Complex Objects in Digital Libraries

Nádia Puchalski Kozievitch
Instituto da Computação
Campinas, Brazil
nadiapk@ic.unicamp.br

ABSTRACT

There are several applications which need support for complex objects, such as new mechanisms for managing data, creating references, links and annotations; clustering or organizing complex digital objects and their components. At this work we present a research proposal to address these issues. The objective is to specify and implement a formal and unified framework to manage multimodal complex objects in digital libraries, using the 5S formalism and Digital Content Component (DCC) aggregation.

1. MOTIVATION

Consider an application that involves the cooperation of researchers from countries whose dominant languages are different. This application aims to support teaching/learning activities in Biology domain, being used by teachers and students. In this application, besides the main information about species, like name, physical description, habitat, etc., images can be partially annotated to describe a specific characteristic. Figure 1 illustrates two parasite images and annotations made by students. The two black squares specify two sub-regions of species *Ancylostoma duodenale* and *Ancylostoma braziliense*, with specific annotations related to their morphological characteristics. This information is important for characterizing the main difference between the two species.

Now suppose that the teacher needs to list all annotations, first separated by student (to indicate how the learning process improved over the semester) and later separated by image (to list all correct annotations made on specific subregions). After reviewing and giving feedback to students, the teacher may share these annotations with other researchers in the project. Later, all the teachers can make comments over the annotations of students, or exchange comments among them.

Suppose that the teacher is interested in executing queries such as: "list the most annotated images; list red images

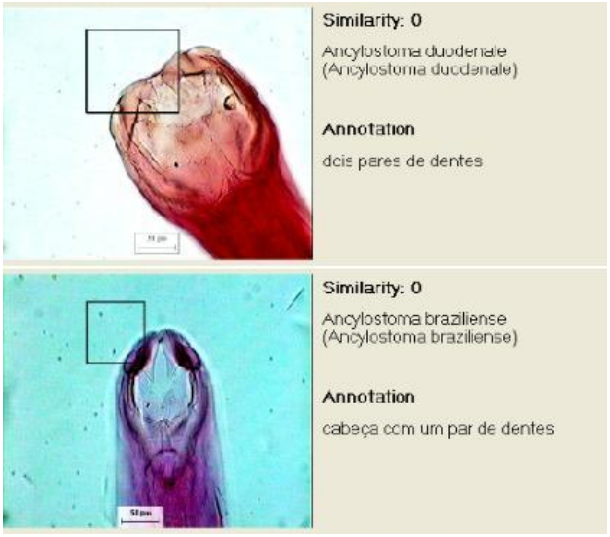

Figure 1: Example of an annotation application in the Biology domain.

that were annotated with the word 'head'"; "identify which species the students had more problems to annotate"; "list images of *Ancylostoma duodenale* with blue blackground whose annotations include the term 'pares de dentes' (pairs of teeth)"; "list images of family *Taeniidae* which annotation includes the term 'par' (pair) or 'pares' (pairs)".

The data available in the application above can be seen in two layers in this scenario. The first layer is the basic information, like metadata, images and annotations. The second layer includes annotations about annotations (for example, comments made by teachers about the annotations of students).

Consider a second scenario, where a PhD student enters a thesis into the Digital Library (DL) of an institution. Suppose that the thesis has 12 chapters; metadata like title, author, area, year; a powerpoint presentation describing chapter one; a recorded video which refers to the content of chapter two; and some annotations about chapter twelve, as illustrated in Figure 2.

The institution DL offers services like browsing, searching, and recommendation. Year after year, more students are encouraged to enter their theses in the system. Suppose

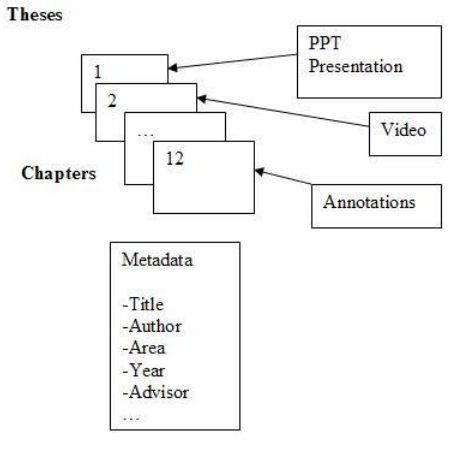

Figure 2: Components of a PhD thesis.

that the institution is interested in executing queries such as: "list the powerpoint files which are associated to theses about parasites with more than ten pages"; "the ten longest videos annotated with the word 'Linux' in computer science area"; "theses with brain images in gray color related to medical area"; "list the number of documents for each document type in the digital library".

In the first scenario, a species image can be seen as a Complex Object (CO) entity, composed of multiple parts (metadata and annotations), each of which is an entity in and of itself. In the second scenario, the thesis is another CO entity, composed of multiple components as pdf file, metadata, powerpoint, annotations, and video.

Unfortunately, existing digital library applications still do not support the management of this kind of complex object. Given the two scenarios, the following questions arise:

- How do these digital objects relate to each other? What structure is used to organize the components of a complex object? How to define/maintain data quality, considering that each CO can be composed from different subparts?
- From the database point of view, how will these digital objects and their metadata be stored and organized (in the disk and in the database)?
- If an application deals with CO whose components include data of different types (like text files, audio, image, video and software), how will the DL services access a CO and its components?
- How to handle/manage annotations over annotations from the database schema point of view?
- How to make the database flexible and tailored to new data types and their standards?

DL applications as those mentioned before need to deal with the integration of different types of data, like text files (doc, pdf, ppt, etc.), video (mpeg, etc.), audio (mp3, wav, etc.), image (GIF, JPG, etc.) or even software. Furthermore, each day is easier to link, browse or aggregate new coding technology (as MPEG-21, for example). On the other hand the database support for integrating these data is still inefficient. Some problems include:

- Waste of time to manage the data since the data is not centralized. Different applications are used to manage/access different components of a CO. Furthermore, from storage perspective, different storage strategies are used to organize CO components (for example, databases are used to store metadata, while images are stored in the file system);
- There are issues related to digital libraries/applications maintenance support with new technologies. Suppose a flexible DL with COs that can be updated (deletion/insertion/update of their components) with new versions and data formats over the years;
- There are issues related to service support. If new data formats are aggregated in the DL, services should be updated for managing the new formats;
- There are issues with database support, to link and organize sub-documents that are related to other sub-documents. Tasks like selection, linking and annotation of sub-parts (activities involved in learning activities) are still not present together in electronic tools.

There are several applications which need support for COs, with new mechanisms for managing data, creating references, links and annotations; or even cluster/organize complex digital objects. Unfortunately, there is no available DL application that supports the management of COs and their components as an integrated tool. At this work we present a research proposal to address these shortcomings. We believe that bringing together the formalism for digital libraries along with data aggregation concepts, we will help to specify in a formal base and implement an unified framework to manage multimodal complex objects in digital libraries.

The rest of this document is organized as follows. Section 2 provides a review of related work. Section 3 provides some preliminary work. Section 4 lists research questions and finally section 5 describes the research methodology, lists expected contributions and describes the scope of this research.

2. RELATED WORK

This section outlines previous works on data models and formalisms for DLs and COs. It presents the main concepts that will be verified in this research proposal for formal characterization of CO, under several perspectives: their definition, integration, flexibility, quality, architecture and services.

2.1 Digital Libraries

Grossman defines DL as a system that is primarily designed for selecting, retrieving, and computing with complex data, built on the model of data mining [14]. The Digital Library Federation (DLF, 1998) define DLs as organizations that provide the resources, including the specialized staff,

to select, structure, offer intellectual access to, interpret, distribute, preserve the integrity of, and ensure the persistence over time of collections of digital works so that they are readily and economically available for use by a defined community or set of communities [8]. Informally, under the 5S formalism perspective [11], a digital library involves a managed collection of information with associated services involving communities where information is stored in digital formats and accessible over a network. Information in digital libraries appear under the guise of digital objects, which can contain textual or multimedia content (e.g., images, audio, video), and metadata. Basic services provided by digital libraries are indexing, searching, and browsing [11].

In the open source community, new digital library software packages were created, as Eprints [3], DSpace [27], Greenstone [30], and Fedora [16]. They provide storage of and access to collections of all kinds of documents, including books, photographs, newspaper images, metadata, audio (MP3 files) and video, as well as mixed collections. Even using databases, it is not possible yet to work with parts of documents, or even make annotations.

2.2 Complex Object

A DL is composed of simple components named digital objects. COs are single entities that are composed of multiple parts, each of which is an entity in and of itself [15].

Several CO formats arise from different communities [18, 21]. In industry, CO can be found in persistent database storage and codecs multimedia. Some standards and formats are used for identifying resources, as Dublin Core, Resource Description Format (RDF) and Extensible Markup Language (XML). Other technologies have been proposed to manage COs, as multimedia framework MPEG-21 (Moving Picture Experts Group - 21 [5]), OAI-ORE (Open Archive Iniative for Object Reuse and Exchange [17]) and digital object formats as MPEG-21 DIDL (Moving Picture Experts Group - 21 Digital Item Declaration Language [4]) and METS (Metadata Encoding and Transmission Standard [7]). Even though there are different standards for CO, there is still incompatibility, motivating solutions for integration and interoperability. As each standard is specialized in a specific domain, their differences and their formats hamper their interoperability for different applications and contexts. On the other hand, it is still possible to match them, as proposed on [24], in their comparative study of IMS CP (IMS Content Package) and RAS (Reusable Asset Specification).

2.3 5S Formalism

Modeling, defining, understanding and describing DLs is still a challenge, as shown in [20, 1, 6]. The 5S [Streams, Structures, Spaces, Scenarios, and Societies] framework was proposed in [11, 9], as a formal theory to describe DLs. It captures the complexity of DLs as shown on Figure 3. The Figure illustrates the supporting layers of definitions: mathematical foundations (e.g., graphs, sequences, and functions), the 5 Ss (Streams, Structures, Spaces, Scenarios, and Societies), and key concepts of a DL (e.g., digital object, collection, repository, catalog). Arrows represent dependencies, indicating that a concept is formally defined in terms of previously defined concepts that point to it.

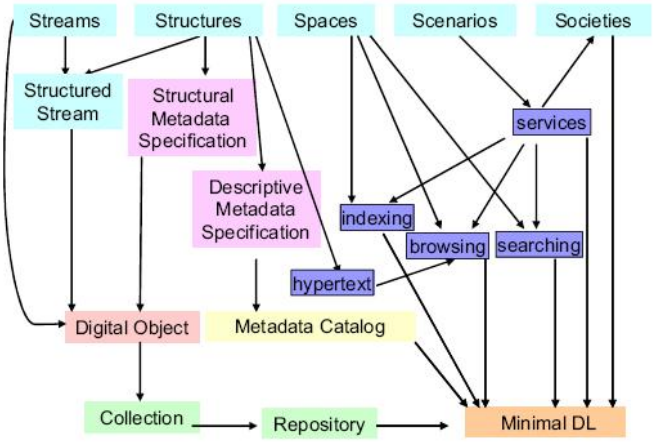

Figure 3: 5S definition structure [26].

In the 5S formalism:

- Streams are sequences of elements of an arbitrary type (e.g., bits, characters, images, etc.). In this sense, they can model both static and dynamic content. The first includes, for example, textual material, while the later might be, for example, a presentation of a digital video, or a sequence of time and positional data (e.g., from a GPS) for a moving object.
- A structure specifies the way in which parts of a whole are arranged or organized. In digital libraries, structures can represent hypertexts, taxonomies, system connections, user relationships, and containment, to cite a few.
- A space is a set of objects together with operations on those objects that obey certain constraints. The combination of operations on objects in the set is what distinguishes spaces from streams and structures. Since this combination is such a powerful construct, when a part of a DL cannot be described well using another of the Ss, a space may well be applicable.
- Scenarios can be used to describe external system behavior from the users point of view; provide guidelines to build a cost-effective prototype; or help to validate, infer, and support requirements specifications and provide acceptance criteria for testing.
- A society is a set of entities and the relationships between them. The entities include humans as well as hardware and software components, which either use or support digital library services. Societal relationships make connections between and among the entities and activities.

Along with the 5S formalization, some tools ([11], [25], [19], [29]) were created for working with the 5S model: 5SL, 5SGraph, 5SGen, 5SQual and 5SSuite. 5SL is a language for declarative specification and generation of DLs. 5SGraph is a domain specific digital library modeling tool that was conceived in the original work with 5SL. 5SGen is a tool

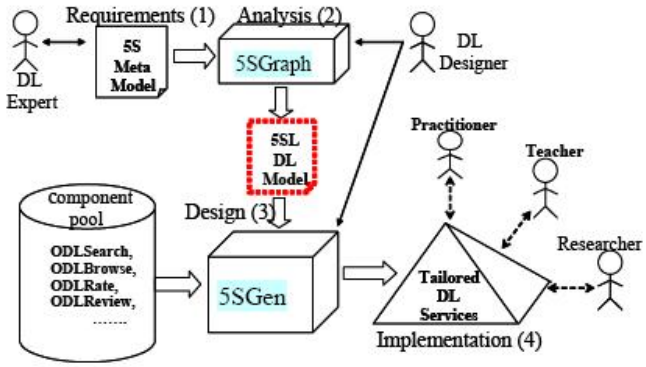

Figure 4: 5S related tools [9].

that generates much of the code necessary for a DL implementation, using as source a digital library defined in 5SL. 5SQual provides ways to verify the quality of digital objects, metadata and services. And 5SSuite toolkit is used to cover the process of union DL generation, including requirements gathering, conceptual modeling, rapid prototyping, and code generation. The 5SSuite tool consists of 5SGraph, 5SGen,and SchemaMapper (a tool which maps a local schema into a global schema for a union DL). But these tools still do not have support for complex objects with subparts and annotations, enabling linking among components. Figure 4 shows how the 5S framework tools support DL analysis and specifications. A similar approach can be adopted for COs.

The 5S framework was also used to elaborate the meaning of quality in DLs [12]. Quality was used in this paper to look for factors/variables involved in measuring concepts like digital object, metadata specification, collection, catalog and repositories. But the DL evaluation was done using other parameters too, as data/collection, system/technology, users and usage [8]. For each of these dimensions, major attributes were described for setting-up a meta-library. The same quality concept can be used in the formal evaluation for CO, regarding their completeness.

Another problem in DL is the one discussed by Murphy [20] - the specification a minimum DL reference model . Murphy used the 5S framework to provide specifications, with the minimum DL as a foundation of various extensions, serving as a base for a DL reference model, as shown in Figure 5. In the first extension there is CBIR (Content-Based Image Retrieval) service, the second one is a meta-model for superimposed information and finally the third extension deals with the DL generation based on the DL software, such as Dspace.

Gordon [13] for example, used 5SL, 5SGraph and 5SGen tools to develop a XML-based model for specifying the nature of DSpace digital libraries, using a part of the CITIDEL (Computing and Information Technology Interactive Digital Educational Library) DL. This case study has successfully generated a working DSpace instance with all the content from the CSTC collection.

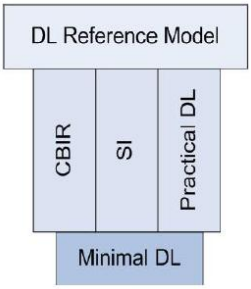

Figure 5: Reference model for a Minimum DL [20].

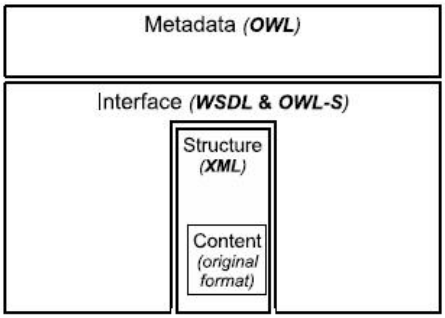

Figure 6: Digital Content Component Representation [22].

The 5SSuite was used in [26], treating another DL problem: the formalization of integration in DLs. A minimal metamodel was developed using the 5SSuite integration toolkit, illustrating the solutions for key problems using archeology DL as a case study to justify and evaluate the integration approach.

2.4 Digital Content Component

Data aggregation facilitates data exchange, annotations and representation. A Digital Content Components (DCCs) [22, 23, 24] is an initiative to aggregate data through encapsulation. A DCC is a digital container that encapsulates any kind of digital component (executable or not, like spreadsheets, workflows, maps, software), exposes its functionality through interfaces and is described and classified through metadata extracted from taxonomic ontologies. It can be used in many ways, like component insertion, content insertion and software product construction. Components can be recursively constructed from composition of other components, based on a model which generalizes reuse content practices of decomposition -storage/retrieval -composition. The connection of two DCCs need other intermediate components, responsible for adaptations and conversions. Figure 6 illustrates a schematic example of a DCC format: (i) the content in its original format (data or code) or a DCC composition; (ii) the declaration, in XML, of a structure defining how DCC components relate to each other; (iii) a specification of DCC interfaces; (iv) metadata to describe functionality, applicability, use restrictions, etc., using OWL.

The DCC format has its conception derived from content

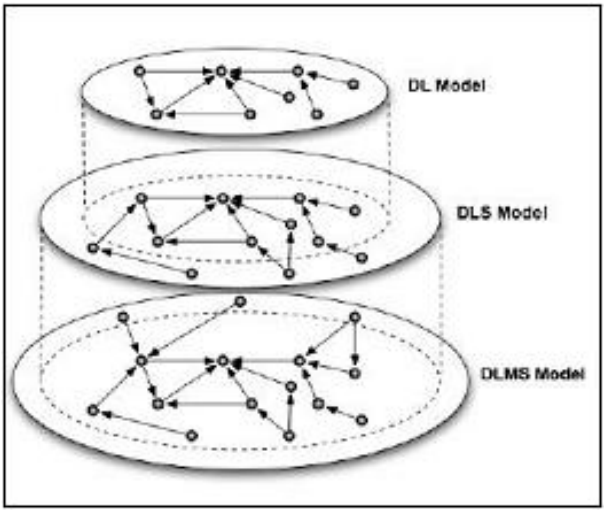

Figure 7: Hierarchy of conceptualizations in Delos [6].

package and reuse standards. Each Deployment DCC is stored inside a ZIP file. The DCC management structure is a generalization of the content reuse manifest file structures (RAS, MPEG-21, METS, and IMS CP). Even if the manifest structures are organized in different ways (like OAI-ORE [17]), they deal with the same problem and the same basic concepts. Therefore, it is possible to map the main structural elements of one manifest to another, if the particularities of each reused content are disregarded.

We plan to use DCC in this research proposal to encapsulate, delimitate, organize and describe the components of CO. And DCC can support 5S in order to organize structures and components, looking for a better interoperability within DL services.

2.5 Other Approaches For Managing COs

A logical data model is proposed in [28] for digital objects, but with repositories. Here several concrete versions of an object are treated as one single digital object. Therefore an object can be of two types: a top object that defines the logical object and shares the globally unique identification, or a component of a logical object. The main problem of this proposal is its storage strategy. The storage of one digital objects consists of several files. We still can not aggregate parts of an object, or annotations, for example, since it is located directly in the repository.

The Delos approach ([6], [2]) uses a framework of three tiers to represent three levels of abstraction: the DL (Digital Library), the DLS (Digital Library System) and DLMS (Digital Library Management System), as shown in Figure 7. The first layer is where the digital objects are kept; the second one manages the software application components, providing useful services to users with the support of a DLMS. Here the three levels are hierarchically related: so are their models, i.e., the DL model is included in the DLS one, and later is included in the DLMS model.

There are object oriented approaches with frameworks too [10]. It includes an hypermedia digital library design methodology that guides the implementation of digital libraries over the object oriented database system. The approach was OOHDM (Object Oriented Hypermedia Design Model), distinguishing four main design phases: 1) conceptual design (specification of hyperbase basic classes); 2) navigation design (specification of navigating paths); 3) user design; and 4) implementation.

The mentioned approaches still do not focus on complex objects, not being suitable for characterizing their multimodal behavior, their management and flexibility with new formats and encodings, including annotations and management of parts of documents.

3. PRELIMINARY WORK

In this section we present prototype applications and initial case studies to to understand existing requirements of applications that need to use COs.

We begin by discussing a Tablet PC image annotation tool, followed by the NOU-RAU system and a multilingual DL prototype.

3.1 Tablet PC image annotation and retrieval tool in parasitology domain

The project Deployment and Assessment of an Image Annotation and Retrieval Tool, including for Biodiversity, with financial support from Microsoft, is a joint collaboration between Institute of Computing at Unicamp and Virginia Polytechnic Institute and State University, USA. This project is funded by Microsoft Research under the Tablet PC Technology and Higher Education program. The main objective of this project is the specification and implementation of an application for image support annotation and search (based on textual and visual description) on biodiversity domain. Please check http://si.dlib.vt.edu/ for further details on the project.

Part of the project was conducted at Unicamp, with parasite images, with the objective of helping the comparison of morphological characteristics in different species, as shown in Figure 8. The comparison of species is important in Parasitology because details like shapes (as round or stick), forms (as pregnant or mature), number of teeth, etc., are decisive to their classification.

The objectives of these experiments were to validate the application tablet PC in the new domain, as an alternative method to teach concepts that dealed with images and annotations. After the deployment of this tool, users were able to:

- Use the pen writing capability of tablet PCs to:
  1. Create marks in images.
  2. Write annotations and associate them with marks and images.
- Link marks to other multimedia information such as text annotations and descriptions, audio/video records,

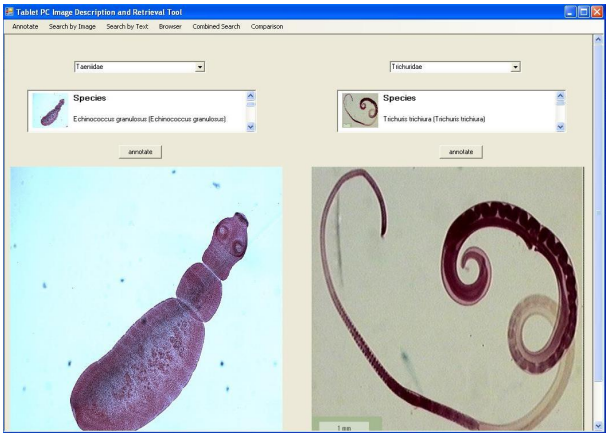

Figure 8: Comparison tab which allows annotations on two different images.

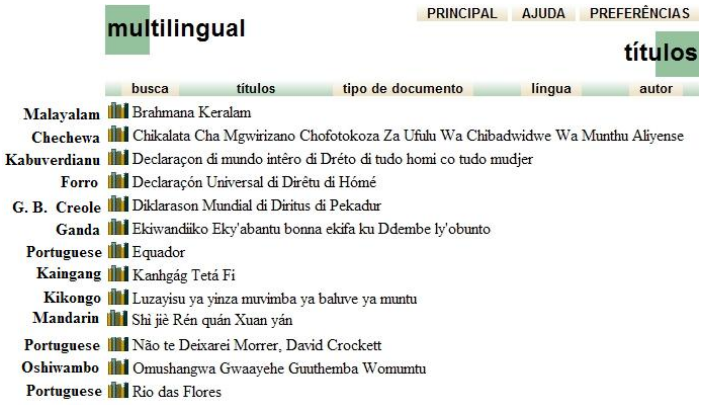

Figure 9: List of titles available at the first prototype with eleven languages.

other images, and marks in other (multimedia) documents. Browse and search over marks and associated data, using text- and content-based retrieval mechanisms.

The project helped to understand the complexibility of the aggregated information and how part of documents can be connected, annotated, and accessed.

3.2 Nou-Rau: Digital Library of Theses and Dissertations of Unicamp

The Nou-Rau system (available at http://libdigi.unicamp.br/) centralizes digital theses of students, with online availability for the community. Using the main information available at theses (like author, title, year, etc.), the user can navigate in the website, search for theses and dissertations, and download theses, after filling some basic informations.

After students graduated (or finish their masters, PhD or specialization), they are required to send the university the digital document resulted of their studies. While more theses are aggregated in the system, some are restricted to user access, due to research rights or restrictions to specific products. Some problems previously mentioned are present in the system, like the quality of metadata and the integration between other DLs. The quality problem is verified for older theses, since it is difficult to get the complete metadata information (like who were the professors that were present for the theses defense). The integration problem is verified when the system connects by harvesting with a brazilian network of libraries.

The Nou-Rau system needs to aggregate new digital objects (like audio, text and images) related to the theses. Furthermore, related services will be updated, and the DL management modified. It helped us to understand the requirements and problems related to complex objetcs.

3.3 Multilingual Digital Library for Portuguese-speaker countries

With the integration of Brazilian digital libraries with harvesting initiative (available at http://bdtd.ibict.br/), Unesco representants at Unicamp started a new project for integrating minority languages from portuguese-speaker colonies in a Digital Library.

An initial prototype was made with 11 languages (chechewa, forro, ganda, guinea bissau creole, kabuverdianu, kaingang, kikongo, mandarin, malayalam, portuguese and oshiwambo), with pdf, doc and mp3 files. Figure 9 shows titles of available digital objects, in eleven languages. The prototype used Greenstone software and is at the initial phase.

The project will help to understand the integration of multilingual CO in DL (since one CO might have several versions, each one in a different language); CO integration (since the universities intend to be integrated in the network); and stand-alone DLs (since some communities have limited support and can not be always online).

4. RESEARCH QUESTIONS

The following research questions will be addressed in this work:

- How to formalize a definition of the connection/aggregation between COs and their services?
- How to formally define an unified framework for managing COs in DLs? How to define a minimum CO?
- How to manage and verify the schema/system/components flexibility (new format types can be created, resulting in new collections, catalogs and services)?
- How to deal with integration problems (like integration of two different complex objects into a single model) and services from a CO perspective?
- How to define quality measures that can be used to evaluate COs? And how to match these measures in a flexible data model?
- and finally, how can available models and formalization can be used with COs, annotations, digital library services, and components flexibility over the years?

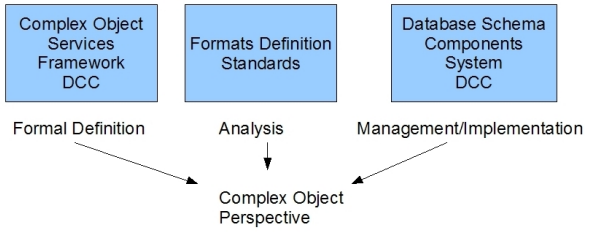

Figure 10: Concepts that will be verified through a complex object perspective

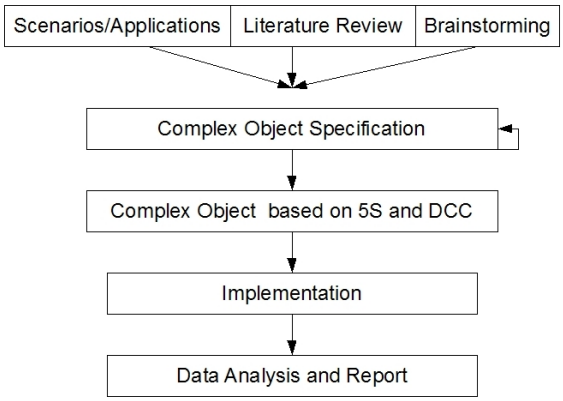

Figure 11: Outline for research methodology

Figure 10 shows the main concepts that will be investigated from the complex object perspective: the formal definition for the complex object, its services, DCC and framework; the analysis of formats restrictions and standards; the management/implementation of the database schema, components, system and DCC.

5. RESEARCH METHODOLOGY

In order to develop an unified framework to manage multimodal COs in DLs, verifying their formal characterization, we need a clear understanding of an SI-DL with different data types. We claim that the 5S framework formalization and DCC encapsulation will provide effective solutions to CO in DL. This hypothesis leads to the following considerations (listed at Figure 11):

- Explore CO from a theoretical perspective, learning from previous work and literature on DL an related fields. We will verify their formal definition, the integration and aggregation of the components, the flexibility with new formats and standards, the quality and completeness of data, architecture and basic services, like browsing and searching. From this we will develop a preliminary specification of a CO model using the 5S framework. Refine these specifications as we apply and verify case studies;
- Specify a theoretical foundation for complex objects and services in digital libraries using the 5S framework. We will extend the 5S model to get precise definitions for components in a CO;
- Evaluate this data model from a practical perspective, with the DCC aggregation, using images, annotation and data encapsulation (under a minimal complex object model evaluation). A prototype will be specified and implemented, with services like browsing and searching, looking for facilities in the data quality and data integration. We also plan to verify the usability of the prototype;
- Verify the data analysis and report outcomes of evaluations.

All approaches will interact with literature available and existing solutions in the DL area. The practical perspective can be improved based on feedback from our studies and the prototype that will be analyzed.

Our contribution is to analyze a digital library through CO perspective, looking for a unified framework to manage multimodal complex objects in digital libraries. We will verify the formal characterization of CO under several approaches, as definition, integration, flexibility, quality, architecture and services.

Many DLs include the manipulation of objects, but not their aggregation (like a complex object), with the use of annotations. There are real applications which need support for COs, with new mechanisms for managing data, creating references, links and annotations; or even cluster/organize complex digital objects. The main contributions of this research will be:

- Identify and abstract information elements and essential funcionality for CO;
- Provide formal and precise definitions for CO and related concepts, using the 5S formal definition, 5S tools, and DCC aggregation;
- Develop ways and verify services to work with aggregation of information over multimedia data, cooperating in one system;
- Case Studies explaining the use of the datamodel;
- Evaluate the use and usefulness of a CO model in a DL and complex application context using a prototype. The specification will have basic services like browsing and searching;
- Analysis and results of case studies.

6. ACKNOWLEDGMENTS

I am very thankful to my advisor Dr. Ricardo da Silva Torres, for guiding me through the development of this research and for giving me comments on this paper. I thank the Parasitology group at the Biology Department (Rubens Riscala Madi, Marlene Tiduko Ueta and Silmara Marques Allegretti) and the Chairholder of the UNESCO Chair Multilingualism in Digital World (Claudia Wanderley) for their collaboration in this work. The author thanks CNPq - BioCORE Project, HP Technology for Teaching, CAPES, and FAPESP for financial support.

7. REFERENCES

[1] Agosti, M., Ferro, N. *A Formal Model of Annotations of Digital Content.* Transactions on Information Systems (TOIS), vol.26, no.1, p.1-55, 2008.

[2] Agosti, M., Ferro, N. *Annotations: A Way to Interoperability in DL.* ECDL '08: Proceedings of the European conference on Research and Advanced Technology for Digital Libraries, p.291-295, 2008.

[3] Allinson, J., Johnston, P., Powell, A. *The Eprints Application Profile: a FRBR approach to modelling repository metadata.* OR '07: Second International Conference on Open Repositories, 2007.

[4] Bekaert, J., Hochstenbach, P., Van de Sompel, H. *Using MPEG-21 DIDL to Represent Complex Digital Objects in the Los Alamos National Laboratory Digital Library.* D-Lib Magazine, vol.9, no.11, 2003.

[5] Burnett, I. S., Pereira, F., R. Van de Walle, and R. Koenen *The MPEG-21 Book.* John Wiley & Sons, 2006.

[6] Candela, L., Castelli, D., Ioannidis, Y., Koutrika, G., Pagano, P., Ross, S., Schek, H.J. and Schuldt, H. *A reference model for Digital Library Management Systems Interim Report.* Delos deliverable, no. 1.4.2, 2006.

[7] Digital Library Federation *Metadata Encoding and Transmission Standard (METS), available at http://www.loc.gov/standards/mets,* accessed on February, 2009.

[8] Fuhr, N., Hansen, P., Mabe, M., Micsik, A., Solvberg, I. *Digital Libraries: A Generic Classification and Evaluation Scheme.* Proceedings of the European Conference on Digital Libraries, p.187-199, 2001.

[9] Gonçalves, M.A. *Streams, structures, spaces, scenarios, societies (5S): A formal digital library framework and its applications.* Ph.D. dissertation, Virginia Tech, Blacksburg, 2004.

[10] Gonçalves, M.A., Medeiros, C.B. *Constructing Geographic Digital Libraries using a Hypermedia.* Multimedia Tools and Applications, vol.8, no.3, p.1-18, 1999.

[11] Gonçalves, M.A., Fox, E.A. *5SL: a language for declarative specification and generation of digital libraries.* JCDL 2002: Proceedings of the ACM/IEEE-CS Joint Conference on Digital libraries. New York, NY, USA: ACM Press, p.263-272, 2002.

[12] Gonçalves, M.A., Fox, E.A., Watson, L.T., Moreira, B.L. *What is a good digital library? - A quality model for digital libraries* Information Processing and Management, vol. 43, no. 5, p. 1416-1437, 2007.

[13] Gorton, D. *Practical Digital Library Generation into DSpace with the 5S Framework.* Master of Science dissertation, Virginia Tech, Blacksburg, 2007.

[14] Grossman, R., Qin, X., Xu, W., Hulen, H., Tyler, T. *An architecture for a scalable, high-performance digital library.* MSS '95: Proceedings of the IEEE Symposium on Mass Storage Systems, p.89, 1995.

[15] Krafft, D. B., Birkland, A., Cramer,E. J *Ncore: architecture and implementation of a flexible, collaborative digital library.* JCDL '08: Proceedings of the ACM/IEEE-CS Joint Conference on Digital libraries, p.313-322, 2008.

[16] Lagoze, C., Payette, S., Shin, E., Wilper, C. *Fedora: an architecture for complex objects and their relationships.* Proceedings of the Joint Conference of Digital Libraries, vol.6, no.2, p.1432-5012, 2006.

[17] Lagoze, C., Sompel, V. de H., Johnston, J., Nelson, M., Sanderson, R., Warner, S. *Open Archives Initiative (Oct. 17, 2008). ORE User Guide Primer.* available at http://www.openarchives.org/ore/1.0/primer.html.

[18] Nelson, M.L., Sompel, H. Van de. *IJDL special issue on complex digital objects: Guest editors' introduction.* International Journal on Digital Libraries, vol.6, no.2, p.113-114, 2006.

[19] Moreira, B., Gonçalves, M. A., Laender, A. H. F., Fox, E. A *5SQual: a quality assessment tool for digital libraries.* JCDL '07: Proceedings of the ACM/IEEE-CS Joint Conference on Digital libraries, p. 513-513, 2007.

[20] Murthy, U., Gorton, D., Torres, R., Gonçalves, M., Fox, E.A. Ad Delcambre, L. *Extending the 5S Digital Library (DL) Framework: from a minimal DL towards a DL Reference Model.* ACM IEEE Joint Conference on Digital Libraries, 2007.

[21] Nelson, M.L., Argue, B., Efron, M., Denn S., Pattuelli, M.C., *A survey of complex object technologies for digital libraries.* Tecnical Report, 2007.

[22] Santanchè, A. *Self Describing Components: Searching for Digital Artifacts on the Web.* PhD thesis, IC-UNICAMP, Campinas-SP, Brazil, 2006.

[23] Santanchè, A, Medeiros, C. B. *A Component Model and Infrastructure for a Fluid Web.* IEEE Transactions on Knowledge and Data Engineering Journal, vol.19, no.2, p. 324-341, 2007.

[24] Santanchè, A., Dourado, P. and Ferreira, P. *Representação Unificada de Objetos Digitais Complexos: Confrontando o RAS com o IMS CP.* III Workshop de Bibliotecas Digitais, 2006.

[25] Shen R. *Applying the 5S Framework to Integrating Digital Libraries.* Phd Thesis, Virginia Tech, Blacksburg, 2006.

[26] Shen R., Vemuri N. S., Fan W., Fox, E.A. *Integration of complex archeology digital libraries: An ETANA-DL experience.* Information Systems, vol.33, Issues 7-8, p.699-723, November-December 2008.

[27] Smith, M., Barton, M., Bass, M., Branschofsky, M., McClellan, G., Stuve, D., Tansley, R., Walker, J. H. *DSpace An Open Source Dynamic Digital Repository.* in D-Lib Magazine. vol.9, no.1, 2003.

[28] Zierau, E., Johansen, A. S. *Archive Design Based on Planets Inspired Logical Object Model* ECDL '08: Proceedings of the European conference on Research and Advanced Technology for Digital Libraries, p.37-40, 2008.

[29] Zhu, Q., Gonçalves, M., Shen, R., Cassel, L., Fox, E.A. *Visual Semantic Modeling of Digital Libraries.* Proceedings of the European Conference on Digital Libraries, vol.2769, p. 325-337, 2003.

[30] Witten, I. H., Bainbridge, D. *A retrospective look at Greenstone: Lessons from the first decade.* Proceedings of the Joint Conference of Digital Libraries, Vancouver, Canada, p. 147-156, 2007.