
# Implementing the Scholarly Works Application Profile in DSpace

*A metadata collision analysis for the MIT Open Access Initiative*

*Robert Wolfe, MIT Libraries*

*8 September 2009*

*Version 2.3*

## Introduction

This project is undertaken in support of an effort to fulfill MIT's mandate to build an open access repository of the peer-reviewed scholarly publications of its faculty. The MIT Libraries intend to deposit publications gathered under this mandate into their DSpace@MIT institutional repository, In preparing the repository for the inclusion of this new material, an opportunity has been recognized to audit and improve the DSpace@MIT metadata tables. The goal of this audit is to ensure that metadata applied to the MIT Open Access Initiative content conforms to applicable metadata standards and application profiles from the open access scholarly publications domain. A number of best practices have been developed and have acquired wide usage in this domain since the DSpace@MIT metadata tables were last audited. It is to the advantage of the DSpace@MIT repository to become compliant with the latest open access scholarly works application profiles. The best candidate application profile for adoption, incorporation, or mapping to DSpace DC metadata is the Scholarly Works Application Profile. The Scholarly Works Application Profile is a Dublin Core Application Profile develop by JISC for use with its Eprints repository software. It is fully conformant with the Dublin Core Abstract Model and all of its extension elements have been declared in an appropriate namespace. It is a robust profile and it is targeted towards the material that we hope to include in DSpace@MIT via this OA mandate.

This document will provide the following information:

- A list of Required, Recommended, and Optional elements for the MIT OA Initiative collection, mostly chosen from the SWAP
- A collision analysis of the mapping from DSpace metadata fields to SWAP with recommendations for amendments to the DSpace DC metadata tables
- A list of Vocabulary Extension Schema to adopt with the SWAP attributes.
- Examples of rights metadata schemas

# Table of Contents


Introduction....1
- Significant Changes from version 2.2....3
  - Additions....3
  - Adjusted Metadata Recommendations....3
  - Adjusted Collision Analyses....3
  - Newly Requested Changes to DSpace Metadata Tables....3
- Important Considerations....4
  - Domain Modeling....4
  - Metadata Element Sources....4
  - Compatibility with current DSpace collections....4
  - Creation and submission of metadata conforming to this recommendation....4
  - MIT Open Access Initiative Metadata Information Model....4
  - Outstanding Implementation Questions....5

Metadata Recommendations....6
- Required Elements....6
- Conditionally Required....6
- Recommended Elements....7
- Optional Elements....7

The Collision Analysis....8
- Elements that Require No Adjustment to DSpace DC Metadata Tables....8
  - Elements....8
- Elements that are Not Declared in the SWAP, but are Declared in DSpace DC....9
  - Elements....9
- Elements that are Not Declared in DSpace DC, but are Declared in SWAP....10
  - Peer-reviewed Flag Element....10
  - Has Translation Element....10
  - References Element....11
- Elements that Pose Interesting Problems....12
  - Version Element....12
  - Materials Description Element....13
  - Grant Number Element....14
  - Provenance Element....14
  - Type Elements....15
  - Journal Element....16
  - MIT Author Element....17
  - Metadata Source Element....18
  - Source....18
  - Approving Author Element....19
  - Rights Elements....20

Requested Changes to the DSpace Metadata Tables....23
- In the existing DSpace DC namespace:....23
  - Additions....23
  - Changes to element names....23
- In a new Eprints namespace:....23
  - Additions....23
- In a new Mets Rights namespace:....23
  - Additions....23

New Encoding Schemes....24

Some Examples of rights metadata schemas in increasing order of complexity....25
- SWAP Elements....25
- DC Elements....25
- METS Rights Elements....25
- PREMIS Rights Entity Metadata....26

## *Significant Changes from version 2.2*

### *Additions*

- Outstanding Implementation Questions – A new subsection of the Important Considerations section
- MIT Open Access Initiative Metadata Information Model – A new subsection of the Important Considerations section

### *Adjusted Metadata Recommendations*

- Depositing Author Element – The name of the Depositing Author element has been changed to Approving Author
- Source Element – The necessity of this element has been changed from Required to Conditionally Required
- Metadata Source Element – The necessity of this element has been changed from Required to Conditionally Required.
- Identifier (of Scholarly Work) Element – The name of this element has been changed to Identifier (of Published Version)

### *Adjusted Collision Analyses*

- Version Element – The approved vocabulary has been added to this element's recommendation.
- Source Element – The recommendation for this element has been adjusted to support its new status as Conditionally required. A recommendation to provide a link to the website for the source has been added.
- Metadata Source Element – The recommendation for this element has been adjusted to support its new status as Conditionally required. A recommendation to provide a link to the website for the metadata source has been added.
- MIT Author and Approving Author Elements – The recommendations for these elements have been adjusted to contain the recommendation for the use of URIs for these elements.

### *Newly Requested Changes to DSpace Metadata Tables*

- MIT Author Element – The name of the DSpace metadata field for this element has been changed to dc.contributor.mitauthor. This change should be made in the DSpace Metadata Tables.
- Identifier – Others Element – Add a dc.identifier.pmid field to the DSpace Metadata Tables in order to distinguish Pub Med identifiers from other kinds.

## *Important Considerations*

### *Domain Modeling*

Mapping the Scholarly Works Application Profile to DSpace begins with the mapping of entities from the SWAP's domain model to the DSpace content model. In this case the most important mapping is:

- SWAP:Expression equals Dspace:Item.

All vital metadata should be attached to the DSpace Item, even if the SWAP assigns it to a different entity in its domain model.

### *Metadata Element Sources*

The mapping table accompanying this analysis contains a list of elements identified and named by MIT Libraries staff independent from any formal metadata standard or application profile. There are some elements in the table that currently are not defined in the SWAP. Similarly, there are elements that are not yet defined in the DSpace Metadata Tables.

### *Compatibility with current DSpace collections*

Any amendments to the DSpace metadata tables must be backwards compatible. They must not overload an element with values that reflect two separate semantic definitions for the element. Also, they must not create a scenario where two elements share the same semantic definition and split the values for what should be a single element between them. No currently declared elements in the DSpace metadata tables will be removed or redefined. This analysis will likely recommend the addition of elements to the DSpace metadata tables.

### *Creation and submission of metadata conforming to this recommendation*

The set of elements that are here recommended for use in the MIT Open Access repository is much larger than are usually provided for a DSpace Item. We will not use every element in this recommendation for every Item, but it is likely that we will want to use metadata fields that are currently not part of the web submission process. Alternative submission processes will likely need to be employed. Much of the metadata will be provided by publishers, also necessitating alternative submission processes. We will need to design metadata workflows and cataloging systems for these processes. It is crucial that workflows for creating or processing metadata be designed so as to provide the cataloger the ability to edit any of the metadata fields listed in this document.

### *MIT Open Access Initiative Metadata Information Model*

More detailed definitions of each element, including brief cataloging instructions have been prepared at:

https://wikis.mit.edu/confluence/display/LIBMETADATA/MITOAInfoModel

## *Outstanding Implementation Questions*

- Element: Date Available – Will the DSpace System automatically provide a value for this field upon ingest? If so, will the process replace the user supplied value with the system value? Will it merely add an additional Date Available element that contradicts the user supplied value?
- Element: Type (DSpace Type Vocabulary) – The value "conference proceeding" needs to be added to the vocabulary.
- Element: Type (Eprints Type Vocabulary) – The Eprint Type Vocabulary provides the ability to more specifically describe the genre of the DSpace Item than is possible via the DSpace Type Vocabulary. Is this increased specificity necessary?
- Element: MIT affiliation – A look-up service needs to be developed so that catalogers can identify and add the MIT affiliated authors' primary affiliations.
- Elements: MIT Author, Approving Author – Ideally, these elements would contain identifiers for MIT affiliated authors. It is undesirable, but currently necessary to duplicate the names of MIT affiliated authors in three elements (Author, MIT Author, Approving Author). A service needs to be developed to assign identifiers (preferably URIs) to authors entered in the MIT Author and Approving Author fields.
- Elements: Rights Holder, Rights Context, Rights Basis, Rights Declaration – These element recommendations need to be discussed with Ellen Duranceau and Richard Rodgers.
- Element: (Identifier of Published Version) – A look-up service needs to be implemented to allow catalogers to identify DOIs for the published version of Items in the collection.

## Metadata Recommendations

### *Required Elements*

| Element Name | SWAP Attribute | DSpace Metadata Field |
|---|---|---|
| Author | Creator | dc.contributor.author |
| Title | Title | dc.title.none |
| Embargo | Date Available | dc.date.available |
| Type (DSpace Type Vocabulary) | Type | dc.type.none |
| Type (Eprints Type Vocabulary) | Type | dc.type.uri |
| Language | Language | dc.language.iso |
| Identifier | Identifier | dc.identifier.uri |
| MIT Affiliation | Affiliated Insitution | dc.contributor.department |
| Peer-reviewed Flag | Status | eprint.status |
| Version | Version Number or String | eprint.version |
| Entity Type | Entity Type | dc.type.uri |
| MIT Author | n/a | dc.contributor.mitauthor |
| Rights Context | n/a | metsrights.contextClass<br>metsrights.otherContextType |
| Rights Declaration | n/a | dc.rights.none<br>dc.rights.uri |
| Rights Basis | n/a | metsrights.rightsCategory<br>metsrights.otherCategoryType |
| Rights Holder | n/a | metsrights.rightsHolder.rightsHolderID<br>metsrights.rightsHolder.rightsHolderName |
| Approving Author | n/a | dc.contributor.approver |

### *Conditionally Required*

| Element Name | SWAP Attribute | DSpace Metadata Field |
|---|---|---|
| Source | n/a | dc.source.none<br>dc.source.uri |
| Metadata Source | n/a | dc.source.metadata |

## *Recommended Elements*

| Element Name | SWAP Attribute | DSpace Metadata Field |
|---|---|---|
| Publication Date (pub. version) | n/a | dc.date.issued |
| Date Submited for Publication | n/a | dc.date.submitted |
| Citation | Bibliographic Citation | dc.identifier.citation |
| Abstract | Abstract | dc.description.abstract |
| Publisher | Publisher | dc.publisher.none |
| Identifier (of Published Version) | Identifier | dc.relation.isversionof |
| Materials Description | n/a | Bitstream Description |
| Journal | n/a | dc.relation.journal |

## *Optional Elements*

| Element Name | SWAP Attribute | DSpace Metadata Field |
|---|---|---|
| Identifier – Others | Identifier | dc.identifier.other<br>dc.identifier.none<br>dc.identifier.govdoc<br>dc.identifier.isbn<br>dc.identifier.issn<br>dc.identifier.sici<br>dc.identifier.ismn<br>dc.identifier.pmid |
| Subjects/Keywords | Subject | dc.subject.none<br>dc.subject.classification<br>dc.subject.ddc<br>dc.subject.lcc<br>dc.subject.lcsh<br>dc.subject.mesh<br>dc.subject.other |
| Description | Description | dc.description.none |
| Has Version | Has Version | dc.relation.hasversion |
| Editor | Editor | dc.contributor.editor |
| Has Translation | Has Translation | eprints.hasTranslation |
| Supervisor | Supervisor | dc.contributor.advisor |
| Sponsor | Funder | dc.description.sponsorship |
| Grant Number | Grant Number | eprint.grantNumber |
| Title of Published Version | Title | dc.title.alternative |
| References | References | dc.relation.references |
| Provenance | n/a | dc.description.provenance |
| Series | n/a | dc.relation.ispartofseries |

## The Collision Analysis

### *Elements that Require No Adjustment to DSpace DC Metadata Tables.*

Though some of the paired elements/attributes have dissimilar names, all pairs have functionally equivalent semantic definitions.

### *Elements*

| **SWAP Attribute** | **DSpace Metadata Field** |
|---|---|
| Creator | dc.contributor.author |
| Title | dc.title.none |
| Date Available | dc.date.available |
| Abstract | dc.description.abstract |
| Publisher | dc.publisher.none |
| Language | dc.language.iso |
| Identifier [of DSpace Item] | dc.identifier.uri |
| Identifier [of Published Version] | dc.relation.isversionof |
| Identifer [other Identifiers] | dc.identifier.other<br>dc.identifier.none<br>dc.identifier.govdoc<br>dc.identifier.isbn<br>dc.identifier.issn<br>dc.identifier.sici<br>dc.identifier.ismn |
| Bibliographic Citation | dc.identifier.citation |
| Subject | dc.subject.none<br>dc.subject.classification<br>dc.subject.ddc<br>dc.subject.lcc<br>dc.subject.lcsh<br>dc.subject.mesh<br>dc.subject.other |
| Affiliated Institution | dc.contributor.department |
| Description | dc.description.none |
| Has Version | dc.relation.hasversion |
| Editor | dc.contributor.editor |
| Supervisor | dc.contributor.advisor |
| Funder | dc.description.sponsorship |
| Title [of Published Version] | dc.title.alternative |

#### Recommendation For These Elements

***Adopt these elements for use in describing Scholarly Works deposited as part of the MIT Open Access mandate. Do not change the metadata tables, use the DSpace DC elements as currently defined.***

*Elements that are Not Declared in the SWAP, but are Declared in DSpace DC*

### *Elements*

| **Element Name** | **DSpace Metadata Field** |
|---|---|
| Publication Date (of published version) | dc.date.issued |
| Date Submitted for Publication | dc.date.submitted |
| Series | dc.relation.ispartofseries |

**Recommendation For These Elements**

***Adopt these elements for use in describing scholarly publications deposited as part of the MIT Open Access mandate. Do not change the metadata tables, use the DSpace DC elements as currently defined.***

## *Elements that are Not Declared in DSpace DC, but are Declared in SWAP*

Recommendations for these elements must be made on an element-by-element basis.

### *Peer-reviewed Flag Element*

- **SWAP Attribute** **DSpace Metadata Field**
- Status eprint.status

This element records whether or not a member of the collection has been peer-reviewed. This information in this element is not a duplication of the information in the Version element. It refers to the status of the intellectual content of the Expression rather than its stage of publication. Only two values in the Version element can be reliably mapped to a Peer-reviewed status—preprint and postprint. The other values (published, reprint, other) may or may not represent peer-reviewed content. Unfortunately, there is no appropriate dublin core term for this element. It should be added to the DSpace metadata tables in the eprint namespace.

#### Recommendation For This Element

***Adopt this element for use in describing scholarly publications deposited as part of the MIT Open Access mandate. Make this a required element for this collection. Add this element to the DSpace metadata tables. Create a new namespace in the DSpace metadata table (eprint) and put the element (status) in that namespace instead of the DSpace DC namespace. See:***
***http://www.ukoln.ac.uk/repositories/digirep/index/Scholarly_Works_Application_Profile#Status***

### *Has Translation Element*

- **SWAP Attribute** **DSpace Metadata Field**
- Has Translation eprint.hasTranslation

There is no DSpace DC element reserved for this information. Translations are currently declared in DSpace via the dc.relation.hasversion element, which is technically correct, but imprecise. Though this element is optional for DSpace Items deposited via the OA initiative, implementation of this new element in the eprints namespace will provide highly valuable to future multi-language collections deposited in DSpace.

#### Recommendation For This Element

***Adopt this element for use in describing scholarly publications deposited as part of the MIT Open Access mandate. Add this element to the DSpace metadata tables. Put it in the eprint namespace (See http://www.ukoln.ac.uk/repositories/digirep/index/Scholarly_Works_Application_Profile#Has_Translation) instead of the DSpace DC namespace.***

## *References Element*

| • **SWAP Attribute** | **DSpace Metadata Field** |
|---|---|
| • References | dc.relation.references |

This element exists in qualified Dublin Core, but not in the DSpace DC tables. It should be added to the tables as soon as possible.

### Recommendation For This Element

***Adopt this element for use in describing scholarly publications deposited as part of the MIT Open Access mandate. Recommend but do not require this element. Add this element to the DSpace DC tables in the DspaceDC namespace.***

## *Elements that Pose Interesting Problems*

Recommendations for these elements must be made on an element-by-element basis.

### *Version Element*

| **SWAP Attribute** | **DSpace Metadata Field** |
|---|---|
| Version Number or String | eprint.version |

First, remember that the MIT Open Access Initiative repository has adopted the domain model of the SWAP, which is based on the FRBR domain model. In the SWAP domain model, a Scholarly Work is defined as, "A distinct intellectual or artistic creation. A ScholarlyWork equates to a FRBR Work." An Expression is defined as, "The intellectual or artistic realization of a ScholarlyWork. A ScholarlyWork may be expressed as several different revisions, translations or other versions. An Expression equates to a FRBR Expression." The MIT OAI has defined a DSpace Item to be equivalent to an Expression.

The Version element was first proposed for use with a controlled vocabulary. That proposed vocabulary (preprint, postprint, published) contains values that identify the type of SWAP Expression the DSpace Item represents. This type vocabulary is distinct from the other type vocabularies defined elsewhere in this recommendation. Those vocabularies are:

- DSpace Type Vocabulary – The nature or genre of a resource. A generic list for the way the intellectual content of the resource is organized and presented (example values: Article, Dataset, Technical Report).
- Eprints Type Vocabulary – Also the nature or genre of a resource, values more specifically categorize scholarly publications (example values: Conference Poster, Journal Article, Thesis or Dissertation).
- Eprints Entity Type Vocabulary – The type of entity the resource represents in the SWAP domain model (example values: Scholarly Work, Expression, Manifestation, Copy).

The types of Expressions that are of interest to the MIT OAI repository represent the particular stages of preparation of a Scholarly Work. The recommended vocabulary for this element is:

- preprint
- postprint
- published
- reprint
- other

The first four values have established, shared definitions. Preprints are commonly defined as not having been referred or peer-reviewed. Postprints are commonly defined as having been referred or peer-reviewed. Published versions are those that a have been included in a commercial publication. Postprints are sometimes identical to published versions. Reprints are also those that have been included in a commercial publication, the reprint version representing inclusion in an additional, different commercial publication from the original. It is conceivable that the MIT OAI will receive examples of each of these types of Expressions. It is also conceivable that the MIT OAI will receive Expressions that are neither preprints, postprints, published, or reprints. These might be versions that are net yet ready to be called a preprint, or that are not destined for the publication or referee process. The "other" value has been added to this vocabulary to allow for the inclusion and categorization of these types of Expressions that are not anticipated or defined.

There is no established practice in capturing this information in DSpace DC. The following elements all represent potential candidates, none of them are optimal:

- dc.identifier.other
- dc.relation.ispartofseries
- dc.relation.isversionof
- dc.relation.hasversion

The dc.identifier.other element contains a melange of values and is inappropriate for this element. The dc.relation.ispartofseries is defined differently from the version element and is also inappropriate. This element is not intended to define a relationship between the DSpace Item and another entity that represents a different version of the Scholarly Work. Therefore, the DSpace Metadata Fields dc.relation.isversionof and dc.relation.hasversion are poor candidates for this element, as they record relationships between the DSpace Item and other entities. There is an attribute of Expressions defined by the SWAP as "A version number or version string associated with the described expression of the eprint." This is exactly the element we want to use with our controlled vocabulary.

### Recommendation For This Element

***Adopt this element for use in describing scholarly publications deposited as part of the MIT Open Access mandate. Make this a required element for this collection. Add this element to the DSpace metadata tables. Put it in the eprint namespace (See http://www.ukoln.ac.uk/repositories/digirep/index/Scholarly_Works_Application_Profile#Version_Number_or_String) instead of the DSpace DC namespace. This information is vital to the success of the OA Initiative and DSpace should seek to insert some clarity into its metadata by separating these metadata values into their own element.***

***A vocabulary for this element has been approved with the following elements:***

- ***Original manuscript***
- ***Peer-reviewed manuscript***
- ***Final published version***

## *Materials Description Element*

| **Element Name** | **DSpace Metadata Field** |
|---|---|
| Materials Description | Bitstream Description |

This element is intended to capture a description for every file that is uploaded to DSpace as part of an Item. This description is displayed in the "File in this item" section of the simple and full records for a DSpace Item. If used properly, it helps identify the correct file to download for a particular Item. This information is entered during the upload section of the DSpace Item submission process. It does not become part of the Item metadata record, rather it is stored in the Bitstream metadata.

### Recommendation for this Element

***Adopt a best practice of providing this information where possible. Recommend but do not require this information.***

## *Grant Number Element*

- **SWAP Attribute** **DSpace Metadata Field**
- Grant Number eprint.grantNumber

This is another element for which there is mixed practice. The following elements have all been used, none of them are optimal:

- dc.relation.ispartof
- dc.description.sponsorship
- dc.identifier.other

### Recommendation For This Element

***Adopt this element for use in describing scholarly publications deposited as part of the MIT Open Access mandate. Add this element to the DSpace metadata tables. Put it in the eprint namespace (See http://www.ukoln.ac.uk/repositories/digirep/index/Scholarly_Works_Application_Profile#Version_Number_or_String) instead of the DSpace DC namespace. Consider migrating all grant number metadata that is currently declared in other fields to the eprint.grantNumber field. The fields currently used to capture this information are catch-all fields that are overloaded with metadata values that ought to belong to separate elements with different semantic definitions. This information is vital to the success of the OA Initiative and DSpace should seek to insert some clarity into its metadata by separating grant numbers into their own element.***

## *Provenance Element*

- **Element Name** **DSpace Metadata Field**
- Provenance dc.description.provenance

DSpace uses the dc.description.provenance field to automatically record changes to the chain of custody of the DSpace Item, including changes to the Item. DSpace record only that part of the history of the scholarly publication that occurs within the DSpace system. Any pertinent content or chain of custody changes that occur outside the DSpace system should also be recorded in this field.

### Recommendation For This Element

***Adopt this element for use in describing scholarly publications deposited as part of MIT's Open Access mandate. DSpace has named this field dc.description.provenance, mistaking provenance for a qualifier of the description element. In fact, provenance does not refine the description element and the field should be renamed to dc.provenance.none. This element is optional.***

## *Type Elements*

| SWAP Attribute | DSpace Metadata Field |
|---|---|
| Type (DSpace Type Vocabulary) | dc.type.none |
| Type (Eprints Type Vocabulary) | dc.type.uri |
| Entity Type (Eprints EntityType Vocabulary) | dc.type.uri |

SWAP overloads the dc:type element with two separate semantic definitions:

1) Entity Type (Values conform to the Eprints EntityType Vocabulary Encoding Scheme)
2) Type (Values conform to the Eprints Type VES)

In addition, DSpace has it's own

3) Type Vocabulary (Custom Controlled Vocabulary – String Values)

that is very similar to the Eprints Type Vocabulary. Currently we have one type element in the DSpace metadata tables--dc.type.none. If we use dc.type.none for each of these elements, then the field could be repeated three times for each DSpace Item, each repetition containing a value from a different encoding scheme. This is a legal use of Dublin Core, but it is not recommended in this case. Using the same field for values from two different vocabularies works well when the values are well-established, persistent URIs. The URIs will indicate the vocabulary encoding scheme to which they belong. It is important that we are be able to deduce which vocabulary a particular value for this element represents. The values of the two Eprints Vocabulary Encoding Schemes are URIs, while the values of the DSpace Type Vocabulary are Strings. It is important then to separate the strings from the URIs.

### Recommendation For These Elements

***Adopt these three elements for use in describing scholarly publications deposited as part of the MIT Open Access mandate. Make them all required. Add a new element, "dc.type.uri" to the DSpace metadata tables, putting it in the DSpace DC namespace. Use dc.type.uri for the two Eprints Vocabulary Encoding Schemes. Constrain the values of the dc.type.none to the DSpace Type vocabulary. It is important to require both the Eprints Type URI and the DSpace Type String. One to conform with the SWAP and the other to make sure that the OA content plays nicely with the other items in DSpace.***

## *Journal Element*

- **Element Name** **DSpace Metadata Field**
- Journal dc.relation.journal

This element records a relationship between a member of the collection and the journal in which another version of the scholarly work was published. This element is similar to the MIT Author and MIT Affiliation elements. The Journal element ought to be an identifier, preferably a URI. An element for the identifier of a publication series or journal exists in the DSpace Dublin Core – dc.identifier.issn. Not every journal will have an issn. There may also be a need to record just the name of the journal as a text string. There are five fields in the DSpace metadata record where series and journal information is placed. In the table below series and journal metadata for three different kinds of scholarly works is compared. Sample content from these fields is shown.

| | Working Paper | Thesis | Published Article |
|---|---|---|---|
| dc.identifier.citation | | | Van Evera, Stephen. "Strategy for the Terror War". Newsday, October 4, 2001 |
| dc.relation.ispartofseries | MIT-CEEPR;09-005WP | | |
| dc.relation.sponsorship | Massachusetts Institute of Technology. Center for Energy and Environmental Policy Research. | | |
| dc.publisher.none | | Massachusetts Institute of Technology | Newsday |
| dc.description.none | | Thesis (B.S.)--Massachusetts Institute of Technology, Dept. of Aeronautical Engineering, 1929. | |

The dc.relation.sponsorship and dc.publisher.none fields contain very similar information for working papers and theses. This is unfortunate, but not relevant as the information is not quite a Series or Journal name. These fields are meant to capture the names of organizations, not their publications. Unfortunately, some published articles have put the names of Journals into the dc.publisher.none field. This is semantically incorrect, though occasionally the name of the publisher and the name of the journal are identical. The dc.relation.ispartofseries and dc.description.none field both contain series statements. These statements include volume and sequence information, which is more information than is wanted for the journal element. The dc.identifier.citation element likewise contains more information than is wanted.

Harvard University has defined a new element for this information in their DASH metadata element set. This element is dc.relation.journal. They have defined this element to contain the name of the journal in which a version of the article was published. We should adopt this element and share practice for its use with the DASH.

### Recommendation For This Element

***Adopt this element for use in describing scholarly publications deposited as part of the MIT Open Access mandate. Recommend this element but do not require it. Add this element to the DSpace DC tables in the DspaceDC namespace. Use authorized headings for entries in the dc.relation.journal field.***

## *MIT Author Element*

| • **Element Name** | **DSpace Metadata Field** |
|---|---|
| • MIT Author | dc.contributor.mitauthor |

The purpose of this field is to be able to distinguish authors of scholarly works that are affiliated with MIT. The need for this field is the desire to provide a service that tracks and publishes information on the publication histories of MIT affiliated scholarly authors. The MIT Open Access Initiative Repository records affiliations in DSpace Item records at the department level.

There are a few considerations for the implementation of this element. One is the fact that affiliations change. An author may join or leave MIT. Is it more appropriate to identify an author as MIT affiliated on all of his publications, or just those that were published during the time that she was affiliated with the Institute?

Another consideration is the relationship of this element to the Affiliated Institution element. That element records the department name of authors that have MIT affilations. It is important that the relationship of author to department be recorded. It is possible that the recording method of this relationship may also be used to distinguish the MIT authors of a scholarly work without the necessity for an additional element in the metadata record.

A third consideration is the fact that another metadata element has already been defined to store author identifications (dc.contributor.author). It is vital that author names appear in that field. Only entries in that element are indexed to enable browsing by author names. It is also very bad practice to use two elements to store metadata that has the same semantic definition. This leaves us with the disagreeable prospect of duplicating some of the names in the dc.contributor.author field in this new field.

This element could be defined by overloading one of the existing contributor fields in the DSpace dc namespace, creating a new field in that namespace (perhaps dc.contributor.mitauthor), or creating a new element in an entirely new namespace (perhaps one for custom fields that Libraries have added to the DSpace metadata tables).

The contents of the field could either be the name of the author, or an identifier for the author. Entering the name in this field duplicates information in two separate fields in the record. It is preferable to use an identifier for each author.

Ideally we would create URIs for MIT affiliated authors as part of an authorities service for agents at MIT, a component of the service to track and publish information on the publication histories of MIT affiliated authors. These URIs would be used to create networked information about the authors including their affiliation to a particular department. The departments would also be given URIs. They are a kind of agent crucial to the function of the authorities and publication history services. It is possible that the SWAP Agent Entity definition could be used as an ontology for these author and department entities.

### Recommendation for this element

***Adopt this element for use in describing scholarly publications deposited as part of the MIT Open Access mandate. Make this element required. Add a new field to the existing DSpace DC namespace (dc.contributor.mitauthor) to capture this information. It is not necessary to include this element in the Item metadata records (brief or full) displayed in DSpace. The values of this field ought to be URIs that are created and maintained for each MIT affiliated author. Until an author identifier service is implemented duplicate the names of MIT affiliated authors from the Author element in this field.***

## *Metadata Source Element*

| • **Element Name** | **DSpace Metadata Field** |
|---|---|
| • Metadata Source | dc.source.metadata |

It is conceivable that origin of content deposited in DSpace will be different from the origin of metadata for that content. In some cases this harvested or received metadata arrives with the requirement for attribution of its source. An element in which to store this attribution does not exist in the DSpace Metadata tables. There is no element of this type in the Dublin Core or Eprint namespaces. In fact, there are few examples of this element type in any metadata schema (LOM being one).

### Recommendation for this element

***Adopt this element for use in describing scholarly publications deposited as part of the MIT Open Access mandate. Make this element conditionally required. The condition for required inclusion of this element is the occasion when metadata for an Item has been harvested or received from an external source that requires attribution. Add a new field to the existing DSpace DC namespace (dc.source.metadata) to capture these attributions. Use authorized headings for the names of metadata sources. Attribution should contain a link to the website for the metadata source whenever possible.***

## *Source*

| • **Element Name** | **DSpace Metadata Field** |
|---|---|
| • Source | dc.source.uri<br>dc.source.none |

The DSpace Dublin Core metadata tables contain two elements for this kind of information-- dc.source.uri and dc.source.none. The scope notes for these elements read "Do not use; only for harvested metadata." This field is currently used for only one Item in DSpace@MIT. The source field for that Item contains publisher and series information that is not appropriate for this element. It is obvious that the original intended use of this element has been abandoned or was never adopted. This field should be redefined to contain the name of the agent that provided the content of the Item for deposit. These agents may be faculty members, publishers, or other organizations that provide content for the repository.

### Recommendation for this element

***Adopt this element for use in describing scholarly publications deposited as part of the MIT Open Access mandate. Make this element conditionally required. The condition for required inclusion of this element is the occasion when metadata for an Item has been harvested or received from an external source that requires attribution. Use authorized headings for the names of faculty members and publishers that provide content for inclusion in the repository. Whenever possible include a link to a faculty member or publisher's website.***

## *Approving Author Element*

| **Element Name** | **DSpace Metadata Field** |
|---|---|
| Approving Author | dc.contributor.approver |

This element is very similar, but not identical, to the Source element. Ellen Duranceau has explained the need for the "Approving Author" element:

> The idea behind "depositing author" is to capture which author approved the idea of submitting the paper to DSpace. (The actual submitter is likely to be someone else entirely, like me or an administrative assistant.)
>
> The idea is that we need to know which of the authors said it's ok for the paper to be submitted. That way if there is any question about how the paper ended up in DSpace, we'll know which author gave it forward for deposit or approved it for deposit -- essentially under whose name it has been made available. We consider this a critical piece of information, so that we can be prepared for any possible issues where a coauthor on the paper says they didn't want it in Dspace, or any of a number of issues that might arise in relation to the paper. We'll want to be able to go back to the author that made the paper available to us.
>
> …
>
> Publications that come from a publisher may not have a depositing author. We'd want to have that publisher provenance expressed in the record.
>
> There would normally be only one depositing author, but there could conceivably be two (or more) depositing authors, if two authors gave us the same paper. I don't know what if any procedures we'd want to put in place to prevent such duplication in DSpace.

Ellen's explanation identifies the content of this field as the name of the agent who approves the deposit. This may or may not be the same agent who provided the content for deposit. For example, we may receive content from one faculty member for deposit, but approval from two or more faculty members that are authors of the content. It is even conceivable that we may receive content for deposit from a publisher that is approved by a faculty member, and vice versa. We may also receive approvals for deposit of the same content from both the publisher and faculty authors. We will encounter serious problems if we attempt to capture the depositor and the approver in the same field.

### Recommendation for this element

***Adopt this element for use in describing scholarly publications deposited as part of the MIT Open Access mandate. Add a new field to the existing DSpace DC namespace (dc.contributor.approver) to capture this information. The values of this field ought to be URIs that are created and maintained for each approving author. Until an author identifier service is implemented duplicate the names of approving authors from the Author element in this field.***

## *Rights Elements*

| Element Name | DSpace Metadata Field |
|---|---|
| Rights User Group | metsrights.contextClass<br>metsrights.otherContextType |
| Rights Declaration | dc.rights.none<br>dc.rights.uri |
| Rights Basis | metsrights.rightsCategory<br>metsrights.otherCategoryType |
| Rights Holder | metsrights.rightsHolder.rightsHolderID<br>metsrights.rightsHolder.rightsHolderName |

There are many ways that rights associated with a scholarly publication can be defined. Some of the more common are by copyright statement, license, statute, or contract. These rights can be very complicated. They can be bought and sold, licensed exclusively and non-exclusively, there can often be many claims on one publication. What is of concern to the MIT Open Access repository are the rights that it has been granted and the rights that are passed on to the users of the repository. It may not always be possible to indicate every fine detail of the rights associated with the publication, but there are a few essentials that will allow users to discover their rights.

- Always clearly distinguish between declarations of the rights of the repository and the rights of the user. Indicate the group of users for which the rights apply.
- Always indicate the basis of the rights.
- Always provide access to the contents of a rights statement, whatever its basis. Either reference the contents via an identifier or include the contents within the metadata record.
- Always provide a reference to a rights holder, some agent that can be contacted for more information.
- Whenever possible try to provide more information about the circumstances in which the rights apply (jurisdiction, dates of applicability)
- Whenever possible try to explain to the user her rights.
- Whenever possible indicate the date upon which the determination of rights was made and the agent who made the determination.

There are four required elements to capture:

- The group of users for whom the rights apply
- The basis of the rights (copyright, license, statue, contract)
- The content of the rights statement or its identifier
- The rights holder

The other three essentials usually require more than one element apiece.

- The circumstances under which the rights apply
- What the user can do
- The circumstances of the determination of the rights

The best practice in capturing these rights metadata elements is to group them into metadata records according to 1) the bitstream to which they apply and 2) the user group for whom they apply. This practice needs to adopted for DSpace. We have already created Items with multiple bitstreams that have different rights declarations. In our thesis collection we put two pdfs in each item, one that is available free to the public, and another that is available free to the MIT community or available for purchase by the public. There should be a rights metadata record for each bitstream. Similarly, the second bitstream described above should have two rights metadata records, one for each group of users for whom rights apply (the MIT community and the general public) The rights for each group differ. This brings us to three rights statements necessary for one DSpace Item. It is not reasonable or recommended to put all of these rights metadata records in the DSpace Item metadata. It would be more appropriate to put these rights statements in the bitstream metadata that is recorded in DSpace. It would also be appropriate to create separate xml files for each rights metadata record and deposit them as bitstreams in the DSpace Item record.

At most, one rights metadata record should be included in the DSpace Item metadata. The rights that are most important to record in the DSpace Item metadata record are those that apply to the general public and whose rights statement is the closest to a statement of Open Access. This metadata describes the rights associated with the DSpace Item that best capture the spirit of the repository. In cases like MIT Theses, where there are two rights statements for the general public, both could conceivably be included by duplicating the dc.rights.none or dc.rights.uri fields. Each statement would need to clearly identify the bitstream to which it applies.

In choosing a rights metadata schema several were considered. Examples are included in the final section of this recommendation. There are two types of rights schemas that were considered. One is simple rights metadata element sets that characterize rights for human consumption. The other is Rights Expression Languages that model rights as networked information with enough detail and complexity that this information can be consumed by Digital Rights Management systems used to enforce access to content. The use of a Rights Expression Language as expressive as ODRL is probably beyond what is possible or necessary for these scholarly publications. It is unlikely that DSpace will soon develop a Digital Rights Management system that would use rights encoded in such a language.

The SWAP and Dublin Core efforts to define rights metadata are inadequate for our needs. The PREMIS and METS communities have developed rights metadata schemas that are remarkably similar and fit our needs. The METS Rights schema is a draft schema whose maintenance is doubtful. The PREMIS Rights Entity is part of the widely adopted PREMIS schema—a stable, well-maintained standard. The PREMIS Rights Entity, unfortunately does not provide methods for identifying the rights holder or the group for whom the rights apply.

### Recommendation for these elements

***The rights of the repository, the DSpace Deposit License, is not the primary concern of this metadata recommendation. DSpace has an established system to record and store this information. The DSpace Item metadata record should contain information about the rights of the users of the repository. There is no need to record any metadata about the DSpace Deposit License.***

***Adopt the following elements from the METS Rights schema. Add these elements to the DSpace metadata tables in a new namespace "metsrights" . Make these elements required:***

| | |
|---|---|
| • *Element Name* | *DSpace Metadata Field* |
| • *Rights Context* | *metsrights.contextClass*<br>*metsrights.otherContextType* |
| • *Rights Declaration* | *dc.rights.none*<br>*dc.rights.uri* |
| • *Rights Basis* | *metsrights.rightsCategory* |
| • | *metsrights.otherCategoryType* |
| • *Rights Holder* | *metsrights.rightsHolder.rightsHolderID*<br>*metsrights.rightsHolder.rightsHolderName* |

*The Rights Context fields should be used together, metsrights.otherContextType only applicable if the value of metsrights.contextClass is OTHER. METS has defined a value set for this element that we should adopt. Those values are:*

- *ACADEMIC USER*
- *GENERAL PUBLIC*
- *REPOSITORY MGR*
- *MANAGED GRP*
- *INSTITUTIONAL AFFILIATE*
- *OTHER*

*The Rights Declaration fields should be used to provide access to the contents of either the Copyright statement, license, statute or contract, whichever is applicable. Either or both fields should be used. One allows for the reference to a rights declaration by its identifier, the other for the inclusion of the contents in the metadata record. These elements exist in the DSpace Dublin Core metadata tables.*

*The Rights Basis fields should be used together, metsrights.otherCategoryType only applicable if the value of metsrights.Category value is OTHER. METS has defined a value set for this element that we should adopt. Those values are:*

- *COPYRIGHTED*
- *LICENSED*
- *PUBLIC DOMAIN*
- *CONTRACTUAL*
- *OTHER*

*The Rights Holder fields should be used together to provide the option to indicate the rights holder by name or by ID. Either or both fields should be used.*

*The other attributes and elements in the METS Rights schema ought to be added to the DSpace metadata tables in their own namespace. These elements should be available, but optional*

## Requested Changes to the DSpace Metadata Tables

*In the existing DSpace DC namespace:*

### *Additions*

- dc.type.uri
- dc.relation.references
- dc.relation.journal
- dc.contributor.mitauthor
- dc.contributor.approver
- dc.source.metadata
- dc.identifier.pmid

### *Changes to element names*

| **Current Element Name** | **Recommended Name** |
|---|---|
| dc.description.provenance | dc.provenance.none |

*In a new Eprints namespace:*

### *Additions*

- eprint.status
- eprint.version
- eprint.hasTranslation
- eprint.grantNumber

*In a new Mets Rights namespace:*

### *Additions*

- metsrights.contextClass
- metsrights.otherContextType
- metsrights.rightsCategory
- metsrights.otherCategoryType
- metsrights.rightsHolder.rightsHolderID
- metsrights.rightsHolder.rightsHolderName

## New Encoding Schemes

- SWAP:Type
  see:
  http://www.ukoln.ac.uk/repositories/digirep/index/Eprints_Type_Vocabulary_Encoding_Scheme

- SWAP:Entity Type
  see:
  http://www.ukoln.ac.uk/repositories/digirep/index/Eprints_EntityType_Vocabulary_Encoding_Scheme

- SWAP:Status
  see:
  http://www.ukoln.ac.uk/repositories/digirep/index/Eprints_Status_Vocabulary_Encoding_Scheme

- SWAP:Access Rights
  see:
  http://www.ukoln.ac.uk/repositories/digirep/index/Eprints_AccessRights_Vocabulary_Encoding_Scheme

## Some Examples of rights metadata schemas in increasing order of complexity

### *SWAP Elements*

- eprint.copyrightHolder
- dcterms.accessRights
- dcterms.license

### *DC Elements*

- dcterms.rights
- dcterms.accessRights
- dcterms.dateCopyrighted
- dcterms.license
- dcterms.rightsHolder

### *METS Rights Elements*

- RightsDeclarationMD (RIGHTSDECID, RIGHTSCATEGORY, OTHERCATEGORYTYPE)
  - RightsDeclaration
  - RightsHolder (RIGHTSHOLDERID, CONTEXTIDS)
    - RightsHolderName
    - RightsHolderComments
    - RightsHolderContact
      - RightsHolderContactDesignation
      - RightsHolderContactAddress
      - RightsHolderContactPhone (PHONETYPE)
  - Context (CONTEXTCLASS, CONTEXTID)
  - UserName
  - Permissions (DISCOVER, DISPLAY, COPY, DUPLICATE, MODIFY DELETE PRINT)
  - Constraints (CONSTRAINTTYPE, OTHERCONSTRAINTTYPE)
    - ConstraintDescription

*PREMIS Rights Entity Metadata*

Rights Entity

- rightsStatement (O, R)
  - rightsStatementIdentifier (M, NR)
    - rightsStatementIdentifierType (M, NR)
    - rightsStatementIdentifierValue (M, NR)
  - rightsBasis (M, NR)
  - copyrightInformation (O, NR)
    - copyrightStatus (M, NR)
    - copyrightJurisdiction (M, NR)
    - copyrightStatusDeterminationDate (O, NR)
    - copyrightNote (O, R)
  - licenseInformation (O, NR)
    - licenseIdentifier (O, NR)
      - licenseIdentifierType (M, NR)
      - licenseIdentifierValue (M, NR)
    - licenseTerms (O, NR)
    - licenseNote (O, R)
  - statuteInformation (O, R)
    - statuteJurisdiction (M, NR)
    - statuteCitation (M, NR)
    - statuteInformationDeterminationDate (O, NR)
    - statuteNote (O, R)
  - rightsGranted (O, R)
    - act (M, NR)
    - restriction (O, R)
    - termOfGrant (M, NR)
      - startDate (M, NR)
      - endDate (O, NR)
    - rightsGrantedNote (O, R)
  - linkingObjectIdentifier (O, R)
    - linkingObjectIdentifierType (M, NR)
    - linkingObjectIdentifierValue (M, NR)
  - linkingAgentIdentifier (O, R)
    - linkingAgentIdentifierType (M, NR)
    - linkingAgentIdentifierValue (M, NR)
    - linkingAgentRole (M, NR)
- rightsExtension (O, R)