1. Executive Summary

The third meeting of the Dryad Consortium Board was hosted at University of British Columbia from 7-9 July 2011. Over 20 representatives from a diversity of journals, societies, publishers and other organizations participated in helping to chart the next steps for the Dryad repository and organization. Major achievements of the meeting were a framework for the governance and revenue model for Dryad as it transitions to an independent nonprofit organization. Participants also provided guidance on a number of important policies, including those affecting software and other supplementary materials, and how to define those publications with data eligible for inclusion in the repository. Partners will have the opportunity to officially vote on the final governance and revenue proposals from the Executive Committee based on the framework emerging from this meeting.

2. Attendees

John Allen (Biol. J. Linnean Society), David Baldwin (Ecological Society of America), Timothy Benton (Nordic Society, Oikos), Theodora Bloom (PLoS), Deborah Ciszek (Systematic Biology), Daphne Fairbairn (Evolution), Liz Ferguson (Wiley-Blackwell), Erica Fleishman (Conservation Biology), Charles Fox (British Ecological Society), Brian Hole (The British Library), Marcel Holyoak (Ecology Letters, Executive Committee), William Michener (Executive Committee), Allen Moore (J. Evolutionary Biology, Executive Committee), Lyubomir Penev (Pensoft), Heather Piwowar (DataONE), Richard Sands (BMJ Open), Peggy Schaeffer (Dryad), Michelle Tseng (Evolutionary Applications), John Wenburg (Journal of Fish and Wildlife Management), Marcy Uyenoyama (Molecular Biology and Evolution), Todd Vision (NESCent, Executive Committee), Michael Whitlock (The American Naturalist, Chair of Executive Committee), Derek Wildman (Molecular Phylogenetics and Evolution), and Laura Zahn (Science). With gratitude to Anna Van Tol (University of British Columbia) for notes.

3. Proceedings

3.1 Opening session

The meeting began with a welcome from Executive Committee Chair and UBC host Michael Whitlock. Following introductions, the rules of order for the meeting were established. The Executive Committee had previously determined that representatives entitled to vote under the interim governance agreement would be those who had attested to the terms of the Partnership Agreement on behalf of one or more journals. This included the following representatives: Baldwin, Bloom, Ciszek, Fairbairn, Holyoak, Moore, Penev, Wenburg, Uyenoyama, and Whitlock.

The voting representatives unanimously assented to the following procedural motions. First, the interim governance agreement [1] was extended through 2011. Second, it was determined that official votes would be conducted by correspondence following the current meeting, thus
enabling all Partners to participate. Third, it was determined for decisions at the meeting to be conducted according to the “sense of the room”, with all persons, regardless of Partnership status, entitled to offer motions and to have exactly one vote on all motions. All votes would be considered advisory to the Executive Committee.

3.2 Project updates

Reports were presented on the current status of the repository by T. Vision, on the JISC-funded DryadUK project by B. Hole, and on recent and ongoing studies of data reuse and data archiving policy by H. Piwowar. Presentations are available online [2].

Vision presented the background behind the recent introduction of strong new data archiving policies by many journals, and increasing expectations for data management and dissemination plans by funding agencies, as exemplified by the National Science Foundation’s recent policy. He presented findings and recommendations from the PARSE Insight study regarding the availability of research data related to publications [3] and compared Dryad to other data archiving options such as publisher-hosted sites, institutional repositories, and specialized disciplinary databases. He described the recently added capabilities of Dryad’s integrated manuscript and data submission workflow, including the ability for journals to provide peer reviewers anonymous access to data submissions. Twelve journals had, by this time, integrated submission with Dryad, and journals from nine other publishers were in various stages of the process. As of the end of June, there were nearly 800 data packages in the repository from over 80 different journals, the repository is receiving ~70 new submissions per month, and the pace of submission is steadily accelerating (Figure 1A). About 75% of new data packages are received through the integrated submission process. The other 25% were introducing data from about five new journals per month. The size of data packages so far is relatively small, with the majority totaling less than 1 MB and consisting of only one data file (Figure 1B). There is evidence for substantial data reuse from download statistics, which are now reported on the website for every data package and data file. Approximately half of datasets had been downloaded at least ten times; six datasets had been downloaded over 100 times (including one from 2011), and one has been downloaded over 1000 times.

Figure 1. (A) Increasing submission rate of data packages through June 2011. The abscissa shows the total number data of packages. (B) The distribution of the size in bytes (left) and number of files (right) in each data package.

B. Hole described DryadUK, a 12-month JISC-funded project at Oxford University and The British Library (BL), with participation from the Digital Curation Centre and Chas. Beagrie, Ltd.
Goals include piloting a mirror of the Dryad repository at the BL, expansion of the journals to UK-based publishers and to new biological and biomedical disciplines, development of an assessment framework, development of a prototype for validation of minimal information standards upon data submission, and sustainability and organizational planning. He presented, in depth, the outcomes of a 1-day workshop held at the BL in April 2011 to get input into Dryad’s value proposition and revenue model from a range of stakeholders [4].

H. Piwowar reported on studies concerned with the reuse of archived data and the effects of data archiving policy [5]. Analysis of microarray studies archived in the Gene Expression Omnibus demonstrate a higher number of citations for articles having archived data. Analysis suggests that publications based on archived data drop off steeply one year after the initial publication, while uses by third parties continue to grow for at least four years. Reuse is distributed broadly across many datasets in the repository. She reported on data suggesting potentially orders of magnitude higher return on investment for data archives relative to traditional grant funding, as measured by publication output. She discussed ongoing studies to collect reuse data from a more diverse sample of repositories, and on early results from a time-course survey of authors publishing in journals that have, or have not, adopted a strong data archiving policy.

T. Vision then gave an update on developments regarding the repository itself. Major features introduced since the last Dryad Consortium meeting include:

- provisioning of DataCite Digital Object Identifiers (DOIs) for all data packages and data files
- allowing authors to save partial submissions for later completion
- the option to allow authors to release data immediately, upon publication, or after an embargo period (one year by default, longer terms available at the discretion of the editor)
- the option for journals to provide peer reviewers with secure and anonymous access to data
- faceted and filtered search of Dryad data from the web interface
- the ability to search for related published datasets in external repositories (currently the Knowledge Network for Biocomplexity and TreeBASE)
- user profile pages showing information on in progress and completed deposits
- display of view and download statistics for all data packages and data files
- deposit handshaking with TreeBASE
- interfaces for machine access (incl. bulk query and download)

Features currently in development for release in 2011 include:

- versioning of datafiles, allowing authors to provide updates post-publication
- a streamlined deposit process for previously published data
- improved indexing of data packages by third parties (including Web of Knowledge)
- improved display of multifile data packages
- enhancement of subject metadata and use of controlled vocabularies, leading to improved data search
- handshaking with GenBank
- landing pages for journals
- distributed replication with DataONE & CLOCKSS
- educational data modules available through DryadLab

An NSF Advances in Bioinformatics grant proposal to be submitted the following week was presented [6]. The grant is intended to build the capacity for growth and expansion of the repository, through greater automation of data curation and providing greater resources for submission integration. It also is intended to provide transitional funding for operations as Dryad implements its sustainability plan. The proposed total budget is $2.4M for four years.
3.3 Policy issues

The group discussed several important aspects of policy regarding software and other non-data content, data citation, peer review of data, data from non peer reviewed publications, and how exceptions are granted to data archiving requirements.

**Software:** Some authors have already been depositing software code in Dryad under CCZero while other authors have reported that they do not wish to use Dryad for code in the absence of open-source licensing options. Some of the issues were discussed in a recent Dryad blog post [7]. Independent of this development, a group of evolutionary biologists recently approached the authors of the Joint Data Archiving Policy (JDAP) [8] to urge a coordinated policy regarding software archiving. These developments have led the Executive Committee to ask whether Dryad is an appropriate repository for code, and/or could it be with modifications? After a frank and open discussion of both the merits and pitfalls, the group resolved that Dryad should aim to provide a repository for code only where it does not otherwise have a better home. This would be expected to apply primarily to “one-off” scripts that would otherwise be lost, rather than ongoing software projects that should be maintained in a public version control system. The issue of whether to continue to require CCZero for new software deposits was not resolved. The group did resolve to undertake an exploratory process among the JDAP journals to discuss a shared software archiving policy. The Ecological Society of America requirement to submit “novel code” was presented as a useful model.

**Other “supplemental materials”:** Some prospective Partner Journals have been conveying a strong desire for Dryad to accept the full range of content that is currently hosted by the journal/publisher as Supplemental Online Material, and not restrict the repository contents strictly to data. While there were concerns expressed about the boundaries of what to allow, and some journals that would rather not archive any non-data supplemental materials at Dryad, the general agreement was that this should be provided as an option to those journals or publishers that wish to take advantage of it. The group resolved that Dryad should allow materials related to a publication even if not clearly data or software, and that authors would be expected to release the materials under CCZero. Dryad curators would be on alert to reject any copyrighted materials from the archive.

**Data citation:** Since researchers have greater motivation to release their data for others to use when they receive professional credit for doing so, Dryad has long aimed to promote the practice of data citation and the technical infrastructure needed to track those citations. H. Piwowar gave a presentation on the background behind Dryad’s current data citation policy, pointed out the major difficulties in tracking data citations given the current diversity of citation practice, and reported on some emerging community standards designed to address that problem. She recommended that the data citation to Dryad in the original article be included in the reference list so that indexing services will be able to discover it, in line with recent workshop recommendations. Concerns were raised about this proposal, and it was not adopted at this time.

**Peer review of data:** As noted above, some journals will now be providing peer reviewers access to the data in Dryad. The group was asked whether Dryad should alert users in any way if the data were (or were not) available to the peer reviewers of the original journal article. There was general concern that information about the availability of data to peer reviewers could provide misleading information about the reliability of the data, and the idea was not pursued further.
Data from non-peer reviewed publications: It was pointed out that the current Collections Policy restricts content to data from peer-reviewed publications, and that exceptions were being made on a case by case basis for data associated with other kinds of publications. The group discussed the value in extending the Collections Policy to include data of equivalent value and reusability associated with a broader range of publications (including dissertations/theses), and voted in the affirmative to allow acceptance of data products associated with non-peer reviewed publications from reputable academic sources. A recommendation was made to appoint a small panel to review borderline cases. The group also expressed general agreement with the recent change in the scope of the repository to include all “basic and applied biosciences”, and with Dryad’s policy of interpreting this scope broadly. It was recommended that a future meeting discuss what Dryad’s policy should be in accepting data in languages other than English.

Exceptions to data archiving policy: A number of the journals that recently adopted the JDAP [8] have been struggling with when and how to allow exceptions to the policy for certain data collections, in particular long-term collaborative field studies. The Executive Committee proposed discussing a common, principled approach to making such exceptions. The group discussed diversity in practice and custom across disciplines, and the relevance of the existing exemptions for data belonging to third parties [9]. While no consensus was reached on how to deal with long-term studies with collaboratively owned data, a number of useful ideas were raised: (1) restricting the archiving requirement to novel data collected by the authors, (2) requesting an alternative data release plan when granting such an exception, (3) permitting long embargoes in return for archiving the data at the time of publication, (4) and having a common understanding of the maximal allowable embargo.

3.4 Governance

The Executive Committee presented the motivation and process for the Dryad Consortium to incorporate as a U.S. tax-exempt not-for-profit organization. This will allow the organization to raise funds from a greater diversity of donors, to manage its own assets, and to oversee operations through contractual arrangements with third parties. Importantly, this provides a way for stakeholders, through membership, to have a say in governance that would not be possible were Dryad entirely embedded within a host institution. Thus, it will help ensure that the organization adheres to its mission while retaining the flexibility to adapt to future circumstances. Over the course of the meeting, participants considered, and considerably revised, a governance proposal that had been put forth by the Executive Committee. The framework that emerged is described in what follows, and broadly consists of a Board of Directors elected by the Membership, and overseeing the executive staff of the repository.

The twelve-person Board of Directors (BoD) would be vested with legal and financial responsibility for the organization, its assets and mission. Directors would serve for staggered three-year terms and could serve multiple terms. Directors would be elected by the Members to represent a diverse range of perspectives and expertise, but need not be drawn from Member delegates. A nomination committee would solicit and vet nominations for the BoD. Directors would elect officers of the board from among their own number. Additionally, the board would be empowered to appoint ex-officio directors, as needed, to complement its expertise.

Dryad would aim to have a broad membership with sufficient power that the organization remains accountable to its diverse stakeholders. Members may include any legitimate organization that supports the mission of the organization and wishes to participate in its governance. Unlike earlier proposals, there is minimal interdependence in this framework between the governance structure and the means by which revenue is raised. Any organization that pays a modest annual fee would be eligible to apply for Membership, including but not limited to journals, societies, publishers, research institutions, and funders. Membership
discounts will be supported for developing countries to the extent possible, and sponsoring memberships will be offered for organizations above some modest base membership level. In addition to voting on Directors, Members would vote on amendments to the Articles and Bylaws, and serve as an advisory body to the BoD (e.g., through participation in an annual meeting). Applications for membership would be vetted by the BoD to ensure that no competing interests have the ability to dominate its numbers.

The executive staff of the repository would report to the BoD and be employees of the independent corporation. This staff would oversee contractual agreements with host institutions, where non-executive repository staff (e.g. software developers, curators) would be employed.

The Executive Committee will use the framework that emerged from this meeting as the basis for drafting Articles and Bylaws, in consultation with legal counsel, to be presented to the Partners for a vote. The process for constituting the initial board will also need to be decided.

### 3.5 Sustainability

A major goal of the meeting was to adopt a revenue plan that ensures Dryad will have stability independent of short-term grant funding and institutional subsidies. The Executive Committee presented some of the background to Dryad’s revenue model [10]. It is informed by a number of guiding principles:

1. Depositors should be assured that Dryad continues to have the resources necessary to protect the integrity and accessibility of the content.
2. Dryad should seek to minimize costs through economies of scale.
3. Spreading the revenue burden for the repository among many contributors reduces the risks of relying on the goodwill of any single organization, and requires only a small contribution from each contributor.
4. The revenue burden should be distributed fairly and simply among contributors, in a manner that is proportional to usage and the costs incurred by the repository, where the primary cost driver is curation at the time of deposit.
5. There should be no charges for access to data by end-users. Instead, the necessary funds for curation, storage, etc. should be in hand when a deposit is accepted.
6. Diverse organizations should be given the opportunity to pay for data deposits on behalf of researchers, including societies, publishers, research institutions, funding agencies, and individual researchers themselves.
7. Dryad should continue to seek grant funds for innovative research and development, while covering operating costs with the base revenue model.
8. Revenue should meet but not exceed expenses over the long term.
9. The revenue model and budget should be reviewed regularly by the Board of Directors.
10. Budget decisions should be made in a transparent manner.

A 3-year revenue target has been set based on a projection that Dryad could, by the end of 2014, be receiving on the order of 10,000 deposits annually. At this scale, a budget consultancy suggested that operating costs could be between $30-50 per data package, depending on how much curation and customer support is provided to depositors, users, members, contributors, etc. Based on discussions at previous Dryad Consortium meetings, and research from a second consultancy, the Executive Committee drafted a revenue plan in May 2010 and circulated it for consideration among the stakeholder community. This was also the focus of the April 2011 DryadUK workshop mentioned previously [4]. Synthesizing this feedback, M. Holyoak, on behalf of the Executive Committee presented a revised revenue plan for further
refinement at this meeting. The framework that ultimately emerged from subsequent discussions at this meeting can be summarized as follows:

A variety of payment plans, distinct from membership in the Consortium, would be offered that would be attractive to different organizations and users: Journal-based, Voucher-based, Pay-as-you-go, and Author-pays.

- **Journal-based**: the journal (or group of journals from one society or publisher) prepays to cover any deposit that might be received associated with that journal or journals. The annual fee is based on the annual number of research articles published in the previous year (rather than the number of deposits anticipated).
- **Voucher**: Under this plan, any organization (not necessarily a journal) may pay in advance for a fixed number of deposits, above some minimum number.
- **Pay-as-you-go**: Under this plan, any organization (not necessarily a journal) pays retrospectively for the number of deposits that occurred during a given time period.
- **Author-pays**: An author may pay for an individual deposit from either an integrated or non-integrated journal that does not have a contract for one of the plans above. Deposits from non-integrated journals would be charged for the additional costs of curation.

Specific price points were suggested, but it was understood that these would ultimately need to be set based on the transaction and uncertainty costs associated with each approach. Membership fees would be in the range of $500 to $1,000, and would provide a 10% discount on deposit costs. Member prices for the different payment plans were suggested at $25/per research paper for the Journal-based plan, $50 per deposit for the Voucher and Pay-as-you-go plans, and $75 per deposit for the Author-pays plan. Authors under all plans would be required to pay a surcharge for unusually large (>10Gb) data submissions. Unlike previous proposals, the group did not recommend discounts for bulk purchases of deposits under the Voucher-based or Pay-as-you-go plans. Furthermore, the group strongly recommended against any charge for the initial setup costs of journal submission integration. Where deposit costs are not covered by another organization, the repository should seek to provide waivers for deposits from authors in developing countries.

This framework will serve as the basis for a revenue plan to be put by the Executive Committee to an official vote by the Consortium Partners.

### 3.6 Strategy and next steps

On the last morning, the participants organized into breakout sessions to develop plans on three fronts:

- How to reach out to new journals, publishers and societies?
- How to articulate the value proposition?
- How to define the roles and identify nominees for the Board of Directors?

#### 3.6.1 Reaching out

This breakout group discussed a number of ideas for reaching out to new journals, publishers, societies, and other members:

1. Systematically target subdisciplines that are rich in irreplaceable and reusable data but lack repositories, such as cell biology (lots of image/video data), animal behavior, physiology, neuroscience, microbiology and quantitative genetics/biostatistics.
2. For disciplines that are close in scope, establishing personal contacts would be most productive; for others, Dryad could seek contacts from data officers at funding agencies.
Also worth speaking to councils of journal editors, and influential members of the publishing community.

3. There is a need to be clear about the two halves of the message. First, why promote open archiving of research data? Second, why use Dryad for that?

4. For commercial publishers, there is a need to motivate them through more than just an appeal based on the benefit to science and scientists:
   a. Increasing citations is a major incentive.
   b. Increased curation may add value such that it provides additional revenue streams. There was discussion of the NIF (in neuroinformatics) and the Entangled Bank project in the UK as potential models.
   c. Include publishers and funders in the upcoming DryadUK outreach workshop and use that workshop to show a good case study for how increased curation can add value to data.
   d. Having a mature revenue plan will be helpful in approaching publishers.
   e. There are opportunities but also significant challenges in recruiting larger publishers as high-level sponsors.
   f. Some publishers who are strong in particular subdisciplines could get behind the repository if they can be convinced that the community is behind it.

5. There is a need for a better term than 'subscribers'. That said, prominently displaying 'subscribers' as well as Members on the Dryad site would be helpful.

6. There was a recommendation to reach out more actively to research institutions for Membership, particularly institutions that have played a leadership role in scholarly communications.

7. Allowing free deposit to new communities for a time period may ease their way in.

8. For reaching out to the medical community, it would be helpful to have medical/ethics expertise on board, good infrastructure for dealing with sensitive human data, and the means to help authors be responsible with those data.

3.6.2 Making the case

This breakout group considered the different stakeholders and sought to articulate elements of the value proposition. Some of benefits noted were:

- Having data that is searchable and 'linked'
- Helping to ensure that credit for data reuse accrues to the original journals and authors
- Providing a way to gather all the data from a publication together regardless of its format and across a broad subject matter
- Handshaking with other repositories
- Enabling authors to provide good documentation (e.g. through a ReadMe)
- Stability and permanence (although the case would need to be made that Dryad is at least as stable as existing publisher-run sites)
- Being community-led, with priorities and policies shaped by the membership
- The ability for authors to update data after publication.
- The ability to search deep-indexed data

The group’s overall recommendations included:

- To work with manuscript processing program vendors to streamline submission integration
- To keep submission integration free
- To provide a benefit to authors by holding data even when an article is not accepted, thus allowing them to resubmit elsewhere without having to re-deposit.
- To seek ISO certification as a Trusted Digital Repository (a potential differentiator relative to other archiving options)
• To more clearly distinguish promotion of Dryad as a repository and data archiving *per se*
• To keep the membership fee low ($300-500)
• To consider allowing members to receive a fee discount for attending the annual meeting
• To provide a deposit discount to members
• To stabilize the membership and revenue structure

3.6.3. Constituting the Board of Directors

This breakout group proposed a list of responsibilities for the Board of Directors, while recommending that the list be reviewed by a qualified attorney:

1. Act on all matters except amendments to the Constitution and By Laws
2. Approve and terminate an organization’s membership
3. Consider proposals from members
4. Receive and act on reports from Executive Director, officers, and committees
5. Designate the time, place, and theme of the Annual Meeting
6. Hire, fire, evaluate, and approve compensation for the Executive Director
7. Appoint committees as needed
8. Approve awards, grants, fees and fee waiver policies
9. Review and provide oversight of assets, budgets and fundraising efforts
10. Instigate and act upon an independent annual audit of the organization’s finances
11. Produce and approve the final slate of Director nominees for voting by the membership
12. Share BoD minutes and financial reports with the membership (and public*)
13. Elect the officers of the Board from among their own number
14. Appoint *ex-officio* Board members as appropriate
15. Consider and respond to member grievances as appropriate
16. Request and act upon recommendations from investigations of ethical misconduct.

The group also recommended some characteristics of the Board of Directors:

• 12 members with 3-yr staggered terms; no limit to number of terms served.
• The Board should include Directors with varied and relevant backgrounds, including publishers (from both small and large organizations), officers of professional societies, journal editors, scientists, representatives from host institutions, and other experts in business, fundraising, sustainability and technical aspects (*e.g.*, from other repositories),
• Membership should be diverse with respect to gender, race, ethnicity, and geography (i.e., international representation and language).
• The Board should set high expectations for ethical and professional standards.

The following recommendations were made regarding the Board Nomination Process:

• A Nominating Committee will solicit nominations from the membership
• The Board will determine the final slate of candidates
• Members will vote by correspondence for the slate of candidates
• Vacancies can be temporarily filled at the discretion of the Board

A list of potential nominees for the initial Board was discussed, and conveyed to the Executive Committee for consideration.
4. Next Steps

The recommendations emerging from this meeting provide a detailed and compelling roadmap for Dryad as an organization. As a next step, the outcomes will be reviewed by DryadUK personnel and by legal counsel. During this time, Dryad staff will confirm the Partnership status for participating journals. Once it is drafted, the Executive Committee will present the governance structure (in the form of Articles and Bylaws) for consideration by the voting Partners, and separately present a nonbinding vote on the revenue plan that the initial Board of Directors will be required to ratify. Following approval of the governance and revenue plans, Dryad staff will develop a marketing campaign and draw up contracts and agreements with host institutions, Members and “subscribers” for implementation as early as 2012. In the meantime, Dryad staff will undertake a fundraising campaign to support implementation of the sustainability plan and build an endowment that can help to defray deposit costs. It is expected that the transition will be well underway in time for the first official Membership meeting, tentatively scheduled for summer 2012.

Notes

2. Presentations from the meeting are linked from the Agenda at http://datadryad.org/wiki/Board_Meeting_2011_Summer