Department of
Information Engineering
and Computer Science DISI

UNIVERSITY OF TRENTO - Italy

DISI - Via Sommarive 14 - 38123 Povo - Trento (Italy)
http://www.disi.unitn.it

FROM SOFTWARE TO ARTIFACTS: SUPPORTING THE CURRENT SCIENTIFIC KNOWLEDGE NEEDS

Ronald Chenu-Abente, Fausto Giunchiglia, and Luca Cernuzzi

December 2010

Technical Report # DISI-10-070

From Software to Artifacts: Supporting the Current Scientific Knowledge Needs

Ronald Chenu-Abente[1,2], Fausto Giunchiglia[1], Luca Cernuzzi[2]

[1] Dipartimento di Ingegneria e Scienza dell'Informazione – Università degli Studi di Trento
Via Sommarive, 14 I-38123 Povo, Italy.
{chenu, fausto}@disi.unitn.it
[2] DEI – Universidad Católica "Nuestra Señora de la Asunción"
Campus Universitario, C.C. 1683, Asunción, Paraguay.
lcernuzz@uca.edu.py

Abstract. Scientific papers are the current de-facto unit for scientific knowledge dissemination. Nevertheless, with the advent of the Web and is related technologies, the popularity of new types of dissemination methods (e.g. blogs, web pages and social networks) has also become difficult to ignore. In fact, several discussions are currently based on the advantages and limitations of both; the traditional paper model and these so called Science 2.0 new web-based opportunities. There is no discussion, however, on the fact that both of these types of documents are in fact used to disseminate and advance scientific knowledge. This paper presents a metadata-enabled model based on Scientific Knowledge Objects that aims to represent and capture the knowledge from both the traditional and the new types of these scientific artifacts. The inspiration for this model is mainly drawn from the comparison between the scientific artifacts and software, as several of the properties and processes currently used for software creation and management are deemed positive and interesting for the scientific publication world.

Keywords: Knowledge representation, Scientific artifacts, Version control, Semantic Web, Document engineering.

1 Introduction

Scientific communities and scientific research in general are described in [1] and [2]. The name scientific artifact, in particular, refers to those creations of humans that convey scientific data and knowledge. The most common of these scientific artifacts is the scientific paper or article, which was introduced during the 17th century when the first academic journals appeared.

Currently, as detailed in [3] and not very unlike those early times, when an author wants to publish a scientific paper he has to submit a physical and or digital copy of it to an academic journal, where it goes through a process of Peer reviewing to determine if its publication is suitable (with similar process occurs in the case of submitting papers in conferences and workshops). This model has remained mostly undisturbed up to now, even with the transition to the electronic era reducing the costs implied in the dissemination process and the advent of the Internet and the Web providing new ways of contact and interaction. As such, several studies (for example [4] and [5]) have been carried out to find the existing limitations on the creation, dissemination, evaluation and credit attribution of these scientific artifacts.

Furthermore, new types of less formal artifacts like web pages, blogs, comments, bookmark sites among, others have also been increasingly popular in scientific environments. While these emerging web-based new types of scientific artifacts are generally less well-regarded than the traditional ones, it is nonetheless, irrefutable that they also are being used to disseminate discuss and structure and, ultimately, advance scientific knowledge. These web-based scientific artifacts have also bridged a little the distance existing with another widely used type of artifact, the software artifacts.

Software artifacts refer to those that contain encoded instructions for the use of a computer and have also evolved over the years in an almost parallel fashion to scientific artifacts. Therefore, one of the key insights of this work is to draw comparisons between the handling of scientific knowledge and the handling of software development. For example, both cases involve collaborative work and exchange of ideas among project members and both normally need several iterations until they reach a final version. Nevertheless (perhaps needlessly) the treatment and evolution for both types of artifacts is very different, as shown in the following examples:

- *Development*: scientific knowledge artifact creation seems to be mostly still based on the Waterfall model, which makes it difficult to adapt to the changes and requirements that are frequently present in the research they represent. While, on the other hand; several other creation, development and distribution models are available and widely used for software artifacts. Creation models, such as extreme programming, and other agile software development models could be also applied to the production management of scientific knowledge as an attempt to bridge several of the delay and collaboration issues existing in the current approach.
- *Distribution*: when considering the great variety of distribution models currently available for software artifacts, it seems like the (still mainly peer-reviewed) scientific artifact distribution has lagged behind in its evolution. A clear example of this is the software-assisted distribution of artifacts through social networks. These artifacts can be created/modified and distributed at the same time, which is a notable change from previous approaches.
- *Versions*: version control is normally used on software artifacts to enable collaboration or manage their contents. Furthermore, version control is also becoming increasingly used on non-software artifacts and is now commonly present on all the major document edition software. Nevertheless, the concepts of fork and branching (separating from a main version to diverge from it) and incremental releases, are also relevant and being increasingly used for scientific artifacts.

As the main contribution of this paper we would like to propose the Scientific Knowledge Object (SKO) model as new way of representing scientific artifacts that would help introduce in them many of the desirable qualities (e.g. composability, evolvability, ease of annotation, etc.) that are currently present in software artifact and the software creation process.

This paper first introduces a short review of the state of the art on scientific artifacts in Section 2, to then proceed in Section 3 with the explanation of the layered approach and evolution considerations used by the SKO model to represent the different types of information from current scientific artifacts. Finally, Section 4 focuses on the validation and test cases for the SKO model and Section 5 presents the concluding remarks with related future work and lines of investigation.

2 State of the Art

This section contains a short review of current existing work related to the creation, evolution, collaboration and dissemination of scientific artifacts.

2.1 General Models and Ontologies

Having an organized knowledge base that is specially constrained and adapted to the target domain as the cornerstone of the approach would help to enable several of its proposed objectives; like providing well-foundedness and reusability at the conceptual level ([6]). Furthermore, particular deference will be given to the modeling structural and conceptual of part/whole relations (as suggested in [7]) because partial or whole sharing and reuse of scientific knowledge is an important objective.

The domain of relevance in this case is the one made of scientific artifacts and their related entities (e.g. authors, conferences, journals). As an interesting approach at scientific domain ontology building, SWRC was devised to empower semantic web services for scientific communities and is introduced in [8].

Finally, the existence and use of the previous and several other scientific-encoding ontologies makes important to devise methods of reuse and communication between them (as explored in [9]).

2.2 Semantics and Discourse Markup

In the area of markup, or aggregation of meta information to content, HTML is the most well-known example. Furthermore, XML (eXtensible Markup Language) is a general-purpose specification that may be used to create custom markup languages that are ultimately used to represent various types of information. Some other more focused approaches include RDF[1] (Resource Description Framework), for modeling information in web resources; and OWL (Web Ontology Language)[2], a family of

[1] http://www.w3.org/RDF/

[2] http://en.wikipedia.org/wiki/Web_Ontology_Language

knowledge representation languages for authoring ontologies. More towards the modeling and specification of scientific publications; a general view is given in [10], with the following deemed among the most relevant specific formats:

- *The ABCDE Format [11]*: provides an open standard and widely reusable format for creating rich semantic structures and annotations for scientific articles. ABCDE stands for Annotation, Background, Contribution, Discussion, and Entities respectively; which are the key semantic annotation types that this model uses to enrich regular artifacts.
- *ScholOnto*: the Scholarly Ontologies Project[3] aims to deploy an infrastructure that is able to represent research articles as 'claims' and situate them within a discourse/argumentation otology. This would facilitate providing services for navigating, visualizing and analyzing the network as it grows (for example, [12] and [13]).
- *SALT*: the Semantically Annotated LaTeX[4] format provides a semantic authoring framework which, much like ABCDE, aims at enriching scientific publications with semantic annotations. SALT is however based on three main ontologies (Document Ontology, Rhetorical Ontology, and Annotation Ontology), which represent the semantics, the linear structure and the small annotations found in scientific publications [14].

Some of these formats have been used with very promising results for the encoding of metadata and semantics of scientific artifacts. As such, providing ways of interacting or understanding them becomes important for any new model.

2.3 Collaborative Authoring Services

Social web services are based on communities of people that are brought together by the use of services like e-mail and forums, among others. Depending on the nature of these social networks the motivation for the creation of their knowledge artifacts also varies. On more communication-oriented networks like Facebook[5] users produce artifacts like photo albums to characterize people or places. Other, more content-oriented social networks like Wikipedia[6], focus on the creation of knowledge artifacts through collaboration of a very large number of persons.

In particular, the previous Wikipedia-model has been criticized because of its restriction of dealing only with well-known subjects (which disallows research) and because it favors consensus and popular information over accuracy. Nevertheless, other similar approaches (e.g. Swiki[7], Knol[8] and Ylvi[9]) introduce modifications to Wikipedia's formula to deal with these limitations and thus become better platforms

[3] http://projects.kmi.open.ac.uk/scholonto/index.html
[4] http://salt.semanticauthoring.org
[5] http://www.facebook.com/
[6] www.wikipedia.com
[7] http://wiki.squeak.org/swiki/
[8] http://knol.google.com
[9] http://www.cs.univie.ac.at/project.php?pid=268

for the collaborative creation of scientific artifacts. These concepts have been applied even for writing complete books (e.g. Wikibooks[10] and Dynamicbooks[11]).

These websites enable collaboration during the creation of the artifacts through version control on complex artifacts (where the subcomponents can be part of several higher-level wholes). The management of the evolution of the artifacts thorough version control has been explored in [15] and [16], and is further complicated by the introduction of semantic annotations and other elements that may also be subject to version control.

2.5 Unified Approaches

Nowadays, more and more online portals and software combine the concepts of both of the previously mentioned areas of social networking and semantic markup for research artifacts. Papyres [17], Cyclades[12], and HypER[13] are some applications for scientific publication sharing and collaborating at both data and metadata levels that are considered relevant for this line of research. On the other hand, Mendeley[14], Galaxy Zoo[15], Cohere[16], and CORAAL[17] are more focused on specific procedures for scientific artifact discovering and disseminations (e.g. indexing and search).

All the information collected from these previously existing approaches will be taken as the starting point and inspiration for the proposal of a model that would enable scientific artifacts to comply with the desirable qualities specified at the introduction.

3 The SKO Model

SKO stands for Scientific Knowledge Object. SKOs allow a variable granularity and multi-faceted representation of today's scientific resources and can be used to capture their evolution/maturity in time and compute credit attribution. In particular the information captured in the SKO may be used to both support standard metrics (e.g. citations) and new social-based metrics emerging from the use of Web2.0 technologies.

As shown in Figure 1, the SKO structural dimension is based on a multi-layered approach mainly aimed at enabling and facilitating the composition, reuse and collaborative creation of scientific resources. Furthermore, it provides the base for

[10] http://en.wikibooks.org/
[11] http://dynamicbooks.com/
[12] http://www.ercim.org/cyclades/
[13] http://hyper.wik.is/
[14] http://www.mendeley.com/
[15] http://www.galaxyzoo.org/
[16] http://cohere.open.ac.uk/
[17] http://coraal.deri.ie:8080/coraal/

related works on improving the evolution, credit attribution, and search/navigation of these artifacts.

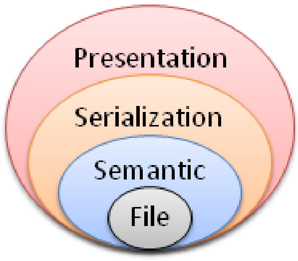

Figure 1. Structural layers of the SKO model

The SKO structural dimension is based on a multi-layered approach mainly aimed at enabling and facilitating the composition, reuse and collaborative creation of scientific resources. Furthermore, it provides the base for related works on improving the evolution, credit attribution, and search/navigation of these artifacts.

3.1 Representing Basic Resources

The file layer, shown at the base of Figure 1, is the foundational layer and main connection to well-established and commonly-used content and standards. This layer contains the actual content or data from the scientific resource. The file layer node is then simply defined as shown in Figure 2.

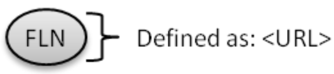

Figure 2. File Layer Node (FLN) definition.

An interesting property emerging from this is that the model is able to handle any scientific resource that has an URL, and not only scientific publications in the conventional sense. The File Layer can then be used to select not only a wide variety of traditional scientific knowledge artifacts (e.g. books and presentations) but it also opens the opportunity for other types of scientific artifacts (e.g. blogs, wiki, web pages), sub parts of these artifacts (e.g. figures, tables) and even experimental data.

3.2 Representing Metadata, Semantics and Structure

Found on top of the file layer in Figure 1, the semantic layer adds attributes and relations to the content from the file layer. This information is used to specify the context and concepts to which the content from the file layer refers and ultimately arrange all the content from a scientific resource into a graph-like structure.

The semantic layer nodes are the approach’s main way to enrich a file node with metadata and semantics that apply exclusively to it. Semantic layer nodes (SNL) are defined in Figure 3

Figure 3. Semantic Layer Node (SLN) definition.

In the definition of SLNs:

- *SURL*: stands for Semantic URL, and is used as a mean to refer to the where semantic information of an object is being stored. For the purposes of the current state of this work these are implemented as regular URLs pointing to a file containing semantic information (in xml or other similar format).
- *FLN*: this refers to the file layer node, as previously defined, that is component of this semantic node. As such, both SURL and FLN are used to define the link between the file and semantic layers.
- *{Attributes}*: a possible empty set used to define information that is common to all semantic nodes. Examples of these attributes include “title”, “authors” and other common metadata for scientific resources. A complete listing of these attributes is beyond the scope of this paper but it will be included in future related works.

Semantic annotations are also introduced in this layer to capture the semantics that exist *between* objects (e.g. object A is better than object B) or that apply to more than one object at the same time (e.g. I like objects A and B) and, finally, to improve the approach’s flexibility by capturing less common and user-defined concepts. These semantic layer annotations (SLA) are defined in

Figure **4**.

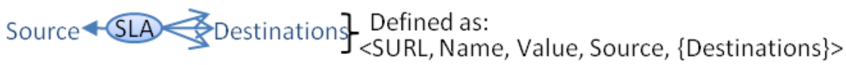

Figure 4. Semantic Layer Annotation (SLA) definition.

Where SURL is defined as before and:

- *Name*: defines the name of the custom attribute that the annotation encodes.
- *Value*: defines the value of such annotation.
- *Source*: contains the locator of the object to which the annotation mainly refers.
- *{Destinations}*: optionally contains the location of other objects to which the annotation refers.

Note that these annotations can be used to represent several types of tags and relations and that additional rules may be attached to each name of semantic annotation (e.g. the “Is related” annotation must have at least one destination).

However for simplicity's sake this specification will be specified in future work. Figure 5 presents an example of these concepts.

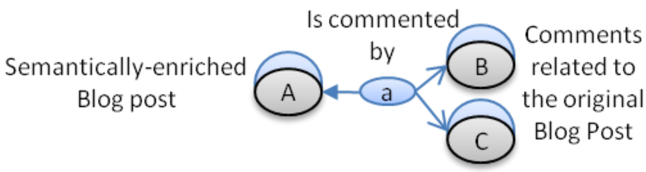

Figure 5. Representation of a commented blog post.

The semantic node *A* from Figure 5 represents a blog post by having the actual text content of the post pointed by its file layer component and the metadata (e.g. author, posting date, etc.) stored in its semantic layer component. At the same time, the annotation *a*, points to the semantic node *A* as its source and to *B* and *C* as its destinations. As *a* has its name value set to "is commented by" this declares both nodes *B* and *C* as being comments of the original blog post represented in *A*.

With semantic layer nodes covering the individual and most common attributes and semantic annotations covering the other cases, it would seem that no other object would be needed to capture semantics. Nevertheless, there is also the (quite common) case in which the attributes to capture emerge from the aggregation of two or more resources.

The Semantic Layer SKOs (SLS) are the structure used by the approach to capture the gestaltic (i.e. that emerge from the aggregation) semantics and are defined in Figure 6.

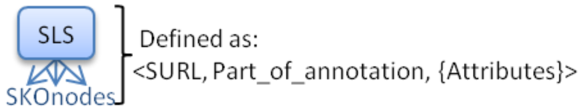

Figure 6. Semantic Layer SKO (SLS) definition.

Where all; SURL, and {Attributes} are as previously defined and:

- *Part_of_SLA*: represents a "part of" semantic annotation which joins the SKO with its immediate components.

It is also worth noting that the {Attributes} optional set would contain the basic semantic information that applies to the aggregation of all the components pointed by the Part_of_SLA component. Figure 7 presents an example of the aggregations of nodes into a SKO.

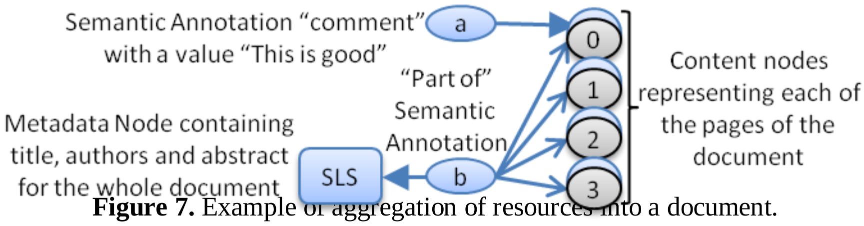

Figure 7. Example of aggregation of resources into a document.

The example of Figure 7 shows how all the structures defined thus far interact, as the "part of" *b* annotation declares the aggregation of resources by joining the *SLS* with the numbered semantic nodes and the *SLS* contains global information about this aggregation. An additional "comment" annotation *a,* was also added to demonstrate further uses of *SLA*s.

The introduction of the semantic layer helps to capture concepts and relations from the scientific artifacts but also, by using the "part of" relation, it structures the resources and the knowledge that are components of the scientific artifacts into a tree-like structure. This structure helps with the navigation of all constituent resources and it will be used and expanded by higher-level layers of the approach.

3.3 Representing Ordering and Style

The third layer in Figure 1, the serialization layer is mainly aimed at selecting and organizing the content and the semantic metadata from the previous layers into a linear sequence of information, which is easier to understand by human consumers. Serialization layer nodes (LLN) are defined in Figure 8.

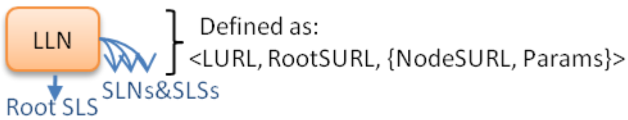

Figure 8. Serialization Layer Node (LLN) definition.

Where:

- *LURL*: similarly to distinction we made for addresses pointing objects of the semantic layer, this LURL (SeriaLization URL) points to a object that is from the serialization layer.
- *RootSURL*: points to the semantic layer SKO that acts as the root of the artifact that wants to be referenced (as shown in Figure 3). As such, both LURL and RootSURL are used to define the link between the semantic and serialization layers.
- *{Node SURL, Params}*: this set of pairs defines which semantic layer nodes are included in this serialization and how they are used in it.

Figure 9 contains an example of these concepts.

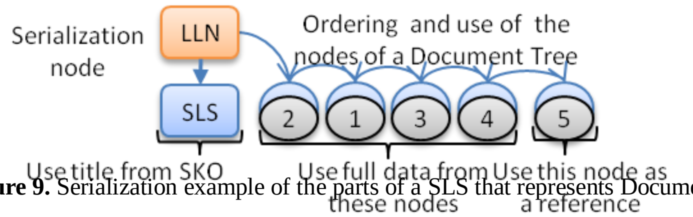

Figure 9. Serialization example of the parts of a SLS that represents Document

Similarly to what was shown in Figure 7, the *SLS* in Figure 9 has as parts the nodes numbered from *1* to *5*. It is through the use of the *LLN*, that the resources in the *SLS* are ordered into a new document by having the data from nodes *2*,*1*,*3* and *4* aggregated in that order and a reference to the data of node 5 is added at the end.

The serialization layer then takes structural tree defined at the semantic layer as the base and by selecting and taking certain components it is able to creates new (or a family of new) artifacts from the same resources. This is in fact similar to what happens in the software word, where several commonly used resources (e.g. libraries) are compiled by the use of a makefile into several different (but related) end applications. In this comparison, it is the serialization that takes the role of the makefile in specifying which concrete resources are going to be used and the way of using them for constructing new artifacts.

3.4 Representing Evolution

A State-based evolution is introduced to abstract away complicated properties from scientific artifacts (e.g. certification, persistence). Through an easy-to-understand metaphor based on the most commonly known states of matter, this evolution model introduces three discrete states for scientific artifacts. Table 1 introduces the Gas, Liquid and Solid states. Much like the physical states of matter, the same object/resource may have very different properties according to the state that it is in.

Table 1. Main properties of the three evolution states.

Property/State	**Gas**	**Liquid**	**Solid**
Maturity	Unfinished, Work-in-progress	Draft, Request-for-comments	Final
Certification	None	Author	Author and Certifying Authoritiy
Persistence	Unwarranted	Web	Web and Digital Libraries

In more detail the main three properties abstracted away in this system are:

- *Maturity*: maturity is the most representative and intuitive of all of state properties. Gas objects are normally used for highly fluctuating work-in-progress and deemed with not enough maturity to be considered serious, while solid objects are considered mature enough for being candidates for extending human knowledge or science. In the middle of this, liquid objects are considered to have some of their basic knowledge or science in place but still undergoing some adjustments.

- *Certification*: refers to the person or entity that takes responsibility and, eventually, credit for the content of the object. For solid objects, both the author and a certifying authority (e.g. a publisher a board of reviewers, etc.) certify and assume responsibility for the artifact.
- *Persistence*: refers to the method used to guarantee availability of the resources over a period in time. In a web-warranted persistence (liquid state) a web repository warrants that all liquid objects will remain available. On the other hand, in a “web and digital libraries” the object may be distributed and duplicated in digital libraries to further improve its persistence.

While published papers, hard-back books, among others are already existing and widely used examples of solid-state objects; gas-state artifacts are also ubiquitous if one also considers the content stored of personal computers (e.g. work-in-progress document, unordered notes). As such, one of the key points of this research is to enable and prove the utility of the introduction of the middle-level liquid state objects (e.g. request-for-comment, wiki-like discussions) as a way to improve the collaboration and early dissemination in the scientific process.

To implement these states, it is possible to use the structures defined in Section 2 to create semantic tags that are useful to identify and keep track of when an artifact is a version of other artifact. As in other version control systems, the same mechanisms could be used to determine whether an artifact was merged or split from previous ones.

4 Validation and Test Cases

This section will introduce two projects in which different levels of the presented theory will be used and eventually validated.

The first of the two is the Liquid Publications[18] (LiquidPub, introduced in [18]) project. LiquidPub is a Future and Emerging Technologies (FET) project that proposes a paradigm shift in the way scientific knowledge is created, disseminated and evaluated. To achieve these goals it aims to developing concepts, models, metrics and tools aimed to better serving people, researchers, publishers and even science as a whole.

Within this project the SKO concepts will be mainly used for providing a common resource and knowledge representation. The SKO theory would then act as a ‘knowledge bus’, allowing the different sources and platforms that the LiquidPub project needs to operate over a known set of attributes and services.

[18] http://project.liquidpub.org

As such, this would represent an important opportunity to validate the SKO structural theory and the power of representation of scientific resources and its metadata. Preliminary results have shown that SKO theory to be expressive, flexible and powerful enough to model any type of scientific artifacts and its evolution over the process of scientific production.

The second project in which the SKO theory is being applied is in the University of Trento's Artificial Intelligence Social Network (AISN), which aims to create a semantically-enabled and artifact-centered social network for the Artificial Intelligence community.

This project involves the conversion and enrichment of a big body of existing papers (in pdf format) to the SKO format, mainly through the use of extractors. Moreover, all this information will be offered within the context of a live artifact-based social network; which would allow users to comment, tag and link different pieces of scientific resources but also to associate them with persons and events.

This platform would enable several interesting semantic-based operations including:

- *Semantic Search*: depending on the level of annotation of the content this would allow to search for the key concepts of the paper and not only on the full text. Furthermore, according to the users' needs, all the metadata collected by any means (e.g. number of views, comments) may also be used to power the semantic search.
- *Relation Navigation*: encoding citation, author, affiliations and other relations existing between the entities related with scientific artifacts allows the creation of navigable graphs. This, in turn, substantially helps having access to the related information of a specific document or paper.
- *Recommendation*: by keeping track of the metadata attributes of a certain type of artifacts, it would be possible to combine search and navigation to offer recommendations of similarly characterized objects.

The first concrete application of this system is already in development and scheduled to go live by the middle of 2011. This is especially relevant to the validation the finer points of the SKO theory as it would cover all the structural layers from the theory explained in this paper and also test some of its evolution-representing aspects.

5 Final Words

In its ultimate purpose of becoming a multi-format aggregation resource that is not only able to access and aggregate content but also meta-content and discourse semantics from various proposed formats (like the ones in [11] and [14]), the SKO model's has evolved from an early version described in [19] to the one introduced in this paper. In addition, much like [20] discussed the SKO model's potential social impact as a scientific inclusion tool, this paper has introduced its motivation of bridging the gap existing between scientific artifacts and software artifacts. Achieving this would not only bring several desirable qualities of software but would also allow some software engineering methods to be applied in the context of documents (e.g. agile development, creation of open data/resource libraries, etc.).

Nevertheless, the SKO conceptual model is currently still evolving to cover the needs of the scientific artifacts. Besides the continued validation and refinement of the model, the following represent interesting options that will be considered for future exploration:

- *Part and wholes version control*: thanks to the clear identification of parts and wholes for each layer that the model includes, it would be possible for the version control system to evolve the individual components and the aggregated artifacts almost independently (thus creating Software family-like groups of related artifacts).
- *Layered version control*: as in the SKO model the artifacts are composed of several objects in different layers, it would be possible to version and change only some layers of a given artifact without introducing any changes to the rest. For example, changing only the presentation-layer objects of a document would create new presentations and styles of the same concepts; on the other hand, changing serialization-layer objects will create a different execution of the original artifact (with more or less the same concepts but different granularity/order).
- *SKO scripting language*: an XML-based specification can be defined to represent each of the proposed model's layers. This could be used as the starting point for the creation of a SKO-based editor for the management of scientific artifacts and artifact-families.

References

1. Latour, B.: Science in Action: How to Follow Scientists and Engineers through. Harvard University Press, Cambridge Mass., (1987) USA.
2. Kornfeld, W. A., & Hewitt, C.: The Scientific Community Metaphor. IEEE Transactions on Systems, Man, and Cybernetics , 11 (1981).
3. European Commission: Study on the Economic and Technical Evolution of the Scientific Publication (2006).
4. Casati, F., Giunchiglia, F., Marchese, M.: Publish and perish: why the current publication and review model is killing research and wasting your money. ACM Ubiquity (November 2006).
5. Parnas, D. L.: Stop the Numbers Game. Communications of the ACM , 50 (2007).
6. Guarino, N.: The Ontological Level. Revisiting 30 Years of Knowledge Representation. Lecture Notes in Computer Science. Volume 5600/2009, pages 52-67 (2009).
7. Artale, A., Franconi, E., Guarino, N., Pazzi, L.: Part-whole relations in object-centered systems: An overview. Data & Knowledge Engineering. Volume 20, Issue 3, pages 347-383 (1996).
8. Sure, Y., Bloehdorn, S., Haase, P., Hartmann, J. and Oberle., D.: The SWRC ontology - semantic web for research communities. In EPIA, pages 218-231 (2005).
9. Staab, S., Erdmann, M., Maedche, A.: Engineering ontologies using semantic patterns. In Proceedings of the IJCAI-01 Workshop on E-Business & the Intelligent Web, Seattle, WA, USA (2001).

10. Groza, T., Handschuh, S., Clark, T., Buckingham Shum, S. and de Waard, A.: A short survey of discourse representation models. In: Proceedings 8th International Semantic Web Conference, Workshop on Semantic Web Applications in Scientific Discourse. Lecture Notes in Computer Science, Springer Verlag: Berlin, 26 Oct 2009, Washington DC.
11. De Waard, A., & Tel, G.: The ABCDE Format Enabling Semantic Conference Proceedings. SemWiki (2006).
12. Buckingham Shum, S., Motta, E., & Domingue, J.: Modelling and Visualizing Perspectives in Internet Digital Libraries. Proceedings of ECDL'99: Third European Conference on Research and Advanced Technology for Digital Libraries (1999).
13. Uren, V., Buckingham Shum, S., Li, G., & Bachler, M.: Sensemaking Tools for Understanding Research Literatures: Design, Implementation and User Evaluation. Human Computer Studies , 420-445 (2006).
14. Groza, T., Handschuh, S., Möller, K., Decker, S.: SALT - Semantically Annotated LaTeX for Scientific Publications. Lecture Notes in Computer Science. Volume 4519/2007, pages 518-532 (2007).
15. Nguyen, T. N., Munson, E. V., Boyland, J. T., & Thao, C.: An infrastructure for development of object-oriented, multi-level configuration management services. ICSE '05: Proceedings of the 27th international conference on Software engineering , 215--224 (2005).
16. Thao, C., Munson, E. V., & Nguyen, T. N.: Software Configuration Management for Product Derivation in Software Product Families. ECBS '08: Proceedings of the 15th Annual IEEE International Conference and Workshop on the Engineering of Computer Based Systems (2008).
17. Amine, N., Hicham, H., & Esma, A: Papyres: A Research Paper Management System. E-Commerce Technology and Enterprise Computing, E-Commerce and E-Services, IEEE Conference and Fifth IEEE Conference (2008).
18. Casati, F., Giunchiglia, F., & Marchese, M.: Liquid publications: Scientific publications meet the web. http://eprints.biblio.unitn.it/archive/00001313/01/073.pdf (September 2007).
19. Giunchiglia, F., ChenuAbente, R.: Scientific Knowledge Objects V.1. Technical Report DISI-09-006, Ingegneria e Scienza dell'Informazione, University of Trento (2009)
20. Chenu, R., Jara, JJ., et al: Supporting Scientific Inclusion: Novel Approaches for the Representation and Assessment of Scientific Knowledge Objects. XXXVI Conferencia Latinoamericana de Informática (CLEI2010), Asunción, Paraguay, 18-22 Octubre (2010).