



Implementing the Scholarly Works Application Profile in DSpace

A metadata collision analysis for the MIT Open Access Initiative

Robert Wolfe, MIT Libraries

13 August 2009

Draft Version 1.0

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 apply a metadata application profiles from the open access scholarly publications domain. A number of best practices and application profiles 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, as well as some elements that should not be used and some that have not been categorized
- A mapping from DSpace DC elements to SWAP attributes.
- A collision analysis of the mapping from DSpace DC 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>SWAP Attribute</i>	<i>DSpaceDC Element</i>
• Creator	dc.contributor.author
• Title	dc.title.none
• Date Available	dc.date.issued
• Type	dc.type.none
• Language	dc.language.iso
• Identifier	dc.identifier.uri
• Bibliographic Citation	dc.identifier.citation
• Affiliated Institution	dc.contributor.department
• Status	eprints.status
• Format	dc.format.mimetype
• Entity Type	dc.type.none

Recommended Elements

<i>SWAP Attribute</i>	<i>DSpaceDC Element</i>
• [Date submitted for Publication] This element is not in the SWAP	dc.date.submitted
• Abstract	dc.description.abstract
• Publisher	dc.publisher.none
• Identifier (of Scholarly Work)	dc.relation.isversionof
• Version Number or String	eprints.version
• [Embargo period] This element is not in the SWAP	dc.date.available
• [Publication Series] This element is not in the SWAP	dc.relation.ispartofseries
• Supervisor	dc.contributor.advisor
• Funder	dc.description.sponsorship
• Grant Number	eprints.grantNumber
• Title (of Scholarly Work)	dc.title.alternative
• References dc.relation.references is currently not in the DSpaceDC metadata tables, it should be added to the tables	[dc.relation.references]

Optional Elements

- | | |
|--|--------------------------------|
| • SWAP Attribute | DSpaceDC Element |
| • [Identifier – Others]
This element is not in the SWAP. | dc.identifier.other + 6 others |
| • Subject | dc.subject.none + 6 others |
| • [Date added to DSpace]
This element is not in the SWAP. | dc.date.accessioned |
| • Description | dc.description.none |
| • [Provenance]
This element is not in the SWAP. | dc.description.provenance |
| • Has Version | dc.relation.hasversion |
| • Editor | dc.contributor.editor |
| • Has Translation | eprints.hasTranslation |
| • Is Manifested As | eprints.isManifestedAs |

Elements That Should Not Be Used

- | | |
|---|--------------------------------|
| • <i>SWAP Attribute</i> | <i>DSpaceDC Element</i> |
| • [Bitstream size]
This element is not in the SWAP | dc.format.extent |

Elements with an Unassigned Metadata Cardinality Value

- | | |
|---|--------------------------------|
| • <i>SWAP Attribute</i> | <i>DSpaceDC Element</i> |
| • [Journal]
This element is not in the SWAP or the DSpace DC. | ? |
| • [Depositing Author]
This element is not in the SWAP or the DSpace DC. | ? |
| • [MIT Authors?]
This element is not in the SWAP or the DSpace DC. | ? |
| • [Supplemental materials description]
This element is not in the SWAP or the DSpace DC. | ? |
| • [Origin of metadata used?]
This element is not in the SWAP or the DSpace DC. | ? |
| • [Origin of paper?]
This element is not in the SWAP or the DSpace DC. | ? |

•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.

• SWAP Attribute	DSpaceDC Element
• Creator	dc.contributor.author
• Title	dc.title.none
• Date Available	dc.date.issued
• Abstract	dc.description.abstract
• Publisher	dc.publisher.none
• Type	dc.type.none
• Language	dc.language.iso
• Identifier (of Expression)	dc.identifier.uri
• Identifier (of Scholarly Work)	dc.relation.isversionof
• Bibliographic Citation	dc.identifier.citation
• Subject	dc.subject.none + 6 more elements
• Affiliated Institution	dc.contributor.department
• Description	dc.description.none
• Has Version	dc.relation.hasversion
• Format	dc.format.mimetype
• Editor	dc.contributor.editor
• Supervisor	dc.contributor.advisor
• Funder	dc.description.sponsorship
• Title (of Scholarly Work)	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. Enforce their cardinality as defined in the metadata recommendations section above. 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

- | <i>Name</i> | <i>DSpaceDC Element</i> |
|--------------------------|------------------------------|
| • [Date Submitted] | dc.date.submitted |
| • [Identifier – Others] | dc.identifier.other + 6 more |
| • [Date added to DSpace] | dc.date.accessioned |
| • [Provenance] | dc.description.provenance |
| • [Embargo period] | dc.date.available |
| • [Publication Series] | dc.relation.ispartofseries |
| • [Bitstream size] | dc.format.extent |

DSpace@MIT has discontinued the recording of bitstream file sizes in Item metadata. Incorporating the SWAP should comply with this practice and the field should not be used.

Recommendation For These Elements

Adopt these elements for use in describing Scholarly Works deposited as part of the MIT Open Access mandate. Enforce their cardinality as defined in the metadata recommendations section above. 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

- *SWAP Attribute* *DSpaceDC Element*

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

Element

- Status `eprints.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. Enforce its cardinality as defined in the metadata recommendations section above. Add this element to the DSpace DC metadata tables. Put it in the eprints namespace (See http://www.ukoln.ac.uk/repositories/digirep/index/Scholarly_Works_Application_Profile#Status) instead of the DSpace DC namespace.

Element

- Has Translation `eprints.hasTranslation`
This SWAP attribute is self-explanatory. 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. Enforce its cardinality as defined in the metadata recommendations section above. Add this element to the DSpace DC metadata tables. Put it in the eprints namespace (See http://www.ukoln.ac.uk/repositories/digirep/index/Scholarly_Works_Application_Profile#Has_Translation) instead of the DSpace DC namespace.

Element

- 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.

Element

- Is Manifested As eprints.isManifestedAs
This element is how SWAP Expressions reference their Manifestations. As we are defining DSpace Items as SWAP Expressions it makes sense to enable that Item to reference those particular bitstreams it contains that are genuine manifestations of Expression represented by the Item.

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 metadata tables. Put it in the eprints namespace (See http://www.ukoln.ac.uk/repositories/digirep/index/Scholarly_Works_Application_Profile#Is_Manifested_As) instead of the DSpace DC namespace. This element is similar to the DSpace Primary Bitstream concept and may be useful in identifying which Bitstreams in DSpace Item represent unique versions of the intellectual content, for instance which PDF is the MIT-community-available Thesis and which is the publicly-available version.

Elements that Pose Interesting Problems

- *SWAP Attribute* *DSpaceDC Element*

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

Element

- Version Number or String eprints.version

This element records not a relationship between the DSpace Item (which is a SWAP:Expression) and some SWAP:Manifestation or SWAP:ScholarlyWork. The DSpace DC elements dc.relation.isversionof and dc.relation.hasversion do not apply. 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. Enforce its cardinality as defined in the metadata recommendations section above. Add this element to the DSpace DC metadata tables. Put it in the eprints 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.

Element

- Grant Number eprints.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. Enforce its cardinality as defined in the metadata recommendations section above. Add this element to the DSpace DC metadata tables. Put it in the eprints 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 eprints.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.

Element

- Entity Type dc.type.none

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

1) Entity Type (Values conform to the [Eprints Entity Type Vocabulary Encoding Scheme](#) – e.g, Scholarly Work, Expression, Manifestation, Copy, Agent)

2) Type (Values conform to the [Eprints Type VES](#) – e.g., Conference Paper, Submitted Journal Article)

***DSpace has it's own Type Vocabulary. This means that the dc.type.none element in DSpace may be repeated for an item, each repetition containing a value from a different encoding scheme. This is a legal use of Dublin Core. It is important that we are be able to deduce which vocabulary a particular value for this element represents. The ideal solution would be to use the URIs established by each VES.

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. Use the DSpace DC element dc.type.none for all three kinds of types (SWAP:entityType, SWAP:type, DspaceDC:type) Constrain the values of the dc.type.none element to one of the three encodings schemes. If it is desired that the values of this element and SWAP:type be URIs, then a new element should be added to the DSpace DC metadata tables in the DSpace DC namespace. The element should be named “dc.type.uri”. This will keep String and URI values separate in the database which will ensure better user interface performance.

Rights Elements

- | | |
|---|-------------------------|
| <ul style="list-style-type: none">• <i>SWAP Attribute</i> | <i>DSpaceDC Element</i> |
| • Copyright Holder | dc.rights.none |
| • Access Rights | dc.rights.none |
| • License | dc.rights.none |
| • License | dc.rights.uri |
|
 | |
| <ul style="list-style-type: none">• <i>Deposit License Info</i> | <i>DspaceDC Element</i> |
| • Permission to deposit license | no element |
| • DSpace deposit license | no element |

Recommendation for These Elements

Let's discuss these elements and their necessity.

Elements that are not declared in either SWAP or DSpace DC

- *Element Name*
- Journal
- MIT Authors?
- Supplemental materials description?
- Origin of metadata used?
- Origin of paper?

Recommendation

Let's discuss these elements and their necessity.

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