Appendix 1: Use cases

The following use cases were considered by the JAV Technical Working Group. Below, we have indicated in bold how our recommended terms would apply. We have used the following abbreviations for simplicity:

AO = Author's Original
AM = Accepted Manuscript
P = Proof
VoR = Version of Record
CVoR = Corrected Version of Record
EVoR = Enhanced Version of Record
C/EVoR = Corrected *or* Enhanced Version of Record, where the corrected or enhanced status is not known from the Use Case

#0: Base case. Author A writes first draft of article [**AO**]. This is circulated amongst colleagues (i.e. not formal peer review yet). Author revises article and submits to Journal Z [**AO**]. The Editors of the journal arrange for the article to be peer reviewed. The peer reviewer's comments are sent to the author, who amends the article, and re-submits to Journal Z [**AO**]. The article is accepted for publication [**AO** becomes **AM** upon acceptance]; the author posts this version [**AM**] on to her personal website and on to a subject or institutional repository. The publisher processes the article (copy-editing, article conversion, and page composition [all versions **P**]) to produce page proofs [**P**], which are sent to the author. The author marks up or annotates the proofs, which are returned to the publisher [**P**]. The publisher corrects the proofs and publishes the article online, without issue pagination, but with a DOI for identification [**VoR**]. This constitutes the official publication date. When the print issue is scheduled to be compiled, the publisher adds volume, issue and page numbers [**VoR –** addition of bibliographical details not sufficient enough to change status of VoR]. The article is published in print and electronic form with pagination and added features (e.g. citation tracking), and the author adds a URL link to this version [**VoR**]. The published version of the article remains available from the publisher's site, or from a third-party aggregator's site, or from an archival site [**VoR**].

#1: Author Smith writes an article manuscript [**AO**]. This manuscript is sent to several individuals to review; they review and provide feedback. Smith submits the article, incorporating the feedback from colleagues, to Journal A for consideration [**AO**]. The journal accepts the article, but with revisions. Smith reviews the article, accepts the reviewers' comments, and makes a number of the revisions requested by the publisher [**AM**]. Smith posts the article pre-print [**AM**] to her institution's institutional repository (hereafter referred to as "IR"), with the approval of her publisher. The article appears in the print version of Journal A [**VoR**]. The article also appears on the publisher's website, but with the data set that Smith used included (i.e. not just linked), so that other researchers can also see how her work was done [**EVoR**].

After three months, researcher Jones uses the data set, and discovers valid and unintentional errors in how Smith processed and interpreted the data. Smith reevaluates the data based on Jones' comments, and makes slight modifications to the conclusions in the initial article, citing Jones' comments and clearly identifying the changes from the published version of the article. The corrected version of the article is posted on the publisher's website [**CVoR**], with a link to the previous version [**EVoR**], as well. Smith also posts the corrected version [**CVoR**], as a post-print, to her institution's IR. The pre-print [**AM**] is removed from that site.

#2: Author Grant composes a manuscript which he does not share with colleagues [**AO**]. He sends a draft [**AO**] to Journal B, which rejects the manuscript. Grant makes some modifications to the manuscript and submits it [**AO**] to Journal C, who accepts it with significant modifications [**AM**]. Journal C posts the modified version [**AM**] to its pre-print site, where it receives comments from other researchers. Displeased with the modifications and the comments, Grant requests that Journal C remove the manuscript from the pre-print site (which it does), and then submits his version of the manuscript, which contains some of the modifications by Journal C and some of his own additional modifications, to his institution's IR [**AO**]. The institution's IR posts that version (without review), and Grant links to that 'final' version [**AO**] from his personal web page. Apart from cached versions of the pre-print version that appeared for some time on Journal C's web page [**AM**], no other version of the article is now available online or in print.

#3: Author Douglas writes a manuscript, which she submits to Journal D [**AO**]. Journal D accepts the manuscript without modification [**AM**], and publishes it in print [**VoR**]. Journal D is not available electronically, but Douglas receives approval to submit a PDFversion of the published article [**VoR**] to her institution's IR. In addition (and without prior approval from the publisher), she posts an HTML and/or text version of her document on her own web page . This version corrects some minor typographical inconsistencies, but does not change any of the data or conclusions presented in the article. [**AO** – although journal D accepted the AO without modifications so it became an AM, author has made her own amendments to her AO]

#4: Author Williams takes the text of an article he published 15 years ago and submits it [**VoR**] to Journal E, which is only published as a free, online product. Editors at Journal E, not realizing this article has already been published, comment on the lack of current citations, but accept the article [**AM -** erroneously] nonetheless, and without modification. The article had previously been published in print only, in Journal F, and is now also available in Journal E [misidentified as **VoR** since not known to be an updated version]. The only difference in the two articles is a fairly significantly different title; the texts of the two articles are identical.

A reader eventually notices the duplication, and the editors at Journal E decide to remove the article, and post an explanatory note, including a citation to the original article [**VoR**] in Journal F. On Williams' faculty website, he cites his publications in both Journal E [**CVoR**] and Journal F [**VoR**]. After Journal E removes the article, Williams posts the article text on his website [**CVoR**] and links directly to it, though he still cites the source as Journal E (reasoning that people searching abstracting databases will be able to confirm that it did appear in the journal).

#5: Conference papers [not considered since out of scope].

#6: Journal G accepts a manuscript [**AM**] from author Davis, and publishes it [**VoR**]. The publisher also distributes the whole volume of the journal (a combined issue of all four quarterly issues for the year) as a monograph [**CVoR** because the bibliographic details will be different for the version in the monograph]; Davis does not have the opportunity to have his article appear only in the journal and not in the monograph. The monograph has its own title but the articles/chapters within are identified as being duplicated from the journal. As a result, Davis' article is a standalone work in the journal [**VoR**], and a part of the whole in the monograph [**CVoR**].

#7: Author makes every stage (from initial draft [**AO**] to published version [**VoR**]) available on his personal website, and on institutional and subject repositories - every version [**AO; AM; P; VoR**] everywhere.

#8: Article [**AO** or **AM**] is taken from author's site and posted elsewhere, without the author's knowledge. (Or, worse still, article is stolen from author's site, and re-posted with alterations, possibly with malicious or plagiaristic intent.) Such fraudulent behaviour will probably lead to malicious misidentifications.

#9: Author updates version on own (or repository) site [author may identify as **CVoR** or **EVoR** but these are *not* formal VoRs**]**], which then diverges over time from the "fixed" version that is the published version [**VoR**].

#10: Corrections to the published version are posted as the equivalent of errata or corrigenda, or the article has these corrections inserted [**CVoR**], or the errors are so serious (technically or legally) that the article is retracted or removed from publication [**VoR**s may still exist on various sites but they are no longer formally recognized; the formal publication site identifies the article as having been retracted or removed].

#11: Publisher goes bankrupt without having made any arrangements for a third party to continue to host or archive the article [may be lost **VoR**].

#12: Article is not formally published in a journal, but the repository takes on all the formal functions of "publication" [**VoR**] - accepting the article for inclusion [**AM**]; processing it for conformance with repository specifications [**P**]; promoting it to the outside world; maintaining access to it; and dealing with subsidiary rights (i.e. use by other parties).

#13: An article is written by three authors, each of whom works in a different institution. Each institution has its own repository, and each institution requires that the article is deposited in its own repository since it reflects work done by its researchers. The research work is also funded by more than one granting body; each granting body also requires posting in a different repository. Over time, the versions in each of these repositories diverge as they are amended by the authors or the repository managers [various **AO**s; one version may be the actual**AM** but others may be author-amended versions of the **AM**].

#14: An article [**AO**] is submitted to Journal C, accepted after peer-review comment and the author deposits the version with the corrections [**AM**] in her IR; the Journal C formally publishes it [**VoR**]. The IR managers convert the version into a different file format suitable for long-term preservation which preserves the layout as well as the information content [remains a **VoR** since VoR is not format-dependent].

#15: User A does a search on Google scholar for a phrase B. Google Scholar retrieves a list of possible records. User A selects article C in Journal D and has the following options:
i) full text of the author's accepted manuscript in TeX, accessible via the author's IR [**AM**];
ii) full text of the published version, in PDF [**VoR**], via a host- access to this version requires that User A's institution subscribes to Journal D;
iii) full text of the published version in HTML format [**VoR],** via an aggregator – access to this version requires that A's institution subscribes to Aggregation E, which includes Journal D;
iv) a link to a preprint site F, which offers a Word version of the manuscript, prior to peer review and acceptance in a journal [**AO**].

#16: User E starts searching his IR and locates a record which links to the publisher's site for article F [**VoR**]. In addition, there is a link to a local (IR-hosted) version of article F. This version has no explicit version information in the metadata, but it appears to be the published PDF since it has the publisher's copyright line and is paginated within a volume and issue of the journal [**VoR**].

#17: User H starts searching herIR and locates a record which links to the publisher's site for article I [**VoR**]. As for Use Case #16, there is also a link to a local (IR-hosted) version of the article, but in this case, the version is in Word, and there is no clear indication of its relationship to the published article [may be **AO** or **AM**].

#18: User J starts searching his IR and locates a record which links to the publisher's site for article K [**VoR**]. As for Use Cases #16 and 17, there is a link to a local (IR-hosted) version of the article, available as a scanned image in TIFF format. Although it is identified as a "pre-publication" version, its relationship to the published version is unclear [may be **AO, AM** or even **P**].

#19: User M belongs to an institution N which has an Open URL resolver which uses its IR as well as publisher resources. User M searches in a journal article database O to locate an article; the Open URL resolver offers M full text in XML from the publisher [**VoR**] and full text, also in XML, from within itsIR. However, although these are both in XML, one has been produced by the publisher as an end-product of the typesetting process [**VoR**], whereas the other has been created by the IR manager in a separate, post-publication process [*not* the formal **VoR** but an unauthorized version of it]. Each XML file complies with a different DTD (one the publisher's and the other the IR's), although the content is identical in each.

#20: User P belongs to an institution which subscribes to the Journal Q, and she uses the publisher's web site to locate the final published version of article R [**VoR**]. The publisher offers both PDF and HTML versions of the article [**VoR**], and also provides links to comments on the article and to related articles [*not* an **EVoR** since the links lie outside the article itself] , some of which appear in the subscribed journal, and others in journals to which the user's institution does not subscribe.

#21: User S does a search with phrase B and locates the article T in author U's IR (with links to the publisher's full text [**VoR**]) and also articles V and W in two different repositories. It is unclear whether V and W are the same or different versions of T, or a different article entirely.

#22: Author A submits an article to Journal Z in LaTeX, Word, PostScript or PDF with separate figure files in PS, JPG, TIFF etc. The Editors of the journal arrange for the article to be peer reviewed and prepare a PDF for the reviewers [**AO**].

The peer reviewers' comments are sent to the author, who amends the article, and re-submits to Journal Z. The Editor prepares a PDF of the amended version and sends it back to the referees. The loop of review and revision can be repeated a number of times [**AO**].

The article is accepted for publication [**AM**]. The accepted version is the last version submitted by the author in his own format. The publisher also creates a PDF version of the article [**AM** since no content processing].

The publisher processes the article: copy-editing, article conversion, and page composition, creating intermediate "versions" [**P**]. This process produces PDF page proofs [**P**], which are sent to the author. The author marks up or annotates the proofs [**P**], which are returned to the publisher. The publisher corrects the proofs and occasionally there will be a second round of proof checking [**P**].

The publisher publishes the article online (without issue pagination, but with a DOI for identification) and in one or more file formats - HTML, PS, PDF [**VoR –** CM: subject to agreement in our WG]. This constitutes the official publication date. When the print issue is compiled, the publisher adds volume, issue and page numbers [remains the **VoR**]. A high-resolution file is sent to the printer. The online files are changed to include pagination and features are added (e.g. citation links, citation tracking, links to

supplementary data) [**EVoR]**. A full-text XML file and figure files exist behind the scenes and could be used online instead of HTML, to regenerate new PDFs [**C/EVoR –** on the assumption that the new PDFs will have some corrections or enhancements compared with the original PDF] or to make other future formats [**VoR** or **C/EVoR**, depending on whether the future format is a new format of the original **VoR** or of the **C/EVoR**].

The online version [**VoR** and then **C/EVoR**] is available from the publisher's site, from a third partyhost, from an aggregator, or from an archival site.

#23: Author (or sanctioned author's agent, such as a department or research unit) deposits herwholly owned working paper [**AO**] into an institutional repository (first converting, if necessary, to one of a limited number of acceptable formats). Depositor supplies an abstract and descriptors. Repository converts the item to PDF, assigns a unique ID, constructs a metadata record, and makes the paper accessible with a "cover page" supplying full bibliographic information and suggested citation [**AO** if not yet **AM**]. Author is encouraged, but not required, to alert the agent and/or repository, and supply details when/if:

- there is a revised version of the item [**AO** or **AM** if accepted for publication in a journal], in which case the repository suggests deposit of the revised version and updates the metadata of both the previous and revised versions and provides forward and backward links to both
- a version of the item is formally published [**VoR**], in which case the repository version's metadata is updated with the bibliographic information and URL of the published version
- the author wishes to remove the item (e.g. due to "prior publication" concerns of a publisher), in which case the item will be replaced bya metadata record indicating that the item has been removed and, where applicable and supplied, the bibliographic information and URL of the published version.

#24: Repository harvests citation data for papers written by institutional authors during time period X, filters them for "postprint friendly" publishers (according to Sherpa/Romeo and locally updated data), and sends a request to the author to deposit an "author's copy" [may be **AO** or **AM**] or, when clearly allowed, a publisher's copy [**VoR**]. Upon deposit the author (or her agent) verifies bibliographic metadata, supplies an abstract and descriptors. Repository double-checks rights status and publisher postprint policy and if all is in order, converts the item to PDF [**AO** or **AM**], assigns a unique repository ID, constructs a metadata record, and makes the paper accessible with a "cover page" supplying full bibliographic information of published version, suggested citation, and required publisher's statement, if any. The author is encouraged, but not required, to:

- provide bibliographic and location information about "preprint" versions [may be **AO** or **AM**] of the item
- alert the repository if the author wishes to remove the item, in which case the item will be replaced by a metadata record indicating that the item has been removed along with the bibliographic information and URL of the published version.

#25: A publisher publishers a short summary version (identified for example as an "abridged PDF") of an article in print, and both the short and full version of the article online. These online versions have the same DOI. Only the long version is regarded as the **VoR** – the short version is regarded as an extended abstract rather than an alternative version of the article.