Comments from JAV Review Group on Recommendations

Subject	Commentator	Comment	Reply
1) *Terminology*			
	Emily Dill (1/4)	Should we use completely new terminology in order to avoid "loaded" terms (the associations that different groups may attach to terms)?	We discussed this possibility early on but decided against it. We tried to avoid loaded terms like pre-print, post-print, authoritative version and chose to use *understandable terms* with more explicit definitions.
	Bruce Rosenblum (4/4)	Top level vocabulary is excellent. However, within each top level category there are many subclasses. Next step may be to define a "similarly brief but powerful" vocabulary of sub-types.	We feel that the next step is to operationalize the definitions. Then we would want to keep the typology simple in order to facilitate work implementing standard citing statements and metadata and to build linking mechanisms.
	Fred Friend (7/4)	Alternative view from BR above: proposals are practical and balanced; likes the simplicity of having only five categories, and we should resist temptation to create sub-divisions. Date-stamping may help users	Thank you – and we agree about the date-stamping.

		navigate through the sub-divisions without the need for new terms	
	Fred Friend (7/4)	Don't follow definitions with "also known as". This just leads to argument over semantic equivalence of various loaded terms	We feel it is helpful to link our standard terms to other terms in common use by particular communities, although they are not always synonymous, and we make that clear in the narrative.
a) *Author's Original*			
	Richard Fidczuk (31/3)	Author may revise AO as part of peer review process. Calling such a version "Author's Original" doesn't seem right – may need an "Author's Revised" stage?	This is a refinement that immediately introduces the need for other refinements as well. The AO may be revised many times by the author on his own or by the author reacting to informal peer review. Value may be added or subtracted from AO, but until Accepted, it is AO in all its iterations. (Version numbers or time stamps might help with all iterations rather than defining sub-types.)
	Kate Sloss (7/4)	Term and definition are fine.	Thank you.
b) *Accepted Manuscript*			
	Sally Morris (24/3)	Are iterations possible? E.g. article withdrawn after acceptance and subsequently accepted by another journal?	The example is rare enough not to be included in high-level semantics. We think it is OK to deal with this type of rare situation as exception handling.

	Sally Morris (24/3)	If an IR Manager changes an AM (e.g. format conversion), is this a "versioned AM"?	We did not generally want to consider formats or media as different versions at the top level. In general, we think these cases would be refinements to major version categories. However, there is some technology dependence here. There are such things as "lossy" conversions, e.g., Google's html'izing of PDFs. These are "defective copies" rather than new versions.
	Chuck Koscher (29/3)	Does the AO designation remain with the "as submitted" article version (use case #0)?	Yes. This is a very important point because it highlights one of the key criteria we considered in differentiating our types, viz. formal public "acts" by a recognized party. So, the re-submitted version *might* incorporate all the changes required by the peer review process, and in such a case the content is identical to the AM, but it is still considered an AO until changes are *externally verified by a formal act of acceptance*. A consequence of this approach is that there must be retroactive version re-naming. If the author deposits the peer-reviewed and corrected AO in an IR at the same time as re-submitting, the "act" that converts the type to AM must apply to all copies.
	Kate Sloss (7/4)	Not clear what is meant by the phrase "takes responsibility" – does it mean "asserts publishers' rights"? The description of the review process is	Good point. We did discuss "takes responsibility" initially in a legal sense, but decided instead use the term in the sense of "publicly lends its imprimatur to the scholarship" or "stands behind and supports the work as a valued contribution to scholarship"

		unclear – the decision taken by an IR manager, for example, is not a review process	
c) *Proof*			
		No-one has made any specific comments about this term, apart from the general one from Bruce Rosenblum about consideration being given to sub-types (e.g. uncorrected proof, corrected proof, revised proof)	At this point, we would like to keep to the "phylum" level. There may be many further subdivisions, but these are more fine-grained than is necessary for our high-level view.
d) *Version of Record*			
	Sally Morris (24/3)	Could we have multiple VoRs, i.e. same VoR but in different locations? (see use case #13)	Yes. Copies of VoRs proliferate online just as in print. Initially, we did consider "location" as a criterion for distinguishing versions, but decided against that. In other words, our typology ignores the fact that one VoR a user finds might be "an unauthorized copy" or infringing IP from a legal standpoint.
	Sally Morris (24/3)	What difference does it make if a VoR is released in various file formats a) simultaneously or b) asynchronously	For us, formats would be sub-types of the progressive staged versions we laid out, not different versions at the same hierarchical level.
	Peter Suber (4/4)	Term is "rather inflated" and doesn't take into account the occasions when	We disagree. Version of Record was chosen to avoid inflation of Authoritative Version or Definitive Version. "Published Version" is a highly ambiguous term. It does

		the published version is inferior to the Author's Original. "Published version" is simpler, more accurate and more neutral than "Version of Record". [But see other views below]	not distinguish between "published" as in "making publicly available" and "published" as the end result of a formal process that changes content and fixes it in a citable stage. Our entirely typology of versions has explicitly built into it the notion of *progression*.
	David Goodman (5/4)	In response to PS above: "There will always be cases that don't quite fit. The author's version [i.e. Accepted Manuscript] may be better in some way but there needs to be at some point a defined final record for working purposes"	We agree.
	Anthony Watkinson (6/4)	AW endorses the use of the term "Version of Record". "The academic community do want one fixed version that will be the default version for citing"	We agree.
	Fred Friend (7/4)	Also endorses term "Version of Record". All terms are loaded.	We agree.
	Bruce Rosenblum (4/4)	There may be multiple VoRs (and UvoRs) floating around. It would be helpful to have some way of	Yes, there will be multiple VoRs (see first VoR comment above). Copies of VoRs abound. There is no need to distinguish them *in this typology* unless one is not a VoR. There is a parallel with printed copies: everyone has their

		indicating which VoR (and UvoR) is the "authentic" one, i.e. the one that is identified as such by the rightsholder. This may also include information about which *format* (e.g. XML, HTML, PDF, print) constitutes the VoR.	own copy. We are not concerned with *copies* but rather with *stages.* Introducing *authenticity* as a function of legal rights to expose, display, or serve a copy of a VoR is, we think, outside the scope of our group's work. Formats are sub-types that we are not concerned with in our high-level typology.
	Kate Sloss (7/4)	Definition is too loose – it says that "any organisation" that makes public can declare a VoR, but this may lead to confusion with IRs	We agree, and we will clarify in the narrative. Our key is the declaration "fit for publication" and the process, implied by the context of the stage, that leads to it. We have a progressive staging model. One cannot skip directly from AO→VoR.
e) *Updated Version of Record*	Sally Morris (24/3)	VoR needs to differentiate between changes to content and more "peripheral" changes, such as pagination, formatting, linked errata, comments or datasets (see also Chuck Koscher below re AVoR and EvoR). Also Richard Fidczuk (31/3), Bruce Rosenblum (4/4) and Janet Halsall (12/4) same basic point.	OK. A number of Review group comments indicate that Updated VoR is too broad a type because it conflates things that are significantly different. We will change the recommendations to distinguish between "Corrected VoR" (which would include all errata and corrigenda) and "Enhanced VoR" (which may contain extra material).
	Sally Morris (24/3)	How does one deal with linked errata and post-	Our scope is the journal article itself. *Links* to supplemental material should not be considered

		publication comments?	changes/updates to VoR. If the content is amended in light of these supplemental materials, then it is either a Corrected VoR (if the amendments are in the nature of a correction) or an Enhanced VoR (if the extra material is added to the VoR, not just linked from it). Corrected VoRs and Enhanced VoRs should always link back to the original VoR.
	Sally Morris (24/3)	What happens when the destination of an embedded link changes? Does that constitute a change to the VoR?	No – the VoR hasn't changed.
	Sally Morris (24/3)	Is the article in use case #10 a "lost VoR" rather than a UVoR?	We agree. A retracted article usually disappears from a site; a change is often made in the Citation Page that remains visible. This could be formalized in a "metadata" editorial note: "Retracted because…". We do not think this affects the versioning scheme that we have. We think it is a rare case that requires non-standardized exception handling. In theory, the extant VoR copies should have a watermark "Retracted" inserted.
	Chuck Koscher (29/3)	The print version of a VoR is a VoR, not a UVoR (use case #0)	We agree and will amend the use case text.
	Chuck Koscher (29/3)	Should we have "AVoR" for "Alternative VoR" (e.g. same content in different locations; maybe also same content in different format)?	We don't think so. In our scheme, location is not a factor in determining VoR and formats would be more refined sub-types.
	Chuck Koscher	If above accepted, would	See above.

	(29/3)	also lead to requiring a UAVoR	
	Chuck Koscher (29/3)	"EVoR" for "Enhanced VoR", e.g. links added but content not changed?	See above re proposal to introduce "EvoR". Added links constitute an Enhanced VoR since the content has changed by the addition of these links.
	Chuck Koscher (29/3)	"EUVoR" for "Enhanced UVoR"?	We will not distinguish between further iterations of EVoRs – as with other iterative versions, version numbering or date-stamping could be used to disambiguate.
	Peter Suber (4/4)	See PS's comments re VoR – by the same logic, prefers "Updated Published Version"	We disagree – see comments re VoR versus "Published Version" above.
	Anthony Watkinson (6/4)	Corrections should be treated as a special case and always be linked to the VoR. Retractions are a special form of UvoR.	See comments above re "Corrected Version of Record" and comments re retracted articles.
	Kate Sloss (7/4)	The VoR should be the one and only definitive version.	Yes but *copies* abound, *Versions of Record* don't.
2) *Metadata*			
	Peter Suber (4/4)	"Practical systems for ensuring that the metadata is applied by authors or repository managers and publishers" may be impediments to authors self-archiving	See below.
	David Goodman	In response to PS above: "It	We agree.

	(5/4)	is not our concern to establish self-archiving or any other model. We are intending to describe what might be found; in a standard way; we are not intending to specify how it might get there, or why one might choose to use a particular form"	
	David Goodman (5/4)	In response to PS above: "no standard can compel. The decision of some body to require the standard is what would compel"	We agree.
3) *Enforcement/compliance*			
	Sally Morris (24/3)	Who will be responsible for making sure that articles are correctly identified and point properly to variants (both backward and forward)?	We agree that this is a key question, but we think it is outside the remit of our WG, and may be covered by the evolution of best-practice guidelines. CrossRef has been working toward practical methods of version identification. One activity is Publisher-to-Author communication of both standard wordings and supply of particular DOIs, see *CrossRef Guidelines for Standard Citations in Author Postings.* The second activity to implement a system for interlinking of versions is CrossRef-IR processes. These have been discussed with a number of IRs; some are

			actively retrieving DOIs for VoR and inserting link into IR versions; and plans are underway to build necessary support for automated look-ups and DOI retrievals into some IR software.
	Sally Morris (24/3)	How would "fraudulent misidentification" (use case #8) be identified?	Well, it certainly won't be embedded in the metadata or full-text by the perpetrator! Once discovered, there could be various resolutions, but we think this situation is rare enough to be dealt with by special case exception handling.
	Emily Dill (1/4)	Will there be suggestions from the group on how libraries/repositories, etc. should adopt and use the standards?	See CrossRef references above. In addition, the JISC-funded VERSIONS project (http://www.lse.ac.uk/library/versions/about.html) may address this issue.
4) *Use cases*			
	Sally Morris (24/3)	Specific comments and wordsmithing	We will address these and amend the use cases text accordingly.
	Chuck Koscher (29/3)	Various specific comments	As above.
	David Goodman (2/4)	Specific comments re Case 10 (as a variant of case 0) and further use cases	As above.
	Kate Sloss (7/4)	#1: Should avoid even mentioning terms "preprint" and "postprint". Despite the SHERPA/RoMEO	See previous comment on utility of referring our standard terms and definitions to other looser and ambiguous uses. The use cases were written prior to the Terms & Definitions so we will re-write the Use Cases using only

		definition, many people (intuitively?) use "postprint" to mean the publisher's PDF. #7: not very realistic. Authors don't want this – but could be an edge case. #9: a bit speculative, but possible. Might be interesting to develop #12: IRs should not be able to give an article "VoR" status. But maybe we need to clarify whether VoR can be assigned to an article that is available only in an IR and not otherwise published? #14: not clear which version the IR manager is converting	our typology. #12: In our typology, a VoR has to have been formally published. It may exist in more than one location (e.g. on a publisher's site, in an aggregator service, in one or more repositories) but its designation as a VoR indicates that it is the version of the article that has been through a process of verification and certification. Even if an IR contains a VoR, it is not the IR that has taken responsibility for declaring the version to be the VoR – this has been done by the entity acting as the publisher. If an IR does take on this declarative responsibility, it is acting as a *de facto* publisher. Other typologies or vocabulary sets take a different view. For example, the CrossRef Glossary (http://crossref.org/02publishers/glossary.html) is not restricted to journal articles and therefore makes a distinction between "Version of Record" understood as a fixed end-point in the life-cycle of a work and "Definitive Work" understood to be a formally published fixed end-point which incorporates processes of community validation and publisher approval. In this way, IR contents can, and will often be "Versions of Record" in CrossRef terminology, when not followed by formal publication. (Although there has been some cross-over in the work of JAVTWG and CrossRef's Institutional Repository Committee that is evident in the Glossary, the perspectives and scope of the two groups differed, so there is also divergence.)
5) *The data model*	Peter Suber	Relationship metadata	OK – we will amend.

	(4/4)	requirements: change "must" to "should"	
	Kate Sloss (7/4)	The diagrams portray a division that doesn't exist from the author's point of view. "Dissemination/publication" and "citation" are not two separate timelines	The conceptual family represents objects that are beyond the scope of our Journal Article Version definitions. Nonetheless, relations exist between these informal, gray literature, works and formal journal articles. Authors recognize the conceptual difference even while wanting to be cited prior to formal journal publication. The left part of diagram might be super-labelled "Formal Literature" and the right side "Gray Literature" and we might add that Conventions and Best Practices have the goal of interlinking the two. Internal, embedded, citations are required but formal pointers external to text are also important. References are not usually called metadata, but are often treated as such in online offerings: they are often extracted from text and exposed external to it with live links. But retroactive forward linking to chronologically later members of the conceptual family (accomplished to a degree by search engine "clustering") should be incorporated in metadata with pointers.
6) *General*			
	Sally Morris (24/3)	Make comment in narrative that date of first "making available" potentially important in claims of priority and patent registration	Articles may be made available without being formally published, e.g. preprints in arXiv. The relevance of these postings to claims of priority and patent registration raises issues that are outside the scope of our scheme.
	Sally Morris	Clarify which article	We will fix the diagram. The narrative states that AMs

	(24/3) and Candy Zemon (27/3)	versions could be iterative. (Narrative implies only AM and VoR are "fixed", but is this true? Graphic in model shows two-way arrows, implying potential iteration.)	and VoRs are not iterative, but all other versions may be.
	Candy Zemon (27/3)	Looking for terminology that will help automated update of holdings, automated resolution of citations, ability to retrieve related materials – check against these requirements	We feel that a standard terminology is at least a start. We do not want to go too far outside our specific scope as a Working Group.
	David Goodman (2/4)	Comment: "for the less important journals, there is now no functional separation between editing to accommodate peer review, and copyediting"	This is a matter of opinion. Peer-review editing and copy-editing may be much the same thing or may differ considerably. Our typology makes no assumptions regarding degree of difference.
	Peter Suber (4/4)	In the preamble, would prefer that we did not refer to various stages as "added value". More accurate and neutral to say "revisions".	We disagree. Value is not a commercial term; it points to something that *may* be commercialized *or not.* Added-value is actually a critical part of our *developmental, progressive stages* model. The VoR is not merely a Fixed Point, but it is a Valued Reference Point that everyone wants identified. All other versions relate to the VoR.
	David Goodman (5/4)	Endorses PS point above	See above. The River Project produced a terrific report. It takes many nuances into account that are purposely

			omitted by the JAVTWG. One of their nuances is to understand that different users will value different versions for different purposes. By trying to place version identification into a relationship with the user, the complexities rise exponentially.
	Fred Friend (7/4)	Commends the balanced approach and avoidance of business model issues. Hopes we can keep this up	Thank you