



Implementing the Scholarly Works Application Profile in DSpace

A metadata collision analysis for the MIT Open Access Initiative

Robert Wolfe, MIT Libraries

27 August 2009

Draft Version 2.0

●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 developed 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.

•*Important Considerations*

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.

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.

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.

● **Metadata Recommendations**

● *Required Elements*

● <i>Element Name</i>	<i>SWAP Attribute</i>	<i>DSpace Metadata Field</i>
● 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 Institution	dc.contributor.department
● Peer-reviewed Flag	Status	eprint.status
● Entity Type	Entity Type	dc.type.uri

● *Recommended Elements*

● <i>Element Name</i>	<i>SWAP Attribute</i>	<i>DSpace Metadata Field</i>
● Publication Date (published version)	n/a	dc.date.issued
● Date Submitted 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 Scholarly Work)	Identifier	dc.relation.isversionof
● Version	Version Number or String	eprint.version

•*Optional Elements*

•Element Name	SWAP Attribute	DSpace Metadata Field
•Identifier – Others	Identifier	dc.identifier.other dc.identifier.none dc.identifier.govdoc dc.identifier.isbn dc.identifier.issn dc.identifier.sici dc.identifier.ismn
Subjects/Keywords	Subject	dc.subject.none dc.subject.classification dc.subject.ddc dc.subject.lcc dc.subject.lcsh dc.subject.mesh 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

•There are more elements that have yet to be fully defined that will appear on the Required, Recommended, and Optional lists. You can find more information about them in the “Elements that Pose Interesting Problems” section of the Collision Analysis.

•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

- Creator
- Title
- Date Available
- Abstract
- Publisher
- Language
- Identifier [of DSpace Item]
- Identifier [of Scholarly Work]
- Identifier [other Identifiers]

DSpace Metadata Field

- dc.contributor.author
- dc.title.none
- dc.date.available
- dc.description.abstract
- dc.publisher.none
- dc.language.iso
- dc.identifier.uri
- dc.relation.isversionof
- dc.identifier.other
- dc.identifier.none
- dc.identifier.govdoc
- dc.identifier.isbn
- dc.identifier.issn
- dc.identifier.sici
- dc.identifier.ismn
- dc.identifier.citation
- dc.subject.none
- dc.subject.classification
- dc.subject.ddc
- dc.subject.lcc
- dc.subject.lcsh
- dc.subject.mesh
- dc.subject.other
- dc.contributor.department
- dc.description.none
- dc.relation.hasversion
- dc.contributor.editor
- dc.contributor.advisor
- dc.description.sponsorship
- dc.title.alternative

- Bibliographic Citation
- Subject

- Affiliated Institution
- Description
- Has Version
- Editor
- Supervisor
- Funder
- Title [of Published Version]

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*

1. Elements

• <i>Element Name</i>	<i>DSpace Metadata Field</i>
•Publication Date (of published version)	dc.date.issued
•Date Submitted for Publication	dc.date.submitted

1. 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 DSpace DC, but are Declared in SWAP*

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

Element

• <i>SWAP Attribute</i>	<i>DSpace Metadata Field</i>
•Status	eprint.status

This element records whether or not an article has been peer-reviewed. It should be added to the tables. Unfortunately, there is no good dublin core element to map it to. It is the first candidate element to be added to the DSpace tables from a namespace other than the DSpace DC namespace.

Recommendation For This Element

Adopt this element for use in describing Scholarly Works 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

Element

• SWAP Attribute	<i>DSpace Metadata Field</i>
•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 Works 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.

Element

• SWAP Attribute	<i>DSpace Metadata Field</i>
•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 Works deposited as part of the MIT Open Access mandate. Enforce its cardinality as defined in the metadata recommendations section above. 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.

Element

•*SWAP Attribute*

•Version Number or String

DSpace Metadata Field

eprint.version

This element is not intended to define a relationship between the DSpace Item (which is a SWAP:Expression) and some other entity such as a SWAP:Manifestation or SWAP:ScholarlyWork. Therefore, the DSpace Metadata Fields dc.relation.isversionof and dc.relation.hasversion do not apply, as they record relationships between the DSpace Item and other entities.. This SWAP attribute is defined as a number or string that names the particular version of a SWAP:ScholarlyWork that the DSpace Item (SWAP:Expression) represents. There is mixed practice in capturing this information in DSpace DC. The following elements have all been used, none of them are optimal:

- dc.identifier.other – I believe that this is current best practice in [DSpace@MIT](#)
- dc.relation.ispartofseries
- dc.relation.isversionof
- dc.relation.hasversion

Recommendation For This Element

Adopt this element for use in describing Scholarly Works 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 enforcing conformity in DSpace by migrating all version metadata that is currently declared in other fields to the eprints.version field. The fields used to capture this information in the past 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 these metadata values into their own element.

Define Vocabulary for this element. (e.g. preprint, posprint, published) – Call Robin Wendler

Element

• <i>SWAP Attribute</i>	<i>DSpace Metadata Field</i>
•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 Works deposited as part of the MIT Open Access mandate. Add this element to the DSpace DC 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.

Type Elements

• <i>SWAP Attribute</i>	<i>DSpace Metadata Field</i>
•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

DRAFT DRAFT DRAFT DRAFT DRAFT DRAFT DRAFT DRAFT DRAFT DRAFT DRAFT DRAFT

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 Works 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.

More Tricky Elements

•*SWAP Attribute*
•[Publication Series] / Journal

DSpace Metadata Field
dc.relation.ispartofseries
dc.identifier.citation
dc.identifier.issn
Some new element in a new namespace

REVIEW AND ADD NEW ELEMENT TO HOLD THE NAME OF THE JOURNAL [ASK Harvard what goes into dc.relation.journal field] Prepare a table of the options discussed for this element. / Recommended

•[Provenance] dc.description.provenance

POSSIBLE USE AS A CATCH-ALL (Submitting Author, DSpace History Events) / Optional

•[Depositing Author / Approving Author]

Ask Ellen about necessity of having this metadata in DSpace Item record / Required?

•[MIT Authors?] 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.mit), or creating a new element in an entirely new namespace (perhaps one for custom fields that Libraries have added to the DSpace metadata tables.

What about a separate field that links an MIT author with their affiliation?

Ideally we would have a field (datatype URI) that points to records for MIT affiliated authors (perhaps use SWAP:Agent), the author record affiliates the author with a MIT organization. / Required

•[Supplemental materials description] Bitstream Description / Optional

•[Origin of metadata used?] ? / Optional

•[Origin of paper?] ?

[What about dc.source.none and dc.source.uri for this and the preceding element – Get export from this field] If the source fields don't work, create new fields in new MIT Libraries namespace / Optional

Rights Elements

- SWAP Attribute*
- Copyright Holder
- Access Rights
- License
- License

DSpace Metadata Field

dc.rights.none
dc.rights.none
dc.rights.none
dc.rights.uri

- Deposit License Info*
- Permission to deposit license
- DSpace deposit license

DSpace Metadata Field

no element
no element

•Elements to Add to the DSpace Metadata Tables

In the existing DSpace DC namespace:

dc.type.uri
dc.relation.references

In a new Eprints namespace:

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

There will likely be more new elements as the last tricky elements are fully defined.

•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