

Appendix A: Publisher Survey results

A questionnaire was sent to 15 publishers/vendors. A total of 12 responses were received, either in writing or in telephone interviews. The organizations who responded are indicated by an asterisk in the list below.

The publishers/vendors included in the survey were: American Chemical Society*; American Institute of Physics; Atypon*; BioMed Central*; EBSCO; Elsevier*; Ingenta*; Institute of Physics Publishing*; Nature Publishing Group*; OUP*; Ovid*; Sage*; Springer*; Taylor & Francis*; Wiley Blackwell. These publishers/vendors are all currently COUNTER-compliant and were selected to ensure that the sample was representative of the industry in terms of scope, size and geographical location

Responses to individual questions

Questions

- 1) Do you think that it will be valuable, in principle, to record and report usage of full-text articles at the individual article level?

The majority of respondents think it would be valuable, in principle, to do this, but most also foresaw problems in practice. Some vendors stated that they already record and report usage at the individual article level, or have the capability to do so. One or two did not think that this level of information would be either useful or necessary for the vast majority of customers.

- a. Do you think that it will be valuable, in principle, for a common standard for measuring such usage to be set for publishers, aggregators, repositories and other entities?

While most could see that such a standard would, in principle, be valuable, several foresaw practical problems with the collation of usage data from different sources and questioned whether a practical mechanism for doing this could be implemented.

Those who answered 'no' to the first question above did not see the value in creating a common standard.

- 2) Unique Article Identifiers:
 - a. Do you attribute the same Unique Article Identifier to a particular full-text journal article, irrespective of format (PDF, html, etc)?

All respondents attribute the same Unique Article Identifier to an article, irrespective of format.

- b. Is this then a permanent attribute of that article?

Yes, in all cases

Formatted: Position:
Horizontal: Right, Relative to:
Margin, Vertical: 0 pt, Relative
to: Paragraph, Wrap Around

3) If so, do you use DOI for this purpose? If not, what is your current practice?

Most respondents use the DOI for this purpose, but not all. In the case of EBSCO, for example, many articles in their full-text database do not have DOIs, so they use an internal Accession Number instead for the article.

4) At what stage in the publishing process do you apply the unique article identifier? Are different article identifiers applied at different stages of the publishing process e.g. pre publication and publication.

There are different scenarios here:

- DOI is applied to the final, accepted author manuscript (most common)
- DOI is applied when content is added to the content management database
- DOIs are assigned and registered at online publication. Internal identifiers are used during the editorial and production process.
- In once case, when an article enters the publisher's production tracking system it is assigned an internal code. This then is the basis for the DOI when the article is published online ahead of print at which point the DOI is registered with CrossRef. When the article subsequently appears in an issue, the article keeps the same DOI, it moves into the issue table of contents, and the associated metadata at CrossRef is updated with page and volume numbers etc.

In all cases a single DOI is used for all versions of a manuscript/article. Several publishers mention that this is required in order to be consistent with the CrossRef guidelines.

5) In which formats do you publish full-text journal articles?

The following formats are used: html, PDF, enhanced PDF, provisional PDF, SGML (not after 2009), XML

6) How do you currently identify the different versions of an article?

There is a diversity of practice here, including:

- appending @just-accepted and @ahead-of-print to the DOI to maintain an internal history. These DOIs are never exposed to the public.
- logging accesses to each version separately.
- creating a format flag
- only the final, published version is made available to users, so no version identifier is needed
- publication of (i) the author manuscript online, followed by (ii) a typeset, edited but unpaginated version. These are then followed by (iii) the final issue-based version. The earlier versions resolve to the most recent version which is the paginated issue version. In common with many HighWire publishers, the 'content box' on the article lists previous versions.
- publication stage is marked in the article metadata (xml A++ DTD)

b. Do you plan to implement the recently published NISO/ALPSP Best Practices for Journal Article Versions document

Formatted: Position:
Horizontal: Right, Relative to:
Margin, Vertical: 0 pt, Relative
to: Paragraph, Wrap Around

(<http://www.niso.org/publications/rp/RP-8-2008.pdf>), which proposes how to identify different versions of articles and how to clarify the relationships between them?

Most were unaware of these new guidelines, but feel that their own applications allow multiple versions and can identify them. Two publishers, however, specifically mentioned that they plan to implement these guidelines.

- 7) Would it be a problem for you if COUNTER specified a new report requiring that the number of successful full-text article requests be reported at the article level, in which the DOI was required as a unique article identifier, to provide a permanent attribute for each full-text article, irrespective of format? (Currently COUNTER only requires that full-text article usage be reported at the journal level, in Journal Report 1)

For the journal publishers this is not a problem in principle, but many are concerned about the extra work involved in creating these additional usage reports. This would be less of a problem if the article-level reports were to be prepared, upon request, at the title level only and did not mimic the existing COUNTER reports, which are prepared monthly, on an institute by institute basis.

Technical challenges include (i) collating the usage of the different versions together and (ii) if a URL structure does not include the DOI and a mapping from URLs recorded in the log files to the associated DOI is required – or the structure of the website is changed.

For full-text aggregators a substantial amount of content is generated from journals and other material types for which there is no online equivalent; in these cases, no DOI has been registered (and it is not reasonable for the aggregator to register the DOI since many aggregators process the same content.) In short, the requirement of DOI by its nature means that such a report will not cover all articles presented through the service.

- 8) Do you have any other comments?

- 'I agree it would be desirable to aggregate such access data from sources such as PMC and IRs to a single location.'
- 'The theory behind these reports is sound; however, very careful consideration needs to be given to the practicality of such reports. The processing power and storage to manipulate millions of transactions per month and retain history for many months also needs to be considered.
A question to consider is if this is if the expectation is for the publisher to maintain 3 years worth of article-level transactions for reporting, or if the desire is for the publisher to offer a feed of article-level transactions for the receiving institution to accumulate and analyze. If the latter, then the requirement could be that the publisher makes the feeds available for 1 to 2 months after the activity has occurred. The latter also assumes the organization receiving will be responsible for providing the technology to store and analyze the data.'
- 'I remain to be convinced of the usefulness of article-level reporting to our customers, as they will be inundated with data. We are a medium-sized publisher and we have 300,000 articles! In order to justify the additional cost of another report I would like to know more about the benefits to customers.'
- 'Breaking out institutional usage into articles will result in huge amounts of data. I'd recommend a thorough cost-benefit analysis including an estimate of implementation and maintenance costs before making this a requirement.'

Formatted: Position:
Horizontal: Right, Relative to:
Margin, Vertical: 0 pt, Relative
to: Paragraph, Wrap Around

|

4

Formatted: Position:
Horizontal: Right, Relative to:
Margin, Vertical: 0 pt, Relative
to: Paragraph, Wrap Around

|

5

Formatted: Position:
Horizontal: Right, Relative to:
Margin, Vertical: 0 pt, Relative
to: Paragraph, Wrap Around

PIRUS – Publisher and Institutional Repository Usage Statistics

a PALS 3 project

Appendix B: Repository Software Applications Table

Software	No. of Research Institutional or Departmental Repositories (from ROAR)		Metadata		Comments
	UK	World	Supported schema	Extendable	
Archimede	-	-	Qualified Dublin Core	No	
Arno	-	3	Dublin Core	Yes	
CDSware	-	6	Standard Marc21	Yes	
Digital Commons (BePress)	4	50	Dublin Core	No	Developments to support qDC are under way
DigiTool	-	1	Dublin Core	Yes	there are additional 'free' fields that can be used to create qualified DC records according to need
DiVA	-	16	Internal	Yes	Sweden. Internal schema can be mapped to a number of standard schemas including marcxml and qDC
DSpace	23	222	Qualified Dublin Core	Yes	qDC out-of-the-box, but, since v1.4, can support other metadata schemes
EDOC	-	1	Internal	Yes	Germany. Metadata based on international standards such as vCard, DC, OpenURL, AMF, LOM, Ariadne, ODRL, OAI, CLD
Eprints	39	163	Internal	Yes	Internal fields generally have a one-to-one mapping to qDC
Fedora	1	5	Any	Yes	
Fez/Fedora	-	3	MODS, Dublin Core	Yes	Has Statistics feature (e.g. Downloads per Author, per Community, per Collection, per Subject etc).
HAL	-	3			France. Contributors are invited to indicate

Software	No. of Research Institutional or Departmental Repositories (from ROAR)		Metadata		Comments
	UK	World	Supported schema	Extendable	
					the relevant bibliographic information and DOI .
i-Tor	-	1	Any	Yes	
MyCoRe	-	3	Qualified Dublin Core	Yes	Germany
Open Journal System	-	1			
Open Repository	4	7	Qualified Dublin Core	Yes	hosted IR service from BioMed Central. Built on the latest DSpace software.
OPUS	-	23	Qualified Dublin Core	Yes	Germany
Other	3	88	Various		
	74	596			

PIRUS – Publisher and Institutional Repository Usage Statistics

a PALS 3 project

Appendix C: UK Research Institutional or Departmental Repositories

Institution	Repository	Software	Citation	DOI	Type	Status
University of Lincoln	Institutional Repository	Eprints	Yes	Yes	Article	Published Y In Press Y Submitted N Unpublished Y
Birkbeck	Birkbeck ePrints	Eprints	Yes	Yes	Article	Published Y In Press Y Submitted Y Unpublished N
University of Birmingham	ePrints Repository	Eprints	Yes	Yes	Article	Published Y In Press Y Submitted Y Unpublished Y
Bournemouth University	Bournemouth University Research Online [BURO]	Eprints	Yes	Yes	Article	Published Y In Press Y Submitted Y Unpublished Y
University of Bristol	Bristol Repository of Scholarly Eprints (ROSE)	DSpace	No	Yes	Journal article Preprint	Published Y In Press Y Submitted N Unpublished N
Brunel University	Brunel University Research Archive (BURA)	DSpace	Yes	Yes	Research Paper Preprint	
Aberystwyth University'	CADAIR	DSpace	Yes	Yes	refereed published journal paper	
University of Cambridge	CUED Publications Database	Eprints	Yes	Yes	Article	Published Y In Press Y Submitted Y Unpublished Y
University of Chester	ChesterRep	Open Repository	Yes	Yes	Article	
University of East Anglia	Digital Repository	Digitool	No, but individual citation elements	Yes	Journal Article	post-print formatted

			are available			
Sheffield Hallam University	Sheffield Hallam University Research Archive	Digital Commons	Yes	Yes	Journal Article	
University of Surrey	Surrey Scholarship Online	Digital Commons	Yes	Yes	Journal Article	(Where appropriate) This document has been peer reviewed

PIRUS – Publisher and Institutional Repository Usage Statistics

a PALS 3 project

Appendix D: JISC-REPOSITORIES Email Survey Responses - Summary

A summary of the responses received to the email survey sent out to JISC-REPOSITORIES list on 2nd Sept 2008. A total of 19 responses were received.

Organisations which responded

London School of Economics
Oxford University
Robert Gordon University
Universiteitsbibliotheek Gent
University College Dublin
University of Aberdeen
University of Bradford
University of Edinburgh
University of Hertfordshire
University of Leicester
University of Southampton
University of St Andrews
University of Stirling
University of Stuttgart
University of Sussex
University of Tasmania
University of Warwick
University of Wolverhampton
University of Twente

Responses to individual questions

1) Do you assign a 'persistent' identifier to individual articles?

All respondents replied that they do assign a persistent identifier to individual articles.

2) If you answered 'yes' to Q1, which persistent identifier(s)? [PLEASE INDICATE ALL THAT APPLY]

Identifiers assigned by respondents included Handle (DSpace), PURL, URN, ItemID (Eprints) and UUID (Fedora)

3) If you answered 'no' to Q1, what identifiers do you assign?

As all respondents claim to assign a persistent identifier this question was non-applicable.

4) If you do add DOIs to your records, which metadata field do you use to hold the DOI (dc.identifier, dc.relation, etc)?

The overwhelming majority of respondents do add DOIs to their records - where they are available.

There was great variation in the metadata element used to store this information, including:

- dc.description
- dc.identifier
- dc.identifier type DOI

- dc.identifier.citation
- dc.relation.isreferencedby
- dc.rights
- DOI
- relation

5) Do you add citations for your full-text articles?

The majority of respondents replied that they do add citations, or that the software generates/synthesises a citation from metadata entered. Others either add citations, if time permits, or don't add a citation at all.

6) What versions of full-text journal articles do you deposit in your IR? [PLEASE INDICATE ALL THAT APPLY]

- a) draft versions (Preprint– pre-refereeing, early version)
- b) submitted versions (Preprint– version submitted for peer review)
- c) accepted versions (Postprint– post-refereeing, pre publishers copy editing)
- d) published versions (Postprint– post-refereeing, publisher version)

Just over half of respondents stated that they deposit all and any versions of articles, subject to copyright and publisher policy restrictions. One in five respondents only deposit accepted or published versions.

7) If you do deposit draft or submitted versions into your IR, what do you do when the accepted or published versions become available?

- a) Replace (overwrite/delete) the earlier version of the article
- b) Add the later version to the existing record in addition to the earlier version
- c) Create a new record for the later version

There were mixed responses to this question, with any of the suggested scenarios being possible.

8) Which software package do you use for your digital repository?

Software packages in use included:

- DSpace (10)
- Eprints (6)
- Fedora (1)
- Open Repository (1)
- Opus (1)

PIRUS – Publisher and Institutional Repository Usage Statistics

a PALS 3 project

Appendix E: Article Report 1 XML Report Example for a Single Article

```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE reports PUBLIC "-//ProjectCounter//DTD reports//EN"
"http://www.projectcounter.org/dtd/2004/reports.dtd">
<reports xmlns="http://www.projectcounter.org/ns/2004/reports">
  <article_report id="cranfield.ac.uk_12345" cop_version="1" cop_report="1">
    <header>
      <title>Number of Successful Full-Text Article Requests by Month and DOI</title>
      <timestamp>2008-12-01-T10:03:47Z</timestamp>
      <vendor>
        <vend_name>Cranfield University</vend_name>
        <vend_imprint>Cranfield CERES</vend_imprint>
        <vend_site>https://dspace.lib.cranfield.ac.uk/</vend_site>
        <vend_contact>cranfieldceres@cranfield.ac.uk</vend_contact>
      </vendor>
      <customer>
        <cust_name>Central Agency</cust_name>
        <cust_ref>123-4567</cust_ref>
        <cust_ip type="cidr">123.456.78.9</cust_ip>
        <cust_username>centralagency</cust_username>
        <cust_criteria>institution</cust_criteria>
      </customer>
    </header>
    <article_data>
      <article doi="10.1016/j.jairtraman.2005.01.007" format="application/pdf">
        <requests start="2008-09-01" end="2008-09-30">108</requests>
        <requests start="2008-10-01" end="2008-10-31">201</requests>
        <requests start="2008-09-01" end="2008-10-31">309</requests>
      </article>
    </article_data>
  </article_report>
</reports>
```