New usage-based measures of impact in the context of altmetrics

- Usage Factor, PIRUS (and IRUS)

Peter Shepherd

COUNTER

April 2013

Altmetrics: the context

‘No-one can read everything. We rely on filters to make sense of the scholarly literature, but the narrow, traditional filters are being swamped… …the growth of new, online scholarly tools allows us to make new filters: these altmetrics reflect the broad, rapid impact of scholarship in this burgeoning ecosystem. We call for more tools and research based on altmetrics.’

From Altmetrics: a manifesto

Jason Priem at al

(http://altmetrics.org/manifesto/)

Filters for the scholarly literature:

- Citations
- Peer Review
- Usage
- Altmetrics

Altmetrics: what are they?

- **Altmetrics: what are they?**
  - Based on a range of measures, including: bookmarks, links, tweets (tweeted half-life!?), Facebook, blog posts and other measures that indicate ways that readers have been influenced by a publication
  - Reflect the fact that scholars are moving their everyday work to the web ((e.g. Mendeley, Zotero)
  - Cover newer forms of publishing, such as datasets, blogging, 'nanopublication', as well as traditional articles
  - Take into account the wider impact of research, e.g. by measuring activity in social media
  - Reflect the 'ongoing conversation' around the outputs of scientific research
  - Tend to focus on individual articles and other research outputs rather than on journals
- **Altmetrics: pros and cons**
  - Pros,
    - Cover all fields of scholarship and all types of publication
    - Take into account activity in social media
    - More immediate measure of impact than citations – 'real time'
    - Scholar and item based rather than journal-based
    - Go beyond counting and emphasis semantic content, e.g. usernames and timestamps
  - Cons:
    - What do the numbers mean? Do they reflect impact or buzz?
    - Less direct link to research quality
    - Easier to manipulate and less transparent than citation data
    - No widely accepted standards or benchmarks. Do 20 tweets = 1 Mendeley upload?
    - Can be gamed

COUNTER usage-based measures in the context of altmetrics

Advantages:

- Usage can be reported at the individual item and individual researcher level
- Usage is more 'immediate' than citations
- Usage potentially covers all categories of online publication
- COUNTER usage statistics are independently audited and generally trusted

Two approaches being pursued:

PIRUS

- Recording, consolidation and reporting of usage at the individual item level
  - Standard applies to publishers, aggregators and repositories

Usage Factor

- Usage-based measure of impact of journals, institutions and individual scholars
  - The Usage Factor for a Journal is the Median Value in a set of ordered full-text article usage data (i.e. the number of successful full text article requests) for a specified Usage Period of articles published in a journal during a specified Publication Period.

Both PIRUS and Usage Factor are based on the recording and consolidation of COUNTER-compliant usage data at the individual article level

PIRUS: mission and project aims

Mission

To develop a global standard to enable the recording, reporting and consolidation of online usage statistics for individual journal articles hosted by Institutional Repositories, Publishers and other entities

Project aims

- Develop COUNTER-compliant usage reports at the individual article level
- Create guidelines which, if implemented, would enable any entity that hosts online journal articles to produce these reports
- Propose a model for a Central Clearing House (CCH) in which these reports might be consolidated at a global level in a standard way.

PIRUS: project outcomes

- **Technical:** a workable technical model for the collection, processing and consolidation of individual article usage statistics, which forms the basis of the PIRUS Code of Practice.
- **Organizational**: an organizational model for a Central Clearing House that would be responsible for the collection, processing and consolidation of usage statistics has been proposed.
- **Economic:** the costs for repositories and publishers of generating the required usage reports, as well as the costs of any central clearing house/houses have been calculated and a model for recovering these costs has been proposed .

For full report on the PIRUS project go to:

http://www.projectcounter.org/News/Pirus2_oct2011.pdf

The draft PIRUS Code of Practice

- The PIRUS Code of Practice has been established as an outcome of the JISC-funded PIRUS (Publisher and Institutional Repository Usage Statistics) project.
- The primary aims and objectives of PIRUS were to assess the feasibility of and develop the technical, organizational and economic models for the recording, reporting and consolidation of usage of *Journal Articles* hosted by Publishers, Aggregators, Institutional Repositories and Subject Repositories.
- The PIRUS Code of Practice builds on the work undertaken by PIRUS, and the work of the JISC Usage Statistics Review and the Knowledge Exchange Institutional Repositories Workshop Strand on Usage Statistics.
- This PIRUS Code of Practice has been developed by COUNTER, which is also responsible for its on-going management and implementation. PIRUS is consistent with the COUNTER Code of Practice.
- To have their usage statistics and reports designated PIRUS-compliant vendors and other services must provide usage statistics that conform to this Code of Practice.

PIRUS:
- the draft PIRUS Code of Practice

The Code of Practice covers the following areas:

- article types to be counted;
- article versions to be counted;
- data elements to be measured;
- definitions of these data elements;
- content and format of usage reports;
- requirements for data processing;
- requirements for auditing;
- guidelines to avoid duplicate counting when intermediary gateways and aggregators are used.

PIRUS:
- the draft PIRUS Code of Practice

The PIRUS Code of Practice provides the specifications and tools that will allow COUNTER-compliant publishers, repositories and other organizations to record and report usage statistics at the individual article level that are credible, compatible and consistent. COUNTER-compliant publishers may build on the existing COUNTER tools to do so, while an alternative approach is provided for non-COUNTER compliant repositories, which is tailored to their systems and capabilities. This Code of Practice contains the following features:

- A list of **Definitions** and other terms that are relevant to recording and reporting usage of individual items
- A **methodology** for the recording and reporting of usage at the individual article level, including specifications for the metadata to be recorded, the content types, and the versions whose usage may be counted.
- **Specifications** for the PIRUS Article Reports.
- **Data processing rules** to ensure that the usage data reported are credible, consistent and compatible
- **Specifications** for the independent auditing of the PIRUS reports
- A **description** of the role of:
  - A Central Clearing House (CCH) in the calculation and consolidation of PIRUS usage data for articles.
  - Other Clearing Houses in relation to the CCH.

PIRUS: article data and metadata

Publisher/aggregator organizations should collect the usage data in the format specified in Article Report 1. The following data and metadata must be collected for each article:

- Either Print ISSN OR Online ISSN
- Article version, where available
- Article DOI
- Online Publication Date OR Date of First Successful Request
- Monthly count of the number of successful full-text requests - counts must remain available for at least 24 months from Online Publication Date OR date of First Successful Request
-
- The following metadata are optional, but are desirable:
- Journal title
- Publisher name
- Platform name
- Journal DOI
- Article title
- Article type

PIRUS: article types

Organizations must be able to record and report usage of the following categories of journal content at the individual article level:

- research articles (full articles and short communications)
- review articles

In addition the usage reports on the following individual items are acceptable, provided they meet the data and metadata requirements listed in 4.2 above:

- editorials
- book reviews
- theses

PIRUS: article versions

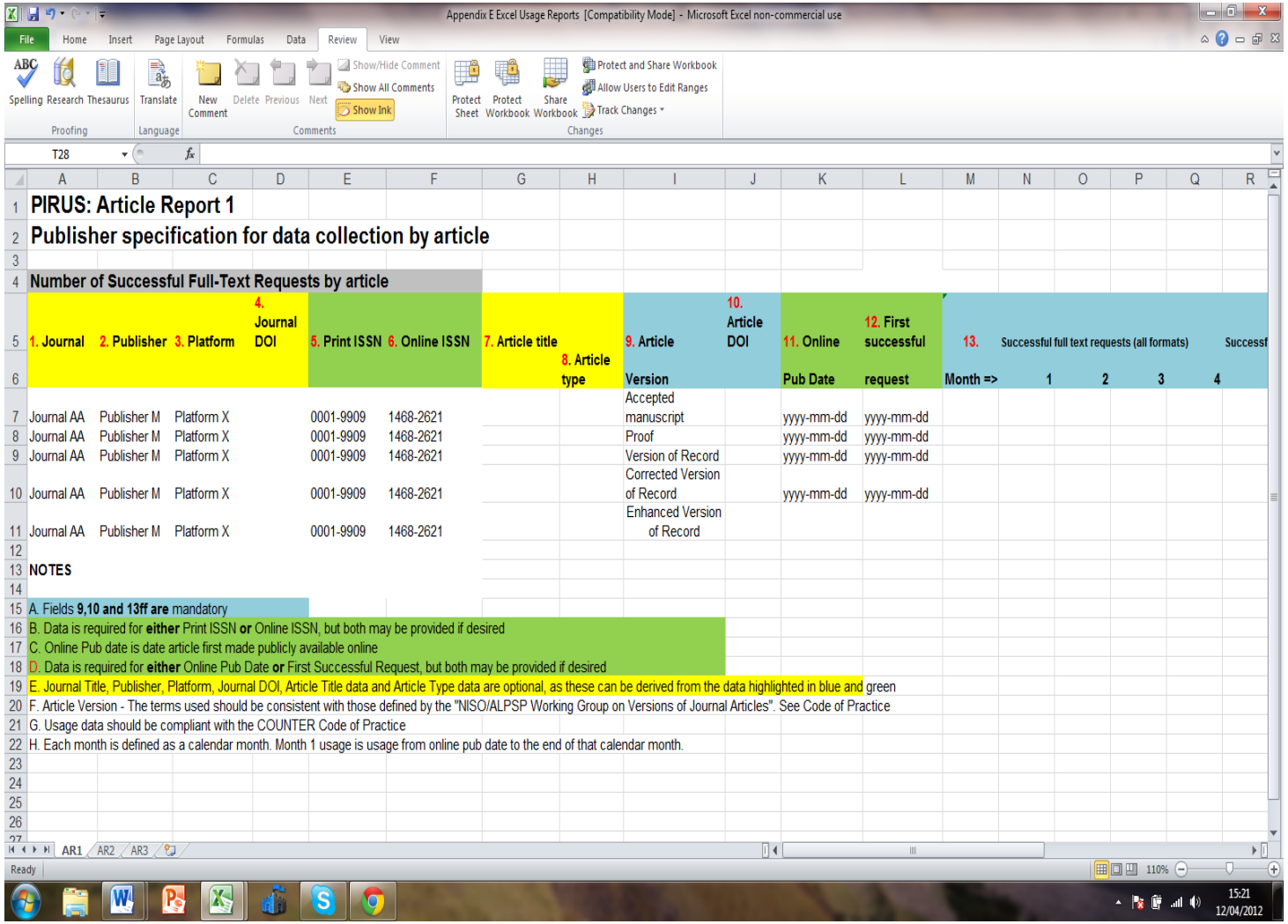
Only usage of the following 5 Article Versions (of the 7 versions defined by the ALPSP/NISO JAV Technical Working Group (http://www.niso.org/publications/rp/RP-8-2008.pdf)) may be counted:

- Accepted Manuscript (AM)
- Proof (P)
- Version of Record (VoR)
- Corrected Version of Record (CVoR)
- Enhanced Version of Record (EVoR)

Usage of the following 2 Article Versions must not be counted:

- Author's Original (AO)
- Submitted Manuscript Under Review (SMUR)

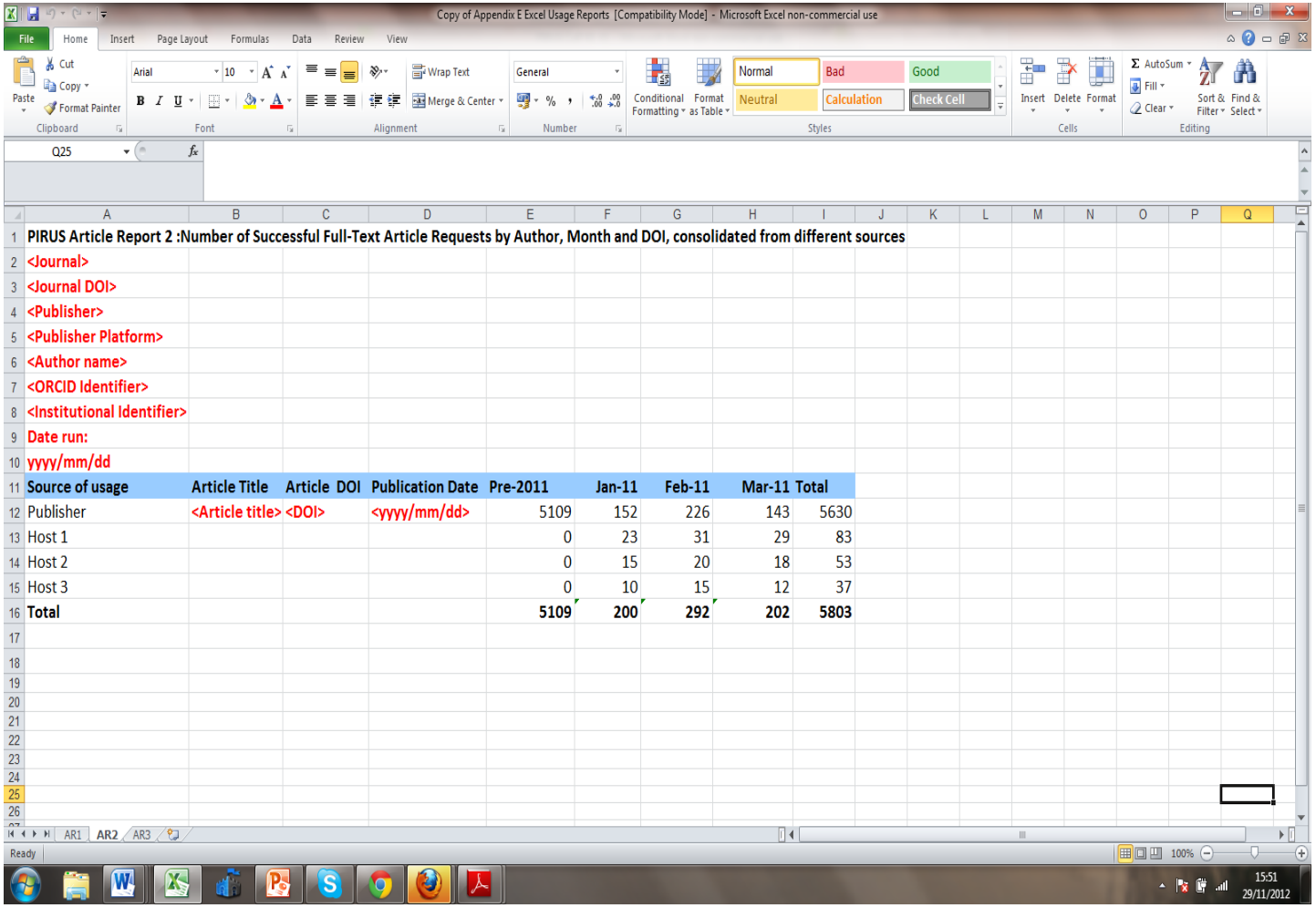
Article Report 1: specification for data collection by article



Appendix E Excel Usage Reports [Compatibility Mode] - Microsoft Excel non-commercial use

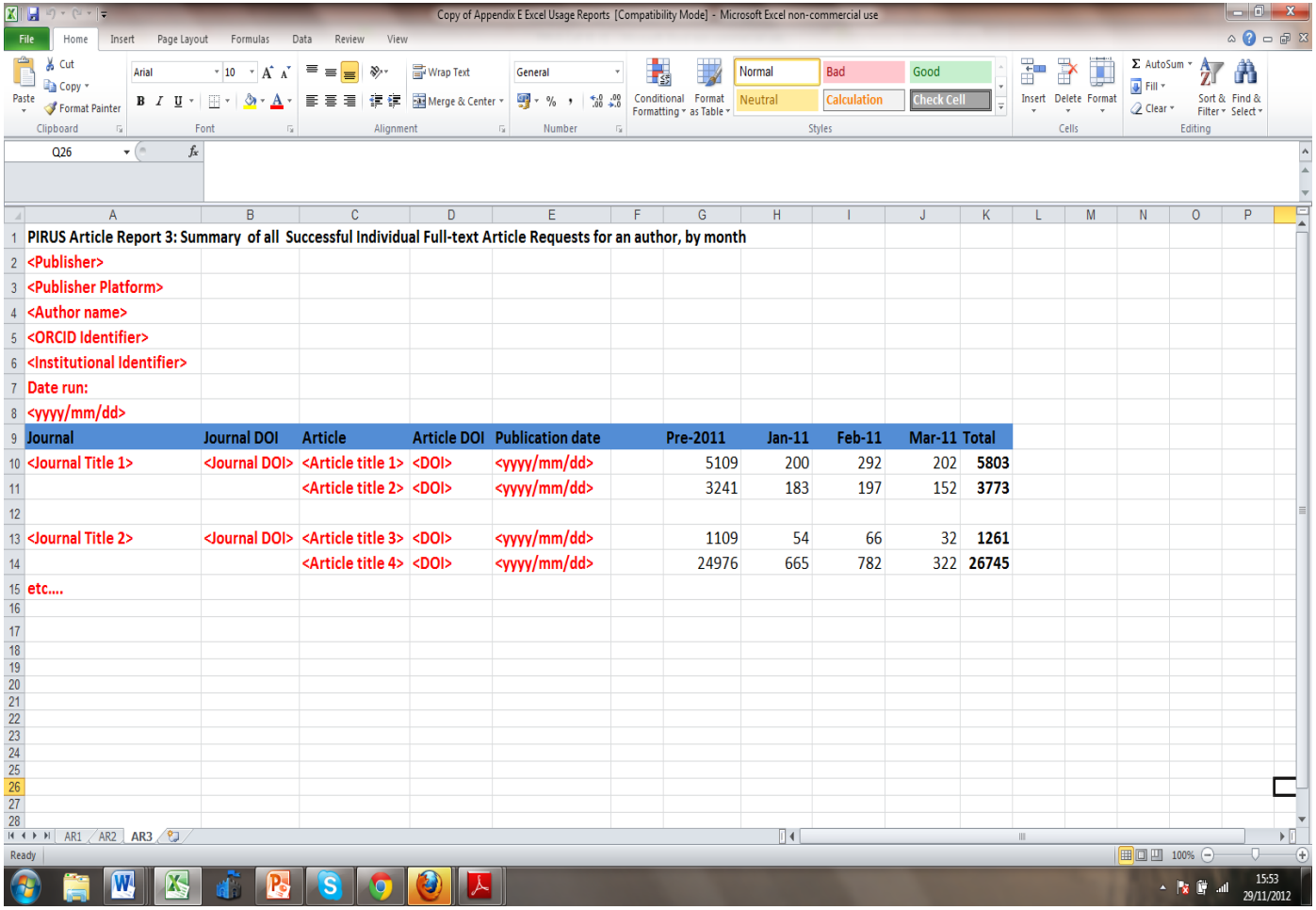
1	**PIRUS: Article Report 1**																	
2	**Publisher specification for data collection by article**																	
3																		
4	**Number of Successful Full-Text Requests by article**																	
5	**1. Journal**	**2. Publisher**	**3. Platform**	**4. Journal DOI**	**5. Print ISSN**	**6. Online ISSN**	**7. Article title**	**8. Article type**	**9. Article Version**	**10. Article DOI**	**11. Online Pub Date**	**12. First successful request**	**13. Successful full text requests (all formats)**					**Successf**
6													**Month =>**	**1**	**2**	**3**	**4**	
7	Journal AA	Publisher M	Platform X		0001-9909	1468-2621			Accepted manuscript		yyyy-mm-dd	yyyy-mm-dd						
8	Journal AA	Publisher M	Platform X		0001-9909	1468-2621			Proof		yyyy-mm-dd	yyyy-mm-dd						
9	Journal AA	Publisher M	Platform X		0001-9909	1468-2621			Version of Record		yyyy-mm-dd	yyyy-mm-dd						
10	Journal AA	Publisher M	Platform X		0001-9909	1468-2621			Corrected Version of Record		yyyy-mm-dd	yyyy-mm-dd						
11	Journal AA	Publisher M	Platform X		0001-9909	1468-2621			Enhanced Version of Record									
12																		
13	**NOTES**																	
14																		
15	A. Fields **9,10 and 13ff are** mandatory																	
16	B. Data is required for **either** Print ISSN **or** Online ISSN, but both may be provided if desired																	
17	C. Online Pub date is date article first made publicly available online																	
18	D. Data is required for **either** Online Pub Date **or** First Successful Request, but both may be provided if desired																	
19	E. Journal Title, Publisher, Platform, Journal DOI, Article Title data and Article Type data are optional, as these can be derived from the data highlighted in blue and green																	
20	F. Article Version - The terms used should be consistent with those defined by the "NISO/ALPSP Working Group on Versions of Journal Articles". See Code of Practice																	
21	G. Usage data should be compliant with the COUNTER Code of Practice																	
22	H. Each month is defined as a calendar month. Month 1 usage is usage from online pub date to the end of that calendar month.																	

Article Report 2: number of successful full-text article requests, by Author, Month and DOI, consolidated from different sources



	A	B	C	D	E	F	G	H	I
1	PIRUS Article Report 2 :Number of Successful Full-Text Article Requests by Author, Month and DOI, consolidated from different sources								
2	<Journal>								
3	<Journal DOI>								
4	<Publisher>								
5	<Publisher Platform>								
6	<Author name>								
7	<ORCID Identifier>								
8	<Institutional Identifier>								
9	Date run:								
10	yyyy/mm/dd								
11	**Source of usage**	**Article Title**	**Article DOI**	**Publication Date**	**Pre-2011**	**Jan-11**	**Feb-11**	**Mar-11**	**Total**
12	Publisher	<Article title>	<DOI>	<yyyy/mm/dd>	5109	152	226	143	5630
13	Host 1				0	23	31	29	83
14	Host 2				0	15	20	18	53
15	Host 3				0	10	15	12	37
16	**Total**				**5109**	**200**	**292**	**202**	**5803**

Article Report 3: summary of all successful individual article requests for an author, by month



Copy of Appendix E Excel Usage Reports [Compatibility Mode] - Microsoft Excel non-commercial use

	A	B	C	D	E	F	G	H	I	J	K
1	PIRUS Article Report 3: Summary of all Successful Individual Full-text Article Requests for an author, by month										
2	<Publisher>										
3	<Publisher Platform>										
4	<Author name>										
5	<ORCID Identifier>										
6	<Institutional Identifier>										
7	Date run:										
8	<yyyy/mm/dd>										
9	Journal	Journal DOI	Article	Article DOI	Publication date		Pre-2011	Jan-11	Feb-11	Mar-11	Total
10	<Journal Title 1>	<Journal DOI>	<Article title 1>	<DOI>	<yyyy/mm/dd>		5109	200	292	202	**5803**
11			<Article title 2>	<DOI>	<yyyy/mm/dd>		3241	183	197	152	**3773**
12											
13	<Journal Title 2>	<Journal DOI>	<Article title 3>	<DOI>	<yyyy/mm/dd>		1109	54	66	32	**1261**
14			<Article title 4>	<DOI>	<yyyy/mm/dd>		24976	665	782	322	**26745**
15	etc....										

PIRUS: Central Clearing House

Publishers will be able to consolidate the PIRUS individual article usage data from various sources themselves, or they may use the Central Clearing House envisaged for this purpose.

- The main features of the Central Clearing House will be:
- A home page that provides summary information on:
  - the number of articles and journals indexed
  - overall totals of successful full-text article requests recorded from publishers, repositories and other organizations
- A Search facility that makes it possible to find individual articles or groups of articles by:
  - DOI
  - Title/Author
- A number of reports can be generated:
  - Article Report 1j (AR1j) – this is a variant of the Article Report 1 (AR1), with usage events restricted to one journal at a time to reduce the report sizes. Its main purpose was to allow easy cross-checking that the data exposed from the PIRUS database matches the original data supplied by publishers
  - Article Report 2 (AR2) – this report is intended for article authors – showing total usage of an individual article, consolidated from publishers and repositories
  - Article Report 3 (AR3) – this report provides an overview of usage of a range of articles from a particular author
- Each of the reports may be viewed in a web page in the portal or downloaded for use locally as MS-Excel/TSV files (See Appendix E for examples of the various reports). These reports must also be available via the SUSHI protocol.

PIRUS: next steps

- Definitive Release 1 of PIRUS Code of Practice
  - Following feedback on draft CoP
  - Publication during 2013
- Invitation to publishers to implement PIRUS Code of Practice
  - A useful service to authors
- Consolidation of usage data from different sources by publishers
  - From IRUS and other COUNTER-compliant services
- Development of the Central Clearing House

For more information on PIRUS:

http://www.projectcounter.org/pirus.html

Usage Factor: aims and objectives

The **overall aim** of this project was to explore how online journal usage statistics might form the basis of a new measure of journal impact and quality, the Usage Factor for journals.

Specific objectives were to answer the following questions:

- Will Usage Factor be a statistically meaningful measure?
- Will Usage Factor be accepted by researchers, publishers, librarians and research institutions?
- Will Usage Factor be statistically credible and robust?
- Is there an organizational and economic model for its implementation that would cost-effective and be acceptable to the major stakeholder groups.

The project is being carried out in **three Stages**:

- Stage 1 (2007-2008): market research
- Stage 2 (2009-2011): modelling and analysis
- Stage 3 (2012-2013): further tests based on draft Code of Practice

Usage Factor metric: recommendations

- Usage Factors should be calculated using the median rather than the arithmetic mean
- A range of Usage Factors should ideally be published for each journal: a comprehensive UF (all items, all countable versions) plus supplementary factors for selected items
- Usage Factors should be published as integers with no decimal places
- Usage Factors should be published with appropriate confidence levels around the average to guide their interpretation
- The Usage Factor should be calculated initially on the basis of a maximum usage time window of 24 months.
- The Usage Factor is not directly comparable across subject groups and should therefore be published and interpreted only within appropriate subject groupings.
- The Usage Factor should be calculated using a publication window of 2 years

Usage Factor metric : recommendations

- Small journals and titles with less than 100 downloads per item may be unsuitable candidates for Journal Usage Factors: these are likely to be inaccurate and easily gamed
- The Usage Factor provides very different information from the citation Impact Factor and this fact should be emphasised in public communications.
- Further work is needed on Usage Factor gaming and on developing robust forensic techniques for its detection
- Further work is needed to broaden the scope of the project over time to include other usage-based metrics

Usage Factor: Journals - the calculation

Publishers will be able to generate Usage Factors using the Code of Practice, but will have to be independently audited for their Usage Factors to be listed in the Usage Factor Central Registry

- The 12 month Journal Usage Factor 2010: all items

  The median number of successful requests during the 12 months following the first successful requests for countable items published in the journal during 2010

  - Different items types have different impacts in different fields

- The Journal Usage Factor 2010: full-text articles

  The median number of successful requests during the 12 months following the first successful requests for full-text articles published in the journal during 2010

- Challenges:
  - Consolidation of usage data from different sources
  - Consistent item type definitions
  - Economic/Organizational model to support the Central Registry

Usage Factor infrastructure: recommendations

- Development of systems to automate the extraction and collation of data needed for UF calculation is essential if calculation of this metric is to become routine
- Development of an agreed standard for content item types, to which journal specific item types would be mapped, is desirable as it would allow for greater sophistication in UF calculation
- Development or adoption of a simple subject taxonomy to which journal titles would be assigned by their publishers
- Publishers should adopt standard “article version” definitions based on ALPSP/NISO recommendations

Stage 3: initial results

- Methodology and process
  - 12 month and 24 month UFs both show a good spread within subject fields, allowing journals to be differentiated on the basis of UF. 12 month Journal UFs range from less than 100 to over 2000 in a given field
  - A fixed UF counting period based on calendar years may be subject to gaming by publishers; a rolling year UF is likely to be more robust
- Subject classification scheme
  - Ringgold scheme works well for the 27 subject fields covered in this study. It corresponds very closely to the subject classifications used by the publishers themselves for their own journals
- Infrastructure
  - Problems with publishers providing data in the required format
  - Aggregation of usage data is time consuming
  - CIBER have developed a tool for the automatic aggregation of usage data from different sources

IRUS-UK: aim

- Enable UK IRs to share/expose usage statistics based on a global standard – COUNTER
  - Produced on the same basis as publishers
  - Filtered to remove robots and double clicks
  - Comparable
  - Reliable
  - Trustworthy
  - Authoritative

IRUS-UK: objectives

- Collect raw usage data from UK IRs for *all item types* within repositories
  - Downloads not record views
  - Not just articles
- Process those raw data into COUNTER-compliant statistics
- Return those statistics back to the originating repositories for their own use
- Give Jisc (and others) a nation-wide picture of the overall use of UK repositories
  - demonstrate their value and place in the dissemination of scholarly outputs
- Offer opportunities for benchmarking
- Act as an intermediary between UK repositories and other agencies
  - e.g. global central clearinghouse, national shared services, OpenAIRE

PIRUS, IRUS and Usage Factor

Common threads

- Article-based metrics
  - Can be rolled up to researcher, institutions and journal level
- Reliable, audited data
  - Based on tested COUNTER standards
- Common process/ infrastructure requirements
  - Similar metadata
  - Efficient, cost-effective processes

Acknowledgements:

- JISC, UKSG and RIN for funding