Versions in the lifecycle of academic papers - user requirements and guidelines for digital repositories

Frances Shipsey, VERSIONS Project

CERN Workshop on Innovations in Scholarly Communication (OAI4)
Geneva, 21 October 2005
Outline

• Issues relating to versions in digital repositories
• The VERSIONS Project
• Other initiatives on version control
• Managing versions in other contexts – examples of approaches
• Possible approaches for versions of eprints
Issues relating to versions

• Copyright
  • generally cannot use publisher versions of PDFs
  • national policies can vary, eg in Netherlands can use pre-1997 author versions without seeking publisher permission
  • slight variations between final author manuscript and published version
  • publisher conducts final corrections via proofs on paper

• Collaboration and co-authoring
  • naturally leads to proliferation of versions especially when collaboration is inter-institutional – paper is placed in working paper series of each institution and on personal websites
  • more effort needed by all authors to keep versions in line with each other and to obtain final version copies from the corresponding author
Issues …

• Administration of different versions – objects and metadata
  • Authors may already be posting their papers in a number of different locations – personal website, departmental website or working paper series, subject-based working paper series, subject collections such as RePEc, university database of research outputs, and now institutional repositories
  • Keeping versions up to date in all of these locations is burdensome – authors may take pragmatic decision to update only personal website
  • Role for libraries and repository managers – streamlining processes

• Informal communication and time lag in publication
  • 3-year wait for publication of journal article is common in economics
  • Preprint and revised versions up to final author postprint are important phases in the communications cycle
Issues …

• Development of a unit of intellectual content
  • Herbert Van de Sompel (et al) (2004) Rethinking scholarly communication: building the system that scholars deserve, D-Lib 10 (9) – ‘Recording the dynamics of scholarship’
  • How many versions to keep
  • Role of comments and discussion in the lifecycle of a paper
  • Journal article definitive and final version, but working papers and technical papers also definitive, can reach a final settled version and have a distinct existence in own right

• Quality issues - peer review
  • Labelling different versions to indicate whether refereed, eg in eprints.org
  • For prominent authors in your field - any version will be worth reading
Issues …

• Authors’ control over versions of research outputs
  • Placing a copy on personal website – authors retain (fairly) full control
  • Posting a copy in a repository – who decides about removing earlier versions, authors, repository managers, or university administration
  • Awareness that once posted on the web, papers cannot easily be withdrawn

• Persistence
  • Pre-prints and postprints commonly cited – possible implications for cited author or citing author if the referenced work changes or disappears
  • Revision vs persistence – a tension

• Impact/visibility and citations
  • Conference, workshop and seminar papers – important for disseminating work quickly and obtaining feedback – but hope for citation of working paper or journal article; hence abstract only conference proceedings

• Digital objects but still mirroring print process
  • Multiple versions – not well linked to each other – still following model of print - snapshot PDF views of evolution of the paper
The VERSIONS Project

• VERSIONS : Versions of Eprints – user Requirements Study and Investigation of the Need for Standards
• Funded by the Joint Information Systems Committee (JISC) under the Digital Repositories Programme
• London School of Economics and Political Science (LSE) - lead partner
• Nereus – consortium of European research libraries specialising in economics – associate partner
• 18-month project – July 2005 to December 2006
A library network
Aims of the VERSIONS Project

• to clarify the position on different versions of academic papers in economics available for deposit in digital repositories, in order to help build trust among academic users of repository content
• to produce a toolkit of guidelines about versions for authors, researchers, librarians and others engaged in maintaining digital repositories
• to propose standards on versions to JISC to inform discussions and negotiations with stakeholders
Focus of the project

• **Economics**
  • Established pattern of using preprints
  • Importance of journals coupled with long lead times for publication
  • Builds on existing experience and partnerships

• **Eprints**
  • Builds on experience of other projects and programmes
  • Not looking at data or other object types

• **Europe**
  • Cultural and IPR differences worth investigating
  • Builds on existing experience and partnerships
Approach of the VERSIONS Project

• User requirements study
  • Talking to authors, researchers, librarians, repository software developers, relevant standards communities, and other stakeholders
  • Attitudes and current practice
  • Interviews, online survey and evaluation of user study

• Publications lists analysis
  • Analysis of publications lists of 70 economists in Economists Online repository, looking at availability of full text
    – By date of publication
    – By country
    – By publisher/self-archiving policy
Guidelines and standards

• Reaching consensus with stakeholders
• Development of guidelines on good practice
• Production of a toolkit for researchers
• Recommendation of standards
Feedback from first set of interviews

- Authors retain many versions of their work
- Most of these are not seen as public versions
- May actively seek to keep some versions out of the public domain or to control their use
  - Very early drafts circulated between co-authors
  - Results that are early or tentative (conference presentations)
  - Evidence of rejected journal submissions
- Delay in publication of peer-reviewed articles contributes to the use of other dissemination outlets
Feedback from first set of interviews

- Posting papers in multiple locations – administrative effort to update in each location
- Collaboration with co-authors requires additional effort to manage versions
- Use of date is crucial; simple way to identify latest version of others’ work would be excellent – a ‘non-obvious task’ at present
- Essential to be able to identify the definitive versions and to point to journal article for citations
- Trade-off between wide dissemination and control over versions
Other initiatives on versions

- NISO/ALPSP Working Group on Versions of Journal Articles
  - Policy paper from February 2005 by Sally Morris
  - Two groups – technical working group and review group
  - Developing use cases

- JISC initiatives
  - Digital Repositories Programme – 21 Projects
  - Scoping Study on Repository Version Identification, commissioned by the JISC Scholarly Communications Group – to report in February 2006
Versions in other contexts

- **Literary criticism**
  - Shakespeare – quarto and folio editions
  - Textual transmission
  - Variorum editions have existed in the print age – new initiatives to produce electronic variorum editions

- **Software development**
  - Concurrent versioning systems used by developers
  - Release numbering systems
  - Divergence of software – local customisation by or for clients

- **Continuous updating**
  - A problem in print environment, eg legal encyclopedia looseleaf publications – earlier versions lost
  - A solution in the fully digital environment? - earlier versions can be retained and compared
    - Cf Wikipedia
Possible approaches

• Labelling
  • Reach consensus on terminology – naming conventions
  • Numbering systems, cf software release
• Describing
  • Author annotations and free-text descriptions in metadata
  • Date
  • Cover sheets – standard template
• Linking
  • MARC21 Linking entry fields (76X-78X)
  • Dublin Core element - relation
  • Dublin Core refinements
    – hasVersion, isVersionOf
• Comparing text
  • Formats
  • Tools
• Signposting
  • To published journal version (for citations)
  • To author-approved latest version of pre-print (for full elaboration of argument, proof, supporting data)
Contacts

• The Library at LSE
  • www.lse.ac.uk/library

• Nereus
  • www.nereus4economics.info

• JISC Digital Repositories Programme
  • http://www.jisc.ac.uk/index.cfm?name=programme_digital_repositories

• VERSIONS Project
  • versions@lse.ac.uk