

Version Identification Framework: requirements and proposed resolutions

JE Brace¹, PL Cave², CM Jones³, D Puplett¹

1. LSE Library

2. University of Leeds

3. Science and Technology Facilities Council

Abstract

The project aims to enhance repository content and functionality in the field of version identification by disseminating:

- a comprehensive understanding of the issues surrounding version identification,
- practical help, advice and tools for repository managers and creators of repository content,
- recommendations for software development community to consider in future development work.

Background

The Version Identification Framework project is a collaboration between the London School of Economics and Political Science, the University of Leeds, the Science and Technology Facilities Council and the Erasmus University in Rotterdam. It is funded through the JISC Repository and Preservation programme and runs from July 2007 to May 2008. It investigated academic and information professional attitudes to versioning across different types of material and has produced a web based information resource to assist in version identification which can be used by authors and institutional repository managers to provide benefits to the end user searching for material. The scope of the project covered digital objects in the following material types: text, images, moving images and individual data instances. Learning objects and large dataset repositories were not in scope.

This project builds in part on work done by RIVER¹; VERSIONS², the NISO/ALSPS working group on Journal Article Versions³ and the Functional Requirements for Bibliographic Records⁴. It is also liaising with the Resourcing Identifiers Interoperability within Repositories (RIDIR) project⁵ as persistent identifiers are a facet of identification.

Version identification is becoming increasingly important as the same, or similar, material is available from many different sources, an important new source being institutional repositories. As the institutional repositories and their contents mature, these issues will become more and more evident. The end user now has a greater problem in being able to identify the differences, if any exist, between different sources. The project aimed to ease this confusion by highlighting the important versioning information that needs to be captured within the object or its metadata.

Stakeholders in the versioning arena

There are five groups of stakeholders identified in the context of this project:

- **Content Creators**, including authors, who produce the material and are in control of the production & retention of the original versions.
- **Repository Managers** who set policy for the content of the institutional repository and manage the system applying these policies.

¹ http://www.jisc.ac.uk/uploaded_documents/RIVER%20Final%20Report.pdf

² <http://www.lse.ac.uk/library/versions/index.html>

³ http://www.niso.org/committees/Journal_versioning/JournalVer_comm.html

⁴ <http://www.ifla.org/VII/s13/frbr/frbr.htm>

⁵ <http://www.hull.ac.uk/ridir/>

- **Software producers and those responsible for creating services on top of repositories** who provide the functionality and common standards for the repositories
- **Wider information environment community** who are interested in strategy & policy within the academic sector.
- **End users** who need to be able to understand their search results and how they relate to each other.

Key to the aims of the project is the end user. The project's focus is to make information about digital objects transparent so that a researcher is able to trust that they have found the right information for their purpose.

Definition of version

The project team used the following definition of a version:

A 'version' is a digital object (in whatever format) that exists in time and place and has a context within a larger body of work.

Following extensive discussion within the team, we agreed that the decision as to what was a version of what and how is very dependant on the designated community who are going to use the information and so the framework does not attempt to suggest what types of relationship might exist, only that sufficient information about the object and relevant relationships should be captured.

To highlight this, see below for some illustrative examples.

Are a pre-publication text document and the published journal article versions of each other?	Are audio recordings of the same piece of music played by different orchestras at different times and in different places versions of each other?
Are a video of an event, a photograph taken at the event and a file containing the presentation given at the event versions of each other?	Are a digitised 18th century map of Hertfordshire and a present day map of the same place versions of each other?

The answers are probably “yes” and “no” depending on who you are!

Key results of the VIF surveys

To get an understanding of the awareness of versioning issues within the wider community and to test the waters for potential solutions, the project team performed two surveys using the BOS⁶ software. These were aimed at academics (both authors & end users) and information professionals to capture the views of two of the stakeholder groups.

The main findings of the survey were:

- Identification of versions is an accepted problem, both for academics & information professionals
- Most academics are happy with the way they organise versions in their personal folders

⁶ <http://www.survey.bris.ac.uk/>

- 60% of academics think that only the “latest” version should be made available but 78% of information professionals are interested in repositories storing all available versions
- No one approach to versioning (taxonomies, chronological or other) will provide a complete solution
- There are no solutions which better fit particular types of material.

The Framework

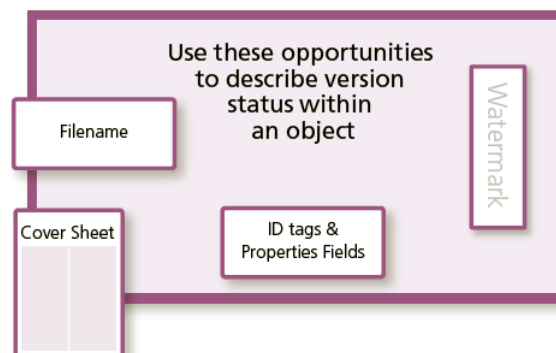
The online Framework addresses versioning issues for the three main stakeholder groups: content creators, repository management and software developers by providing guidance and by making recommendations. Underpinning these recommendations is the recognition of five key pieces of versioning information and some object solutions. It is available at <http://www.lse.ac.uk/library/vif>.

Key versioning information

There are five key pieces of information, in addition to the title & creator, that will enable distinctions to be made between versions of digital objects:

- Defined dates related to the object
- All relevant identifier(s) for the object
- Version numbering
- Version labels or taxonomies
- Text description

It must be noted that most of these pieces of information are known to the content creator but are not necessarily accessible after deposit and not all are appropriate to every type of material. Where ever possible these key pieces of versioning information should be associated with the object as well as in metadata and the framework provides extensive and detailed guidance on how where and when to make this information transparent. The diagram below outlines some of the opportunities to do so within the object itself.



Repository Management recommendations

Repository managers can set the local environment for versioning issues and can influence the wider landscape by adopting common standards. VIF recommends that Repository Managers:

1. are clear about what the repository is for and clarify the versioning requirements within software (upgrading software if necessary)
2. formulate wider strategy; set and promote clear policies
3. understand the FRBR model and how it relates to the software

4. think about how to manage different types of object and ensure at least one object solution is embedded consistently within the repository
5. include some version information in metadata as a minimum although use full DC application profiles if possible
6. ensure versioning is part of the ingest process and
7. advocate good versioning practice to content creators.

Software developers recommendations

Software developers for repositories and other services can shape the way versioning is done by the conceptual design and functionality of their systems. VIF recommends that software developers:

1. ensure the system can cope with and link more than one version – the different ways of linking are not necessarily mutually exclusive
2. use a FRBRised approach – it is consistent with application profiles and also allows the end user to understand relationships between versions
3. ensure the key five pieces of version information are available for capture and display and support the object versioning solutions
4. enhance deposit workflow to assist in upfront capture of versioning information
5. facilitate searching for similar titles within the local repository and other external repositories to prevent duplication
6. support the harvest/export of metadata using application profiles and
7. broaden the language and interface used in repository software to work better with multimedia.

Content creators recommendations

The seven recommendations for content creators are about clarity of versioning while the object is being created before it is deposited in repository. Content creators should:

1. ensure all versions of work state the author, title and date last changed
2. keep track of which versions have been made publicly available and where
3. ensure filename is meaningful and relevant to version
4. consider using version numbering
5. make use of any opportunity to record version information within the object
6. give repository staff all versioning information available at deposit stage
7. give details to all repository staff, if work is deposited in more than one repository.

Conclusions

Version identification is an important area for the end users of repositories and federated search engines. As content grows it is becoming more likely that the same piece of multi-authored work may be found in the repositories of each of the authors, leading to the need to be clearer about what is held and how it might differ from other versions located by the same search.

The VIF recommendations aim to assist this area by providing practical solutions which can be adopted by our three key audiences.