

February 2007

# Institutional Repositories Workshop Strand Report

## Strand title: research paper metadata

Author: Neil Jacobs (in collaboration with the participants)

### 1. Executive Summary

The Knowledge Exchange organised a workshop in January 2007 on the interoperability of institutional repositories, at which the exchange of research paper metadata was one of six thematic strands. The group covering this topic included representatives from Denmark, Germany, the Netherlands and the UK. Current practice was reviewed, and found unsatisfactory, although it is difficult to be precise about exactly where the shortcomings are without a clearer idea of the services that are intended to use the metadata. The group also concluded that the other thematic strands at the workshop were likely to recommend work that would be relevant to research paper metadata. Therefore, the group recommended further work that makes up an iterative programme, that should be considered in conjunction with those from other strands, and which should both clarify the aims and drivers toward greater interoperability, and achieve some tangible steps along the way.

### 2. Summary of recommendations

The group recommends that:

1. a small piece of work be done to build a framework within which the business drivers can be identified for improved sharing of richer<sup>1</sup> descriptions of research papers.
2. a range of scenarios be agreed and documented, which would be made possible or easier or more effective by the sharing of richer descriptions of research papers.
3. A piece of work should be scoped and commissioned that:
  - a. compares the metadata formats in common use within KE countries with the scenarios previously collected, to identify the strengths and weaknesses of each with respect to those scenarios and the services they might imply.
  - b. based on this analysis, makes recommendations on how metadata exchange and interoperability may be improved; the recommendations should refer to both "quick wins" and steps toward longer term goals.
  - c. as well as looking at metadata structures, also looks for opportunities with respect to:
    - i. time-stamped factual authority relations
    - ii. dependable (resolvable, persistent) identifiers for entities described
    - iii. cataloguing rules
    - iv. shared vocabularies for common elements such as "resource type" and the maintenance of such vocabularies
    - v. shared understanding on the use of xml wrappers or containers

This work should be based on the more detailed discussion in the main text of this report, and should not be undertaken in isolation from similar work following from other strands of the KE workshop.
4. an interoperability demonstrator / testbed should be established, in collaboration where appropriate within existing initiatives, to provide a realistic environment in which to monitor and evaluate the problems and progress of the area.

### 3. Discussion (including recommendations and items of interest)

#### 3.1 Introduction

Within repositories, research papers are described in a variety of ways. These differ not only in how they structure the description (eg, granularity of metadata) but also in what they assume about why research papers are being described.

---

<sup>1</sup> By "richer" we mean fuller as well as more precisely structured descriptions that may be parsed and interpreted with precision by computer programmes

Most of them are far better (more detailed, better structured) than the simple format in which these descriptions are generally shared (Simple Dublin Core). Simply sharing these better descriptions as they stand would not help much, because they are so different to each other. Nevertheless, there is sufficient common ground between them to be optimistic that – with some effort – improvements can be made in the extent to which descriptions of research papers can be shared.

### 3.2 What is the problem?

Repositories are to manage content and to share it. It is conventional to use the Open Archives Protocol for Metadata Harvesting (OAI-PMH) to share descriptions of content, and then point any interested user back to the source repository for the content itself. Therefore, it is important that the descriptions that are shared are good enough that, when brought together, they enable useful services to be created. Many people believe that much better services would be possible if only the descriptions of research papers that are shared were richer and more consistent than Simple Dublin Core allows. The challenge is, then, to find a way to use the richness of the descriptions held within repositories to improve those that are shared, without the variety of the descriptions held within repositories preventing this sharing from happening at all. Improving the descriptions held within repositories, and their consistency between repositories, will contribute to both more potential for sharing descriptions, and an increased common understanding of the challenge.

### 3.3 The current situation

Because this is mandated by the OAI-PMH standard all repositories support the Simple Dublin Core metadata format for harvesting by service providers. In many cases, this is a "dumb down" of the richer internal format used by the repository to Simple Dublin Core. Most of the services offered on the basis of the simple Dublin Core metadata that is shared are cross-search services, that is, they offer users the ability to search across a number of repositories. While valuable, the Simple Dublin Core metadata does not offer either the detail or the structure to allow more sophisticated services that would offer more benefits to users.

Within the Knowledge Exchange there are at least four proposals for exchanging metadata of research papers in richer formats than Simple Dublin Core. These are XMETADISS (Germany), DDF-MXD (Denmark), MODS (Netherlands) and the Scholarly Works DC Application Profile (SWAP) in the UK. In Germany, the widespread OPUS system uses Simple Dublin Core, but with particular extensions. In addition, subject based approaches such as that of the Nereus Consortium (in the field of economics) use other formats (qualified Dublin Core in the case of Nereus). Summaries of these formats can be found online and so are not given here<sup>2</sup>. Some of these rich formats (or implementations of them) devolve particular parts of the descriptive work to other standards. For example, both SWAP and Nereus qDC use the OpenURL ContextObject to capture bibliographic information about a research paper. The Dutch implementation of MODS will devolve certain link and timestamp information to the MPEG 21 DIDL wrapper that is used to contain both the description and content files<sup>3</sup>.

Within the Knowledge Exchange, there is no agreement on how to share these richer descriptions of research papers.

### 3.4 What do we want to achieve?

The range of useful services that may be created using rich descriptions of research paper is potentially huge. It is therefore necessary (i) to know why this effort is worthwhile, and (ii) to focus on a selection of potential services in order to begin to see the scope and scale of the challenge.

Regarding (i), the group **recommends** that a small piece of work be done to build a framework within which the business drivers can be identified for improved sharing of richer descriptions of research papers.

Regarding (ii), the group **recommends** that a range of scenarios be agreed and documented, which would be made possible or easier or more effective by the sharing of richer descriptions of research papers.

The following three scenarios summarise some of the functionality that we thought would be useful<sup>4</sup>, and on which we based our discussion:

---

<sup>2</sup> Metadata formats proposed as exchange formats in KE countries include:

XMETADISS (used in Germany): <http://www.d-nb.de/standards/xmetadiss/xmetadiss.htm>

DDF-MXD (used in Denmark): <http://mx.forskningssdatabasen.dk/mxd/1.0.0/>

MODS (proposed in the Netherlands): <http://www.loc.gov/standards/mods/>

As proposed in the Netherlands, MODS is combined with an MPEG 21 DIDL xml wrapper:

<http://xml.coverpages.org/mpeg21-didl.html>

Scholarly Works Application Profile (SWAP, proposed in the UK):

[http://www.ukoln.ac.uk/repositories/digirep/index/Eprints\\_Application\\_Profile](http://www.ukoln.ac.uk/repositories/digirep/index/Eprints_Application_Profile)

The Nereus qDC xml schema is shown at: <http://drcwww.uvt.nl/~place/nereus/nereusxsd>

<sup>3</sup> The use of such wrappers (MPEG 21 DIDL, METS, etc) is, of course, yet another source of variety in the way descriptions of research papers are structured and expressed. This strand did not address this issue directly, though it could usefully be addressed.

<sup>4</sup> In developing the Scholarly Works Application Profile (SWAP), a large number of scenarios were captured and documented at: [http://www.ukoln.ac.uk/repositories/digirep/index/Functional\\_Requirements#Scenarios\\_and\\_Use\\_case](http://www.ukoln.ac.uk/repositories/digirep/index/Functional_Requirements#Scenarios_and_Use_case)

1. Create a publication list for an author, where her papers are in several repositories across KE countries. (user: an author)
2. Find and download a copy for which full text is available open access, where potentially relevant copies are available in several repositories across KE countries. (user: a reader)
3. Provide information on publications (held in several repositories across KE countries) produced from a particular research project. (user: a research manager)

The group also **recommends** that an interoperability demonstrator / testbed is established, as done by the JISC-SURF-DiVA collaboration on e-theses, to provide a realistic environment in which to investigate very practically what interoperability issues exist. If possible and appropriate, this should be done in collaboration with existing initiatives such as DRIVER, and in partnership with national work such as DARE/SurfShare and UK Intute Repository Search.

### 3.5 What would be necessary to achieve this?

Until there is an agreed range of scenarios to define the interoperability challenge, then it is difficult to define what is necessary to address this challenge. On the other hand, while there is limited interoperability, it is hard to imagine the potential benefits (eg, new services) that might be possible. We therefore need to adopt an iterative approach, whereby improvements in interoperability run in parallel with ongoing review of the requirements. As a first step, on the basis of the expertise available to the group and a brief exploration of the three scenarios given above, it is possible to outline some work that would improve the situation.

We assume (on the basis of a small experiment, see Annex 2) that – with some small changes – any of the main metadata formats in use in repositories in Knowledge Exchange countries could offer a complete and sharable description of research papers that is sufficiently detailed and structured to support a range of services that may want to refer to them. However, further work could usefully be done, and the group **recommends** that this work be scoped and commissioned.

#### ***The group defined this piece of work further as follows:***

The work should compare the metadata formats noted in 3.3<sup>5</sup> with the scenarios previously collected, to identify the strengths and weaknesses, semantic and structural commonalities and divergences of each with respect to those scenarios and the services they might imply<sup>6</sup>. Based on this analysis, recommendations should be developed on how metadata exchange and interoperability may be improved. These recommendations may specify how existing formats and practices may be improved and aligned to achieve a higher level of interoperability and they may also suggest a new common exchange format, if that is considered practical. The recommendations should refer to both “quick wins” and steps toward longer term goals.

On the basis of briefly working through the scenarios noted above we identified the following issues that would be in scope for the piece of work described above:

- (i) time-stamped factual authority relations (for example “this paper was written March 2005, *at which time the author was a researcher at the Max Plank Institute*”);
- (ii) dependable (resolvable, persistent) identifiers for entities described (people, organisations, research papers, etc);
- (iii) cataloguing rules, that is, accepted ways in which particular pieces of information are entered into metadata fields where the format / schema does not (for whatever reason) specify separate fields.
- (iv) shared vocabularies for common elements such as “resource type” and the maintenance of such vocabularies.
- (v) shared understanding on the use of xml wrappers or containers, such as MPEG 21 DIDL, METS, FOXML or abstract models such as the DCMI Abstract Model.

There is work to be done in all these areas:

(i) Other strands at the workshop (especially “author identification” and “exchanging research information”) might between them address (i).

- We noted that the CERIF standard for research management information (people, projects, organisations, publications) offers long term potential for expressing the kind of relational information implied by (i), and so might usefully be included – if only as a point of reference at this time – in the comparison of metadata formats noted above.
- We also agreed that it would be valuable to establish author identifiers, either a single accepted name space or, at least, a way of resolving a particular identifier to a particular namespace.

<sup>5</sup> CERIF might also usefully be included – if only as a point of reference at this time – in this comparison.

<sup>6</sup> In the DFG-Project KIM (Competence Centre for Interoperable Metadata) the KIM Working Group will analyse the XMETADISS-Format (Format of the metadata set of the German National Library for online dissertations and post-doctoral theses) with regard to the level of interoperability and in compliance with the Dublin Core Abstract Model. This work could provide one basis for the comparison of the metadata formats in the four KE countries.

- We also agreed that an authoritative list of research funders (and other organisational entities) could support scenario 3.4.3 above, for example as a prefix to project codes to ensure they are unique, and work to explore requirements in this area would be in scope.
- (ii) We did not discuss identifiers further, other than to note that unique and persistent identifiers for research projects would be necessary to support. Scenario 3.4.3 above.
- (iii) We did not discuss cataloguing rules further.
- (iv) We suggested that agreement on the following vocabularies would be a practical and real step forward in interoperability:
- Publication (resource) type<sup>7</sup>
  - Version<sup>8</sup>;
  - Review status (eg peer reviewed)<sup>9</sup>
  - Person role (eg author, editor, translator)
  - Access rights<sup>10</sup>
- (v) We did not discuss xml wrappers further, other than to note that it would be helpful to agree on when search/discovery services best identified the wrapper and when the individual file. This is similar to the question of whether (in FRBR<sup>11</sup> / SWAP terms) the record exposed relates to the 'work' or a 'copy / manifestation' of the work. An exploration of the implications of these choices would be in scope.

### 3.6 Related activities

#### Exchanging Research Information

There is considerable overlap in the data held by CERIF-based research information systems and research repositories. However, the two communities have very different traditions and business drivers. Work will be required to develop a shared understanding of how these two systems can best work together.

#### ETheses

Etheses are very like research papers in terms of their metadata. The few differences can easily be accommodated by supplements to most schemas, and could be best identified by analysing the few existing thesis-specific schemas.

#### Author identification

We agreed that, in the long term, it was important to work toward the expression of time-stamped factual authority relations more generally, and to identify/build authoritative sources to populate this information.

## 4. Outcomes

1. Any of the main metadata formats in use in repositories in Knowledge Exchange countries could – with small changes – offer a description of all types of research paper that is sufficiently detailed and structured to support fully a range of services that may want to refer to them.

---

<sup>7</sup> There are a number of resource type vocabularies, including:

- OpenURL genres (probably insufficient for purpose)
- Both the DC Type and MODS 'typeOfResource' vocabularies are limited to 'text', covering papers, articles, theses, etc.
- The DC-based Scholarly Works Application Profile (SWAP) has developed a type vocabulary at: [http://www.ukoln.ac.uk/repositories/digirep/index/Eprints\\_Type\\_Vocabulary\\_Encoding\\_Scheme](http://www.ukoln.ac.uk/repositories/digirep/index/Eprints_Type_Vocabulary_Encoding_Scheme)
- PRISM are in 2007 developing a "vocabulary specifically designed for web-based article content": <http://www.prismstandard.org/>
- Commercial companies have also developed vocabularies, for example, Thomson Scientific's Document types vocabulary is extensive and highly relevant. It is unclear whether the same vocabulary is used in their 'Web Citation Index' product.

<sup>8</sup> There is also a lot of work in the 'version identification' area, including:

- JISC-sponsored work in the UK, such as
  - the RIVER study: [http://www.jisc.ac.uk/uploaded\\_documents/RIVER%20Final%20Report.pdf](http://www.jisc.ac.uk/uploaded_documents/RIVER%20Final%20Report.pdf)
  - the Versions project: <http://www.lse.ac.uk/library/versions/> - which is likely to have follow-up work funded from 2007
- A NISO working group, sponsored by the ALPSP: [http://www.niso.org/committees/Journal\\_versioning/JournalVer\\_comm.html](http://www.niso.org/committees/Journal_versioning/JournalVer_comm.html)

<sup>9</sup> The SWAP has defined a two-element vocabulary:

[http://www.ukoln.ac.uk/repositories/digirep/index/Eprints\\_Status\\_Vocabulary\\_Encoding\\_Scheme](http://www.ukoln.ac.uk/repositories/digirep/index/Eprints_Status_Vocabulary_Encoding_Scheme)

Other similar vocabularies exist. It is likely that there will be some relation with a version vocabulary.

<sup>10</sup> Again, the SWAP has defined a simple access rights vocabulary:

[http://www.ukoln.ac.uk/repositories/digirep/index/Eprints\\_AccessRights\\_Vocabulary\\_Encoding\\_Scheme](http://www.ukoln.ac.uk/repositories/digirep/index/Eprints_AccessRights_Vocabulary_Encoding_Scheme)

However, such record-level descriptions may need to work closely with automatic access and rights management/enforcement arrangements.

<sup>11</sup> Functional Requirements for Bibliographic Records: <http://www.ifla.org/VII/s13/frbr/frbr.htm>

2. Further work could usefully be done to identify steps toward greater interoperability (in terms of metadata structures) within the KE countries, and the group began to scope this work.
3. In addition to work toward better interoperability in terms of metadata structures, further work is also needed on factual authority, identifiers, vocabularies, cataloguing rules and xml wrappers.
4. Much of this work overlaps with that from other strands at the Knowledge Exchange IR workshop, and should not be undertaken in isolation.

Recommendations are summarised in section 2.

## 5. Annexes

### ANNEX 1. List of participants

lead coordinator	Neil Jacobs	JISC
moderator 1	Mogens Sandfær	Technical Knowledge Center of Denmark
moderator 2	Julie Allinson	UKOLN , University of Bath
Knowledge Exchange/JISC	Louisa Dale	JISC
Participants DFG		
	Ulla Tschida	Max Planck Digital Library
	Justine Haerberli	State & University Library Goettingen (SUB)
Participants JISC		
	Linda Kerr	Intute, Herriot Watt University
	Ross MacIntyre	MIMAS@ University of Manchester
Participants SURF		
	Thomas Place	University of Tilburg
	Jos Odekerken	University of Maastricht
Participants DEFF		
	Adrian Price	Danish Veterinary and Agricultural Library
	Liv Fugl	Institute for Product Development IPU

## **ANNEX 2: An example of how to approach encoding a research paper record according to SWAP, MODS and DDF-MXD**

**The original record to be encoded**

**Classical research paper, peer-reviewed, presented at a conference, published in a paper proceedings volume and on the net.**

Global warming and the new ice age. Hot news and cold facts.

Proceedings of the 3<sup>rd</sup> World Meteorological Union Conference

Tromsø, Norway, July 3-7, 2006

Edited by Sophie Ellis Baxter (Head of Institute of Precipitation, London School of Weather, seb@lsw.ac.uk) and Herbert Müller Thürgau (President of WMU, herbie@wmu.org)

WMU Press, Berlin, January 2007

<http://wmu.org/conferences/wmuc3/proceedings/>

Volume 2

Cream of the crop - Full scientific papers, fully peer reviewed

Special editor: Sir Spencer F. Chapmann-Jones (President of Sweet Water Corp., Jersey, drinkme@hotmail.com)

546 pages

<http://wmu.org/conferences/wmuc3/proceedings/vol2/>

Page 234-245

Crying Wolf ??? - Recently Observed Changes in Temperature and Salinity within Antarctic Intermediate Water. Findings of the WMU AAIW project Phase 2.

Schneider, Wolfgang, Univ. de Concepcion, Chile, (wos@gmail.com)

Masao Fukasawa and Takeshi Kawano, Japanese Marine

Science and Technology Center (JAMSTEC), Japan (mf, tk @jamstec.jp)

Lian Yichin, Weather Center of WISTIC, P. R. China (lianyic@wistic.cn)

DOI: 10.1000/182

<http://wmu.org/conferences/wmuc3/proceedings/vol2/schneider1/>

## An approach to encoding the record according to SWAP

Please note that for human readability this example is not encoded in the eprints dc xml format, or in any XML format. Tags (e.g. <type>) correspond to the metadata element names used in the application profile. For further information please consult the full documentation at

[http://www.ukoln.ac.uk/repositories/digirep/index/Eprints\\_Application\\_Profile](http://www.ukoln.ac.uk/repositories/digirep/index/Eprints_Application_Profile)

```
<descriptionSet>
  <description> <!-- Description of a Scholarly Work -->
    Resource URI: http://dx.doi.org/10.100/182
    <type>ScholarlyWork</type> <!-- Value taken from a vocabulary encoding scheme
-->
    <title>Crying Wolf ??? - recently observed changes in temperature and salinity
within antarctic intermediate water : findings of the WMU AAIW project phase
2</title>
    <identifier>10.1000/182</identifier>
    <creator>Schneider, Wolfgang</creator>
    DescriptionRef: Agent1
<!-- For each Agent, a value string can be used to represent the static name as
it appears on the item, with an accompanying Agent description which may record a
dynamic name (i.e. this may change over time) -->
    <creator>Fukasawa, Masao</creator>
    DescriptionRef: Agent2
    <creator>Kawano, Takeshi</creator>
    DescriptionRef: Agent3
    <creator>Yichin, Lian</creator>
    DescriptionRef: Agent4
    <affiliatedInstitution>Univ de Concepcion</affiliatedInstitution>
    DescriptionRef: Agent5
    <affiliatedInstitution>Japanese Marine Science and Technology Center
(JAMSTEC)</affiliatedInstitution>
    DescriptionRef: Agent6
    <affiliatedInstitution>Weather Center of WISTIC</affiliatedInstitution>
    DescriptionRef: Agent7
    <isExpressedAs>Expression1</isExpressedAs>
</description>

<description> <!-- Description of an Expression of the Scholarly Work -->
  DescriptionId: Expression1
  <type>Expression</type> <!-- Value taken from a vocabulary encoding scheme -->
  <dateAvailable>2007-01</dateAvailable>
  <status>PeerReviewed</status> <!-- Value taken from a vocabulary encoding
scheme -->
  <type>ConferencePaper</type> <!-- Value taken from a vocabulary encoding
scheme -->
  <bibliographicCitation>
    Cream of the crop : Full scientific papers, fully peer reviewed in, Volume 2
of Global warming and the new ice age. Hot news and cold facts : Proceedings of
the 3rd World Meteorological Union Conference Tromso Norway, July 3-7 2006. WMU
Press, Berlin, January 2007, pp 234-245
  </bibliographicCitation>
  <bibliographicCitation>
    &ctx_ver=Z39.88-
2004&rft_val_fmt=info:ofi/fmt:kev:mtx:book&rft.genre=proceeding&rft.btitle=Cream+
of+the+crop++Full+scientific+papers,+fully+peer+reviewed.+Volume+2&rft.series=Gl
obal+warming+and+the+new+ice+age.+Hot+news+and+cold+facts++Proceedings+of+the+3r
d+World+Meteorological+Union+Conference+Tromso+Norway,+July+3-
7+2006&rft.spage=234&rft.epage=245&rft.date=2007-
01&rft.place=Berlin&rft.pub=WMU+Press
  </bibliographicCitation>
<!-- Both text citation and OpenURL can be captured -->
  <editor>Chapmann-Jones, Spencer F.</editor>
  DescriptionRef: Agent8
  <editor>Baxter, Sophie Ellis</editor>
  DescriptionRef: Agent9
  <editor>Thurgau, Herbert Muller</editor>
```

```

    DescriptionRef: Agent10
    <isManifestedAs>Manifestation1</isManifestedAs>
</description>

<description> <!-- Description of a Manifestation of an Expression of the
Scholarly Work -->
    DescriptionID: Manifestation1
    <type>Manifestation</type> <!-- Value taken from a vocabulary encoding scheme -->
->
    <publisher>WMU Press</publisher>
    DescriptionRef: Agent11

<isAvailableAs>http://wmu.org/conferences/wmuc3/proceedings/vol2/schneider1</isA
availableAs>
</description>

<description> <!-- Description of a Copy of a Manifestation an Expression of the
Scholarly Work -->
    ResourceURI: http://wmu.org/conferences/wmuc3/proceedings/vol2/schneider1/
    <type>Copy</type> <!-- Value taken from a vocabulary encoding scheme -->
    <dateAvailable>2007-01</dateAvailable>
    <isPartOf>http://wmu.org/conferences/wmuc3/proceedings/</isPartOf>
    <isPartOf>http://wmu.org/conferences/wmuc3/proceedings/vol2/</isPartOf>
</description>

<description> <!-- Description of an Agent the Scholarly Work -->
    descriptionID: Agent1
    <type>Person</type> <!-- Value taken from a vocabulary encoding scheme -->
    <givenName>Wolfgang</givenName>
    <familyName>Schneider</givenName>
    <mailbox>wos@gmail.com</mailbox>
</description>

<description> <!-- Description of an Agent -->
    descriptionID: Agent2
    <type>Person</type> <!-- Value taken from a vocabulary encoding scheme -->
    <givenName>Masao</givenName>
    <familyName>Fukasawa</givenName>
    <mailbox>@jamstec.jp</mailbox>
</description>

<description> <!-- Description of an Agent -->
    descriptionID: Agent3
    <type>Person</type> <!-- Value taken from a vocabulary encoding scheme -->
    <givenName>Takeshi</givenName>
    <familyName>Kawano</givenName>
    <mailbox>tk@jamstec.jp</mailbox>
</description>

<description> <!-- Description of an Agent -->
    descriptionID: Agent4
    <type>Person</type> <!-- Value taken from a vocabulary encoding scheme -->
    <givenName>Lian</givenName>
    <familyName>Yichin</givenName>
    <mailbox>lianyic@wistic.cn</mailbox>
</description>

<description> <!-- Description of an Agent -->
    descriptionID: Agent5
    <type>Organisation</type> <!-- Value taken from a vocabulary encoding scheme -->
->
    <name>Univ de Concepcion</name>
</description>

<description> <!-- Description of an Agent -->
    descriptionID: Agent6
    <type>Organisation</type> <!-- Value taken from a vocabulary encoding scheme -->
->

```

```

    <name>Japanese Marine Science and Technology Center (JAMSTEC)</name>
</description>

<description> <!-- Description of an Agent -->
  descriptionID: Agent7
  <type>Organisation</type> <!-- Value taken from a vocabulary encoding scheme -->
->
  <name>Weather Center of WISTIC</name>
</description>

<description> <!-- Description of an Agent -->
  descriptionID: Agent8
  <type>Person</type> <!-- Value taken from a vocabulary encoding scheme -->
  <givenName>Spencer F.</givenName>
  <familyName>Chapmann-Jones</givenName>
  <mailbox>drinkme@hotmail.com</mailbox>
</description>

<description> <!-- Description of an Agent -->
  descriptionID: Agent9
  <type>Person</type> <!-- Value taken from a vocabulary encoding scheme -->
  <givenName>Sophie Ellis</givenName>
  <familyName>Baxter</givenName>
  <mailbox>seb@lsw.ac.uk</mailbox>
</description>

<description> <!-- Description of an Agent -->
  descriptionID: Agent10
  <type>Person</type> <!-- Value taken from a vocabulary encoding scheme -->
  <givenName>Herbert Muller</givenName>
  <familyName>Thurgau</givenName>
  <mailbox>herbie@wmu.org</mailbox>
</description>

<description> <!-- Description of an Agent -->
  descriptionID: Agent11
  <type>Organisation</type> <!-- Value taken from a vocabulary encoding scheme -->
->
  <name>WMU Press</name>
</description>
</descriptionSet>

```

#### **ASSUMPTIONS:**

- The DOI will de-reference to a 'splash page' rather than the actual full-text; the URL for the full-text is <http://wmu.org/conferences/wmuc3/proceedings/vol2/schneider1/>
- No other assumptions have been made, for example, the document format is not specified, nor is the workplace homepages of the authors so are not included.

#### **ISSUES IDENTIFIED:**

- Dislocation of affiliated institution from author means that it is not possible to tie an author to their institution at the point of creation. The institution identified by the Agent description may change over time.
- Use of space colon space as a separator for sub-titles is problematic as it introduces the potential for human error.
- Use of workplace homepage and organisation homepage are problematic as URIs may change.
- It is difficult to say precisely what the DOI identifies. If additional expressions were added, the DOI is likely to relate to one expression only.

#### **OTHER COMMENTS:**

- This record describes the single paper and NOT the volume or proceedings. Information about the proceedings is captured in the bibliographic citation insofar as it is needed to identify the research paper.
- It would be easy to add additional expressions, manifestations or copies, e.g. if this were a record in a repository pointing to an external copy, a pre-print expression might also exist and would also become part of the description set.

## Encoding the record according to MODS

```
<mods
  xmlns:xlink="http://www.w3.org/1999/xlink"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xmlns="http://www.loc.gov/mods/v3"
  xsi:schemaLocation="http://www.loc.gov/mods/v3
    http://www.loc.gov/standards/mods/v3/mods-3-2.xsd">

  <!-- research paper published in multi volume conference proceedings -->
  <titleInfo>
    <title>Crying Wolf ???</title>
    <subTitle>
      Recently Observed Changes in Temperature and Salinity within Antarctic
      Intermediate Water. Findings of the WMU AAIW project Phase 2
    </subTitle>
  </titleInfo>

  <!-- authors -->
  <name type="personal">
    <displayForm>Wolfgang Schneider</displayForm>
    <namePart type="given">Wolfgang</namePart>
    <namePart type="family">Schneider</namePart>
    <role>
      <roleTerm type="text">author</roleTerm>
      <!--
        authority="marcrelator" refers to the MARC relators list,
        see http://www.loc.gov/marc/sourcecode/relator/relatorsource.html -->
      <roleTerm authority="marcrelator" type="code">aut</roleTerm>
    </role>
    <affiliation>Univ. de Concepcion, Chile (wos@gmail.com)</affiliation>
  </name>
  <name type="personal">
    <displayForm>Masao Fukasawa</displayForm>
    <namePart type="given">Masao</namePart>
    <namePart type="family">Fukasawa</namePart>
    <role>
      <roleTerm type="text">author</roleTerm>
      <roleTerm authority="marcrelator" type="code">aut</roleTerm>
    </role>
    <affiliation>
      Japanes Marine Science and Technology Center (JAMSTEC), Japan
      (mf@jamstec.jp)
    </affiliation>
  </name>
  <name type="personal">
    <displayForm>Takeshi Kawano</displayForm>
    <namePart type="given">Takeshi</namePart>
    <namePart type="family">Kawano</namePart>
    <role>
      <roleTerm type="text">author</roleTerm>
      <roleTerm authority="marcrelator" type="code">aut</roleTerm>
    </role>
    <affiliation>
      Japanes Marine Science and Technology Center (JAMSTEC), Japan
      (tk@jamstec.jp)
    </affiliation>
  </name>
  <name type="personal">
    <displayForm>Masao Fukasawa</displayForm>
    <namePart type="given">Masao</namePart>
    <namePart type="family">Fukasawa</namePart>
    <role>
      <roleTerm type="text">author</roleTerm>
      <roleTerm authority="marcrelator" type="code">aut</roleTerm>
    </role>
    <affiliation>
```

```

    Japanes Marine Science and Technology Center (JAMSTEC), Japan
    (mf@jamstec.jp)
</affiliation>
</name>
<name type="personal">
  <displayForm>Lian Yichin</displayForm>
  <namePart type="given">Lian</namePart>
  <namePart type="family">Yichin</namePart>
  <role>
    <roleTerm type="text">author</roleTerm>
    <roleTerm authority="marcrelator" type="code">aut</roleTerm>
  </role>
  <affiliation>
    Wheater Center of WISTIC, P.R. China (lianyic@wistic.cn)
  </affiliation>
</name>

<!--
  authority="marcgt" refers to the MARC 21 list of genre terms, see
  http://www.loc.gov/marc/sourcecode/genre/genrelist.html -->
<genre authority="marcgt">conference publication</genre>

<!--
  authority points to vocabulary/classification with terms for
  expressing the status of a publication -->
<classification authority="...">fully peer reviewed</classification>

<!--
  There is no controlled list of identifier types.
  "doi" is one of the suggested values -->
<identifier type="doi">10.1000/182</identifier>

<location>
  <url>http://wmu.org/conferences/wmuc3/proceedings/vol2/schneider1</url>
</location>

<!-- volume 2 of multi volume proceedings -->
<!--
  In relatedItem all MODS elementes can be used including relatedItem.
  Possible values of type-attribute of relatedItem are: preceding,
  succeeding, original, host, constituent, series, otherVersion,
  otherFormat, isReferencedBy.
  'otherVersion' or 'otherFormat' could be used to refer to the copies
  in an (institutional) repository. Or alternatively, the MODS record can
  be included in an XML container that describes the object files -->
<relatedItem type="host">
  <titleInfo>
    <title>Cream of the crop</title>
    <subTitle>Full scientific papers, fully peer reviewed</subTitle>
  </titleInfo>

  <!-- editor -->
  <name type="personal">
    <displayForm>Sir Spencer F. Chapmann-Jones</displayForm>
    <namePart type="given">Spencer</namePart>
    <namePart type="family">Chapmann-Jones</namePart>
    <role>
      <roleTerm type="text">Special editor</roleTerm>
      <roleTerm authority="">edt</roleTerm>
    </role>
    <affiliation>
      President of Sweet Water Corp. Jersey, drinkme@hotmail.com
    </affiliation>
  </name>

  <!--
  To describe the volume as completely as possible originInfo is copied
  from the top level of the multi volume proceedings.-->

```

```

<originInfo>
  <place>
    <placeTerm>Berlin</placeTerm>
  </place>
  <publisher>WMU Press</publisher>
  <dateIssued encoding="iso8601">200701</dateIssued>
</originInfo>
<physicalDescription>
  <extent>546 pages</extent>
</physicalDescription>
<location>
  <url>http://wmu.org/conferences/wmuc3/proceedings/vol2/</url>
</location>
<part>
  <extent unit="page">
    <start>234</start>
    <end>245</end>
  </extent>
  <date encoding="iso8601">200701</date>
</part>

<!-- proceedings of the conference -->
<relatedItem type="host">
  <titleInfo>
    <title>Global warming and the new ice age</title>
    <!--
      Not possible to describe the conference as an event -->
    <subTitle>
      Hot news and cold facts. Proceedings of the 3rd World Meteorological Union
      Conference. Tromsø, Norway, July 3-7, 2006
    </subTitle>
  </titleInfo>

  <!-- editors of the proceedings -->
  <name type="personal">
    <displayForm>Sophie Ellis Baxter</displayForm>
    <namePart type="given">Sophie</namePart>
    <namePart type="family">Baxter</namePart>
    <role>
      <roleTerm type="text">Editor</roleTerm>
      <roleTerm authority="">edt</roleTerm>
    </role>
    <affiliation>
      Head of Institute of Precipitation, London School of Weather,
      seb@lsw.ac.uk
    </affiliation>
  </name>
  <name type="personal">
    <displayForm>
      Herbert Müller Thürgau (President of WMU, herbie@wmu.org)
    </displayForm>
    <namePart type="given">Herbert</namePart>
    <namePart type="family">Thürgau</namePart>
    <role>
      <roleTerm type="text">editor</roleTerm>
      <roleTerm authority="">edt</roleTerm>
    </role>
    <affiliation>
      Head of Institute of Precipitation, London School of Weather,
      seb@lsw.ac.uk
    </affiliation>
  </name>

  <originInfo>
    <place>
      <placeTerm>Berlin</placeTerm>
    </place>
    <publisher>WMU Press</publisher>

```

```
<dateIssued encoding="iso8601">200701</dateIssued>
</originInfo>
<location>
  <url>http://wmu.org/conferences/wmuc3/proceedings/</url>
</location>
<part>
  <detail type="volume">
    <number>2</number>
  </detail>
</part>
</relatedItem>
</relatedItem>
</mods>
```

## Encoding the record according to DDF-MXD

```
<ddf_doc format_version="1.1.0" doc_type="dcp" doc_lang="eng" doc_year="2007"
doc_review="pr" doc_level="sci" rec_source="xxx" rec_id="xxx" rec_upd="2007-11-
02" rec_status="n">
```

```
<title>
<original>
<main>Crying Wolf ??? - Recently Observed Changes in Temperature and Salinity
within Antarctic Intermediate Water</main>
<sub>Findings of the WMU AAIW project Phase 2.</sub>
</original>
</title>
```

```
<person pers_role="pau" aff_no="001">
<name>
<first>Wolfgang</first>
<last>Schneider</last>
</name>
<email>wos@gmail.com</email>
</person>
```

```
<person pers_role="pau" aff_no="002">
<name>
<first>Masao</first>
<last>Fukasawa</last>
</name>
<email>mf@jamstec.jp</email>
</person>
```

```
<person pers_role="pau" aff_no="002">
<name>
<first>Takeshi</first>
<last>Kawano</last>
</name>
<email>tk@jamstec.jp</email>
</person>
```

```
<person pers_role="pau" aff_no="003">
<name>
<first>Lian</first>
<last>Yichin</last>
</name>
<email>lianyic@wistic.cn</email>
</person>
```

```
<organisation org_role="oaf" aff_no="001">
<name>
<level1>Univ. de Concepcion</level1>
</name>
<country>cl</country>
</organisation>
```

```
<organisation org_role="oaf" aff_no="002">
<name>
<level1>Japanese Marine Science and Technology Center</level1>
<acronym>JAMSTEC</acronym>
</name>
<country>jp</country>
</organisation>
```

```
<organisation org_role="oaf" aff_no="003">
<name>
<level1>Weather Center of WISTIC</level1>
</name>
<country>cn</country>
</organisation>
```

```

<event event_role="ep">
<title xml:lang="eng">
<full>Global warming and the new ice age. Hot news and cold facts : Proceedings
of the 3rd World Meterological Union Conference.</full>
<number>3</number>
</title>

<dates>
<start>2006-07-03</start>
<end>2006-07-07</end>
</dates>
<place>Tromsø, Norway</place>
<uri>http://wmu.org/conferences/wmuc3/proceedings/</uri>
</event>

<publication>
<in_book pub_status="p">
<title>Proceedings of the 3rd World Metereological Union Conference : Cream of
the Crop</title>
<part>Volume 2</part>
<editor>Sir Spencer F. Chapmann-Jones</editor>
<place>Berlin</place>
<publisher>WMU Press</publisher>
<year>2007</year>
<pages>234-245</pages>
<uri>http://wmu.org/conferences/wmuc3/proceedings/vol2/</uri>
</in_book>

<inetpub>
<uri>http://www.wmu.org/conferences/wmuc3/proceedings/vol2/schneider1/</uri>
</inetpub>
</publication>
</ddf_doc>

```

**COMMENT:**

Areas not currently covered in the example are:

- the series editors, their affiliation and emails
- the proceeding editor's affiliation and e-mail
- the full number of pages of the proceedings
- the DOI of the paper

---

**DFG**

**JISC**

**DEff**

**SURF**