteeing an appropriate 'alignment' between business requirements and software functionality. What is needed is a structured approach for developing service-oriented software solutions, in which consistency with business requirements, (de-)composition of application services, and mapping onto (alternative) technology platforms can be systematically and separately addressed [11][14]. Hence, our third conclusion is that a business-software alignment is needed particularly in the SOA context.

Taking into account those 3 challenges, we formulate the following 3 (corresponding) desired properties concerning the SOA-driven application development: 1. Application architecture must allow usage of a SOA infrastructure; 2. 'Loose coupling' should be enforced; 3. Application architecture must fit within the business context.

Even though each of those 3 desired properties has received "attention" in past and current (R&D) advances, an exhaustive solution is still missing to date, as seen from corresponding obstacles in particular application domains, such as e-Voting [5]:

- There is no "widely recognized" SOA infrastructure currently, even though there are standards concerning Cloud Computing, Internet-of-Things, and Interoperability. Still, infrastructure enabling depends on the technology platform at choice and mis-alignment is often observed.
- Loose coupling is often possible; still, often implementations de facto obstacle this.
- Since the real-life (business) context is one thing and the technical implementations (of a software components) is another thing, it was expected that, by considering web services at higher level would help bridging this gap. We argue that this did not happen - we currently observe often a semantic gap between the service description and the corresponding service functionality. Hence, we claim that enterprise models are needed as basis when developing software no matter if service-oriented or not.

Learning from the above observations and conclusions, we have formulated several service-orientation-related RECOMMENDATIONS that concern the CH digitization:

1. We need a "reference model", namely a CH-DIGITIZATION ENTERPRISE MODEL, that: Exhaustively specifies the possible actor-roles and interactions; Formulates the decision-making mechanisms; Defines Quality-of-Service criteria (Only such a reference model can be a basis for distributed service-oriented solutions).
2. We suggest minimizing the processing at the user end as a way to easily standardize interactions, such that loose coupling is realistic.
3. Adequate traceability is needed between the reference enterprise model and the corresponding software models, to allow controlling the business-IT alignment.

Hence, we expect that applying SOC for the benefit of CH digitization is possible, if following the proposed recommendations. As for e-Brokerage, we would refer to a generic brokerage functionality, as presented in [4], where there are two actor-roles, namely: DISTRIBUTOR (the one who is delivering a digitized item) and USER (the one who is using the item). There are several use cases concerning that: ADD DATA and REMOVE DATA, CHECK USER'S INFO, CHECK DATA ACCURACY, REQUEST ADDITIONAL INFORMATION, and PERFORM MATCH-MAKING. The Use Case model is not presented in the current paper in more detail, for the sake of brevity and for more information, interested readers are referred to [4]. Realizing such a brokerage functionality in a service-oriented way is possible and then each of the above-mentioned use cases would point to a corresponding web service.

3 Enforcing Regulations Through Semiotics

In this section, we consider Organizational Semiotics and Workflow Management.

3.1 Organizational Semiotics and Semiotic Norms

Organizational Semiotics (OS) considers a number of concepts, such as *sign* and *affordance*, as essentially useful in modeling a (real-life) system and adequately considering relationships and meanings. Often what we observe goes beyond the primary "appearance" – for example, one could hold a *Rolex* pen not only as a means of writing but also as a way to demonstrate wealth (this is a sign). As for the affordance concept, it relates to potential abilities (for example: a book affords to be borrowed). Those concepts and other OS concepts, allow for building complex models that reflect both *OS* and *norms* (rules), and that is reflected in the widely popular semiotic norm pattern:

```
whenever <condition>
if <state>
then <agent>
is <deontic operator>
to <action>
```

The norm pattern is considered useful in modeling relationships among entities. For more information on OS, interested readers are referred to [15].

3.2 Workflow Management

It is claimed that a business process can be viewed as a collection of processes, where a process can be described as "*a set of identifiable, repeatable actions which are some way ordered and contribute to the fulfilment of an objective*"; typical process patterns are sequence, parallelism, split, and so on. Workflows are useful in modeling business processes and those models can be enriched in terms of semiotic norms. For more information on Workflow Management, interested readers are referred to [16].

3.3 Combining Workflows and Semiotic Norms

As mentioned before, whatever is done (behaviorally and/or technically), it needs to be based on underlying rules/regulations. Those would stem from the business logic and need to be reflected in corresponding enterprise models. For the sake of brevity, we are no going in much detail here and would consider a workflow model that is methodologically derived on the basis of a case briefing (for more information on how this can be done, particularly according to the SDBC approach, interested readers are referred to [4]). Referring to the same source, we consider a typical scenario according to which a user is uploading digital images (that reflect CH artefacts) to sys-

tem that we call "DDMS" – Digitized Data Management System. The scenario (that follows below) is simplified since the purpose is just illustrative:

The user is willing to upload images reflecting CH artefacts, to DDMS and for this reason, the user needs to be registered with the system. Then, after accessing his or her profile in the system (based on credentials entering), the user is ready to upload images. Before doing that nevertheless, the system should check if the user has upload rights – such "rights" would "guarantee" that the person has adequate professional expertise and is part of a legitimate relevant institution and/or project. Said otherwise, only those system users who have upload rights would be allowed to upload images to DDMS (the others would only be allowed to browse through existing images). Further, within one upload session, the user can upload one or more images and after each image upload, the image is temporary stored. Then the system would check the quality of each of the uploaded images – this is an automatic check where a number of parameters are considered. Each image that appears to be of poor quality would be automatically deleted (and the system would notify the user of each image being deleted). Then the system would be inviting the user to upload metadata for each of the uploaded and not deleted images and this is done via a template that is offered to the user to fill in electronically – the user would need to indicate when the image was taken, who is the photographer, who is the copyright holder, what is the CH artefact reflected in the image – type or artefact, origin, and so on. Then the system would check each metadata record against an existing knowledge based on all metadata records that show inconsistency would be automatically deleted as well as their corresponding images (and the system would notify the user of each metadata+image being deleted). And in the end, all not deleted images + corresponding metadata would be permanently stored.

The workflow reflecting the above scenario is expressed using the notations of the UML Activity Diagram [17] and we are using two "swim lanes", namely User, and DDMS, for better organizing the actions.

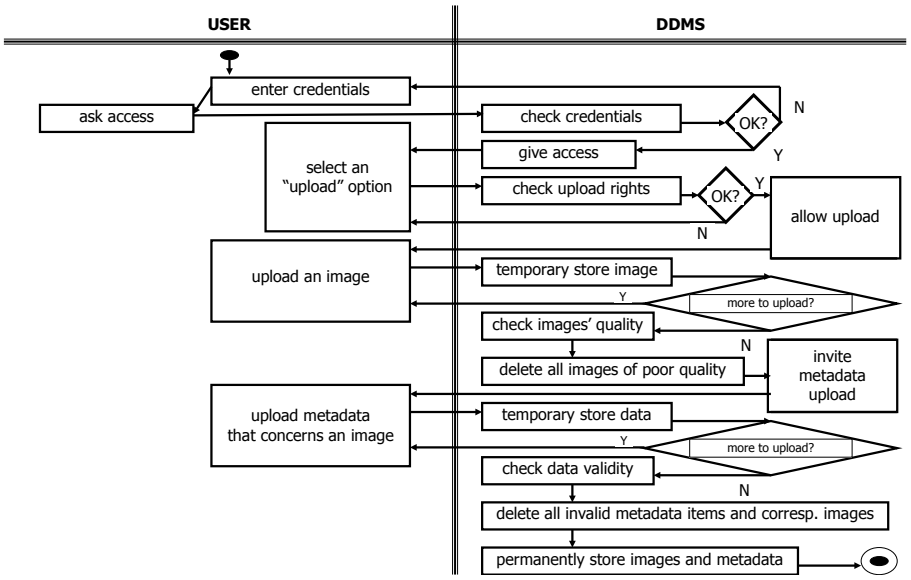

Fig. 1. Workflow Model for the Image Upload Scenario

As seen on the Figure, the user needs firstly to access the system and then it is to be checked if the user has upload rights (the credentials check and the upload rights check are reflected in decision blocks). Then all uploaded images are firstly temporary stored and then those images that would not pass the quality check would be deleted. Then the same with the corresponding metadata. Hence, DDMS is simple for the user – the user is to only enter credentials, indicate the (s)he is willing to upload images + metadata, and is then firstly uploading images and secondly – metadata. All checks are done automatically and the system maintains a simple "dialogue" with the user. Further, the user cannot challenge what the system does – it does it automatically. Finally, after the upload session is finished, the only option the user has is to browse through existing images – this is not reflected in the workflow model.

The workflow model gives a useful perspective on the processes and the relations among actions. This is nevertheless insufficient with regard to a regulatory/normative elicitation and this we do, by extending workflow "fragments" by means of semiotic norms. Because of the limited scope of this paper, we will not exhibit all norms that are to elaborate the workflow – we are just presenting two norms, to illustrate the usefulness of complementing workflow fragments by norms. The first one is the following:

Whenever John has "upload rights"
`If` John has uploaded an image
`then` the Digitized Data Management System (DDMS)
`is` obliged to check the quality of the image

This norm reflects the part of the workflow that represents the need for the system to quality-check all uploaded images that have been uploaded by a user who in turn must have passed the upload-rights check. The second norm is the following:

Whenever DDMS is checking the quality of an image
`if` the image quality is poor
`then` DDMS
`is` obliged to delete the image

This norm reflects the part of the workflow reflecting the images' quality check that the system should conduct.

In summary, in this section, we have shown how workflow models and semiotic norms can be combined to establish and represent regulations and rules with respect to the Cultural Heritage digitization process.

4 Bringing Together Web Services and Semiotics

In the previous sections we have not only discussed SOC and OS but we have also studied and partially exemplified their relevance with regard to the digitization of CH. In particular, we have discussed the role of web services and have particularly considered e-Brokerage in this regard (from the perspective of *technical enabling*) and we have also discussed semiotic norms (from the perspective of *regulations enforcement*). In this Section, we discuss ways for bringing the two together, basing the discussion on several key (web) services (Fig. 2) considered relevant to the CH digitization, namely:

- A *coordination service* that orchestrates the overall work of the system (for example: a digitized data management system), invoking other services at the right moment and offering them also the right input.
- Hence, this service would need a supporting information service that takes care of all data delivered at the right moment to the right entity.
- The other services (S) are: S1 (Mediation Service), S2 (Broker Service), S3 (User Agent Service), and S4 (Resource Discovery Service).

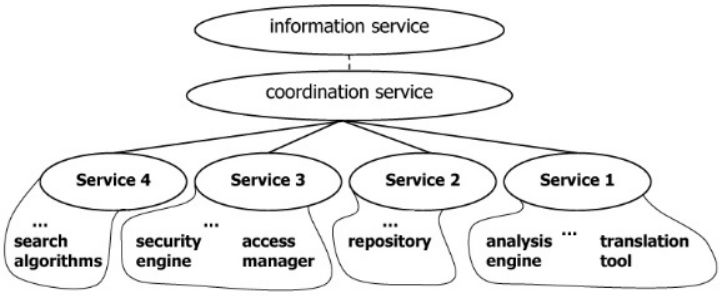

Fig. 2. Identified Services for Cultural Heritage Digitization

S1 concerns the advise delivery, by directing the user for making the next steps in using the environment – for this, an ANALYSIS ENGINE (AE) would be required (linked to a knowledge-base), such that the situation of the user is established and hence – an adequate advise is delivered; also, a TRANSLATION TOOL (TT) may be required, depending on the user's language of use. AE and TT may be used by S1 as "sub-services" (then S1 is only concerned with the results delivered) or be part of it (all is integrated in S1 as an overall functionality).

S2 concerns *match-making*, for example between what artefact image is being searched for and what artefact images are available, which match-making is to be based on strictly-defined criteria and pointing to a corresponding REPOSITORY.

S3 concerns *request processing* issues, such as what is required by the user (for example, to upload an image) and what is the particular user allowed for (for example, allowed to upload images or not allowed; what is needed for that is not only an ACCESS MANAGER (that is responsible for managing user's access) but also a SECURITY ENGINE (since this is matter of personal data that is to be reliably treated); similarly to AE and TT (with regard to S1), those two could either be considered as sub-services or as part of the functionality of S3.

S4 concerns the data searching that is to be based on SEARCH ALGORITHMS.

The relevance of those particular services is justified by previous work [4] and we do not claim exhaustiveness with regard to the (identified) services. Still, we find this useful with regard to illustrating the use of web services for the benefit of CH digitization and combining this with regulations enforcement brought forward through OS. With regard to this, we are using semiotic norms, as already discussed, and the norms used are categorized in three main categories: *Constitutional Norms*: regulating the digitization of Cultural Heritage in general (for example – prohibiting the photography and photos distribution of artefacts without the needed permission(s)); *Governing Norms*: regulating clusters of services (for example: what pattern recognition technology to be used in data searching, match-making, and so on); those norms are hence crosscutting with regard to services; *Service Norms*: regulating a particular service (for example: how to deliver advise to the system user).

It is claimed that if combined in such a way, web services and semiotic norms could be useful with regard to Cultural Heritage digitization.

5 Conclusions and Related Work

CH digitization has an interdisciplinary essence assuming distributed collaboration and distant communication, subject to rules and regulations. It is a challenge bringing together the domain experts (who need to use technology) and the technology deliverers (who specify technical solutions for the domain). We have shown that web services in combination with semiotic norms can be useful in this regard – the web services reflect the technical enabling while the semiotic norms reflect the regulations enforcement. We have also shown how those two can be brought together, especially in the context of CH digitization. Limitation of this work is the consideration of an illustrative example only and we plan as future work case study research.

References

1. B. Shishkov, M. van Sinderen, R. Cavalcante, 2010: Employees' Upskilling through Service-oriented Learning Environments. ICEIS'10 – Proc. of 12th Int. Conf. SciTePress.
2. M. Papazoglou, 2013: Web Services and SOA: Principles and Technology. Pearson
3. LiLa Project, Lib. of Labs, http://www.lila-project.org (2016)
4. B. Shishkov, 2005: Designing a Cultural Heritage Sector Broker Using SDBC. Int. Journal "Information Theories & Applications" Vol.11
5. B. Shishkov and . Janssen, 2016: Towards a Service-Oriented Architecture for eVoting. BMSD'16 – Proc. of 6th Int. Symp. on Business Modeling and Software Design. SciTePress.
6. Pasley, J., 2005. How BPEL and SOA are changing Web Services development. IEEE Internet Computing, Vol.9, Iss.3 (2005). IEEE Press
7. World Wide Web Consortium, 2016. Web Services Description Language 1.1, W3C Note, http://www.w3.org/TR/wsdl
8. Newcomer, E., 2002. Understanding Web services, XML, WSDL, SOAP and UDDI, Addison-Wesley. Boston.
9. Papazoglou, M.P., Kratz, B., 2006. A Business-Aware Web Services Transaction Model. In ICSOC'06, Int. Conf. on Service-Oriented Computing. Springer Berlin/Heidelberg.
10. Bosworth, A., 2001. Developing Web Services. Proc: 17th Int. Conf. on Data Engineering.
11. Alonso, G., F. Casati, H. Kuno, V. Machiraju, 2004. Web services, concepts, architectures and applications, Springer-Verlag. Berlin Heidelberg
12. B. Shishkov, J. L. G. Dietz, M. van Sinderen, 2007: Closing the Business-Application Gap in SOA - Challenges and Solution Directions. ICSOFT'07. INSTICC Press.
13. Wang, H., Zhang, H., 2006. Enabling enterprise resources reusability and interoperability through Web services. In ICEBE'06, IEEE Int. Conf. on e-Bus. Engineering. IEEE Press
14. Shishkov, B., Van Sinderen, M.J., Quartel, D., 2006. SOA-driven business-software alignment. In ICEBE'06, IEEE Int. Conference on e-Business Engineering. IEEE Press
15. Liu, K., 2000. Semiotics in Inf. Systems Engineering. Cambridge Univ. Press, Cambridge
16. Van der Aalst, W., 2011. Process Mining - Discovery, Conformance and Enhancement of Business Processes. Springer-Verlag, Berlin Heidelberg
17. Object Management Group, 2016. Website: http://www.omg.org