



**Dryad Consortium Board
December 2009 Meeting
London, UK**

Overview

The second meeting of the Dryad Consortium Board was held at the British Library in London, England from 14-15 December 2009. Organizational goals of the meeting included agreement on an interim governance structure, developing a strategy for expansion of the consortium, and development of a sustainability plan for the repository. The partners shared progress on the adoption of data archiving policies and the integration of data submission into the publication process. Finally, the consortium adopted policies for data citation and data identifiers.

Attendees

Journal representatives:

John Allen (*Biological Journal of the Linnean Society*)
Scott Baker (*Journal of Heredity*)
Erica Fleishman (*Conservation Biology*)
Marcel Holyoak (*Ecology Letters*)
William Michener (*Ecological Applications, Ecological Monographs, Ecology, Frontiers in Ecology and the Environment*)
Allen Moore (*Journal of Evolutionary Biology*)
Richard Nichols (*Heredity*)
Mohamed Noor (*Journal of Evolutionary Biology*)
Rod Page (*Systematic Biology*)
Marcy Uyenoyama (*Molecular Biology and Evolution*)
Tim Vines (*Molecular Ecology*)
Peter Wagner (*Journal of Paleontology*)
Michael Whitlock (*The American Naturalist*)

Dryad personnel:

Todd Vision (Executive director, Dryad)
Peggy Schaeffer (Communications coordinator, Dryad)
Ryan Scherle (Repository architect, Dryad)
Kathleen Smith (Director, US National Evolutionary Synthesis Center)

Host:

Sarah Kemmitt (British Library)

Guests

Neil Beagrie (Charles Beagrie Limited)
Julia Chruszcz (Charles Beagrie Limited)
Ed Pentz (CrossRef)
Max Wilkinson (British Library, DataCite)

Executive summary of outcomes

- *Governance*: The Board adopted an interim structure which specifies that journals are Interim Partners, based on the intention to become full Partners; an Executive Committee has responsibility for setting routine repository policy and business goals, approving consortium members, and bringing resolutions to the larger Board for discussion and vote.
- *Sustainability*: The Board heard the results of consultants' work on cost models for Dryad, and provided input on the development of the business plan, an initial version of which is scheduled to be completed by May 2010. The Board discussed strategies for growing the membership of the consortium, public relations to different stakeholder groups, maintaining the emphasis on internationalization, interacting with publishers, and investigating prospective funding sources.
- *Data citation and data identifiers*: The Board voted to assign Digital Object Identifiers to Dryad data packages. the Executive Committee was charged with making a recommendation on the use of identifiers for individual files, and on versioning policy. The Board also approved a framework for a data citation policy, the final form of which will depend on the outcome of the above two decisions.
- *Submission interface*: The Board reviewed the early experience of submission integration with partner journals. Attendees discussed possible improvements to Dryad's data submission process and the information provided to authors.

New representatives

Two new representatives were present at the meeting. Marcel Holyoak, representing *Ecology Letters* (published by Wiley-Blackwell and the CNRS) and Richard Butler, representing *Heredity* (a journal of the Genetics Society published by the Nature Publishing Group). *Ecology Letters* has not formally adopted JDAP yet, but is willing to do so provided other ecology journals do, particularly those of the BES and ESA. *Heredity* has recently adopted a requirement for archiving of data, but is not yet enforcing it. Both journals wish to explore joining the consortium.

Governance

The Executive Committee came to the meeting with a proposal for an interim governance structure that would be in effect while Dryad's formal business structure

remains under development. The proposal defines the Dryad Consortium Board as: "... consist[ing] of a representative from each journal that expects to join Dryad as a Full Partner. Full partnership will consist of one or more of the following criteria: (1) Formal adoption of the Joint Data Archiving Policy, or an equivalent policy requiring submission of data as a condition of publication; (2) Commitment to developing a self-sustainable business model for Dryad; (3) Appointing a representative to the Dryad Consortium Board with full voting authority. Interim Partners of the Dryad Consortium Board are those journals that have been recommended by the Executive Committee on the basis of a good faith intention to ultimately join as a Full Partner. The membership of the Interim Board consists of one representative from each journal designated an Interim Partner. Each Interim Partner has one vote in matters brought to the Board. Other journals are welcome to join meetings of the Board and express their opinion in discussion, but do not have formal voting rights on the Board."

The structure includes an Executive Committee "...of five members who are representatives of Interim Partners, with the Executive Director serving as an *ex officio* member... The executive committee is responsible for setting routine repository policy and business goals, and has the discretion to bring these to discussion and vote by the Board. It is also responsible for overseeing the development a full governance model and long term business plan. The governance model and business plan must be brought to the full Consortium Board for discussion and vote. The executive committee is responsible for developing the criteria for Full Partners and determining when it is appropriate to require that members of the Board meet them."

The interim governance model was approved by those present, effective immediately, and will remain in effect for one year, at which time it must be either renewed or replaced.

Discussions included the need to investigate international incorporation (if/when we go nonprofit), and the desirability of a board of advisors or trustees with more diverse expertise, e.g. legal, financial, technical and not just scientific.

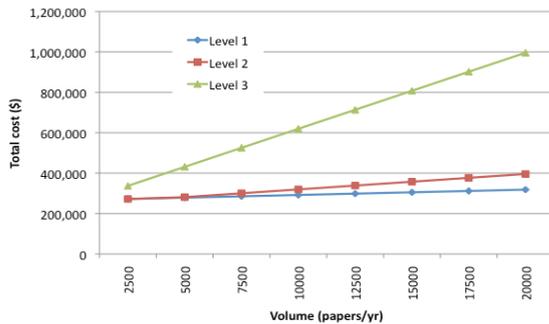
Following the meeting, the Executive Committee solicited nominations and held a vote for the 2010 Executive Committee officers. The new Executive Committee consists of Marcel Holyoak, William Michener, Allen Moore, Mohamed Noor and Michael Whitlock. [Subsequent to the meeting, Michael Whitlock was selected as chair of the Executive Committee].

Sustainability planning

A major goal of the meeting was to get the input of the Board into development of a sustainability plan to ensure the health and longevity of the repository beyond the life of the current project period. The preliminary cost-modeling results from two consultancies were presented, and the Board discussed the outlines of a sustainability plan presented by one of those consultancies.

The first consultant, Lorraine Eakin-Richards, was asked to focus on estimating future repository costs on a per paper basis, with two major factors: paper volume and curation level. Cost models were based on exemplar repositories and organizations for which cost data is available, although no exact parallels for Dryad could be found. The findings suggest that the operating costs of a repository such as

Total operating costs: Effects of curation level and volume

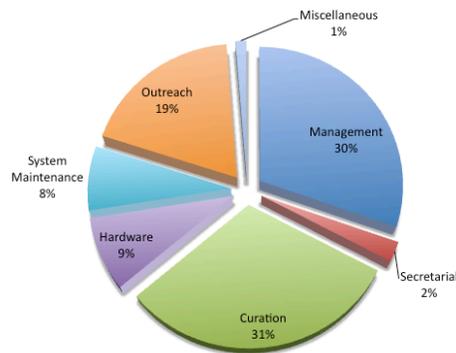


Dryad could be met, with either a low or an intermediate level of data curation, if the costs were recovered at a rate of under \$30 per paper and the volume of papers curated were >12,500 per year. If the sizes of the present partner journals are representative, this would require the consortium to include at least 70-75 journal partners. The report also identified the hidden, but substantial cost savings, provided by the repository

organization having an institutional host (as opposed to being an independent nonprofit).

The second consultancy, Charles Beagrie, Ltd., was charged with recommending an outline for a sustainable business plan for Dryad. Neil Beagrie and Julia Chruszcz presented their draft report to the Board for clarification and feedback. In the report, key issues affecting sustainability are identified, benefits to different stakeholders are enumerated, potential revenue sources assessed and ranked in terms of potential, and risks assessed. Highly ranked revenue sources include subscriptions, in-kind institutional support, fees, and grants. This work continues in Q1 2010 with interviews of journals, societies, publishers and data repositories about the costs of current practices for maintaining Supplementary Data and other materials, as well as extension and refinement of the recommendations in light of input from the Board.

Cost categories (assuming 15K papers/yr, level 2 curation)



Several issues found resonance with the Board. One, success with fund-raising will hinge on how well the idea of data archiving is accepted by the scientific community and the ability to demonstrate community support. The Board is concerned about the double challenge of simultaneously negotiating a cultural shift in data sharing practices while at the same time asking for financial support from societies and

journals. It would be easier to have demonstrated value first before selling the repository as a business proposition. The "Catch 22" is that it is difficult to convince scientists that it is of value unless there is a sufficient amount of archived data that is being reused so that the point can be demonstrated empirically. One solution proposed was that greater, and more widely-publicized, research effort to show that data archiving can lead to increased citations and other benefits is needed. [This research will be pursued in the context of the DataONE project, starting in June 2010]. Another suggestion was to ask only for "start-up capital" in the short term, and avoid seeking long-term commitments until after success has been demonstrated, and there is community acceptance; however, this approach might make it more risky to secure transitional project funding (see below).

The question arose as to what viable alternatives there are to using a digital repository such as Dryad. The two major alternatives discussed include data repositories hosted by national libraries (such as GenBank or GEO at NCBI) and datasets hosted by publishers together with other Supplementary Materials. Since this latter mechanism for archiving orphan data already exists, appears to be provided by publishers at no cost to journals, and at least superficially offers many of the same benefits as Dryad, the Board felt that it will be important for Dryad to make a specific case for how it provides more value. [To that end, Mohamed Noor drafted a chart comparing the features of Dryad to a publisher-hosted repository, and Dryad staff will refine this chart for use in communicating with journals, societies and publishers].

Another issue is that as journals require more data archiving, publishers will presumably incur greater costs that will ultimately be passed on to those journals, but it is not clear if publishers are fully aware of what those costs are, or how they are being passed on. Thus, research on this topic is clearly needed, and Beagrie, Ltd. has been charged with conducting a study to ascertain costs of journals & publishers' current practices with Supplementary Data in order to make a more informed business case; the report is due in April 2010.

Finally, publishers may have a concern about being tied to one repository, and individual journals/societies may be reluctant to dictate that publishers work with Dryad. To address this, it was suggested that partner journals from the same publisher should approach that publisher as a group in order to get a more favorable response. [Allen Moore volunteered to approach Wiley-Blackwell on behalf of the Board].

Dryad's long-term financial viability will be dependent on stakeholders, such as journals, societies and publishers, considering the repository's services to be of sufficient value that they are willing to make an ongoing financial commitment to it. Todd Vision expressed the desire to begin cost return early enough (i.e. by summer 2011) so if NSF needed to be approached for transitional funding to cover a budget gap, contracts would be in place and there would be a trajectory toward stable funding. To that end, Board members were canvassed regarding estimated levels of potential future support possible from each journal (or, in some cases, their society). There were a wide range of responses. Many Board members suggested that an ongoing level of support of 20-40 USD/paper, or 5,000 Euros (~6,700 USD) per year, would be possible, possibly including an additional one-time contribution of 2000-5000 USD. The largest amounts proposed were in the neighborhood of 10K USD per year, while some members indicated that no level of financial support could

presently be asked of their organizations. Different organizations had varying opinions on the relative merits of a one-time 'joining fee', a flat subscription charge, and per-paper subscription charges. Another possibility raised by the Board was to develop a tiered cost model, with contributions scaled to the revenue of the society or publisher.

The sense of the Board was that Dryad should not immediately seek to recover full operating costs from a small number of journals, but instead to plan for consortium growth, even if that requires a transitional funding phase. Both transitional and ongoing funding should be sought from a wide variety of sources, including scientific societies, non-U.S funding agencies (such as the Wellcome Trust), foundations (through institutional development officers), information technology companies (such as Google), and large publishing organizations (such as Elsevier). One way to enable international participation would be distribute/replicate some of the repository infrastructure (such as LOCKSS sites) in other countries (e.g. at universities or national libraries).

The Board discussed strategies and targets for consortium expansion. A larger and more diverse consortium would dampen concerns that there are plausible competitors to Dryad. Directions of possible expansion were discussed. It was pointed out that new journals, if approached at start-up, would be good targets for involvement.

The conversation summarized above will inform the cost-return models to be outlined by Beagrie, Ltd, and reviewed by the executive committee at a meeting planned for April 2010. Beagrie, Ltd. was charged with developing revenue estimates covering growth over the next five years, and exploring a variety of specific scenarios for cost-return to understand the budget implications. The executive committee, with assistance from Dryad staff, is then charged with drafting a specific proposal for a cost-return model, so that it can be presented to the full Board for discussion at the business meetings held during spring and summer society conferences. The business plan will be summarized in the form of a 2-3 page, easily-digestible vision document detailing Dryad's goals and business plan, and versions customized for each of the different categories of stakeholder.

Data citation and data identifiers

A key incentive for data archiving is the ability for authors to receive credit for data reuse by others, above and beyond what they would receive from reuse of the facts or ideas already reported in their published articles. In order to ensure the effective promotion data citations through Dryad, the Board considered both the choice of technology used for data identifiers, and the adoption of a specific data citation policy.

Dryad currently assigns Handles to its data objects, and a proposal was made at the May 2009 meeting to replace these with Digital Object Identifiers (DOIs). The executive committee was split in its decision on the proposal, and so chose to bring the issue back to the Board for further discussion. In order to make an informed decision, Ed Pentz (of CrossRef) and Max Wilkinson (of the British Library and DataCite) and were invited to discuss the use of DOIs by their organizations and to answer the Board's questions about the technology.

CrossRef is the premier DOI registry for journal articles. 90% of DOIs are used for articles themselves, the remainder for components of articles (such as graphs). CrossRef's focus is on content hosted publishers, and not content in data repositories. This is related to the fact that CrossRef's membership consists of scholarly publishers. Publishers provide publication metadata upon DOI registration (similar to Dryad's journal integration). CrossRef offers an optional "Cited By" citation tracking service.

DataCite is an international data citation initiative that was officially launched a few days prior to this meeting. Participants include the Technische Informationsbibliothek (TIB) in Germany, The British Library, the L'Institut de l'Information Scientifique et Technique (INIST) in France, the Library of the ETH Zürich in Switzerland, the Canadian Institute for Scientific and Technical Information (CISTI), the Australian National Data Service (ANDS), the California Digital Library, and Purdue University. The goal is to "enable organisations to register research datasets and assign persistent identifiers to them, so that research datasets can be handled as independent, citable, unique scientific objects." The long term vision of the partnership is to support researchers by providing methods for them to locate, identify, and cite research datasets with confidence. Policies and procedures are still being developed, and Dryad has an opportunity to contribute to the design of the services that DataCite will provide. DOIs are likely to be of negligible expense to Dryad if DataCite is selected as the registration agency.

Following discussion, the Board chose to adopt DOIs as data identifiers for data packages, and charged Dryad staff with selecting between DataCite and CrossRef as the registration agency. The Executive Board was charged with deciding the following related issues:

1. Whether to use DOIs for individual datafiles or to continue using Handles.
2. Whether and how to use identifiers to track the different versions of datafiles and data records (files plus metadata), should they undergo changes.
3. How the above is to be reflected in the data citation guidelines.

Repository policy and interface issues

Recent changes to the technical infrastructure of the repository were reviewed, and the Board was given the opportunity to provide guidance on future developments. The major decisions made by the Board were as follows.

High-priority repository development goals:

1. Authors need the ability to save deposits in progress and return to complete them at a later time
2. The corresponding author of each journal article should be provided with a unique private identifier number (PIN) so that co-authors or other designees, given that PIN, can access the data package record and add/modify data.
3. It will be necessary to modify the data deposit interface to make explicit at the outset that all data is deposited under the terms of the Creative Commons Zero waiver and that it is being deposited on behalf of all co-authors. The Board would like the repository to accept non-data material (i.e. such as

software and descriptions of supplementary methods) under other IP terms, but further investigation is required to determine how this is to be done.

Policy decisions:

1. The embargo period is to be defined as starting with the online publication date.
2. Journals will ask publishers to provide a link to the Dryad data package in the electronic version of the article along with the list of other standard links (i.e. the HTML version of the abstract, the full text, the PDF version of the paper, and the publisher-hosted Supplementary Materials).
3. Dryad should emphasize the relationship to the journal article by using a title of the form "Data From: [Article Title]" wherever the data package title is used.
4. Dryad should provide a clearly labeled link in the data package/data file display that users can click on to report a problem with a data file. The message should be conveyed to the curation staff, and a 'flag' should appear on the website that "an issue has been reported with this data file." However, Dryad should not reveal the content of the message on the website. Curators should refer concerns to data depositor where needed and seek to resolve the issue so that the flag can be removed.
5. Access and download statistics for each data package and datafile can be made publicly available on the Dryad website.
6. Dryad should implement a two-tier model of curation, in which most papers receive minimal curation, but an enhanced level of curation is applied to the most valuable data packages, as determined by access & download statistics, paper citations, help requests, or other automated metrics.
7. The best practices for data management and data citation should include the citation to the original paper in the example, and should autogenerate the superset of authors for data package citations (since it may differ from the journal article). It should also instruct users to reference the original source where the data are not themselves original.