

| Project Information | | | |
|--|---|-----------------|----------------|
| Project Acronym | RIOJA | | |
| Project Title | Repository Interface for Overlaid Journal Archives | | |
| Start Date | 01 March 2007 | End Date | 15 August 2008 |
| Lead Institution | UCL (University College London) | | |
| Project Director | Paul Ayris | | |
| Project Manager & contact details | Martin Moyle | | |
| Partner Institutions | Cambridge University, Cornell University, Glasgow University, Imperial College London | | |
| Project Web URL | http://www.ucl.ac.uk/ls/rioja | | |
| Programme Name (and number) | Repositories and Preservation April 2006 | | |
| Programme Manager | Amber Thomas | | |

| Document Name | | | |
|-------------------------------------|---|-----------------|--|
| Document Title | Final Report | | |
| Reporting Period | | | |
| Author(s) & project role | Martin Moyle, Project Manager Antony Lewis, Co-Investigator (Technical) | | |
| Date | 10 September 2008 | Filename | |
| URL | http://eprints.ucl.ac.uk/12562 | | |
| Access | | | <input type="checkbox"/> General dissemination |

| Document History | | |
|------------------|----------|-------------------------------------|
| Version | Date | Comments |
| 0a | 25/07/08 | Draft approved by Programme Manager |
| 1a | 10/09/08 | Final draft |
| | | |

RIOJA

RIOJA (Repository Interface to Journal Archives)

Final Report

Martin Moyle
Digital Curation Manager, UCL Library Services, UCL (University College
London)

Antony Lewis
Advanced Fellow, Institute of Astronomy, Cambridge University

September 2008

Table of Contents

| | |
|--|-----------|
| Acknowledgements | 4 |
| Executive summary | 5 |
| Background | 6 |
| Aims and objectives | 7 |
| Methodology | 8 |
| Implementation | 9 |
| Outputs and results 1: overview | 11 |
| Outputs and results 2: illustration of the toolkit | 12 |
| Outcomes | 16 |
| Conclusions | 16 |
| Implications | 17 |
| References | 17 |
| Appendix A: the RIOJA toolkit: technical introduction | 18 |
| Appendix B: the RIOJA APIs | 19 |
| getRepositoryInfo: Get information about the repository | 19 |
| validateAuthor: Author validation | 19 |
| getMetadata: Metadata exchange | 20 |
| setStatus: Journal status change notification | 22 |
| getStatus: Get journal status from repository | 22 |
| getTrackbacks: Get repository trackbacks | 22 |
| getStatistics: Get repository statistics | 23 |
| getCurrentPaperInfo: Get current state of paper | 23 |
| getJournalInfo: Get information about the journal | 24 |
| getStatus: Get submission status within the journal | 24 |
| submit: Submit to journal from repository | 25 |

Acknowledgements

The JISC (Joint Information Systems Committee, UK) provided funding towards the RIOJA project.

The following members of the Project Team contributed their time:

- Martin Moyle (UCL Library Services, Principal Investigator)
- Dr Antony Lewis (Institute of Astronomy, Cambridge, Co-investigator, technical work packages)
- Dr Sarah Bridle (UCL Physics and Astronomy, Co-investigator, supporting work packages)

Funded participants in the project:

- Dr Panayiota Polydoratou (UCL Library Services, Researcher, 0.5 FTE)
- Dr Simeon Warner (Cornell - arXiv implementation)
- Ramkumar Nandakumar (MetaOme software, OJS customisation)

The following individuals also made significant contributions to the work of the project:

- Dr Paul Ayriss (Director of UCL Library Services and UCL Copyright Officer, Project Director)
- Dr David Prosser (Director, SPARC-Europe)
- Professor David Nicholas (Director of UCL SLAIS, Project Evaluator)
- Professor Andrew Jaffe (Imperial College, London)
- Dr Martin Hendry (University of Glasgow)
- Dr Oya Rieger (Cornell University).
- Alec Smecher (Public Knowledge Project).

Executive Summary

RIOJA (Repository Interface to Overlaid Journal Archives) was a 18-month partnership between UCL (University College London), Imperial College London, and the Universities of Glasgow, Cambridge and Cornell. The project worked with the Astrophysics community to investigate aspects of overlay journals. For the purposes of the project, an overlay journal was defined as *a quality-assured journal whose content is deposited to and resides in one or more open access repositories.*

The project had both technical aims and supporting, non-technical aims. The primary technical deliverable from the project was a toolkit for the creation and maintenance of overlay journals. The toolkit supports the exchange of data between a repository and a piece of journal software. It supports functions such as author validation, metadata extraction from the source repository, and submission tracking. The toolkit is platform-neutral and could, in theory, be employed by any journal using content from any number of repositories, in any discipline. The project also implemented a demonstrator overlay journal, applying the RIOJA toolkit to the arXiv subject repository, and a demonstrator implementation of the RIOJA tool for GNU EPrints.

Aside from creating the demonstrator and its underlying tools, the project aimed to test the acceptability and feasibility of the overlay model. First, a large-scale survey of the Astrophysics community was undertaken. The survey collected data about research and publishing practices within this community, and probed its reaction to the principle of overlay publishing. Second, the views of editors and publishers in this discipline were sought through interviews. These views were added to findings from the literature and summarised in a more general report on issues around the sustainability of an overlay journal.

The survey confirmed the everyday importance of the arXiv repository in the working lives of astrophysics researchers. Moreover, the project found that researchers are, in general (and with very little variation between those with different first languages, career lengths and other demographics), sympathetic to the overlay model. Their main concerns about the model were that the long-term accessibility of the research material should be guaranteed - surprising, perhaps, in such a fast-moving, repository-dependent discipline - and that the process of quality certification should be robust. Researchers' career concerns also informed their reaction to the overlay model, and it was clear that to attract submissions, an arXiv-overlay journal would need to be able to demonstrate academic acceptability and a substantial readership. All of these concerns are generic issues, which would be faced by any new journal whether or not overlaid on repository-housed content.

In the interviews, publishers and editors showed a certain willingness in principle to experiment with new models, but not to lead the academic community in this respect. The impression received was that change in this sector, particularly within the established journals managed by commercial and professional society-based publishers, is generally driven by the consumer. (Of course, these interviews were largely informal, and cannot be said to be in any way representative of the publishing community.) Meanwhile, ascertaining the detailed costs of the various parts of a publishing operation proved to be beyond the reach of the project team - unsurprisingly, given the element of commercial competition in the field, and the very substantial range of journal business models in use - although the interviews and the literature did support the project's initial contention that, in general, the costs of parts of the publishing process, notably submission and storage (and, in cases where responsibility lies with the repository, archiving and digital preservation), could be significantly reduced by the incorporation of repository overlay workflows.

The toolkit, demonstrator and supporting reports, together with copies of all the papers and presentations prepared by the project team, are available through the RIOJA Web site.

Background

The project was initiated by its academic partners, researchers in Astrophysics who expressed dissatisfaction with the machinery of formal publication and the costs to institutions of journal subscriptions and page charges. Through discussion with their peers, they had arrived at a very clear concept of an open access 'journal' based on the quality-stamping of content submitted to, stored in and publicly available through the arXiv repository. The UCL Library Services partners recognised this concept as that of an overlay journal - discussed for some time in the literature, but with few exemplars to date. The primary goals of the project were to develop the first openly available overlay toolkit, and to build an implementation for arXiv to demonstrate the overlay concept.

Modern Astrophysics research offered an interesting environment in which to put an overlay model to the test, because of the ever-growing importance of the arXiv repository to the discipline. Anecdotal evidence suggested not only that depositing papers with arXiv is the norm for this community, but also that arXiv alone satisfies the current awareness needs of most astrophysicists. The project team undertook a survey to test these assumptions and to learn more about the information-seeking habits of astrophysicists and their requirements from a journal.

In the RIOJA model, a journal 'overlaid' onto an open access repository such as arXiv adds quality assurance (whether through peer review or other means) to papers deposited to and stored in the repository. The main function of the journal is to guide researchers to "accepted" papers - those awarded its quality stamp. The journal achieves this guidance through its Web site, through a presence in third party indexing services, and at the repository or repositories on which it is overlaid, by permitting papers to be quality-marked. Accepted papers continue to reside in the repository, where the accepted version (and only the accepted version, notwithstanding that earlier versions and versions later updated with corrections and annotations may also be held in the repository) carries details within its metadata of its acceptance by the overlay journal as an indicator of academic merit. The RIOJA toolkit would facilitate the automated exchange of data between journal software and repositories.

It was felt that the RIOJA overlay model could help to deliver fast certification of research, with much lower costs than those associated with the traditional publication model, and without publication delays, re-formatting requirements, and other unpopular (certainly with some authors in Astrophysics) trappings of the traditional publication model. Given a serviceable repository, a good editorial board, and a willing community of depositors, readers and reviewers, the RIOJA toolkit could potentially enable quality-assured open access publishing at minimal cost. Overlay journals offer an opportunity for maximising the efficiency of research repositories and for increasing return on the academic and research funding sectors' increasing investment in building a corpus of repository content.

Aims and Objectives

The original aims of the RIOJA project, in detail, were as follows:

1. To build the RIOJA tool, a generic module enabling interoperability between journal software and public repositories in support of the overlay of quality certification.
2. To implement the RIOJA interoperability tool for the arXiv subject repository.
3. To construct a demonstrator journal, incorporating the RIOJA tool, illustrating interaction between arXiv and the DPubS software.
4. To define the ideal functional requirements of a community-led journal in Astrophysics and Cosmology, and to implement these in the demonstration journal as far as is feasible.
5. To identify factors critical to the successful academic take-up of a journal founded on the principle of overlaid quality certification in the field of Astrophysics and Cosmology.
6. To recommend a Digital Preservation strategy for content accepted for an arXiv overlay journal, supported by life-cycle costing techniques.
7. To identify a continuation plan for the journal, including a fully-costed business model of proven acceptability to the Astrophysics and Cosmology community.
8. To disseminate widely the outputs and outcomes of the project.

There were some changes to these aims in the course of the project.

The core aims, 1-3, remained essentially intact, although PKP's OJS software was used for the demonstrator instead of DPubS (it transpired at the beginning of the project that DPubS's editorial and review functions were still in development, whereas OJS offered a more mature platform for experimentation). The second aim was augmented to encompass an implementation of some of the RIOJA APIs for GNU EPrints, in addition to the arXiv implementation which was initially envisaged.

Aims 4,5 and 8 remained unchanged.

Aims 6 and 7 were adjusted in the course of the project. The long-term accessibility of its archive is clearly important to the sustainability of any journal, and this implies that a strategy and a budget for digital preservation should be in place. However, lifecycle costing techniques (Aim 6) could not be applied: these depend on real data, and the overlay model is still essentially at the 'conceptual' stage. Aim 7 also proved to be rather ambitious, in that hard data on costings was not readily available to the project researcher.

With the agreement of the funders, therefore, Aims 6 and 7 were conflated into the single aim of producing a single, general report on aspects of the costs and sustainability of a repository-overlay journal model. This report was driven by a review of the literature and supplemented by relevant findings from the survey and from the discussions with editors and publishers.

Methodology

The RIOJA project aimed to create a generic interoperability toolkit to enable the overlay of certification onto papers housed in subject repositories. The project also sought to create a demonstrator overlay journal, using the arXiv repository and PKP's OJS software, as an implementation of the toolkit; and a modification of GNU EPrints also to create application-level support for the RIOJA APIs. A supporting strand of non-technical work also took place to test the reaction of the astrophysics community, including its editors and publishers, to the overlay journal concept, and to try to determine the factors which would contribute to the sustainability of such a model.

The initial requirements for the technical work were shaped through discussions within the Astrophysics community through an open bulletin board (CosmoCoffee). This allowed a set of APIs to be scoped: some for implementation by a journal, some for implementation by a repository, with both optional and desirable APIs on either side. They were clarified and refined through iterative discussion within the project team; new optional APIs, such as trackback support, were agreed at this stage. A full API specification was then created and published through the project Web site.

To create the demonstrator - a dummy overlay journal using the RIOJA APIs to communicate with arXiv - it was necessary to ascertain the possibilities for repository-side implementation, and associated timescales, from the arXiv team. The scope of the arXiv implementation helped to define the requirements for OJS customisation, which was also informed by feedback from the survey (for instance, the respondents showed an attachment to copy-editing which had not initially been anticipated, but which needed to be reflected in the workflows of the demonstrator). Development work on OJS was then outsourced: a freelance team worked to a schedule of deliverables and payments drawn up and monitored by the technical Co-investigator; early release candidates were tested by the academic members of the project team and any outstanding issues fed back to the developers in an iterative process.

The modification of GNU EPrints to showcase the RIOJA repository APIs in conjunction with OJS, an aim agreed relatively late in the course of the project, was managed along the same lines as the initial OJS customisation.

The survey of Astrophysics and Cosmology researchers took the form of an online questionnaire survey, targeting scientists in the international top 100 academic and other institutions in these disciplines. The survey design involved wide consultation, not least with the project evaluator and his members of Research Group, and several stages of refinement. The scientists were individually contacted with an introduction to the project and a request to participate. The questionnaire survey was supplemented by 6 semi-structured discussions with publishers and members of editorial boards, chosen from commercial publishers and learned societies, including both open access and toll access publishers. A literature review was also undertaken.

In constructing a dissemination plan, the RIOJA team recognised that the project had a wide range of potential stakeholders - among them researchers in astrophysics and other arXiv users, the publishing community, libraries, and research funders. The project dissemination was designed to follow a straightforward path of initial awareness-raising, interim results, and final deliverables, but care was taken not to restrict exposure of the work of the project to the library and JISC communities.

Implementation

The arXiv team at Cornell implemented a RIOJA interface to arXiv incorporating the RIOJA repository APIs for author validation and metadata exchange. It became apparent that arXiv did not issue a stable identifier immediately on receipt of a deposit, which would have prevented the immediate submission of an arXiv-deposited paper to an overlay journal. The arXiv team already had development work planned in this respect, and the issue was resolved towards the end of the RIOJA project.

The work on customisation of OJS and EPrints was outsourced to a freelance company. The project team felt that the project budget did not stretch to a full-time developer, and neither did the quantity of development work warrant it. The technical Co-investigator had a track record of overseeing successful outsourced development projects. However, progress on these aspects of the project was much slower than was originally anticipated. This was largely due to the unfamiliarity of the freelancers with the academic setting, which appeared to make it difficult for them to interpret the team's requirements. The Project Plan had anticipated slippage here to some extent, but in fact it was only possible to sign off the demonstrator at a relatively late stage in the project.

Three issues around standards and interoperability arose in the course of the development work.

Firstly, use of OAI-PMH was rejected. There were several reasons for this: OAI is designed for metadata harvesting, not as a dynamic, immediate-response API to be used interactively. The RIOJA model allows for the extraction of information before a paper is public on a repository, that is, before such data would be available through the repository's OAI interface. RIOJA needs to track the different versions of papers carefully, and detailed version information is out of the scope of OAI-PMH. The overlay scenario requires a well-defined format for text and equations in titles and abstracts, so that they can display the same way on the journal and repository without further editing; OAI's flexibility means that it is vague in this regard. Finally, RIOJA requires several other functions in addition to metadata extraction.

Later, the adoption of SWORD, which had not been commissioned when work on RIOJA began, was considered. There is clearly some commonality between the two - in the RIOJA APIs, a repository 'deposits' metadata to a journal - but RIOJA as specified goes further than SWORD in managing post-deposit exchanges of data between the journal and the repository (eg updates from the journal to the repository about the 'acceptance status' of the deposited paper). The relationship between SWORD and RIOJA is possibly worth further investigation in future.

Finally, the RIOJA specification process also highlighted the desirability of a names registry for authors.

The implementation of the questionnaire survey was hampered by the laborious nature of author identification. Researchers' contact details are often not easily discovered in institutional Web sites; moreover, while an institution's astrophysics researchers know who they are, the institutional Web site usually (with good reason) does not classify its staff to such a helpful level of detail. The commitment to making individual contact with the sample also slowed down the process. However, it is felt that the good response - 17%, 683 respondents - was at least in part down to the decision to target individuals rather than simply to broadcast the survey.

The survey was also hampered by the fact that it necessarily (because of the overall project timescales) was circulated before the demonstrator was ready. The absence of a

demonstrator journal left rather a lot to the imagination of the respondent and created some room for interpretation. It is not felt that this invalidates the feedback gathered through the survey, but a demonstrator would nonetheless have been helpful in providing the respondents with a consistent understanding of the overlay model.

The informal interviews with editors and publishers provided some interesting anecdote, discussion and pointers; but they were of limited use: it was known at the outset of the project that limitations of time and human resource meant that it would not be possible to make any systematic survey of this group, and the results could not be said either to be representative of the sector or statistically meaningful. Moreover, for understandable commercial reasons, very little hard data was made available to the researcher.

The project dissemination activities succeeded in finding a wider audience beyond the Library and JISC communities: publishers were particularly well targeted (for instance through the Charleston and EIPub conferences). Members of the astrophysics community were kept up to date on the project through the survey and by follow-up contacts with interested respondents. In an attempt to reach academics from other disciplines, as well as to disseminate to other interested parties, the project team ran its own one-day end-of-project meeting on the theme of 'new publishing models'. In general, it is probably the case that the lucidity of the project dissemination would have been enhanced had the demonstrator journal been available at an earlier stage in the project.

Outputs and results 1: overview

The technical strand of the project succeeded in creating a set of reusable tools to support overlay journals. These may be implemented by any journal software and any repository to support overlay publishing.

Implementations of the tool, both for demonstration and for redistribution, were also developed. A permanent RIOJA interface to arXiv to support author validation and metadata export was implemented. arXiv is now enabled for closer integration with journal publishing software.

Modifications to OJS to support overlay publishing were developed, as were modifications to GNU EPrints 3 for the same purpose. These outputs will facilitate the quick implementation of an overlay model in Astrophysics or any discipline. A demonstration journal, using the arXiv implementation of RIOJA and RIOJA's OJS customisation, has also been made available.

All the RIOJA technical outputs are publicly accessible through the RIOJA Web site. An illustration of the toolkit is given below.

The supporting survey gave a snapshot of the working practices and attitudes of one, very repository-orientated, research community, based on 683 responses from Astrophysicists. The results confirmed the importance of arXiv to Astrophysics researchers. 93% deposit papers into arXiv; 53% access arXiv daily, and another 24% do so weekly; and after arXiv discovery, only 7% always prefer to seek the final published version of a paper. arXiv use is not to the exclusion of other resources: 65% may use journal Web sites to follow up interesting titles/abstracts, alongside arXiv which is used by 610 (89%) for this purpose. 97% of the respondents publish in refereed journals, at an average of 6.5 papers per researcher per year, in titles whose high impact factor, perceived quality, and updates throughout the refereeing process they consider to be important. They were comfortable with the overlay model: 53% were very supportive, and 35% interested; 80% would referee for an overlay journal; 26% were willing to serve in Editorial capacity; 33% would submit papers without hesitation. Their concerns about a hypothetical arXiv-overlay journal were the quality of the accepted papers, the community standing for the title, the robustness of long-term archiving arrangements, and the quality and speed of the peer review process: these are concerns which one might imagine could easily apply to any academic journal, regardless of publishing model.

A full survey report is available through the RIOJA Web site. The project also delivered an up-to-date review of the literature around journal costs and sustainability, also available through the RIOJA Web site.

Outputs and results 2: the toolkit (illustration)

Fig.1 summarises the ways in which the RIOJA APIs support the movement of data between a repository and a journal.

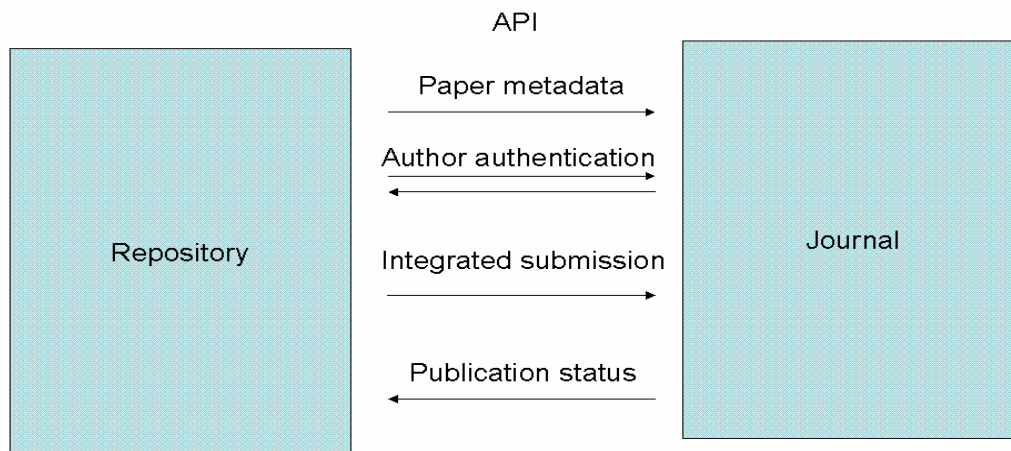


Fig.1 Overview of RIOJA APIs

The following sequence illustrates aspects of the RIOJA demonstrator.

The screenshot shows a web browser window with the URL <http://arxiv.org/abs/0801.0554>. The page header includes 'arXiv.org > astro-ph > arXiv:0801.0554' and a search bar. The main content area displays the title 'Likelihood Analysis of CMB Temperature and Polarization Power Spectra' by Samira Hamimeche and Antony Lewis. The abstract text begins: 'Microwave background temperature and polarization observations are a powerful way to constrain cosmological parameters if the likelihood function can be calculated accurately. The temperature and polarization fields are correlated, partial sky coverage correlates power spectrum estimators at different ℓ , and the likelihood function for a theory spectrum given a set of observed estimators is non-Gaussian. An accurate analysis must model all these properties. Most existing likelihood approximations are good enough for a temperature-only analysis, however they cannot reliably handle a temperature-polarization correlations. We give a new general approximation applicable for correlated Gaussian fields observed on part of the sky. The approximation models the non-Gaussian form exactly in the ideal full-sky limit and is fast to evaluate using a pre-computed covariance matrix and set of power spectrum estimators. We show with simulations that it is good enough to obtain correct results at $\ell \gtrsim 30$ where an exact calculation becomes impossible. We also show that some Gaussian approximations give reliable parameter constraints even though they do not capture the shape of the likelihood function at each ℓ accurately. Finally we test the approximations on simulations with realistically anisotropic noise and asymmetric foreground mask.'

Comments: 35 pages, typos corrected
Subjects: **Astrophysics (astro-ph)**

On the right side, there is a 'Download:' section with links for PostScript, PDF, and Other formats. Below that is a 'References & Citations' section with links for SLAC-SPIRES HEP, NASA ADS, and CiteBase. At the bottom right, there is a '1 [trackback \(?\)](#)' and navigation links: '<< astro-ph >>', 'new | recent | 0801'.

First, a paper is uploaded to the arXiv repository. It is allocated an identifier by the repository (in the example above, the ID is arxiv:0801.0554).

http://arxivjournal.org/index.php/cc/author/submit/1

Step 1. Starting the Submission

Home > User > Author > Submissions > New Submission

1. START 2. PAPER DETAILS 3. KEYWORDS 4. CONFIRMATION

Authors submit items to this journal through this web site in a four-step process. The manuscript must be already be present in a repository. The submitter must be one of the authors and must enter a valid repository identifier and email. If difficulties are encountered in this process, contact [Admin](#) by email or phone for assistance.

Repository Details

Host Repository*

Enter a valid repository identifier.

Repository Id*

Submitter Email*

Journal Section

Select the appropriate section for this submission (see Sections and Policies in [About the Journal](#)).

Section*

Comments for the Editor

Enter text (optional)

* Denotes required field

Journal Help

USER

You are logged in as...
author

- [My Journals](#)
- [My Profile](#)
- [Log Out](#)

JOURNAL CONTENT

Search

All

Browse

- [By Issue](#)
- [By Author](#)
- [By Title](#)
- [Other Journals](#)

INFORMATION

- [For Readers](#)
- [For Authors](#)
- [For Librarians](#)

The identifier is notified to the OJS-based RIOJA demonstrator journal, along with the email address of the submitting author. A validation process takes place: the journal queries arXiv to ensure that the email address is associated with that paper.

le Edit View History Bookmarks Tools Help

http://arxivjournal.org/index.php/cc/author/submit/2?articleId=370

Step 2. Checking Paper Details

1. START 2. PAPER DETAILS 3. KEYWORDS 4. CONFIRMATION

This journal gathers metadata about each submission, according to the standards set by the Open Archives Initiative, which is the emerging online standard for identifying research resources. This enables collective and accurate searching across OAI databases, increasing the likelihood that readers will find submissions such as this one.

Metadata

| | |
|----------------|---|
| Title | Likelihood Analysis of CMB Temperature and Polarization Power Spectra |
| Author List | Samira Hamimeche, Antony Lewis |
| Abstract | Microwave background temperature and polarization observations are a powerful way to constrain cosmological parameters if the likelihood function can be calculated accurately. The temperature and polarization fields are correlated, partial sky coverage correlates power spectrum estimators at different ℓ , and the likelihood function for a theory spectrum given a set of observed estimators is non-Gaussian. An accurate analysis must model all these properties. Most existing likelihood approximations are good enough for a temperature-only analysis, however they cannot reliably handle a temperature-polarization correlations. We give a new general approximation applicable for correlated Gaussian fields observed on part of the sky. The approximation models the non-Gaussian form exactly in the ideal full-sky limit and is fast to evaluate using a pre-computed covariance matrix and set of power spectrum estimators. We show with simulations that it is good enough to obtain correct results at $\ell > \sim 30$ where an exact calculation becomes impossible. We also show that some Gaussian approximations give reliable parameter constraints even though they do not capture the shape of the likelihood function at each ℓ accurately. Finally we test the approximations on simulations with realistically anisotropic noise and asymmetric foreground mask. |
| Repository Id* | arXiv:0801.0554 |
| Categories | astro-ph |
| Repository | 35 pages, typos corrected |
| Comments | |

USER

You are logged in as...
author

- [My Journals](#)
- [My Profile](#)
- [Log Out](#)

JOURNAL CONTENT

Search

All

Browse

- [By Issue](#)
- [By Author](#)
- [By Title](#)
- [Other Journals](#)

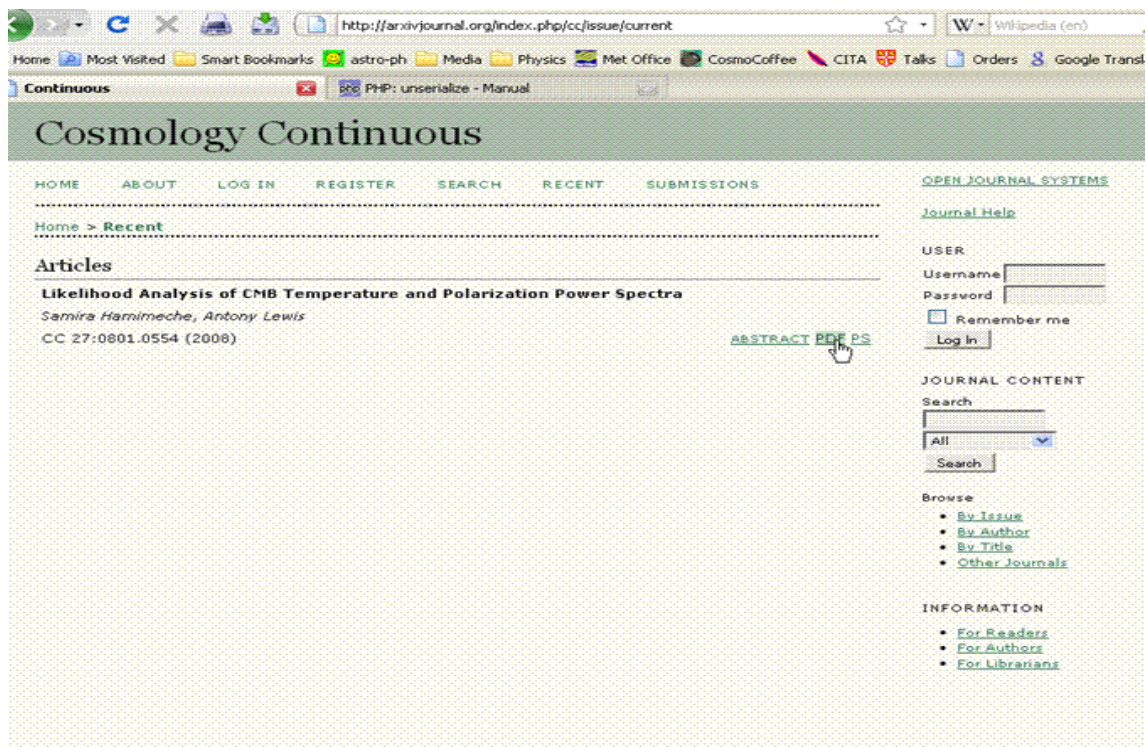
INFORMATION

- [For Readers](#)
- [For Authors](#)
- [For Librarians](#)

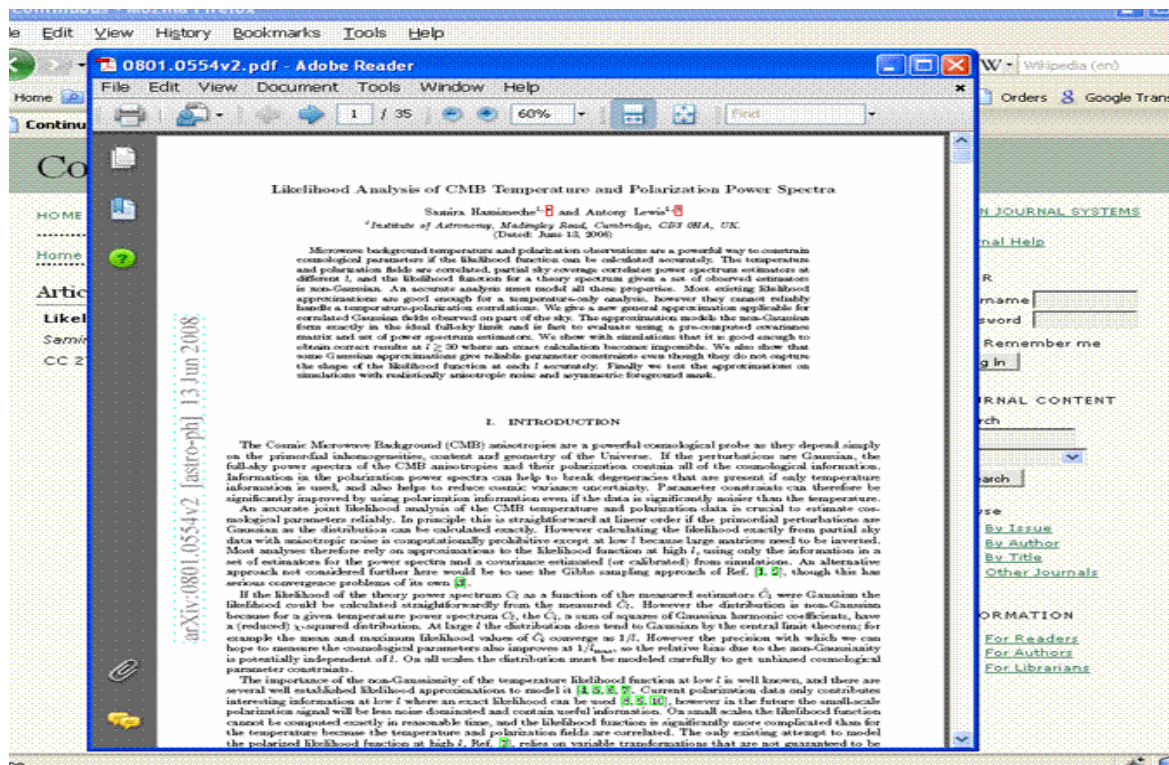
If validation is successful, metadata about the paper is transferred to the journal by the repository.



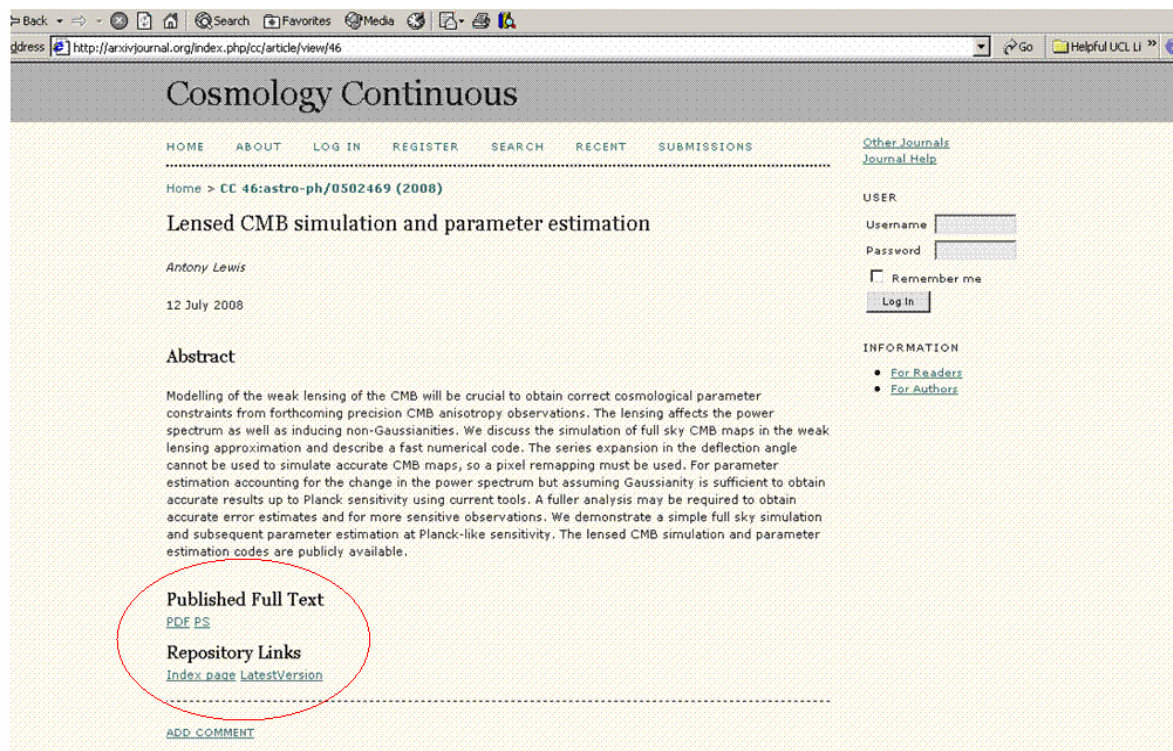
The author may add keywords to the paper, to help the journal to assign an appropriate editor and referees. The paper is then submitted for consideration by the journal.



The process of certification is not managed by the RIOJA tools, but the RIOJA APIs support the tracking of the status of a paper in the submission process with updates from the journal to a repository. If and when the paper is accepted, it is listed on the journal web site with an identifier allocated by the journal. The journal maintains links to the full text.



The full text of the accepted paper resides in the repository, from where it has remained openly and freely available throughout this process.



The demonstrator's Abstract pages support post-publication correction, annotation and update by the author. The journal's identifier for any paper is directly associated only with the accepted version. However, the journal also supplies a link to the latest version of the paper at the repository.

Outcomes

RIOJA's most significant contribution to the wider community was to build the interoperability tool, which will facilitate automated interaction between journal software and any repository in support of overlay journals. To accompany that core deliverable, the team successfully raised awareness of the overlay journal concept, and the general fact of there being alternatives to the 'traditional' publishing model, through its publications, presentations and interviews, and the RIOJA Meeting. The one-to-one discussions with editors and publishers also helped to raise awareness within the scholarly publishing community, as well as providing a very helpful steer to the project team.

The survey report gives a detailed snapshot of the working practices and requirements of one research discipline in which a repository has an integral role; this is of interest to the publishing and repository communities at large, as well as to the community of Astrophysics editors and publishers. The survey itself is in the public domain, and may be repeatable for different communities, with some tailoring. The comments were particularly elucidatory. Again, the project team managed to carry out a relatively large amount of dissemination of the survey results.

The proposed work around cost projections and business analysis for the development and maintenance of a journal founded on overlay certification had to be reined in for reasons of time and general practicability, as noted above. However, the literature review component of this report, supplemented by some discussion of costs and sustainability. is a useful survey, and it may help to inform future repository-overlay undertakings, including those in disciplines other than Astrophysics and with different repositories, as well as contributing to the corpus of general research into publishing processes and costings.

Conclusions

RIOJA confirmed that, in principle, journals founded on the principle of certifying repository content can work. Technically they are feasible: RIOJA has delivered tools which will support them. The important issues facing any start-up overlay journal are generic ones: how to build an audience and a reputation, what model of certification to follow, how to assure the future integrity of the archive, how to ensure a high impact factor, and so on.

It is also possible to assert that the running costs of a journal, especially the costs of content hosting and archiving, can be minimised by using overlay technology. Integration with an overlay journal would also help to maximise the efficiency of the underlying repository or repositories. However - and again, as with any other journal - the running costs will ultimately be determined by the features, functions and services which the journal provides: the apparatus for certification, the volume of submissions and rejections, any supplementary archiving arrangements, supplementary data management, copy-editing, and so on. Overlay could lead to lower costs, but to succeed a journal will have to meet the needs of its community. Where a 'no-frills' approach is tolerated by the readership, costs could indeed be minimal; but the running costs of an overlay journal will vary from discipline to discipline and title to title.

The readiness of the majority of survey respondents to embrace the notion of the overlay journal is encouraging for the prospects of overlay publishing. Having said that, the researchers clearly value some of the services which are added by publishers. They were sanguine about the use of library subscriptions to purchase versions of their work and subsidise those services. An overlay journal in this discipline would have to be a strong offering in order to challenge the few professionally-published titles which already thrive.

Moving away from astrophysics, there are other fields in which an overlay journal model might be attractive, notably in the traditionally cash-starved disciplines of the Arts, Humanities and Social Sciences. A 'no-frills' overlay journal model might be more likely to succeed where the alternative is 'no journal'. It is not clear, however, whether the repository network in these disciplines is sufficiently well-developed to support overlay publishing, and whether there is sufficient willingness and technical expertise in these fields to take the steps necessary to turn public repositories and open source journal software into operational quality-assured and sustainable journals.

Implications

In terms of technical development, it may be rewarding to build on the commonality between RIOJA and SWORD. As noted above, the SWORD deposit API potentially fulfils part of the remit of the RIOJA APIs, while the RIOJA tools also manage post-'deposit' data exchanges between a journal and a repository. It may serve the interests of standardisation for a small project to examine the feasibility of overlay extensions to SWORD based on the requirements identified and specified by RIOJA.

More work could certainly be carried out in embedding the RIOJA model in the publishing community. In astrophysics, this might involve converting an existing journal title to use an overlay model, if a willing publishing partner can be found; or setting up a new journal title with the clear aim of implementing the overlay model and making the most of the community's already active engagement with the arXiv repository. Alternatively it might involve working with a new discipline, again on a start-up title. Further investigation of the prospects of overlay publishing in the traditionally cash-starved Arts and Social Science disciplines in particular might be rewarding.

The RIOJA survey and demonstrator assumed, for the sake of argument, that certification would be by peer review, with other methods of certification explicitly out of scope. The survey, however, elicited a great deal of comment on the subject of peer review itself - speed, quality, efficiency, methods. It is clear that few researchers are entirely satisfied with the way in which peer review is currently managed in this field; it is also clear that there is no consensus within the field on how best to move peer review forward. However, there was a strong body of opinion which called for a more open and community-based approach, taking advantage of social networking technologies. With a willing journal partner, there is an opportunity for further exploration of new mechanisms for review and certification.

References

RIOJA Project Web site: <http://www.ucl.ac.uk/ls/rioja>

Technical outputs

RIOJA Software Products: <http://www.arxivjournal.org/rioja/>

API documentation <http://cosmologist.info/xml/APIs.html>

Demonstrator journal site: <http://www.arxivjournal.org>

Other outputs

Questionnaire: <http://www.ucl.ac.uk/ls/rioja/project-docs/RIOJA-SurveyJune07.pdf>

Survey report: <http://eprints.ucl.ac.uk/5102/>

Costs and sustainability report: <http://eprints.ucl.ac.uk/11927/>

Collected project dissemination outputs: <http://www.ucl.ac.uk/ls/rioja/dissem/>

Appendix A. The RIOJA toolkit: technical introduction

Introduction

The RIOJA toolkit is a collection of XML-base APIs for the exchange of data between digital repositories and journals to facilitate the overlaying of academic journals on separate digital repositories. The repository is assumed to provide the registration, awareness and archiving functions of a journal. The journal is assumed to provide only the certification (peer review) and additional awareness functions. All copies of a paper are stored in the repository, from the original submission to the published version and beyond. The repository can tag papers with their status, so end users can if desired filter papers so that they only see submitted, accepted or published papers as they prefer. The journal tracks different versions of the repository paper, and applies its final "published" quality stamp to one particular "final" paper version. The repository may however allow updates to a paper after publication, allowing easy-access to a corrected version as well as the "published" version.

Typical workflow

- Author submits paper to repository
- Repository optionally offers link to submit to a registered RIOJA-enabled journal, or author manually visits journal website
- Author provides repository ID to the journal (or this is provided by optional API)
- Journal extracts metadata from repository, displays summary to author, and confirms submission
- Journal checks repository paper status at regular intervals; once accepted by the repository the journal continues as below; if the paper instead becomes rejected or withdrawn by the repository the journal automatically rejects the paper.
- Optional API informs the repository that the paper is under consideration by the given journal (so the repository can report the status, and/or prevent submission to other journals)
- Journal proceeds with normal peer-review process: appointing editors, referees, etc, who write reports based on links to the paper on the repository
- Author may update paper on the repository in response to referee/editor feedback and journal re-submission can take place
- Journal finally either accepts or rejects the paper. If rejected optional API informs the repository.
- On paper acceptance the journal assigns a journal ID to the paper, lists it in its lists of accepted papers, optionally notifies the repository of the publication information, and makes metadata available by standard OAI interface
- Paper has page on the journal site giving metadata of published version, version-specific link to the published paper on the repository, and optional links to any future modified versions on the repository.

In addition there are optional APIs allowing the journal to display trackback and download statistics for the paper.

Two general information APIs provide information about the repository or journal. This should make it easy to support additional repositories from journal software by providing just one URL - the URL of the RIOJA API on the repository. Similarly additional journals can be supported by the repository by supplying just the URL of the RIOJA API on the journal.

Appendix B. The RIOJA APIs

The RIOJA APIs operate by HTTP POST or GET calls to a URL which returns an XML file. The input contains an **apiName** parameter that determines which API is being called. The response contains information in various formats depending on the API, plus a standard **resultStatus** element, which is either **OK**, **notImplemented**, **invalidID** or **error**. The optional element **errorMessage** can specify any text error message. Some APIs must be implemented, some are optional (can return **notImplemented**).

The APIs are summarized below along with sample xml. See also the full .xsd schema files:

- [rioja-types.xsd](#) - various general type definitions
- [riojaRepOut.xsd](#) - repository API output
- [riojaJrnlOut.xsd](#) - journal API output

Repository APIs

getRepositoryInfo: Get repository information

- Input: none, e.g.
- `http://someurl.com/rioja?apiName=getRepositoryInfo`
- Output: **repositoryInfo** structure with information about the repository. e.g.
- `<rioja-repository-output>`
- `<resultStatus>OK</resultStatus>`
- `<repositoryInfo>`
- `<repositoryId>arXiv.org</repositoryId>`
- `<repositoryName>arXiv</repositoryName>`
- `<supportedAPI>getRepositoryInfo</supportedAPI>`
- `<supportedAPI>getMetadata</supportedAPI>`
- `<supportedAPI>getStatus</supportedAPI>`
- `<supportedAPI>setStatus</supportedAPI>`
- `<APIVersion>1.0</APIVersion>`
- `<paperIDformat>[0-9]{4,6}\.[0-9]{4,6}</paperIDformat>`
- `<homepage>http://arxiv.org</homepage>`
- `<submitPage>http://arxiv.org/help/submit</submitPage>`
- `<paperIDbase>arXiv:</paperIDbase>`
- `</repositoryInfo>`
- `</rioja-repository-output>`

RIOJA-R1: Author validation

Repositories typically fulfil the preliminary filtering and author validation requirements of a journal. The RIOJA validation API is designed to ensure that someone submitting a repository paper to a journal is actually the original author of the paper. The chosen method is via email validation: the journal author account will typically validate a submitter's email, and hence know that it is valid. The repository will likewise have at least one validated email associated with each submitted paper. For privacy/spam avoidance issues the RIOJA API does not include output of associated email addresses; instead the repository can validate an email and then ask the journal "is this email associated with that paper?". This is sufficient to prevent malicious third-parties or spam-bots submitting a paper of which they are not an author.

validateAuthor Required repository API

Input: comma-separated list of **email** addresses and repository **paperId**. e.g.

- http://someurl.com/rioja?apiName=validateAuthor&paperId=1234.5678&email=email_1,email_2,email_3

Output: **validEmail** list of the email addresses associated with that paper in the repository. e.g.

- `<rioja-repository-output>`
- `<resultStatus>OK</resultStatus>`
- `<validEmail>email_1</validEmail>`
- `<validEmail>email_2</validEmail>`
- `</rioja-repository-output>`

If no paper is associated with the paper the returned list can be empty. **resultStatus** can be **invalidID** if the paper does not exist.

RIOJA-R2: Metadata exchange

Although all copies of the paper reside on the repository, the journal will require the metadata: authors, title, abstract, etc. In general it will need this for several versions of the paper: the version originally submitted to the journal, versions after author-responses to referee reports, the final version accepted by the journal. The published-version metadata will ultimately also be available from the journal via standard [OAI](#) APIs. However any OAI implementation on the repository will not in general provide enough version-specific information for overlaying a journal, so the RIOJA **getMetadata** repository API is provided to supply this information. It should be available from the moment paper submission to the repository is complete, even if the paper is not yet validated or made public on the repository. The metadata includes as **status** tag for each version so that the journal can use to ascertain the status of a paper. The journal should update metadata once **status** of the latest version changes to **public** (typically checked once or twice a day after submission), as repository may not provide valid paper URLs until this point. **version** metadata must include title and author information and at least one paper URL once the **status** is public. Some repositories may make all submissions from validated authors immediately **public**.

The repository can use non-public internal paper IDs for use by RIOJA. In this case the **publicId** metadata element should be set to the public ID as soon as the first version of the paper **status** changes to **public**. Subsequent calls to RIOJA API can use either temporary ID or public ID; they should therefore be distinct. The journal may want to use the publicId as part of its journal paper ID once the paper is accepted.

Title, abstract and comments fields can contain latex (contained in **latex** element), or standard text with optional html-compatible embedded formatting elements **b**, **i**, **u**, **ovl**, **sub**, **sup**, **scp**, **tt** (compatible with the [Crossref](#) standard).

getMetadata Required repository API

- Input: repository **paperId**. e.g.
- <http://someurl.com/rioja?apiName=getMetadata&paperId=astro-ph/0702600>
- Output: **paperMetadata** structure with metadata for all versions of the paper in the repository, links to repository file version, etc. e.g.
- `<rioja-repository-output>`
- `<resultStatus>OK</resultStatus>`
- `<paperMetadata>`
- `<submitDateTime>2002-02-02T14:10:02Z</submitDateTime>`
- `<publicId>astro-ph/0702600</publicId>`
- `<latestPaperURL format="PDF">http://arxiv.org/pdf/astro-ph/0702600</latestPaperURL>`

- `<latestPaperURL format="PS">http://arxiv.org/ps/astro-ph/0702600</latestPaperURL>`
- `<latestPaperURL format="repository-summary-page">http://arxiv.org/abs/astro-ph/0702600</latestPaperURL>`
- `<version number="1" status="public">`
- `<dateTime>2007-02-22T14:10:02Z</dateTime>`
- `<title>The 21cm angular-power spectrum from the dark ages</title>`
- `<abstract><latex>At redshifts $z \gtrsim 30$ neutral hydrogen...</latex></abstract>`
- `<submitter type="author-person" repositoryAuthorID="bcadf518">`
- `<name>Antony Lewis</name>`
- `<affiliation>IoA, Cambridge</affiliation>`
- `<contactPage>http://cosmologist.info/</contactPage>`
- `<repositoryIndexPage>http://arxiv.org/find/astro-ph/1/au:+Lewis_A/0/1/0/all/0/1</repositoryIndexPage>`
- `</submitter>`
- `<author type="author-person" repositoryAuthorID="bcadf518">`
- `<name>Antony Lewis</name>`
- `<affiliation>IoA, Cambridge</affiliation>`
- `<contactPage>http://cosmologist.info/</contactPage>`
- `<repositoryIndexPage>http://arxiv.org/find/astro-ph/1/au:+Lewis_A/0/1/0/all/0/1</repositoryIndexPage>`
- `</author>`
- `<author type="author-person">`
- `<name>Anthony Challinor</name>`
- `<affiliation>IoA, Cambridge</affiliation>`
- `<affiliation>DAMTP, Cambridge</affiliation>`
- `<repositoryIndexPage>http://arxiv.org/find/astro-ph/1/au:+Challinor_A/0/1/0/all/0/1</repositoryIndexPage>`
- `</author>`
- `<comments>29 pages..</comments>`
- `<paperURL format="PDF">http://arxiv.org/pdf/astro-ph/0702600v1</paperURL>`
- `<paperURL format="PS">http://arxiv.org/ps/astro-ph/0702600v1</paperURL>`
- `<paperURL format="repository-summary-page">http://arxiv.org/abs/astro-ph/0702600v1</paperURL>`
- `<keyword>21cm</keyword>`
- `<keyword>dark ages</keyword>`
- `<keyword>CMB</keyword>`
- `<keyword>power spectrum</keyword>`
- `<category>astro-ph</category>`
- `</version>`
- `</paperMetadata>`
- `</rioja-repository-output>`

See the fully qualified [sample XML output file](#).

RIOJA R-01: Journal status change notification

The purpose of this optional API is for the journal to inform the repository of changes in the status of the paper within the journal submission process. If implemented, the repository is automatically informed via this API when the journal submission status changes. To ensure the notification is genuine the repository should call the journal **getStatus** API, which can also be used to obtain more information.

setStatus optional repository API

Input: **journalId** (string), repository **paperId** (string) and **paperVersion** (positiveInteger), and **paperStatus** (submit-status-enum) describing paper status in journal submission process. e.g.

- `http://someurl.com/rioja?apiName=setStatus&paperId=1234.5678&paperVersion=1&journalID=OJAC&paperStatus=submitted`

paperStatus can be one of **accepted**, **published**, **rejected** or **withdrawn**.

Output: **resultStatus**. e.g.

- `<rioja-repository-output>`
- `<resultStatus>OK</resultStatus>`
- `</rioja-repository-output>`

RIOJA R-02: Get journal status from repository

Since a repository may host papers on various different journals, this optional API allows a journal to obtain the status of a paper in other journals, for example to prevent submission to more than one journal.

getStatus optional repository API

Input: repository **paperId** e.g.

- `http://someurl.com/rioja?apiName=getStatus&paperId=1234.5678`

Output: **allJournalStatus** with list of **journalID** and status of the latest repository **paperVersionID** within that journal. Usually only one, but may be re-submitted to another journal after rejection. e.g.

- `<rioja-repository-output>`
- `<resultStatus>OK</resultStatus>`
- `<allJournalStatus>`
- `<journalID>OJAC</journalID>`
- `<paperVersion>2</paperVersion>`
- `<status>submitted</status>`
- `</allJournalStatus>`
- `</rioja-repository-output>`

RIOJA R-03: Get repository trackbacks

This optional repository API allows journals to get information about trackback links to versions of a paper on the repository. The list of trackbacks can be empty.

getTrackbacks optional repository API

Input: repository **paperId**. e.g.

- <http://someurl.com/rioja?apiName=getTrackbacks&paperId=0705.3980>

Output: list of **trackback** information

- `<rioja-repository-output>`
- `<resultStatus>OK</resultStatus>`
- `<trackback>`
- `<name>CosmoCoffee</name>`
- `<URI>http://cosmocoffee.info/viewtopic.php?t=891</URI>`
- `<dateTime>2007-05-29T00:00:00Z</dateTime>`
- `<paperVersion>1</paperVersion>`
- `</trackback>`
- `<trackback>`
- `...`
- `</trackback>`
- `</rioja-repository-output>`

RIOJA R-O4: Get repository statistics

This optional repository API allows journals to get information about paper download statistics from the repository.

getStatistics optional repository API

Input: repository **paperId**

- <http://someurl.com/rioja?apiName=getStatistics&paperId=1234.5678>

Output: **statistics** e.g.

- `<rioja-repository-output>`
- `<resultStatus>OK</resultStatus>`
- `<statistics>`
- `<downloads-week>32</downloads-week>`
- `<downloads-month>123</downloads-month>`
- `<downloads-year>343</downloads-year>`
- `<downloads-total>363</downloads-total>`
- `</statistics>`
- `</rioja-repository-output>`

Get current paper information

This is a wrapper API to return the results of `getStatus`, `getTrackbacks` and `getStatistics` all in one go.

getCurrentPaperInfo optional repository API

Input: repository **paperId** , e.g.

- <http://someurl.com/rioja?apiName=getCurrentPaperInfo&paperId=1234.5678>

Output: statistics, trackback list, journalStatus.

getJournalInfo: Get journal information

Input: journalId, e.g.

- <http://someurl.com/rioja?apiName=getJournalInfo&journalId=OJAC>

Output: **journalInfo** structure with information about the journal. e.g.

- `<rioja-journal-output>`
- `<resultStatus>OK</resultStatus>`
- `<journalInfo>`
- `<journalID>OJAC</journalID>`
- `<journalName>Open Journal of Astrophysics and
Cosmology</journalName>`
- `<ISSN>1234-5678</ISSN>`
- `<OAIurl>http://arxivjournal.org/sample/oai2.php</OAIurl>`
- `<supportedAPI>getJournalInfo</supportedAPI>`
- `<supportedAPI>getStatus</supportedAPI>`
- `<supportedAPI>submit</supportedAPI>`
- `<APIVersion>1.0</APIVersion>`
- `<homepage>http://cosmologist.info/ojs-
2.0/index.php/astrocosm</homepage>`
- `<submitPage>http://cosmologist.info/ojs-
2.0/index.php/astrocosm/user/register</submitPage>`
- `</journalInfo>`
- `</rioja-journal-output>`

Journal APIs

RIOJA-J1: get journal status

This API returns the status of a given repository paper within the journal.

getStatus Required journal API

Input: **repositoryId** and repository **paperId**. e.g.

- <http://someurl.com/rioja?apiName=getStatus&repositoryId=arXiv.org&paperId=1234.5678>

Output: **journalStatus** structure with status in the journal

- `<rioja-journal-output>`
- `<resultStatus>OK</resultStatus>`
- `<journalStatus>`
- `<journalPaperId>1234</journalPaperId>`
- `<journalReference>OJAC 1234.2345
(2008)</journalReference>`
- `<journalPaperURL>http://cosmologist.info/ojs-
2.0/index.php/astrocosm/article/view/1</journalPaperURL>`
- `<repositoryPaperVersion>2</repositoryPaperVersion>`
- `<submitDate>2007-02-02T14:10:02Z</submitDate>`
- `<status>submitted</status>`
- `<submitDetail>withEditor</submitDetail>`
- `</journalStatus>`
- `</rioja-journal-output>`

Note that normally `journalReference` and `journalPaperURL` would not normally be provided unless status is published. Further information about published version can be obtained from journal OAI interface.

RIOJA-JO1: submit from repository

This optional API allows the repository to re-direct to a page on the journal website immediately after repository submission. The API takes the repository paper details and outputs a re-direct page to continue the submission process. Typically the repository would display a list of supported journals after confirming paper submission, and provide Submit button that would call this API and then re-direct to the journal website.

submit Optional journal API

Input: **repositoryId**, repository **paperId**, optional (set of) **email** associated with the paper.
e.g.

- `http://someurl.com/rioja?apiName=submit&repositoryId=arXiv.org&paperId=1234.5678&email=email_1`

Output: **submissionURL** to re-direct to. e.g.

- `<rioja-journal-output>`
- `<resultStatus>OK</resultStatus>`
- `<submissionURL>http://arxivjournal.org/submit.php?author=id-of-email_1&rep=arXiv&paper=1234.2345</submissionURL>`
- `</rioja-journal-output>`