

## **Version Identification Framework: requirements and proposed resolutions**

J Brace<sup>1</sup>, PL Cave<sup>2</sup>, CM Jones<sup>3</sup>, D Puplett<sup>1</sup>

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 is investigating academic and information professional attitudes to versioning across different types of material and aims to produce a web based information resource to assist in version identification which can be adopted by authors and institutional repository managers to provide benefits to the end user searching for material.

Version identification is becoming increasingly important as the same 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 aims to ease this confusion by highlighting the important information that needs to be captured within the object or its metadata.

This project builds on work done by RIVER<sup>1</sup> ; VERSIONS<sup>2</sup> , the NISO/ALSPS working group on Journal Article Versions<sup>3</sup> and the Functional Requirements for Bibliographic Records<sup>4</sup>. It is also liaising with the Resourcing Identifiers Interoperability within Repositories (RIDIR) project<sup>5</sup> as persistent identifiers are a facet of identification.

The scope of the project covers digital objects in the following material types: text, images, moving images and individual data instances. Learning objects and large dataset repositories are not in scope. The project is exploring the versioning issues,

---

<sup>1</sup> [http://www.jisc.ac.uk/uploaded\\_documents/RIVER%20Final%20Report.pdf](http://www.jisc.ac.uk/uploaded_documents/RIVER%20Final%20Report.pdf)

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

<sup>3</sup> [http://www.niso.org/committees/Journal\\_versioning/JournalVer\\_comm.html](http://www.niso.org/committees/Journal_versioning/JournalVer_comm.html)

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

<sup>5</sup> <http://www.hull.ac.uk/ridir/>

will promote awareness of these within the repository management community and is creating a framework to provide practical tools and guidance to repository managers and content creators and recommendations to the software development community.

### **Stakeholders in the versioning arena**

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

- **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.
- **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 decided on the following definition of versions:

*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 18<sup>th</sup> 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!**  
However there are key pieces of information that will enable a distinction to be made between versions; not all of these are appropriate to every type of material:

- Attributes/labels describing the object’s place in a workflow
- Any relevant dates in the creation/deposition workflow
- Any relevant identifier for the object
- The location where the object is stored
- What has changed since the last version
- The relationship between objects

It must be noted that most of these pieces of information are known to the content creator but are not necessarily accessible after deposit.

### **Survey outlines and results**

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 BOS<sup>6</sup> 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 personally
- 60% of academics think that only the “latest” version should be made available.
- 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.

### **Framework outputs**

The framework is intending to address the following issues and will report on the key findings in each of these.

- Policy issues to consider
- Practical suggestions for version description
- The use of DC application profiles
- Suggestions for functional changes
- Publicity material for end users (authors & searchers)

The information resource underpinning the framework has three sections aimed at our different audiences: software developers; repository managers and content creators. These sections will provide practical advice and guidance by April 2008.

### **Conclusions**

---

<sup>6</sup> <http://www.survey.bris.ac.uk/>

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.

VIF is aiming to assist this area by providing practical solutions to the issues raised specifically tailored to our key audiences.