Metadata for Advanced Structures of Learning Objects in Mathematics

An Approach for TRIAL-SOLUTION[1]

(Extended) Version 2.0

Wolfgang Lenski, e-mail: lenski@informatik.uni-kl.de

Elisabeth Wette-Roch, e-mail: wette@informatik.uni-kl.de

The primary objective of the TRIAL-SOLUTION project is to use open platforms and tools to develop and validate a powerful new way of creating, delivering, and managing personalized education and training. This led to a new structure of educational material developed by the project.[2] This document concentrates on the problem of structuring such material by means of suitable meta-data descriptions.

As common meta-data element sets turned out to be insufficient for the material dealt with, a principle reflection on meta-data issues is performed. This transcends any ad-hoc solutions and ends up with a well-founded meta-data description element set preventing to construct a proprietary system without impact on general meta-data issues.

This paper introduces the theoretical fundaments for the formal structure for the new kind of material the project is going to establish and provides the corresponding meta-data element set. And it is at the same time meant to constitute the grounds for the functionalities to be developed in the course of the project.

I. Meta-data challenges in TRIAL-SOLUTION

To pursue its goal to develop and validate a powerful new way of creating, delivering, and managing personalized education and training the project especially designs new types of resources which are intended for personalized (re-)use. This involves two main procedures:

1. the decomposition of (existing) documents into learning objects[3]
2. the creation of new resources based on a library of learning objects.

In order to build new resources out of such a library, a meta-data description of learning objects and material created out of these is fundamental.

As the materials the project deals with are protected work, we are especially faced with the challenge to handle intellectual property and citation issues of a new kind. It turned out that we have to exhibit a theoretical background for the general citation problem as well.

I.1 Meta-data standards and TRIAL-SOLUTION

Each new compilation of a set of meta-data elements has to observe the already existing standards. The final meta-data structure for TRIAL-SOLUTION is compliant with these standards where appropriate but deviates where enforced by the new type of material we are going to deal with.

---

[1] funded by the EU under the Information Society Technologies (IST) program.

[2] An overview of the project, its scope, aims, techniques may be provided through http://www.trial-solution.de.

[3] for principles of this cf. [19]

Among the most prominent candidates for meta-data descriptions there are certainly the Dublin Core Metadata Element Set [1] (DC) and the IMS Learning Resource Meta-data Specification [4] (IMS) based on [2]. Though the approaches of DC and IMS differ in vocabulary as well as in structure, they both cover a kernel of meta-data elements which may be mapped to each other. In particular, both standards describe the necessary features to identify an object (published material) and at the same time meet the formal requirements needed to give an account of intellectual property.

Now considering composed documents of the type to be developed in TRIAL-SOLUTION, the current issues of DC and IMS meta-data turn out to be insufficient. Therefore the Trial-Solution project has designed its own meta-data specification[4]. But references of DC/IMS-type are of course still required as "the" model for dealing with ownership and originality issues where appropriate or necessary.

Whereas DC and IMS provide a general standard for reference and citation of documents, there are other, further specified standards, which are relevant in the context of the TRIAL-SOLUTION project. The IMS Content Packaging Specification [5] defines an open standard for the design of exchange formats with a focus on the aggregation and disaggregation of resources. This specification will be applied for the exchange of learning objects between the tools of the TRIAL-SOLUTION tool set. On the other hand there are particular schemes and best practice lists recommended for the description of mathematical resources. Based on the IMS meta-data specification, the American Mathematics Metadata Task Force (AMMTF) has developed a "Mathematics Metadata Base Scheme" [7] including a "Subject Classification for School and College Mathematics" [8]. TRIAL-SOLUTION primarily uses the Mathematics Subject Classification MSC2000 [6] which is the leading system for professional mathematics, but the meta-data model enables an integration of other existing classification schemes and controlled vocabularies as well. For German language the Schlagwortnormdatei (SWD) of the Deutsche Bibliothek Frankfurt am Main [9] provides a thesaurus of mathematical terms which is taken into account as well.

I.2 Basics of the TRIAL-SOLUTION approach

For developing our approach, we start with the concept of a document as an abstract entity. Such an abstract entity

- is created and composed out of individual units
- under a unifying idea and
- by utilization of work at hand, possibly created by others.

A collection volume may be viewed as a simple model for this document paradigm: It groups together several units which are meant to contribute to the overall topic illustrated by the title (and possibly detailed in the introduction) of the collection. The general principles of acknowledging intellectual property rights require reference to the individual contributors who are called authors (of subparts) as opposed to the editors of the whole volume. The work of the former ones is usually cited with reference to the collection volume whereas the work of the latter ones is cited with reference to the publisher of the volume.

In our approach, however, a more concise description methodology is necessary which should result in a finer granularity of meta-data elements. Our meta-data model is intended to capture the principles of document composition and decomposition within a general scenario. This

[4] see [12] and part IV of this document

scenario also includes higher generation composition when new documents are composed from parts of other documents that have been previously composed by others. While techniques supporting an iteration of the composition process are out of the scope of TRIAL-SOLUTION, the meta-data model we have designed supports more general structures and may constitute the basis for future developments.

I.2.1 Ontology of learning objects

The composition of given entities under a unifying viewpoint (which thus results in an abstract entity as mentioned above) may formally be considered as the construction of an *ontology of learning objects.* An ontology in this sense is characterized by organizing entities according to explicit relationships between them which results in a structured network of interrelated entities.[5] The meta-data description has to reflect this.

Starting from an analysis of typical book formats we end up with two different structures in the ontology of a document. The first one is describing its formal organizational structure (the *document skeleton*), consisting of chapters, paragraphs, sections, etc. along with textual passages denoting the respective (sub-) titles. This organization principle basically results in a tree-structure and may -- from an abstract point of view -- be regarded as the result of neglecting all dependencies and relationships but the PartOf-relation applied to the pure textual structure and enriched by annotations (the section, etc. titles). The second type of constituents of the ontology are the learning objects themselves, which are to be described by attributes particularly specifying their properties and interrelations.

It should be mentioned that according to the TRIAL-SOLUTION terminology the notion of a learning object (the technical term is "slice") is used in a rather broad sense. In principle, any document or document part which is identifiable and therefore can be isolated and might be subject for reuse may have the status of a learning object. As a consequence, learning objects are not restricted to elementary units like single text paragraphs or mathematical equations but may form complex structured resources on their own --- even complete books.[6]

As meta-data is a description methodology which does not include the reasons leading to a specific organization, meta-data must specify the *results* of an organizational process and not the principles of the process itself. This means that the structural part of the ontology is only to be described; the composition tools to enforce the dependency structure are not a meta-data issue. Consequently, we have to install a methodology to *describe* the ontology of a document (and not a procedure to *create* or *construct* it). Especially, this means that a precise definition of a learning object is out of the scope of a meta-data element set. Instead the meta-data descriptions rely on entities that are supposed to be given, either created by an automated procedure or manually by a person.[7]

I.2.2 Skeleton of a document

Typically the skeleton tree of a book is represented by its table of contents preceding the expository text and prescribing a subdivision of this text into single subparts. From a general point of view, all nodes in the tree correspond to entities (learning objects) representing the units a document is composed of. Hence all these entities are to be described by specific sets of meta-data. In particular, the leaves of the skeleton tree are supposed to be identified with

[5] A theoretical background for possible representations of ontologies may be found in [15].
[6] Note that this terminology differs from version 1.0 of this paper
[7] Some more considerations on this topic may be found in [19].

the atomic units of a document, whereas the inner nodes are interpreted as (pre-)composed subparts, which may have document status on their own right. In view of the functionalities, the learning objects at the leaves may also be considered as short full-text documents to which appropriate indexing methods could be applied (cf. [13], [14]).

A meta-data description for the entities in the skeleton tree of a document requires DC/IMS-style references in all stages of composition. Moreover, meta-data are needed to specify the relative position of objects within the source they are taken from and within the skeleton tree of the actually composed document. The first type of reference concerns correct citation of units used in composed material, whereas the second type hints more precisely at the exact subpart inside an entity that is actually referenced. Both types of reference are intended as a specification of an abstract locator function[8].

I.2.3 Learning object subtypes

In general, learning objects appear as entities of different types depending of the specific field. As TRIAL-SOLUTION deals with mathematics, the learning objects differentiate into several subtypes reflecting the character of the entities found in mathematical literature. Accordingly, this structure is based on a set of (binary) relations reflecting mathematically relevant dependencies and allows (external) descriptions (cf. the DTD-specification for TRIAL-SOLUTION in section IV.2). The learning object subtypes provide a categorization for units that have a certain elaborate status within the context of a document. These include mathematical categories like *definition*, *theorem*, *proof*, or *example*. The type list for TRIAL-SOLUTION has taken into account the list of learning resource types of the IEEE Draft Standard for Learning Objects Meta-data[9]. However, major changes have been performed to meet the requirement of the particular subject.

The document skeleton structure and the dependency structure of learning objects are intended to support different functionalities and thus differ in structure and kind of their basic relations. The skeleton of a document will typically be constituted by a PartOf relation and is hierarchically organized. Based on this hierarchy the TRIAL-SOLUTION project combines meta-data assignment with an application of inheritance mechanisms. This is a new approach to improve the efficiency and flexibility of meta-data assignment as a condensed description of each single unit is achieved. The default mechanism is top-to-bottom inheritance where each node in the hierarchy inherits the meta-data from its parent node unless values will be overwritten on a certain stage. But there are also applications of an inverse bottom-to-top inheritance procedure where meta-data at a node are accumulated from descendant nodes. In particular this applies to keyphrase assignment.[10]

Dependency relations between learning objects are described as relation meta-data within a *relation* element. Through additional attributes the *relation* element specification allows to distinguish relations of different kind and status. Examples are the *referred* and *required* relation which may be proposed, assigned or inherited. It is possible to add relations between different slices of one document or relations between different documents. All relations of this kind are intended to be used as guidelines for coherent document composition. The definition of an operational semantics for dependency relations is subject to experiments within the project and is reported elsewhere.

---

[8] For a further details on citation and identification/localization see part III of this document.
[9] This list is entailed in [4].
[10] For further details on keyphrase assignment see [13]

I.2.4 Generalized document types

It should be mentioned that this paper pursues a universal approach which is not bound to the kind of material TRIAL-SOLUTION delivers at present. Instead, it exceeds the problems raised in the present context of TRIAL-SOLUTION and prepares for future developments as well including higher generation document preparation. Let us just spread out the scenario which this paper addresses in addition to the project's present aims.

TRIAL-SOLUTION starts with a repository of sliced books and provides tools for the creation of personalized documents out of these slices. After generation TRIAL-SOLUTION delivers the documents in formats like postscript, PDF, or HTML. However, TRIAL-SOLUTION does not insert these documents into its stock of material; it is not intended to successively extend the offerings of material by including all document that have been generated out of the basic documents as new learning objects. Instead, TRIAL-SOLUTION creates the documents on the fly and does not store it permanently. However, these documents may circulate in the public and may thus well be basis for subsequent modifications or improvements by others -- and they may be referred to by other persons.

Since this paper is on meta-data and intellectual property issues which will be understood in a rather broad sense, it will consider such documents as material subject to (intellectual) property as well. Hence this paper will explicitly include more advanced problems arising from (re-)use of (parts of) such documents as another kind of learning objects in a different context. So arbitrary iterations of the generation process will be admitted for the meta-data description that is envisaged in this paper. We will call documents *virtual documents* in case that they are not created out of the basic offerings of learning objects of the TRIAL-SOLUTION repository but are generated on the basis of iterations of the new construction processes introduced by the TRIAL-SOLUTION methodology. This is motivated from the fact that these documents do not exist as precompiled entities ready for download from a server, e.g., but may possibly only exist as a description of their generation process. This overall position meets usual practice and prohibits our approach from being bound to a specific document generation technology.

II. Foundations

In the previous chapter we have sketched the general aims of TRIAL-SOLUTION. Before we turn to technical details of the TRIAL-SOLUTION meta-data structure we concentrate on some fundamental aspects which deserve special attention.

When dealing with the creation and dissemination of composed documents, questions of copyright, authenticity, and correct citation play a major role. Hence for TRIAL-SOLUTION special attention has been paid to design a meta-data system that enables a correct handling of these in documents that are composed from parts of other documents.

In general, copyright issues may be considered from at least three different points of view. The first one refers to the legal situation where rights governed by laws of a certain country are owned by a person or company. The second aspect concerns the actual protection and administration of rights by the rights holders. And the third viewpoint is regarding acknowledgement of intellectual property within the context of referencing and correct citation.

Within the TRIAL-SOLUTION approach, the legal situation of specific rights ("copyrights") is described by a particular set of rights meta-data extending the IMS Specification of rights. These meta-data provide information to trace the conditions of use for the individual parts of a sliced document. The respective element set is listed in part IV of this document and is not the main concern of this paper.

Instead, the following chapter addresses the problem of maintaining references to content that is subject to intellectual property in connection with sliced documents and document composition. Our approach is focused on a meta-data struture which is applicable to the general scenario including higher generation document composition. As common metadata element sets turned out to be insufficient in a way such that some local amendments would not meet the requirements, an overall reflection on meta-data in general is demanded which should go beyond ad-hoc improvements and rather request foundational considerations.

Starting from an abstract point of view we are looking for the principles which govern an appropriate meta-data structure. The central problem is to find a solid ground on which a systematic treatment may be based. We claim that these principles can be determined through the way we are (responsibly) dealing with external material. Hence an analysis of the principles underlying the usage of material will exhibit the basics for the desired meta-data structure. From a methodological point of view we develop the structures for a well-founded metadata element set from an analysis of the overall situation in which meta-data are involved. As a consequence, the structure and assignment of meta-data is no larger arbitrary but subject to abstract principles.

II.1 Background considerations

Our approach is committed to the pragmatic philosophy in the succession of C.S. Peirce (see [16],[17]). *Pragmatism* (or pragmaticism as it was called later on) tries to react on deficiencies identified in the philosophical tradition where especially the gap between pure epistemology and practice could not be bridged in a philosophically sufficient way. While even a rough outline of the philosophical insights would be out of the scope of this paper, we will instead briefly mention some features that may illustrate the influences this position has shown for the design of the meta-data structure we are developing in the sequel. It is intended to introduce into the main concepts and to evoke the spirit of the position laying behind.

According to the philosophical tradition originating from the (idealistic) position of Kant, knowledge is composed via perception and subsumption of concepts under categories. While sharing this general approach (at least in principle) so far, Peirce emphasizes at this point that the actual establishment of knowledge is performed by individuals and in addition results in an orientation in the ‘world’. This understanding gives rise to the idea that some even more fundamental structure or categories prior to the application of categories in the classical sense must be present to enable subsumption processes. These are the Peircean categories *firstness*, *secondness*, and *thirdness* [16]. They are thought of as *universal* categories, i.e. categories that are always present in every attempt to establish knowledge.

Because of their general epistemic status, they are also the basis of the semiotical relationships. According to the so-called *pragmatic semiotics* we have to deal with *signs*, *concepts*, and *objects* together with their relations. Concepts on the other hand must not be separated from the process of *establishing* knowledge and according to the pragmatic maxim

[16] thus may not be thought without possible consequences or actions associated with them in our imagination.

Philosophical ideas pre-structure the domain of things that may be dealt with. So these structures implicitly underly all phenomena and thus have to be taken into consideration in order to establish a cognitively adequate (information) model. But since the main focus of this paper is not on philosophical issues, we will interrupt the discussion of the philosophical background at this point. Instead, figure 1 is meant to serve as an 'icon' to represent the underlying ideas.

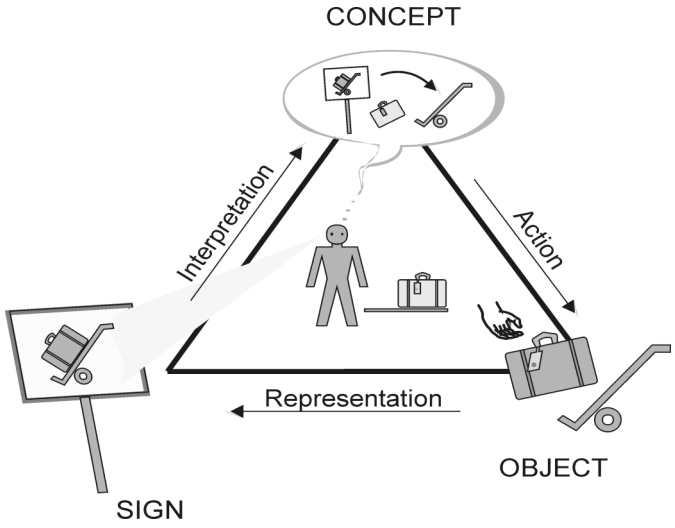

Fig. 1: An illustration of pragmatic relationships

In the sequel a purely analytical reinterpretation of the semiotical relationships on the basis of the three categories by Morris (cf. [17]) led to the distinction of *syntax*, *semantics*, and *pragmatics*. In the following we will adopt this analytical reduction of the philosophical background for our considerations.

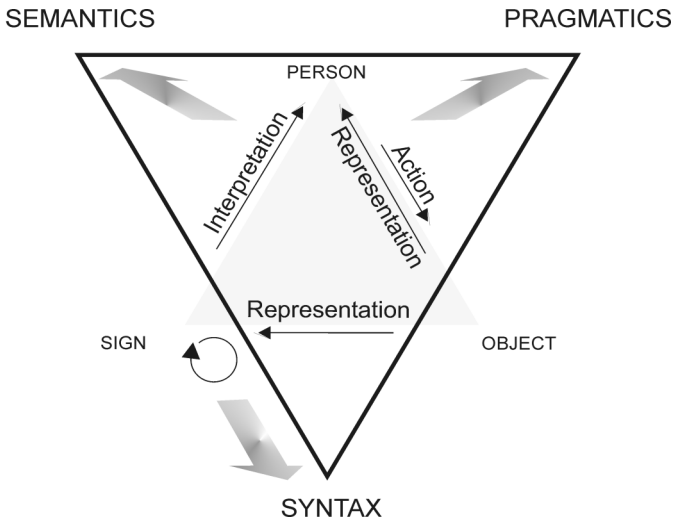

Fig. 2: Semiotic dimensions

Now interpreting this philosophical setting in the context of dealing with material in the sense of TRIAL-SOLUTION the main focus is on handling reference to given material. In an abstract sense this includes the question how the usage of material is communicated between

individuals. Such references may be loosely communicated using individual shortcuts in case a shared understanding of the persons involved is guaranteed or may at least be assumed giving the chance of subsequent inquiries. However, this is in the structure of private conversations. According to the philosophical background it is on the pragmatic level.

Untying these issues from the their bounds to concrete situations amounts to substituting the reference processes which have been settled between individuals beforehand by universal structures which are no longer bound to subjective understanding but provide a general reference methodology. This is commonly understood as *citation*. Hence *citation* is the semantic term; it is considered as a general resolution structure for requests for the underlying material -- independent of actual requests raised in individual communication processes. Its corresponding formal expressions are the meta-data. Hence meta-data structures are the syntactic counterparts of the semantic citation problem.

Finally, *principles* found in the context dealing with questions of adequate citation provide the required abstraction of concrete handling of external material and may thus constitute the desired background for establishing meta-data structures . Therefore we claim that once fundamentals of citation have been made explicit they will provide a justification for a well-founded meta-data description.

III Fundamentals of citation

In the first step we analyse the landscape of concepts dealing with citation issues and exhibit the posts that mark the field. The abstract structure of this field should then condense into requirement specifications for a well-founded meta-data structure. To prevent our approach from ending up in a highly artificial construct we demonstrate that our approach

(1) does not introduce a totally new paradigm but is in conformity with actual practice when applied to classical material,
(2) establishes the necessary refinements enforced by the elaborated treatment of learning objects developed by the TRIAL-SOLUTION project.

We start with the very elementary description of this landscape whose borders may be marked by the most abstract description of what we are dealing with:

Something generated by somebody and communicated to others. (1)

From this abstract description we then derive the whole topography of this landscape; we pave the ways and install connections between claims in an increasing granularity.

We first derive an immediate consequence by concentrating on the aspect of communication being bound to *something*. If emphasis is indeed on something , it must provide a (universal) background structure beyond subjectivity on which all communicational issues prove success and which admits universal inspection. This raises an organizational issue; so at first hand instances are required to perform this task. Thus the initial marks in our topography are concepts to implement these requirements. For this purpose we introduce the concept of a *repository*. This is meant — in an abstract sense — as an instance to which requests for material that has been referred to can be addressed and which in turn is able to provide the respective material for inspection.

Definition 1:
A *repository* is an entity which is able to deliver documents upon specific requests.

A repository is understood in a broad sense; it denotes any instance - be it an organization, a technical device like a Web-server or even a person -- that is able to resolve well-stated requests and to deliver the respective item. For example a publisher is able to deliver a book for which I have found a citation and which I want to inspect. From a functional point of view the publisher thus converts an abstract citation into an actual piece of data.[11] But a repository in a broad sense may also be my colleague next door who may hand over the manuscript she wrote in case I am able to explain it to her in a sufficiently specific way.

Note that from a methodological point of view a request is nothing than a citation used in an action to acquire the denoted entity. According to Morris (semiotic) reduction of the Peircean categories [17] a request may thus be considered just as the pragmatic relationship for which citation is the respective semantic term. In this context, the given definition also implicitly hints at the second task an adequate meta-data set must provide, namely to unambiguously denote and identify a given item inside a repository.

For a well-founded meta-data description we have thus to deal with *semantic* instead of *pragmatic* relationships. This means that we have to be independent from every actual pragmatic aspect of a request and instead focus on situations not depending on the involvement of (specific) persons. So my colleague next door may well be able to give me the document I had lent last week or, more generally, her manuscript titled My most recent work , but — as we have seen — this cannot be a model for general citation purposes. Instead, such a model would indicate to the public which sources had been used -- such that everybody can access the mentioned material and check the statements included in it. Transforming this methodological aspects into practice we end up with a first guideline.

Guideline for citations I:
Cite an entity in a way such that everybody can obtain it through a repository!

As a rather trivial consequence this requires that a citation must include a reference to a repository. It should also be mentioned at this point that an account of what a repository may actually resolve and deliver is mandatory. In general a publisher does not deliver a conference contribution contained in one of its proceedings volumes (and is not necessarily able to identify the proceedings volume for a specific conference paper at all[12]). So for a general discussion of citation we have to consider the nature of documents contained in a repository as well. This determines what the repository may be able (or willing) to deliver.

At this point we want to interrupt the course of our considerations and introduce a postulate which cannot necessarily be derived from abstract principles but we claim that it will yield assent as well. We call it a postulate as opposed to principles which inherit some aspect of necessity:

[11] Note that a sort of ‘handle’ is installed to organize this 'conversion' namely in most cases a bookshop is involved which intermediates between the request and the desired result.

[12] This is rather a task for structures surveying scientific developments such as a bibliography which organizes and structures the landscape of publications; cf. the *Bibliography of Mathematical Logic* [18] for the field of logic, the *Fachinformationszentrum Karlsruhe* FIZ Math running the *Zentralblatt für Mathematik und ihre Grenzgebiete/Mathematics Abstracts* or the *Mathematical Reviews* of the *American Mathematical Society*.

Postulate for citation:

Cite as simply as possible!

We now turn back to our main line of argumentation and concentrate on another aspect contained in our abstract description of what citations are dealing with. Citations bear two categorial aspects which focus on *somebody* and *something* in statement (1). These are grasped by the following concepts:

(I) *Authentication*:
Citations have to ensure that existing intellectual property rights are respected.

(II) *Identification*:
Citations have to ensure that the cited unit may be uniquely recovered.

These aspects are categorial in the sense that both should necessarily be incorporated in all systematic approaches to managing and referencing material. Categories claim to have a fundamental status on their own; they do not describe the way how they are to be realized. Instead, this is a question of a functional implementation in a concrete setting. Following this requirement, standard rules of adequate citation are derived from these abstract principles in the sequel to enforce their validity.

III.1 Authentication

We first discuss questions concerning *authentication*. Observing the obligations expressed by this principle we have to specify which parts of a document are subject to which copyrights or intellectual property and in which way they have to be referenced. As the learning objects of TRIAL-SOLUTION are copyrighted material it is most important for the project to develop solutions that meet the demands of the copyright's owners. But as the legal aspect of copyrights is tied to the terms and conditions determined by the repository the handling of this part is already settled and can be routinely dealt with by suitable meta-data tags — which is definitely not the case for the intellectual property rights. As a consequence we may concentrate in this section on intellectual property rights solely.

Now the meta-data expression of *authentication* namely that intellectual property rights are to be respected is met by providing references to the right's owners: In a technical sense acknowledging intellectual property rights is performed by referring to the (copy)rights' owners via suitable meta-data tags. However, the presence of virtual documents entails some technical complications. We start with some elementary considerations on which the following argumentation will then be based.

Citation guidelines in the context of TRIAL-SOLUTION are faced with the following situations:

1. citation of complete works like classical books or articles:
   this is common praxis and thus well understood,
2. citation of slices of complete works:
   this is open but might be not too hard to solve
3. citation of documents composed out of slices of other documents:
   this is the crucial case.

In view of these basic situations, we state two different principles underlying *authentication*.

(III) The *Principle of Creation (Authorship)*:
The principle of creation respects the *origin* of intellectual work, i.e. it respects the authorship of a person.

(IV) The *Principle of Composition (Editorship)*:
The principle of composition respects that a *combination* of material under a unifying intention is also work being subject to intellectual property, i.e. it respects the editorship of a person.

The first principle acknowledges that indeed a concept of originality beyond utilization of work of others may be maintained when reference to pre-existing work is of minor importance - despite the (undoubted) fact that every creative act cannot be understood without a cultural background with all its traditions. On the other hand, the second principle claims that even utilizing work of others comprises a creative aspect: it must be interpreted as the evocation of insight into a phenomenon through an intentional *arrangement* of work which is thus meant to express an underlying idea that has not been performed as far. This is considered as the creative intellectual act of interpreting and judging the possible contribution of given work with respect to some abstract goal a person wants to represent. In this sense the principles of creation and composition also provide a justification for the existence of author/creator and editor/creator tags in all meta-data descriptions of DC-style (typically the *contribute* element of the IMS Specification with a role author or editor, respectively)!

An example may illustrate these principles: Assume you start with an empty sheet of paper and two printed pieces of textual material. Basically, you have then two options to fill the empty sheet of paper: by writing something onto it or by gluing the two pieces onto the paper. In the first case you have created something starting from scratch (which might well be trivial -- but this is actually an external judgement like statements concerning its truth or originality). As in the first case, performing the second alternative may well evoke a new insight into some phenomenon just by confronting the two pieces with each other (which is especially well-known from the field of political cartoons). In this sense such combinations must also be acknowledged as a creative act -- which should be allocated to the person having performed the work. It should be mentioned that these examples address two extreme settings in which certain reductions and simplifications apply; any concrete intellectual activity involves both aspects: creating something new from scratch and utilizing other work.

Now for TRIAL-SOLUTION these settings can be given a precise meaning: TRIAL-SOLUTION has at its disposal smallest addressible (and hence referentiable) units (citational atoms) namely the atomic slices. Now the most elementary form of a new creation is the addition of a newly written atomic slice to the repository of slices. As mentioned above these slices may well include references to other work such as, e.g., citations, but from the viewpoint of TRIAL-SOLUTION they are considered as atomic. Similarly, in TRIAL-SOLUTION we can perform the simplest possible form of composition: build up a new slice by putting together some selected slices of the repository.

Within TRIAL-SOLUTION, creation is manifested by assigning authorship to a slice which has not existed previously. Technically, it amounts to the assignment of a name for a respective meta-data tag. But note that the justification of an authorship assignment is not an issue here and is definitely out of the scope of an abstract theory of citation. Instead, questions of this kind are to be handled informally within a community or treated formally by appropriate authorities responsible for copyright issues, respectively. Accordingly, the

correctness of the authorship assignment of the slice is not to be addressed in this scenario of dealing with the citation of existing material -- it is simply not a meta-data issue!

Observe that the principle of composition requests that *a unifying intention* be connected with the combination act. [13] These intentions have to be made explicit through a decisive act, e.g. via the assignment of a title in a canonical system. In TRIAL-SOLUTION corresponding meta-data tags including a special title tag must be supplied. The principle of editorship then legitimates the composer of a set of slices to fill his/her name into the creator (editor) meta-data tag.

In any case we end up with the construction of a unifying view of some parts of the world. To be able to deal with issues of this kind we have to identify the manifestations of these construction processes. However, for these aspects we may just rely on the creators own decisions: they determine the formal frame for their construction processes when releasing their work.

Definition 2:

A *root type document* is a unit which has been released by the intellectual property holder(s) as a whole.

As a consequence of this definition, root type documents are closed units which only depend on a decisive act of the intellectual property holders to publish them in the present form.

According to their specification, root type documents will include monographs and single web documents on a server (specified by a URL) as well as collection volumes or periodicals like journals and conference or monograph series. Moreover, contributions to a collection or proceedings volume are also considered as root type documents. As a consequence of this view, even periodicals may be considered as special cases of a composed document. Note that the definition explicitly admits that virtual documents be of root type.

After all, root type documents may occur as parts of other root type documents, e.g. monographs or collection volumes as parts of a series, so they may be considered as forming a structure which is hierarchically ordered by a broader / narrower relation. Note that the assumption of such a structure does not withstand the requirement that root type documents are self-contained entities. You may, e.g., order a book series as well as individual volumes from a bookshop.

Root type documents are considered as (intellectual) work that develops a coherent view on some aspect in its entirety. Now in TRIAL-SOLUTION we are specifically faced with the situation that pre-existing documents within a repository may be subdivided into pieces which then may be rearranged and combined with other material to form a new personalized document. In this situation we have to observe that a unifying intention results in the construction of a unifying view of some parts of the world that is at least implicitly shared by all its subparts. This means that even isolated units inherit some commitments made in the overall exposition which ties them closely to the root type document from which they have been taken. This implies that the contents of sub-units is always bound to the overall perspective developed in the work in which it is contained. So the relatedness of all work to root type documents must not be neglected: it finally determines whether re-use in another

[13] A coherent intention is assumed for manifestations of the principle of creation anyway.

context may actually contribute to the intended result or even compromise it. For example consider the following mathematical statement:

All functions f from the interval [0,1] into the reals are uniformly continuous.

Whether this statement is true or not depends on the context in which it is established: Whereas it is obviously false in classical mathematics, it is true in constructive mathematics (cf., e.g., [20]). However, this statement may be contained in a classical textbook on calculus as, e.g., an exercise which has to be checked. This clearly demonstrates that origination is absolutely necessary to judge the truth (and hence the possible reusability) of a statement: it must be interpreted in the context established by the root type document from which it is taken.

This adds a new facet to the authentication problem: reusability has to consider more aspects that just acknowledging the generationship of individual slices. Instead, reusability must also take into consideration what the intellectual property rights holders have considered as being a whole in itself, because this determines the overall meaning of the topic being developed. This view especially honours that a whole work is more than a mere aggregation of isolated parts. Note that this acknowledgement does not withstand the fact that the authors may have released parts as reusable items of their work (for details cf. III.2.2) -- or that other persons may concentrate on subparts for other reasons.

These considerations give raise to another principle to be observed in the context of reusability — and hence in the context of citation as well. As introduced above it may be viewed as a principle of consistent understanding supporting canonical views instead of disparate understandings on a topic.

Principle of acknowledgement:
Respect whole works!

It should be mentioned that the principle of acknowledgement entails the postulate of simplicity as a concrete guideline for the kind of material to be cited: refering to the whole work supersedes all references to subparts and may thus substitute a whole bunch of citations by just a single reference. Formally, this means that we have to prefer shorter citations (which still have to preserve all intellectual property rights) over more detailed ones. So the principle of acknowledgement can be implemented as a demand of minimality in the meta-data descriptions. In this sense it especially admits a technical counterpart which, however, can only be explained on the basis of a technical description methodology for (virtual) documents that will be set forth in the next section.

III.1.1 Guidelines for citations

In the sequel we will derive concrete citation guidelines from the above principles which up to now are merely expressions of an (abstract) understanding of authentication. This step requires a manageable implementation of the given principles: we have to specify a methodology which will be based on the document model of TRIAL-SOLUTION and at the same time captures the underlying ideas of the principles.

Now the principles of authentication especially acknowledge the creation of a whole document as opposed to a mere aggregation of isolated items. They respect a holistic view on

some topic as expressed by a self-contained work. This does not withstand the fact that the property rights owner has admitted to reuse *parts* of his work in another context but each *part* explicitly being marked as part of his *entire work.* Thus there is a subtle difference between the fact that an author has released *slices* for reuse and the fact that he or she released *parts of his/her work* for reuse. Accordingly, parts taken from an overall work have to be cited with respect to the whole work.

A possible alternative would consist in the fact that authors just create slices, i.e., small, isolated pieces independent of each other. As a result we would end up with just an aggregation of small entities instead of a whole work. But this does not establish a holistic view. It depends on the decision of the intellectual property rights owners what has to be considered as a whole work: these are exactly the entities which have been released by the generator as a whole. In the classical sense this is the manuscript the author submits to the publisher for publication.

So the crucial question is what has been released as a whole by the copyright owner: these are the entities that have to be cited for *contents* reference. These are by definition documents which are authenticated. However, as we have seen above a variety of root type documents appear as candidates for reference respecting the principle of authentication. Here we may apply the postulate of simplicity to provide a choice.

Definition 3:

A *source* for a unit U is a most specific root type document containing U.

We immediately associate a guideline with this notion[14]. This will turn the postulate of simplicity in the context of the intended usage of this notion into practice.

Guideline for citations II:

For citation use sources!

For example, a single proceedings volume may be referenced as a source instead of the proceedings series which may have editors by itself and may certainly be mentioned in addition but not substitute the citation of the single proceedings volume.

III.1.1.1 The witness citation model

The guideline provides an abstract description of the entities that should be used for reference. But it does not provide a manageable guideline on *what* to cite in concrete settings. For a concise formulation of this aspect we introduce the concept of a *witness document.* It is meant to capture the idea of a *reference suited for authentication* in case that parts are taken from another document. The idea of using witness documents for citation specifically acknowledges that a complete work is more than just an aggregation of its individual parts.

The principles of creation and composition care about authentication as assignment of generatorship to slices. They will, e.g., already be served, if the meta-data tags of slices reflect creation or composition of the atomic slices correctly. On the other hand this means they do

[14] In anticipation of a definition in a later section a root document may also be called a source. However, the notion of a source will be introduced lateron in a more general setting in which a root document will turn out to constitute a special case of a general kind of material only. But since the notion of a source is well established in citation practice, we already hint at this connection at this place and postpone a specific definition till the general setting has been established .

not care about the holistic view of a substantial new insight into a topic developed in the course of an extended work. The concept of a witness document reflects this position in practice insofar it tries to refer to the most contiguous part of work of an author or editor that has been reused in a given context. Hence the concept of a witness document reflects a document-oriented perspective in the context of authentication while the concept of a root type document rather represents a creator-oriented view.

The principles of creation and composition, however, provide the basis for a clear definition of a respective concept. Its functional description given by the subsequent definition is designed to detect a change in authorship/editorship assignment. Herein the *Principle of Creation* primarily concerns the citation of single (atomic) slices -- but may be extended to larger and more complex units as well, whereas the *Principle of Composition* addresses citation in case of utilizing material created by other persons independent of original authorship.

Definition 4:

Let *S* be a slice in a document *D*.
Then the *witness document W* for *S* (*W(S)* in short) is the maximal subtree in the hierarchy of *D* satisfying the following conditions
(i) *W(S)* contains the slice *S*
(ii) All subtrees between *S* and *W(S)* (including S and W) share the same author or editor.

Note that all nodes in the skeleton of a document will be enriched with meta-data tags (either directly or via an inheritance procedure) such that definition 4 actually results in a valid meta-data description of the witness document.

Our approach is intended to provide a canonical extension of common citation practice which is well applicable to situations dealing with sliced documents. In this context, the notion of a witness document just lays the grounds for specifying certain structures in the landscape of citation issues and thus serves to improve conceptual clearness. Moreover, the concept of a witness document is thought as to provide a functional base to derive guidelines for concrete citations.

Note that condition (ii) contains an *all*-quantifier which prescribes the direction of search for a witness document. The procedure to detect the witness document for a given node *S* starts with the node *S* within the skeleton tree of the document containing *S*, determines the authorship/editorship meta-data tag belonging to *S* and performs a bottom-up search for authorship/editorship changes. We end up with a third guideline for concrete citations which condenses the holistic view into concrete demands:

Guideline for citations III:

For citing a slice specify its witness document for authentication!

It should be emphasized that the notion of a witness document provides a new structure in the intellectual landscape of citation issues. It implements principles of authentication for concrete situations which arise in the presence of iterated composition of documents.[15] It generalizes citations of units in , e.g., a collection volume which should be related to the author of the contribution and not to the editor of the whole volume: the witness document is

[15] In cases of conflict (two possible sources for a slice) choose the root which entails the largest extension of the given unit.

just the full contribution of a person in an overall work and is thus in conformity with common citation practice.

III.2 Identification

We turn back to our abstract description (1) of what citation in general is about and concentrate on the aspect of *something*. This aspect condenses into the following abstract relationship

Unit U *in document* A (2)

unit *U* being an arbitrary subpart of the (possibly virtual) document *A*. Unit *U* is considered to be given without further specification and constitutes the unspecified and unspecifyable initial point of our considerations.

Further substantiating statement (2) then amounts to a description of *A* as well as a clarification of the relationship *in*. This demands a description methodology suitable to identify document *A* along with an instruction on how to detect unit *U* inside *A*. Remember that citation has to ensure that anybody may check its contents. Consequently, any procedure implementing the principle of identification has to provide a method prescribing how a cited unit may actually be recovered. Citation has to support and to reflect this. Remember that citation as a general operation essentially refers to a semantic context. This means that everybody should be able to check the source of a citation given in the course of an argumentation just by interpreting its meta-data descriptions. This should in a canonical way result in a concrete procedure to obtain unambiguously a copy of the desired material. It is thus the task of citation to provide a description of a document in such a way that it can be used to materializing its contents.

Thus the overall task to resolve a citation into its contents essentially incorporates three different facets:

(1) To determine the document in which the cited entity is enclosed (source document).
(2) To get hold of a copy of this document for subsequent inspection (accessibility).
(3) To identify the subpart inside the document that is specifically referred to.

The resulting general resolution process is denoted as *localization*. Only satisfying these facets altogether will result in a concrete, justified method implementing the principle of identification.

For *A* being a root type document the required description methodology may rely on already available meta-data descriptions such that no further investigations are necessary. This conforms with the TRIAL-SOLUTION maxim to rely on open standards where appropriate and possible. Hence it remains to concentrate on virtual documents.

We start investigating the third facet, because it may be treated independently from the others and is needed for an explicit description of the first one. On the other hand it will become immediately clear how to generalize these considerations to virtual documents whereas more efforts are needed for the other facets.

III.2.1 Source locator functions

The third point of the localization process expresses that most citations only refer to a subpart of a hugher entity on which special interest in a given context is focused. Simply referring to the whole document would be considered as a sloppy style of citation and would be out of scope for serious and liable intellectual work. This means that instructions have to be provided to identify and locate a piece of work inside a more comprehensive unit in which it is enclosed. In classical books page numbers are available for this purpose -- and are well acknowledged in citation practice as the smallest referential entity. It is common citation practice that reference to single text passages inside a book is indicated by providing the page number. "*An interesting book*, p. 13" is an example of such a reference: "*An interesting book"* being the comprehensive document (for which a complete reference may be presupposed), and "p. 13" denoting the actual position of the item in question in the comprehensive work.[16]

In the presence of electronically available material a comparable localization method for subparts may be not at hand in a canonical way (cf., e.g., HTML-files on a Web-server). Hence the question for counterparts for this citation facet comes up. This amounts to a specification which subparts of a document are actually addressable and what the corresponding method for locating then could be. However, this is highly depended on the concrete setting and cannot be solved in generality.

From an abstract point of view we have to specify a *source locator* function which is able to identify subparts of a document in a sufficient granularity and at the same time hints at a procedure how to locate the respective entity inside the comprehensive work. Referencing page numbers is an example for a concrete implementation of this functionality.

The following guideline for citation emphasizes that using a suitable source locator function is indeed an integral part of the localization process. One might argue that one could do without. However, for huge documents scanning through the whole document would be the only alternative for a citation method neglecting a respective functionality. This may certainly be considered as acceptable for short texts but is not acceptable for voluminous work such as complete editions -- and is definitely not acceptable for a general investigation of citation principles which aims at reasonable procedures to inspect statements on hand.

Guideline for citations IV (principle of specificity**):**
Use the most specific source locator for identification!

So the citation guideline postulates that in the context of classical citation issues page numbers are given as part of a citation hinting at specific subsections of an exposition. Particular variations are certainly possible in cases when the whole spirit of an entire work is referred to instead of a single unit in the course of an extensive argumentation.

In general we may observe that we have the fine structure of a document at hand which is provided by the organization structure of the document, i.e. the document skeleton (cf. chapter I). Hence one sort of locator functionality may be directly derived from the document skeleton structure which assigns a relative position inside the whole document to each unit. A respective functionality may thus be gained via suitable meta-data tags which specify the

---

[16] This does not withstand the fact that for subtle inquiries a further differentiation may be necessary which is then given in a canonical way by referring, e.g., to the line numbers on a page. However, it seems that the (public, scientific) community well lives with the slightly coarser identification method via page numbers.

relative position of objects within the skeleton tree of the document under consideration and which then have to be associated with each unit (cf. the TRIAL-SOLUTION meta-data tags described in IV.1). A suggested functionality for the question of localization may, e.g., use the path from top of the document skeleton to the object under consideration.

III.2.2 Advanced description methodology for virtual documents

We now turn to the first facet. So far we were only concerned with the identification of single slices. However, as we have stated above we will also have to deal with virtual documents which cannot simply be accessed through a (single) repository as a whole. Moreover, all methods presented so far turn out to be insufficient for the description of composed documents which consist of (iterated) combination of slices.

Accepting the semantic character of citation requires that a complete reference to used material is given such that everybody is enabled to identify generationship of a document - independent of his or her personal awareness of the document. Hence especially for virtual documents there is urgent demand for a description that meets the principles stated above.

To start with compare this situation with the citation of a collection volume issued by an editor. Instead of mentioning all its contributions it suffices to cite the title and the editor (thus indicating that it is some work generated by making use of material created by others) together with the publisher (being the repository). By accessing the repository everybody may easily identify those parts of the collection volume that have been taken from other sources. We conduct that the given citations suffice to provide a complete overview of the contents of the volume and consequently a complete description of its contents may be established. This is then in accordance to the principle of identification as well as to common practice.

Now what if the collection volume is not available as a whole, but it is nevertheless well-known in a community as it is the case with some 'grey' literature: every member of the (core) community is aware of and able to describe its contents upon request? How should it be introduced to a newcomer such that he or she may be able to refer to its contents with a similar kind of certainty? According to current trends this situation is comparable to the problem of citation of a personalized document used as, e.g., some course material for a lecture. Citation just by title and the name of its composer (possibly together with nothing more than the phrase manuscript) prohibits general inspection and thus compromises the principles we have stated so far.

The situation which we are faced with is the problem of a description methodology for citation of a document which is composed out of slices on the fly, i.e. such that exactly this single document is not available on a repository as such. This results in a far more complex citation methodology in order to maintain conformity with the given principles. As a classical example consider the case that a complete work consists of two volumes which had been written one after the other by the author. Now suppose that the publisher of the first volume is no longer interested to publish the second volume as well. Hence the author has to look for another publisher to issue volume II. Nevertheless the whole work is devoted to one single topic and thus must be considered as whole. Since there is no common repository to get the entire work, the meta-data description cannot simply be of the form Author: Title of the work I, II. Publisher, year but must describe the single volumes individually. However, as in the case of the collection volume mentioned above the idea of a whole work must still be maintained, i.e. a complete description of its contents should be derivable from the citation. This may lead to a citation of the form Author: Title of the work I, II. Publisher1, year1 (Vol.

I), Publisher2, year2 (Vol. II) . Hence one way out is to develop a description methodology which allows a complete reconstruction of the items through its meta-data description — by relying on the repositories finally. This idea is generalized for virtual documents in the sequel.

To start with, suppose we have a newly composed collection *D* of different pieces of content taken from external documents. We want to cite this collection in accordance to our principles. The solution for this problem is to reflect the composition process of the document on the syntactical level. So we will first have to identify its different parts and refer to these as building blocks.

We may concentrate on the case of virtual documents only, because this is the most general case — and it is at the same time the main point why DOI (see III.2.3.1) turned out to be insufficient for the TRIAL-SOLUTION approach. On the other hand a solution for this problem will yield a solution for all other cases as well.

Now the only chance to obtain a copy of such a document is by providing an *unambiguous description of the structure and contents* of the document. This especially includes

(1) describing the document structure
(2) information where its subparts are from

such that a copy of the document *can be constructed* -- as opposed to simply accessing it via a repository. So identification for virtual documents can only be guaranteed through a detailed construction plan which starting from actually accessible subparts (its atomic building blocks) unambiguously demonstrates how to arrange these to end up with an actual copy of the document in question.

Identification implies differentiation - to identify a document one must provide criteria to differentiate between two virtual documents which are claimed to be not the same. So *defining* a document amounts to *describing* it which in case of virtual documents may only be thought of as a procedure how to *construct* it. Accordingly, we describe them via an identification of the parts it is composed of together with a construction plan prescribing how to combine them. However, in order that this process is well founded it has to be based on documents that are immediately accessible which in turn may only be guaranteed by repositories. So the very basis of this construction process must be so-called *root* documents that guarantee basic accessibility..

Definition 5:

A *root* document for a unit U is a most specific document among all possible root type documents that contain some subpart of U and which may be obtained from a repository as such.

Hence information where slices may be taken from amounts to relating them to the root documents in which they are contained, respectively. Observe that the definition of a root document just relies on arbitrary *subparts* of the document U. This specifically covers the case of a virtual document U that is not contained in any repository as a whole. In this case U is composed out of slices taken from at least two documents which in turn, however, are accessibly through a repository, respectively. These slices then give raise to two corresponding root documents for U being the basis for a composition of U.

It is worth emphasizing at this point the differences between a root document and a source document. While the root document indicates where an entity may actually be taken from, the concept of a source is meant to describe the intellectual background the material hinges on.

As a consequence these two concepts are independent of each other; a source may be contained in a root document or may be a virtual document on its own.

Now in order to identify the whole document we will have to describe how these parts are combined within the new collection, i.e. provide a detailed specification where the slices are located inside the document together with a description of the skeleton.[17]. To capture this idea the notion of a *composition unit* is introduced to describe the building blocks that is, material that relies on pre-existing entities subject to intellectual property. As composition units will be slices of a special form, citation of these units will be already clarified via guidelines II-IV.

It should be noted at this point that a composition of a new document can only be based on *known* material of which the generator is especially aware. This does certainly not prevent from re-constructing a document that has been constructed by someone else before. However, it may happen in the classical case, too, that a collection volume, e.g., has been erroneously issued twice by different persons not knowing of the work of each other. This might be quite trivial at the first glance but it is worth mentioning in the context of the following definition to motivate the parameter D*: you can only relate to what you know and not to what might exist somewhere! In this sense the parameter D* explicitly represents the available background material.

Definition 6.

Let D and D* be documents.
A *composition unit U* of a document *D* with respect to document D* is a maximal subtree in the hierarchy of *D* which is contained in D*.

In other words a composition unit is the most extended part taken from a document D* that is included in a new document. We now have introduced two concepts: witness documents and composition units which at the first glance look similar. The differences are thus worth mentioning. In short, the witness document is authentication-oriented whereas the composition units are document-oriented. While the witness document denotes the most contiguous contributions of one person, the composition unit characterizes the most extensive parts which have been completely taken over from another document. Witness documents are irrelevant for the construction process of documents whereas composition units are irrelevant for authentication; they may well comprise different authorship attributions.

Observe that document *D* will contain several composition units which put together constitute the entire contribution of work by others. Collecting its composition units will result in a decomposition of *D* into the maximal contiguous parts that are already contained in pre-existing documents. However, as the decomposition process via composition units only results in an (unstructured) set, it does not supersede the complex description of virtual documents as it is reflected by their meta-data description, i.e., a specification telling how this (flat) set can be transformed into a structure including skeleton information.

It may be annotated that this definition is a special requirement for virtual documents, since in 'classical' cases the composition units are all trivial: By definition, a given part of a root document has exactly one composition unit with respect to the root document namely the whole part itself. Insofar the notion of a composition unit shows effects only in case of virtual documents.

---

[17] In TRIAL-SOLUTION this is performed by attaching suitable meta-data tags to slices which include data about structure as well as data about containment in root documents.

The introduction of composition units via definition 6 especially observes the postulate of simplicity stated at the beginning of chapter III and thus supports a holistic view. Compare this to the definition of a witness document for which a similar principle has been applied. In addition, if we could take arbitrary slices (even atomic ones) as composition units as well, this would unnecessarily increase the list of references to be made and thus would make citation rather inefficient. So the practical aspects may be considered as insightful consequences of the principles we have stated.

III.2.2.1 Identifying composed documents

It remains to show how virtual documents have to be cited in accordance to the principles we have stated. Remember that citation for a virtual document must allow identification which is only possible on the basis of a disclosure of its generation process. To constitute this final step first observe that the process of generating documents is an inductive one, namely constructing a succession of parts one after another which finally results in the completed document. Accordingly, the meta-data able to identify a document unambiguously can be composed in an inductive procedure as well mimicking the construction steps of the document.

Suppose that $\{A\}$ is a meta-data description of a composition unit A. Note at this point that $\{A\}$ includes all the necessary meta-data according to our principles including the root documents and the source locators. As we may suppose that it is well-known how to create the meta-data for composition units (refer to the previous sections for details), the only problem that remains open is to show how meta-data are aggregated for the composition steps during document generation. This is described in the following composition principle for meta-data:

Meta-data composition principle for virtual documents

Suppose $\{A_1\},..,\{A_n\}$ are meta-data descriptions for given units $A_1,..,A_n$, suppose a composition process out of these resulted in a unit B.
Then $\{B\} = \{<Sk>,\{A_1\}<l_1>,..,\{A_n\}<l_n>\}$ is a meta-data set for the unit B composed out of $A_1,..,A_n$, where Sk is the meta-data description of the skeleton of B and $l_{1,...,}\ l_n$ are the source locators for the units $A_1,..,A_n$ in B, respectively.

Remarks:
1.) l_i may be trivial for units that have been created in an intermediate stage of construction - they are just hinting to *top* of the skeleton, i.e., the whole unit.
2.) This composition principle may give raise to a formal definition of a *document*. Remember that a document is some unit that has been released by the generator. From the viewpoint of this principle a document is thus just a unit for which the construction step ended.

Meta-data descriptions which observe all aspects of this definition may be called *valid* descriptions. Note that a composition is in general associated with additional meta-data: the name of the composer together with additional descriptions such as a (section-) title, etc. in case they cannot be simply taken over from a pre-existing document. All these are usually contained in the skeleton information. Hence a valid meta-data description of a complex unit composed out of composition units has to include skeleton information besides the corresponding source locators according to guideline IV.

It may also be annotated that iterations of this composition process are explicitly admitted. These iterations simply require dissolving the braces and recompiling the skeleton and source

locator function. This recompilation in turn just amounts to an extension of the source locators by adding a prolongation of the path description leading up to the root of the tree. Similarly, the skeleton is extended by just putting one skeleton tree on top of the other in a suitable way.[18]

Furthermore, the composition principle denotes an abstract technique for the meta-data construction of complex, virtual documents which is neutral in relation to concrete meta-data element sets; it just prescribes the construction process. As a consequence every concrete meta-data description needs an implementation of the composition principle in a specific setting. Such a setting may consist of a commitment to a meta-data element set as, e.g., Dublin Core or the IMS meta-data element set. These are required for the description of composition units. Note that once these descriptions are fixed, the composition principle may yield a minimal description methodology by just $\{<Sk>,\{A_1\}<l_1>,..,\{A_n\}<l_n>\}$ implementing as $[Sk, \{A_1\},l_1,..,\{A_n\},l_n]$ (bracket-notation). This is a 'conservative' extension of citations of the form "[5]" where '5' just links to a complete reference like "Author: Title, in: Root, p. n" usually given at the end of the document, "Author: Title" constituting the witness and "p. x" denoting the source-locator. This may be rewritten as [Author: Title, in: Root, p. n] dissolving the link and illustrating the analogy more concretely[19]. Hence the composition principle hints a canonical way for a concrete citation based on a description methodology for composition units. In addition to guidelines II-IV we now end up with the following rule:

Guideline for citations IVa (composition rule):
Use the meta-data composition principle for identifying virtual documents!

III.2.3 Accessibility

Let us now concentrate on the second facet. We may assume that we are able to identify the basic material under consideration sufficiently enough such that the retrieval engine of the repository should be able to detect it inside its stock. W.l.o.g. we may suppose that we are dealing with root documents only in this section, since we have already stated how composed documents can be generated on the basis of root documents, source locators, and meta-data descriptions including the skeleton. This construction process is completely based on root documents and composition rules. It can be utilized to construct an access methodology for virtual documents as soon as an access methodology for root documents is given: it just suffices to delete access information which states how a root document may be addressed from a virtual document. This amounts to establish shortcuts for the path information that underlies the access functionality. Then conventions are required for the description of access methodologies for root documents only.

To deal with the aspect of accessibility in general we introduce an *abstract* document locator function which we assume to be associated with a feasible functionality for accessing the resource (a *handle*). *Handles* are supposed to implement the functional characteristics of the document locator function in concrete. A handle may be a URL for direct electronic access.In classical models for citation as, e.g., in the Dublin-Core approach as well in the IMS metadata specification, the functionality of 'handles' is taken by bookshops which intermediate between a contents description provided by a DC-style characterization of the

[18] It may be annotated that this may cause some re-arrangement e.g. for chapter numberings in case they are not automatically calculated.

[19] A more detailed description is provided in the sequel.

root document and the requirement to get actually hold of a copy of the referenced material for check the cited passage which is provided by the publishers.

In the age of digitization substitutes for such 'handles' are required.The most prominent being the DOI approach which addresses exactly this problem. As we have stated above TRIAL-SOLUTION uses open standards and methods preventing a proprietary system whenever possible. So we have to investigate already existing procedures for usability and suitability in TRIAL-SOLUTION.

III.23.1 Remarks on the DOI approach

DOI is concerned with the organization of intellectual properties and for this purpose maintains a registry. In detail this comprises the following:

(1) a system of fixed identifiers which may resolve to Internet accessible data like URLs (achieved by a handle system)
(2) a description model required to find identifiers for a given object (structured by a set of meta-data)

The DOI approach uses a fixed kernel of meta-data elements which essentially forms a subpart of the Dublin Core and IMS scheme, respectively, except from a few additional elements.

DOI adopts the notion of an entity being something that is identified (inside its registry) [10, p. 14]. Moreover, DOI allows the identification of entities of finer granularity than a completed document (independent of the size or form). However, the decision as to what is going to be registered and how this will be done has to be taken by the registrant.

Though a registry of single parts of a document is possible within the DOI system in principle, the DOI handbook does not tell any details on how to procede in this case. In particular it tells nothing about

(1) how to perform the slicing of a document
(2) how to describe the document slices
(3) how to register/cite a document composed out of slices.

From the viewpoint of TRIAL-SOLUTION DOI treats every item as a root type document as its fine structure is not processed and determined by DOI itself. Instead it seems that DOI would accept any description of slices and composed documents which conforms to its canons of meta-data. Especially, the main problems we are concerned with in this paper:

- the reference to documents created on the fly (virtual documents):
- the relation of a part to the whole from which is has been taken as well as the specification of the relative position in the underlying document

are not addressed by DOI itself and must be determined elsewhere.

Actual localization methods are usually based on a kind of handle system thus establishing a DOI like organization.[20] But these approaches all turn out to be insufficient for the type of documents we are dealing with. Firstly, in the presence of document slices, we have to set up criteria for identifying a resource within a certain context. So for citation we first have to specify a suited context as the *source* of the resource and then to describe the actual position

[20] Details on the DOI organization may be found in [10].

of the cited unit with respect to this source. But information like this, relating parts to a whole, is not manageable by approaches like DOI. Secondly, we have to identify documents which have been composed out of pre-existing parts on the fly (virtual documents). This means that we first have to identify these parts and then have to specify the way in which they have been organized, i.e. we have to describe the new document skeleton. DOI however does not foresee the description of such ontologies either.

So we cannot rely on already available solutions and hence have to develop more advanced description methodologies inside the project. From the viewpoint of TRIAL-SOLUTION DOI — and similarly any other registry of this kind — may be considered as an organizational issue which eases the *localization* process by providing a centralized handling methodology for root type documents but give no clues how our principles are obeyed.

III.2.3.2 References

The postulate of simplicity (which is derived from the principle to acknowledge whole work as stated above) demands a description methodology as simple as possible which still meets all the abstract requirements. This means that both, authentication and identification, have to be observed. It proposes to combine parts of the document which share the same specification besides the source locator specification, i.e. they have the same witness document but hint to different places therein. The idea is to concentrate on such slices whose meta-data descriptions under the principles of authentication and identification are identical: the so-called *meta-data indiscernibles* which may only be differentiated by their locator function.

These (maximal) building blocks constituting the basis for the description are the *references* which are introduced in the sequel. They will generalize the concept of root documents which in turn will appear as sort of 'atoms' in the meta-data description.

So far we have settled the problem of citation of arbitrary units U which are actually contained in some root document as well as citation of virtual documents composed out of these. In this case a suitable reference for accessibility of U is simply provided by the corresponding root document and we are through. It remains to show how arbitrary units U have to be cited according to our principles. Hence let us suppose in the sequel that U is not contained in a single composition unit. Then U must necessarily stem from some virtual document D.

Now observe that a virtual document solely consists of a composition of material taken out of root documents, i.e. stemming from a repository. If this would be not the case, there would exist some units that may not be considered as published in the strict sense. We would then end up with sort of a 'manuscript' which is out of scope for a foundation of the citation problem. So we may confine ourselves to virtual documents composed out of root material as described above.

We distinguish two cases. First suppose that U may be decomposed into composition units without losing insight into the phenomenon developed in the course of the argumentation in which the citation is involved. Then citation of U simply reduces to a sequence of citations consisting of single composition units and we are through. So the crucial case is a unit U which refers to some context that especially emphasizes the compositional aspect performed by somebody on (at least) two non-contiguous parts. This is especially the case for some joint subparts of the example given above where individual parts of a work have been issued by different publishers.

Now accessibility requires that we are enabled to actually get hold of a copy of the document in question. We have already seen that in case of a virtual document a construction plan shows how to compose the document ourselves. So it remains to clarify where the basic material for the construction can be obtained. This material have to be taken from repositories at the end. So we need a generalization of the concept of a *root* document clarifying where the data to be composed later on may be taken from. Consider the meta-data composition principle once more: it generates documents on the basis of pre-existing documents by adding source locators specifying which parts of those documents have been used, and adds a new skeleton. Remember that the meta-data are aggregating into a set. The inverse process then immediately leads to the root documents which contributed to the unit: just drop the newly added skeleton as well as the source locators. This inverse dissolving process is well-founded, i.e. the meta-data of the root documents contributing to the document under consideration will remain!

If the set which arises from this dissolving process is a 1-element set, then we just have the meta-data description of a single root document which is a valid meta-data description. If this set contains more than one element, then the unit U is composed out of several root documents. But then the resulting set is no longer a valid meta-data description, because at least the skeleton information is missing which on the other hand is mandatory. The definition of a *reference* -- which looks a little bit artificial -- just describes this dissolving process and generalizes the concept of a root to composed documents. Its 'classical' counterpart may be a reference list of cited documents as it is usually given at the end of a document and is oftenly called a 'Bibliography' or just 'References'.

Definition 7:

Let *U* be an unit in a document *D*.
A *reference* for U is the list of root documents occurring in the meta-data description.

The references may be compiled from the meta-data description of a document by repeated application of the following steps as long as skeleton data or source locator information is contained in the resulting set:

- dropping the skeleton information
- dropping all source locator information.

Finally collect the resulting data in a set.

This procedure just reverses the construction of meta-data according to the meta-data composition principle: Assume that the document B has been composed out of documents $A_1,..,A_n$. Then the meta-data description of B must look like $\{<Sk>,\{A_1\}<l_1>,..,\{A_n\}<l_n>\}$. Now dropping all skeleton and source locator data results in the structure $\{\{A_1\},..,\{A_n\}\}$. Now $\{A_i\}$ have been valid meta-data descriptions by assumption. Hence $\{A_i\}$ is of the form $\{<Sk>,\{C_1\}<l_1>,..,\{C_n\}<l_n>\}$ and we may repeat the dissolving process once more. Since the construction is grounded in root documents, this dissolving procedure is well-founded and hence stops after finitely many steps. Collect the remaining meta-data descriptions in a set. This set is then called the *reference* for the unit under discussion. It lists all root documents that have been processed in the course of the composition of the said unit. Now for a unit already contained in a root document the dissolving process just results in the root document itself. Hence the definition of a reference generalizes the definition of a root for documents composed out of material stemming from different root documents.

Now one final step might be reasonable: how shall source data be related to the references in citation practise? In common citations we often find relations of the following kind: "Author:

Title, in: root, source locator", e.g. "Nobody, Noname: An interesting paper, in: Everybody, Al (ed.): Sense and Non-sense. Importance Press, Everywhere, 2001, p. 1-2." In this citation method the 'coppola'*", in:"* in the previous example relates source to root information. So we might be suggested to introduce a formal canonical extension which could, e.g., look like "Author: <source>, taken from <references>" where <source> is the meta-data description of the document according to our principles and <reference> is the list of root documents. The coppola *", taken from:"* could be called a *origination indicator* and the usage of an appropriate such indicator might be added to a guideline for citation as well. However, this seems quite a bit too formal and too overregulated. So we leave that aspect to the reader.

III.2.3.1 Citation guidelines

The final step towards a concluding guideline for citation must realize the principle of identification in full. As a result of the previous sections the addressability of the root documents can be taken for granted by existing methods. Following the overall policy to rely on open standards whenever available, TRIAL-SOLUTION will adopt these for the root documents and extend it in a canonical way which is minimal in a certain sense. Hence for the citation of given units it suffices to specify its unique relative position with respect to (one of) its reference documents. This position is determined by the source locator associated with the unit. In addition to the preceding guidelines for citation we now end up with the following

Guideline for citation V:

To cite a given unit S in a document D specify

1. *the witness document of S for authentication*
2. *the source-locator of S for localization*
3. *the source document for identification*
4. *the references for accessibility.*

Remember that identification requires reference to a root document, and authentication requires that the work of a person as a whole is to be acknowledged. But note that in the meta-data composition principle for complex units reference to root documents is already contained! This approves that our solution is in accordance with these abstract principles by definition of its composition technique already!

Note that this guideline supersedes the guideline III stated for witness documents. It may be mentioned though that reference to witness documents should be the standard way of acknowledging work of others. Using reference documents will be necessary in very exceptional cases only. This may justify that we have established two competing guidelines -- one for rather practical reasons the other comprising the general citation problem in full.

It remains to provide a guideline for the citation of virtual documents as a whole. Based on the construction methodology for virtual documents given above this now simply mimics the construction process as follows:

Guideline for citations VI:

To cite a document newly composed out of slices, proceed as follows:

1. *identify the composition units that are contained in the document*
2. *cite these units according to guideline V*
3. *cite complex documents according to an implementation of the meta-data composition principle!*

Observe that the maximality condition in the definition of a composition unit then guarantees a minimality of a complex meta-data description and at the same time preserves existing author rights. Also note that a reference for a virtual document always results in a reference to some documents contained in a repository, i.e. root documents. So citations of parts of virtual documents do not pose new unsettled questions!

We conclude by giving some examples to illustrate our citation method.

1. For a table within an article of a collection, we have the collection as root document and the article as witness document; source locators within the root document are the page numbers.
2. For an article within a journal, we have the journal (a periodical) as root document and the article itself as witness document; source locators are the number of the journal volume together with page numbers.
3. For an article of a proceedings volume published within a conference-series, we have the single volume (not the periodical) as root document. From a changed point of view however, the periodical will serve as a root for the whole volume; this is a case similar to example 2.
4. For a complete work issued in two parts by two different publishers we have the whole work as a source describing the two parts plus the way how they are to be combined together with the list of the two volumes (together with their meta-data description) as reference.
5. For a joint subpart of the two volumes, respectively, we have this new entity as the witness document and the complete work as a source together with the references as in example 4.

It is worth mentioning at the end of our theoretical exposition that the citation guidelines established in this paper completely rely on repositories and transparent constructions built on top of these. So besides the abstract principles established in the course of our investigations the guidelines for citations all acknowledge the copyright protection issues as much as possible in the presence of techniques to generate personalized documents that possibly will not remain in the private but will circulate in the public as it is the case, e..g., for material made available during a lecture. In this sense our approach may indeed be considered as a most conservative extension of existing issues insofar it always relies on simpler cases and at the same time hints everything to possibly existing copyrights.

Finally we observe that citation yields that all subparts of a document have to be described by respective meta-data tags which is guaranteed by the TRIAL-SOLUTION system functionality. Our approach thus results in a justified, extended citation method which at the same time provides the grounds for a corresponding meta-data description methodology. This meta-data specification will be described in the following chapter of this paper. We end up with some examples which demonstrate our approach in a concrete setting.

III.2.3.2 Examples

In this section we illustrate consequences of our approach in the following scenarios

(1) Suppose a (physical) book A = Aut/Tit/Pub has been cut down to pages. We end up with a pile of single pages preserving the original order. This pile is easily identified by everybody as the original book (which is presented in a non-standard form, though).

(2) Now suppose we build a heap of the individual slices (pages), since we want to combine them with other material. Hence we have to add meta-data to the slices (being the real, physical pages) which is done by the expression "Aut/Tit/Pub p. x". (Note that according to principles of meta-data assignment all slices inherit the creator and root document).

(3) Suppose, in our attempt to create a composed book according to our specific intentions, we incidentally (re-)arrange all slices (i.e. pages) according to increasing page numbers. We then (re-)bind them using a new cover which consequently mentions ourselves as the creators. (Note that we are unaware of having reconstructed an original book!) We thus only have at our disposal the citations of the slices, i.e. the pages. How to cite this newly composed document? Since we only know the pile of slices (pages), we seem to end up with a collection of individual page numbers like
"{Aut/Tit/Pub p1, Aut/Tit/Pub p2, Aut/Tit/Pub p3, ..., Aut/Tit/Pub pn}"
which altogether indeed might constitute a fair citation of our newly composed documentthough neglecting the whole work created by Aut.

As we want to be honest, we mention ourselves as the creator of the collection "{Aut/Tit/Pub p1, Aut/Tit/Pub p2, Aut/Tit/Pub p3, ..., Aut/Tit/Pub pn}" of slices and end up with the following citation
"{myself/B/<new skeleton>,Aut/Tit/Pub p1, Aut/Tit/Pub p2, Aut/Tit/Pub p3, ..., Aut/Tit/Pub pn}" read as
"B composed out of {Aut/Tit/Pub p1, Aut/Tit/Pub p2, Aut/Tit/Pub p3, ..., Aut/Tit/Pub pn} by myself".

The underlying principle of identification detects that all slices stem from only one book which is by definition also the witness document of the collection of slices. So the principle of simplicity demands to resolve the citations of the form
"{Aut/Tit/Pub p. 1, Aut/Tit/Pub p. 2, Aut/Tit/Pub p. 3, ..., Aut/Tit/Pub p. n}"
to simply "{myself/B/<new skeleton>,{A}}" to be read as
"B composed out of A by myself" (which sounds ridiculous in some sense but it's a possible scenario we have to deal with!). Now the principle of simplicity demands a citation of the whole work, i.e. of the form "A". We thus end up with the citation of the witness document which accidentially turns out to be identical to the original book! Note that now this is governed by simply applying the mentioned principles and hence no longer arbitrary which is highly important from a foundational point of view!

(4) Next suppose we have a "Handbook on X" which groups together contributions from several authors in, say, four parts A-D each part being complemented by an introduction of the editor ED, respectively. What is the equivalent to a listing of all the slices involved in the presence of our citation principles?

Let us consider the following examples:

a) We want to cite part C of the Handbook, which one would expect to look like "Handbook on X, part C, edited by ED". This is now justified by the *principle of composition* in case the handbook is published by a publisher. If not, the handbook is actually a virtual document and we have to describe its contents in full together with information where the contributions are taken from (which does actually only make sense for a collection of papers previously publishers elsewhere)..

b) We wish to refer to a single slice, e.g. one table in the Handbook which is part of the contribution TIT created by AUT. We first check the meta-data description of this slice and find that it is created by AUT. According to the definition of a witness document, we look for the maximal subtree containing that table. In our example this is the complete contribution TIT of author AUT in part B of the Handbook. Hence we cite "TIT by AUT in (part B of) Handbook on X, edited by ED". (Note that by the principle of creation we might also drop "part B" in this citation - provided a corresponding source locator function is gievn.)

c) Finally, we want to create a new document consisting of part B of the Handbook on X together with one single article TIT of part C. According to the *principle of composition* we may claim some property rights on exactly this combination. But instead of listing all slices we again construct the witness documents and end up with a reference like "Part B of the Handbook on X together with TIT from AUT in part C of the Handbook on X edited by ED".

III.3 The TRIAL-SOLUTION locator functions

TRIAL-SOLUTION starts with a list of books which thus constitute the root type documents. Following a DOI-like approach, these source documents are accessible via identifiers from a registry that is maintained by the project. Moreover, extending DOI, the approach of TRIAL-SOLUTION provides an additional identifier for each slice *S* of a document that uniquely characterizes this document together with the relative position of *S* within the hierarchical structure of the document skeleton. This is achieved by an accumulated list of *sourcereference* elements determined by the path leading from the root of the skeleton to *S*. In TRIAL-SOLUTION this is implemented by the *sourcereference* meta-data tags which specify the traces from the root document to the slice.

In general, when iterated composition of document slices is taken into account, this approach together with meta-data inheritance leads to the construction of a *citation history* for all slices. From this history and the *resources type* metadata attached to the slices it is easy to determine the root document of a slice in the sense of our first guideline stated above.

IV. The TRIAL-SOLUTION Meta-data Specification

In the sequel we give an overview of the meta-data approach of TRIAL-SOLUTION for the description of learning objects. Existing meta-data standards are extended by the TRIAL-SOLUTION system in some different respects. One of them concerns the particular needs for most flexible and detailed resource descriptions. For this, the project has designed its own meta-data specification including among others the following features:
(1) conditions of use for the individual parts of a sliced document
(2) precision in content descriptions
(3) support for the use of structured thesauri
(4) assignment of specific roles to specific data

Moreover, the TRIAL-SOLUTION meta-data system combines meta-data with document structures by an inheritance mechanism. But details on this meta-data inheritance approach are closely related to the TRIAL-SOLUTION document composition model and they will be shown in chapter 5. The present chapter however is concerned with the description of the TRIAL-SOLUTION meta-data element set and its structural representation. Before a formal specification is given in section 2 by the TRIAL-SOLUTION meta-data DTD, we first will

point out in a more informal way the particular structure of the element scheme, its relations to other schemes, as well as the intended meaning and use of its elements.

IV.1 Meta-data Elements and Structure

Meta-data in TRIAL-SOLUTION are organized by a hierarchical structure of elements shown in Fig. 3. The root of this meta-data hierarchy is given by a *record* element with descendants for a further specification and grouping of meta-data. As learning objects are considered as resources to be handled within particular composition and decomposition scenarios, we will distinguish two main groups of meta-data that are required for the project, one for resource description in general and the other for a description of the specific application environment.

The first group is covered by the existing standards of DC and IMS, so for this part a kernel of IMS meta-data elements has been extracted which is suited to the particular needs of the project. The additional, TRIAL-SOLUTION specific elements hold the information required to identify single decomposition units, to trace the sources of these units, to give details on relations between them, and to describe their content by using contolled vocabularies. The intended interpretation of all these meta-data elements is given at the end of this section.

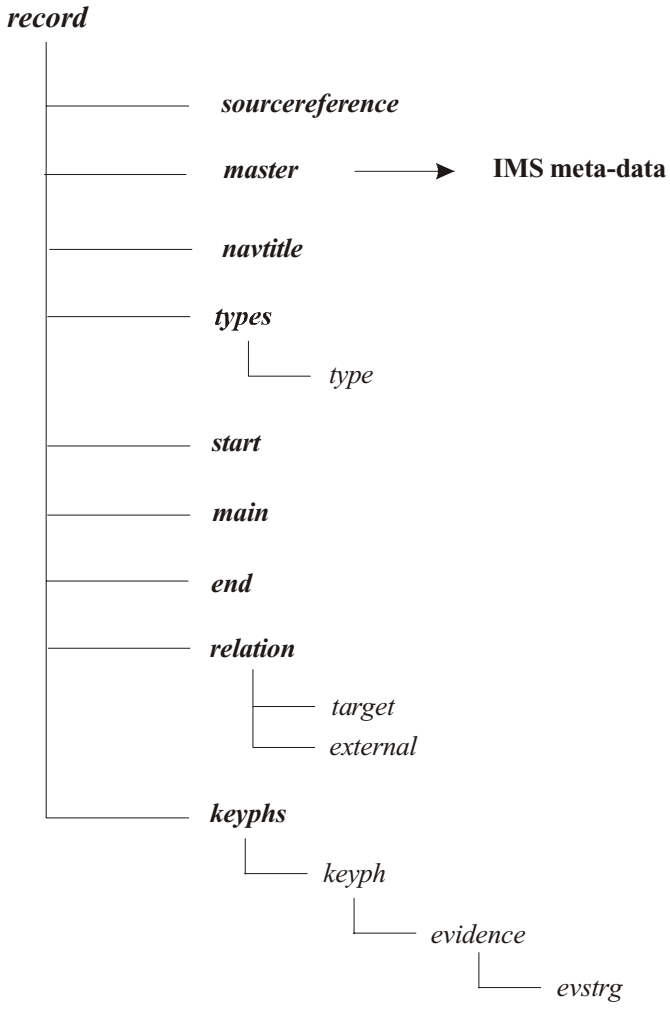

Fig.3 The TRIAL-SOLUTION meta-data record

Each element in the meta-data hierarchy has a specific definition, datatype, and allowable value. Moreover, some of these elements provide information that is basically required for

system design and are therefore marked as mandatory by the *record* element specification of the TRIAL-SOLUTION Meta-Data DTD:

<!ELEMENT record (sourcereference, master?, navtitle, types, start+, main+, end+, relation*, keyphs*)>

The *sourcereference* element allows to construct an identifier that uniquely characterizes the source document of a slice *S* as well as the position of *S* within this source. The content of the *navtitle* element is to be used for navigation purposes and thus must be suitable for inclusion in HTML documents. The *type* element assign slices to different categories taken from a restricted vocabulary which is domain dependent. The TRIAL-SOLUTION project has compiled a respective vocabulary for the field of mathematics. It contains domain specific types like *definition*, *theorem*, *proof* , *example*, *figure*, or *exercise*.[21] Note that multiple types are allowed so that a slice can be an *example* as well as a *figure*. *start* and *end* files are used to make sure that the content of slices is presented within a well-formed technical environment. For example they can contain headlines or start/end-tags of list environments. *main* files contain the content of the leaves. These files may contain references to other auxiliary files.

IV.1.1 The TRIAL-SOLUTION master element

The *master* element describes the TRIAL-SOLUTION core of meta-data fields for resource description that are taken from the IMS Meta-Data Specification:

<!ELEMENT master (contribute*, ts_title?, language*, classification*, rights?)>

The decision to use IMS meta-data for the project rather than Dublin Core was not least taken because of the choice of the formal model for document composition and decomposition. For exchange between the components of the TRIAL-SOLUTION Tool Set sliced documents are encoded as IMS content packages as described in [5]. Now on the one hand the IMS Content Packaging Best Practice Guide[22] recommends IMS meta-data to be used as a default standard, on the other hand this document also defines a mapping of the Dublin Core element set to the IMS meta-data structure. In particular the existence of such a mapping implies that part of the *master* subelements can be also expressed by Dublin Core but IMS meta-data may be more specific in some respects.

An overview of the *master* element substructure is given in Figure 4.[23] The definition of all of its subelements coincides with their definition in the IMS Learning Resource Meta-data Information Model, Version 1.1, [24] with the restriction that extensions are not allowed, except for the *rights* element. The IMS *title* element has been renamed to *ts_title* to distinguish it from the *title* element of the IMS Content Packaging specification. Since the project is concerned with mathematical materials, the *classification* element will be used to store MSC2000 classifications only but the IMS meta-data specification also enables an integration of other existing classification systems. The extension of the *rights* element allows to indicate the legal situation of specific rights:

<! ELEMENT EXTENSION rightsowned* >

---

[21] The complete type list may be found in [13]
[22] see http://www.imsproject.org/metadata/mdbestv1p1.html#IMSCore
[23] Underlining indicates that subelements belong to the IMS core of the IEEE LTSC LOM Base Scheme, Version 3.5, TRIAL-SOLUTION specific extensions or changes are marked in italics.
[24] see [5]

<! ELEMENT rightsowned (typeofright?, underwhatlaw*, rightowner*)>

1	**general**	°	°	°	
1.2	°	title [*ts_title*]		°	
°	°	°	langstring	°	
				language	
°	°	°	°	string	
1.4	°	language	°	°	
°	°	°	°	°	
2	**lifecycle**	°	°	°	
2.3	°	contribute	°	°	
2.3.1	°	°	role	°	
				langstring	
2.3.2°	°	°	centity	°	
				vcard	
2.3.3°	°	°	date	°	
°	°	°	°	datetime	
				description	
					langstring
°	°	°	°	°	
6	**rights**	°	°	°	
6.1	°	cost	°	°	
			langstring		
6.2	°	copyrightandotherrestrictions	°	°	
			langstring		
6.3	°	description	°	°	
°	°	°	langstring	°	

				language	
°	°	°	°	string	
		extension			
			rightsowned		
				typeofright	
				underwhatlaw	
°	°	°	°	*rightowner*°	
9	**classification**	°	°	°	
9.1	°	purpose	°	°	
°	°	°	langstring	°	
°	°	°	°	string	
9.2		taxonpath			
9.2.1			source		
9.2.2			taxon		
9.2.2.1				id	
9.2.2.2				entry	
					langstring
9.3	°	description	°	°	
°	°	°	langstring	°	
				language	
				string	
9.4	°	keywords	°	°	
°	°	°	langstring	°	
				language	
°	°	°	°	string	

Fig. 4 IMS Meta-data Elements from the TRIAL_SOLUTION Master Meta-data Set ◦

It should be mentioned that the TRIAL-SOLUTION *master* element does not include any special meta-data for explicitely describing educational or pedagogic features of a resource resp. slice as this is provided by the IMS *educational* category.[25] Instead information about the user's preferences and interests and about his/her previous knowledge is processed by the system based on settings in a separate internal user model. This way of treating educational issues as user dependent rather than document dependent takes into account that the special focus of the project is not on previously existing learning resources but on personalized documents.

IV.2 The TRIAL-SOLUTION Meta-data DTD [26]

```
<!ELEMENT record    ((sourcereference, master?, navtitle,
types, start+, main+, end+, relation*, keyphs?)|(navtitle,
logical_id))>

<!ELEMENT sourcereference (#PCDATA)>

<!ELEMENT master (contribute*, ts_title?, language*,
classification*, rights?)>

<!-- The definition of the subelements of the master element
coincides with their definition in the IMS Metadata
Specification with the restriction that extensions are not
allowed, except for the rights element
title has been renamed to ts_title to easily avoid clashes
with title element from IMS Content Packaging specification --
>
<!ELEMENT CONTRIBUTE (role?, centity*, date?)>
<!-- Type of contribution defined by Role -->

<!ELEMENT role (langstring?)>

<!-- langstring: A string in a particular language -->
<!ELEMENT langstring (#PCDATA)>
<!ATTLIST langstring lang CDATA #IMPLIED>

<!ELEMENT centity (vcard?)>

<!ELEMENT vcard   (#PCDATA)>
<!-- STRING per vCard specification -->

<!ELEMENT date (datetime?, description?)>

<!ELEMENT datetime (#PCDATA)>
```

[25] The *type* element however may be viewed as an analogon of the *learning resource type* element of IMS which has been adapted to the domain specific needs.

[26] This part is taken from Appendix C of [12]

```
<!-- Per  W3C, e.g., 1999-08-07 -->

<!ELEMENT description (langstring*)>

<!ELEMENT ts_title (langstring*)>

<!ELEMENT language (#PCDATA)>
<!-- Human Language -->

<!ELEMENT classification (purpose?, taxonpath*, description?,
KEYWORDS*)>
<!-- classification is a category element -->
<!-- Description or cataloging of a characteristic. -->

<!ELEMENT purpose (langstring?)>

<!ELEMENT taxonpath (source?, taxon*)>
<!-- A taxonomic path in a specific classification. There may
be different paths, in the same or different classifications,
that describe the same characteristic. -->

<!ELEMENT KEYWORDS (langstring*)>

<!ELEMENT source (#PCDATA)>

<!ELEMENT taxon (id?, entry?)>
<!-- Ordered list -->

<!ELEMENT id    (#PCDATA)>
<!-- Alphanumeric identifier -->

<!ELEMENT entry (langstring*)>

<!ELEMENT rights (cost?, copyrightandotherrestrictions?,
description?, extension?)>
<!-- rights is a category element -->
<!-- Conditions of use of the resource. -->

<!ELEMENT cost (langstring*)>
<!-- Boolean, yes|no -->

<!ELEMENT copyrightandotherrestrictions (langstring*)>
<!-- Boolean, yes|no -->

<!-- The following elements extend the IMS Specification of
rights -->

<!ELEMENT extension   (rightsowned*)>

<!ELEMENT rightsowned (typeofright?, underwhatlaw*,
rightowner*)>
```

```
<!ELEMENT typeofright (#PCDATA)>
<!-- Rights that are owned -->

<!ELEMENT underwhatlaw (#PCDATA)>
<!-- Laws of which country govern these rights -->

<!ELEMENT rightowner (#PCDATA)>
<!-- Who owns the right -->

<!ELEMENT navtitle    (#PCDATA)>
<!ATTLIST navtitle    isvisible    CDATA    #IMPLIED>

<!ELEMENT logical_id    (#PCDATA)>


<!ELEMENT types      (type+)>
<!ELEMENT type       (#PCDATA)>

<!ELEMENT start        EMPTY>
<!ATTLIST start        href   CDATA    #IMPLIED
                       type (
                         latex |
                         latex_ps |
                         latex_pdf |
                         latex_ps_slide |
                         latex_pdf_slide |
                         html |
                         all
                       ) "all"
>

<!ELEMENT main         EMPTY>
<!ATTLIST main         href   CDATA    #IMPLIED
                       type (
                         latex |
                         latex_ps |
                         latex_pdf |
                         latex_ps_slide |
                         latex_pdf_slide |
                         html |
                         all
                       ) "all"
>

<!ELEMENT end          EMPTY>
<!ATTLIST end          href   CDATA    #IMPLIED
                       type (
                         latex |
                         latex_ps |
                         latex_pdf |
                         latex_ps_slide |
                         latex_pdf_slide |
```

```
                        html |
                        all
                      ) "all"
>

<!ELEMENT relation    ((target | external)*)>
<!ATTLIST relation    kind     (References  |
                                Requires |
                                IsSubsumedBy |
                                Before |
                                NotWith |
                               IsAugmentedBy
                                ) #REQUIRED
>

<!ELEMENT target    EMPTY>
<!ATTLIST target    href    CDATA  #REQUIRED
                    author    CDATA #IMPLIED
                    status (assigned | inherited | proposed)
"assigned"
                    inherits (none | up | down | updown)
"none"
>

<!ELEMENT external  #PCDATA>
<!ATTLIST external  href CDATA     #REQUIRED
                    author    CDATA #IMPLIED
                    status (assigned | inherited | proposed)
"assigned"
                    inherits (none | up | down | updown)
"none"
>
<!ELEMENT keyphs    (keyph)*>

<!ELEMENT keyph     (evidence?)>
<!ATTLIST keyph     href      CDATA    #REQUIRED
                    author    CDATA    #IMPLIED
                    status (assigned | inherited | proposed)
"assigned"
                    inherits (none | up | down | updown)
"none"
>
<!ELEMENT evidence    (evstrg)*>
<!ATTLIST evidence    perc    CDATA    #REQUIRED>
<!ELEMENT evstrg      (#PCDATA)>
<!ATTLIST evstrg      weight    CDATA    #IMPLIED>
```

V. References

[1] Dublin Core Metadata Element Set, Version 1.1: Reference Description. http://dublincore.org/documents/1999/07/02/dces

[2] IEEE Draft Standard for Learning Object Metadata. LOM Scheme Version 3.5, July 1999 http://ltsc.ieee.org/doc/wg12/scheme.html

[3] IEEE Draft Standard for Learning Object Metadata. IEEE P1484.12/D6.0, February 2001 http://ltsc.ieee.org/doc/#LOM Draft

[4] IMS Learning Resource Meta-data Specification V.1.2. http://www.imsproject.org/metadata/index.html#version1.2

[5] IMS Content Packaging Specification. V. 1.1.2 http://www.imsproject.org/content/packaging/index.html#version1.1.2

[6] MSC2000 - Mathematics Subject Classification Scheme. http://www.zblmath.fiz-karlsruhe.de/MATH/msc/index

[7] AMMTF Mathematics Metadata Base Scheme Recommendations. http://mathmetadata.org/ammtf/basescheme

[8] AMMTF Level II Draft Classification. http://mathmetadata.org/ammtf/taxonomies/level2_draft_0_5.html

[9] Schlagwortnormdatei (SWD). http://www.ddb.de

[10] International DOI Foundation (IDF): The DOI Handbook, Version 1.0.0* February 2001. http://www.doi.org/handbook_2000/index.html

[11] Norman Paskin: E-Citations: actionable identifiers and scholarly referencing. Paper version in: *Learned Publishing* (ISSN 0953-1513) 2000, vol 13 July (no. 3) pp. 159-168. Electronic version under: http://www.doi.org/citations4.pdf

[12] B. I. Dahn: Data and Meta-data in the Trial-Solution Project (V.1.1). http://www.math-logic.org/trial/TS-IMS_V_1_1.pdf

[13] M. Hazewinkel: Key words and key phrases in scientific databases. Aspects of guaranteeing output quality, In: Proc. ISI meeting Statistical Publishing Warsaw August 1999, ISI, 2000, 44-48

[14] M. Hazewinkel: Topologies and metrics on information spaces, CWI Quaterly 12 (1999), 93-110

[15] Wolfgang Lenski, Elisabeth Wette-Roch: Pragmatische Aspekte der Wissensmodellierung in wissenschaftlichen Informationssystemen. In: H. P. Ohly , G. Rahmstorf, A. Sigel (eds.): Globalisierung und Wissensorganisation: Neue Aspekte für Wissen, Wissenschaft und Informationssysteme, ERGON Verlag, Würzburg 2000, pp. 363-374

[16] Charles S. Peirce: Collected Papers, Vol. I-VI, ed. By Ch. Hartshone and P. Weiss, Harvard University Press, Cambridge 1931-1935

[17] C.W. Morris: Signs, Language, and Behaviour. New York, 1946

[18] G.H. Müller,.W. Lenski et al.: Omega-Bibliography of Mathematical Logic, Vols. I-VI. Springer-Verlag Heidelberg, 1987

[19] W. Lenski, E. Wette-Roch: The TRIAL-SOLUTION approach to document re-use. Principles and realization. In: Electronic Media in Mathematics (EMM ’01), CD-ROM issued by the Departamento de Matematica da Universidade de Coimbra, Coimbra 2001, 2_3.pdf

[20] A. Troelstra, D. van Dalen: Constructivism in Mathematics I. North-Holland, Amsterdam, 1988, p. 22.