

April 2011

# Are These Two Versions the Same? Functional Equivalence and Article Version

Todd Carpenter  
NISO, [tcarpenter@niso.org](mailto:tcarpenter@niso.org)

Follow this and additional works at: <https://docs.lib.purdue.edu/atg>



Part of the [Library and Information Science Commons](#)

---

### Recommended Citation

Carpenter, Todd (2011) "Are These Two Versions the Same? Functional Equivalence and Article Version," *Against the Grain*: Vol. 23: Iss. 2, Article 6.

DOI: <https://doi.org/10.7771/2380-176X.5783>

This document has been made available through Purdue e-Pubs, a service of the Purdue University Libraries. Please contact [epubs@purdue.edu](mailto:epubs@purdue.edu) for additional information.

# Are these Two Versions the Same? Functional Equivalence and Article Version

by **Todd Carpenter** (Managing Director, NISO, One North Charles Street, Suite 1905, Baltimore, MD 21201; Phone: 301-654-2512; Fax: 410-685-5278) <tcarpenter@niso.org> [www.niso.org](http://www.niso.org)

How we determine whether two things are the same or different depends on how we define “same.” This question dates back to **Aristotle** and **Plato** and the differences between universal forms and instantiated forms of objects, and understandings of how items are grouped together in classification systems. Without getting too deeply philosophical, how we group and classify things is at the heart of librarianship. It is also core to the question of identification and description in the context of published information.

In our digital world, producing copies of an item is as easy as pressing the <F12> keys to “save as” on a PC, and distributing that information worldwide is only a matter of saving that item to a Web-accessible server. We desperately need a common understanding of the differences that might exist between the original file and the copy. However, to do so with every minutely changed file — even to the level of its creation metadata, e.g., the newly saved file’s date of origin is different from the original’s — is an unmanageable and sometimes unnecessary task. In an era where duplication is easy, managing versions most certainly is not.

One of the principles of determining the differences between items is to consider their functional equivalents, a concept that has its origins in literature translation, but was developed into a metadata theory described in detail in **Godfrey Rust’s** and **Mark Bide’s** <indec> *Metadata Framework* ([http://www.doi.org/topics/indec/indec\\_framework\\_2000.pdf](http://www.doi.org/topics/indec/indec_framework_2000.pdf)). In this context, the distinction between when it is useful to identify a thing as being different from another thing should be done only when it is useful to do so. By focusing our attention on when it is valuable to maintain separate “records” of a version change, it addresses some of the problems of limiting the scope of the problem. For example, we needn’t design systems to track every possible change, if doing so is not something that people derive value from. For example, there may be multiple draft versions that an author might write, but for the overwhelming majority of users and uses, those versions are not useful, nor are they worth the expenses of identifying, describing, and preserving them.

This issue formed the basis for the joint project between **NISO** and **ALPSP** on Journal Article Versions, led by **Cliff Morgan** at **Blackwell-Wiley** and **Bernie Rous** at **ACM**, which resulted in a **NISO Recommended Practice on Journal Article Versions (JAV)** (NISO-RP-8-2008) (<http://www.niso.org/publications/rp/RP-8-2008.pdf>). In addition to the recommended practice, a detailed article about the JAV project was published in this magazine in January 2007. ([www.against-the-grain.com/TOCFiles/v18-6\\_ToddCarpenter.pdf](http://www.against-the-grain.com/TOCFiles/v18-6_ToddCarpenter.pdf)).

*pdf*). **The recommendations consisted of these seven stages that correspond to stages in the publication process:**

- AO = Author’s Original
- SMUR = Submitted Manuscript Under Review
- AM = Accepted Manuscript
- P = Proof
- VoR = Version of Record
- CVoR = Corrected Version of Record
- EVOR = Enhanced Version of Record

The rationale for choosing these stages was that each provides unique contributions to the content by one or more players — creator, editor, or publisher — in the scholarly publication process. There are a variety of sub-stages that could also be included, but the JAV working group excluded other stages because they did not add substantial value over the previous stage and might be too complicated to clearly identify. For example, once a manuscript is submitted, it might go through a series of revisions and resubmissions prior to being finally accepted. The differences among the first, second, or various other iterations of a paper might not be significant enough to track, and not all papers will go through multiple revisions. Some papers could go directly from author’s original (AO) to accepted manuscript (AM) without any changes, or even straight to version of record (VoR). **While metadata structures for each of the stages were discussed, it was considered out of scope for the initial group.**

Other organizations have also considered the issue of journal article versions and have added their own perspectives. The **Joint Information Systems Committee (JISC)** in the UK conducted a Scoping Study on Repository Version Identification (RIVER) followed by the Version Identification Framework (VIF) project. One important distinction between the **NISO/ALPSP** work and the **JISC** work is that the former focused solely on the issues surrounding journal article versions in the scholarly publication chain, whereas the **JISC** work focused on broader issues of a variety of content objects that primarily resided in repositories, although they often had been submitted for formal publication as well.

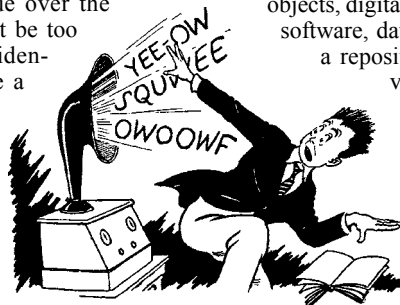
The initial study funded by the **JISC** — The **VERSIONS** (Versions of Eprints – a user Requirements Study and Investigation of the Need for Standards) — was conducted by the **London School of Economics and Political Science** and the **Nereus Consortium** of European research libraries in economics.

The goal of the study was to “address the issues and uncertainties relating to versions of academic papers in digital repositories.” The recommendations included a set of definitions for: Draft, Submitted Version, Accepted Version, Published Version, and Updated Version. These track closely with the **NISO/ALPSP** recommendations and were issued at approximately the same time. The final toolkit from the project ([http://www2.lse.ac.uk/library/versions/VERSIONS\\_Toolkit\\_v1\\_final.pdf](http://www2.lse.ac.uk/library/versions/VERSIONS_Toolkit_v1_final.pdf)) included suggestions for authors and repository managers to improve the identification, use of, and recognition of version terminology.

The **RIVER** ([http://www.jisc.ac.uk/uploaded\\_documents/RIVER%20Final%20Report.pdf](http://www.jisc.ac.uk/uploaded_documents/RIVER%20Final%20Report.pdf)) project developed another recommendation based on a set of use cases that included a variety of content forms, such as learning objects, digital images, wikis, documents, software, data files, or search results in a repository. The project outlined various scenarios where the content could be collocated, might need to be disambiguated, or might need version control. The project, after reviewing the use cases and existing industry practice, put forward a number of data elements that might be used to identify versions of the content in question. The project group recommended some follow-up work, which was done in the subsequent Version Identification Framework (VIF) project.

The **VIF** project (<http://www2.lse.ac.uk/library/vif/Project/index.html>), which built on both the **VERSIONS** and **RIVER** projects, looked at the broad range of content forms that researchers use in their work and how that content is managed and stored, with an eye toward the role of repositories in that process. The project team conducted surveys of how researchers, teachers, students, and others are using digital objects and managing their personal digital resources. The project focused a lot of attention on the workflow issues of creation, revision, dissemination, storage, and especially the issue of how versioning terminology can be integrated into version control. The team then produced a framework for version control, which included information both on metadata that should accompany and be embedded in objects to maintain good version control. The Framework developers identified “Essential Versioning Information” consisting of: Defined Dates, Identifiers, Version Numbering, Version Labels or Taxonomies, and Text Description. Other information was recommended to be embedded within an object including: ID Tags

*continued on page 18*



# What's JAV Got to Do with It? Indicating Versions of Record with CrossMark

by Carol Anne Meyer (CrossRef)

Scholarly content exists in a multi-channel environment. Journal articles are available from authors' Websites, from institutional repositories, from subject archive repositories, as pre-publication manuscripts from publishers Websites, as peer-reviewed, accepted, and copyedited manuscripts on publishers' Websites, and as licensed, redistributed content from aggregator journal vendors. Articles also increasingly live as PDFs on researchers' hard drives or in manuscript management systems.

Just as the text of the document has many homes, so too do bibliographic metadata about that document. There may be a separate secondary record for different versions of the text.

So what happens to content in all of these channels when something important changes? Maybe an author's name was misspelled. Maybe a table was missing a caption. Perhaps a figure was mislabeled. Or an editing error changes the interpretation of the results. The author could have discovered a calculation error. An individual may disavow knowledge of the research and ask to be removed as an author. Or evidence that part of the content was plagiarized could surface. Occasionally, cases of academic fraud require that a paper be retracted.

How, as scholarly publishers and academic librarians can we ensure that all of the con-

sumers of scholarly information have simple, prompt access to important information about status and changes?

**NISO's** recommended Journal Article Versions (JAV)<sup>1</sup> represents one attempt to classify and label the stages of journal articles to provide important information to readers about exactly what it is they are looking at.

Independently, **CrossRef**, a not-for-profit organization for scholarly publishers that made scholarly reference linking a reality, has been working on a new initiative of publishers to clearly label their content. **CrossMark**, which will launch in mid-2011, will provide a way to clearly mark versions of record and communicate information about their current status.

## How will it work?

An article that is part of the **CrossMark** service will sport a distinctive logo — which will be the same across all participating publishers. The presence of the logo tells a reader two things: 1) publisher of this document has made a commitment to label it, maintain it, and communicate any changes that it may undergo; and 2) further information is available about the status of the document. In most cases, the presence of the **CrossMark** logo actually indicates that the document is a version of record, though there may be exceptions due to individual publisher practices, as we will see below.

When a reader sees the **CrossMark** logo, whether on an HTML page or a PDF copy of a document, he or she may click on it. After doing so, a box pops up containing important information: 1) the current status of the document, 2) if this particular copy is being maintained by the publisher, 3) where to find the copy of the document that is being maintained by the publisher (the **CrossRef** DOI link), 4) the version of the document, and 5) additional important publication record information.

In this article, we are mostly concerned with the version of the document, but I will spend a little bit of time discussing the other parts of **CrossMark** to make it clear what the service provides.

**Status** — Most of the time, when a reader clicks on a **CrossMark** logo, the status will be "This document is current." Occasionally and when appropriate, the status will be "An update is available for this document." If that is the case, the **CrossMark** status box will display the **CrossRef** DOI link to the updated document. This feature is especially powerful for PDFs that may have been sitting around on a researcher's hard drive for a considerable time.

**Version** — The **CrossMark** status box will also display the version of a document. Here's where JAV comes into play. The version field that **CrossRef** publishers will indicate as part

of **CrossRef** will have some flexibility, and will not require that they use JAV terminology. However, **CrossRef** will encourage publishers to look at the **NISO** JAV recommendations in creating their version labels. For many, the JAV recommendations will work fine. In the majority of cases, publishers will use the term "Version of Record" for peer-reviewed, published articles.

For other publishers, the JAV statuses may not fit as well, so they may need to use their own terminology. For example, though a few publishers may make corrections to articles *in situ* by replacing the previous version, it is not a common practice, so the JAV term "Corrected Version of Record" will probably not be necessary for most publishers. We expect a more common scenario to be an additional entity with its own metadata that is the correction to the original "Version of Record." In order for the scholarly record to remain clear, it may be important for the publisher to retain the (now) incorrect version, while clearly labeling it as such.

As an organization based on the network advantages of the DOI standard, and as a long-time supporter of **NISO**, it almost goes without saying that **CrossRef** would recommend that its members adopt the JAV terminology. Yet, we are also practical, and we understand that not every publisher's workflow fits nicely into the JAV definitions.

A more fundamentally important reason why **CrossRef** is not "hardwiring" JAV terminology into **CrossMark** is that **CrossMark** may be used for content other than journals. Of the more than 45 million DOIs assigned at **CrossRef**, over 13 percent of them now come from books and book chapters (including reference entries), conference proceedings, components, database records, and other non-journal content.

Another important relationship between **CrossMark** and JAV is that, as part of the rules of participation, **CrossMark** logos may not be displayed on pre-acceptance versions. In fact, if a document is not eligible to get a **CrossRef** DOI, the purpose of which is to ensure persistent linking, then it may not have a **CrossMark** either. Publishers who make Accepted Manuscripts or Proofs (both JAV terms) available publicly may wish to display **CrossMark** logos on those, and use those terms in the **CrossMark** Version Field.

**Publication Record** — **CrossMark** can also communicate valuable publication record information about the document to which it applies. Though not directly related to the



Are these Two Versions the Same?  
from page 16

and Properties Fields, a Cover Sheet, Filename, and a Watermark.

While these different structures and approaches have different specifics, the core of the problem remains clear: users have to be able to understand the differences between different instances of what appears to be the same content. At their core, the different structures proposed by the **JISC** and the **NISO/ALPSP** recommendations are not so dissimilar as to require much distinction. Where the **JISC** has pushed forward is in developing a more robust system, extending beyond journal articles into other content forms. The **VIF** project has also proposed a more robust metadata framework, which will be particularly useful. As with all standards projects, pushing the adoption of these recommendations in the community and making them *lingua franca* among the scholars who use these content forms are the biggest challenges. Hopefully, as more attention is focused on the issue, researchers and systems managers will adopt the existing terminology and require the necessary metadata to ensure clarity. 🐾