

# *Dryad Consortium Management Board*

## *Meeting Report*

*May 21-21, 2009*

*NESCent, Durham NC, USA*

### **1. Overview**

Dryad is a digital repository for the publication of data associated with scientific publications. It is governed by a consortium of journals, each of whom has one representative on the Dryad Consortium Management Board (DCMB). This goals of this first meeting of the DCMB were to establish rules for governance, set policy, plan for sustainability, plan community engagement activities, and decide on priorities and a timeline for development of different features of the repository.

### **2. Background**

Dryad was conceived as a home for the many "small science" datasets, in fields such as ecology and evolution, that are used to support the findings in a publication, but then are not made available in a form that can be revisited and repurposed by other investigators. The Dryad project was launched in early 2007 in conjunction with an effort to draft a joint data archiving policy that would require deposition at publication among participating journals. The strategic goals for the repository emerged from various workshops during 2005-2007. These goals include:

- To preserve data at the time of publication, when there is the greatest incentive and ability for authors to share their data
- To allow journals and societies to pool their resources in the support of one shared, trusted, and financially self-sustainable repository.
- To lower the burden of data sharing by integrating with the journal submission process and providing one-stop data deposition.
- To assign persistent, resolvable identifiers to datasets so they can be cited in print and linked on the web.
- To allow end-users to search for data across journals, and by fields such as taxon, geography, geological age, biological concept, etc.
- To enable bidirectional search and retrieval, and ultimately data integration and synthesis, with other relevant datasets.

In 2008, the National Science Foundation awarded a 4-year grant to NESCent to develop the Dryad repository. A key goal in that grant is to establish stakeholder ownership and governance of Dryad, in the form of a Management Board, to promote the adoption,

*Page 1 of 15*

*Dryad Consortium MB report Spring FINAL.doc*

*by Todd Vision & Peggy Schaeffer, June-August, 2009*

*[https://www.nescent.org/wg\\_dryad/Management\\_Board\\_Meeting\\_Spring\\_2009](https://www.nescent.org/wg_dryad/Management_Board_Meeting_Spring_2009)*

sustainability and utility of the repository. Journals that had been previously supportive of the development of Dryad and the Joint Data Archiving Policy were invited to appoint a representative to the meeting, as well as several additional journals that were seen as reaching out to major stakeholder communities. In several cases, these representatives were appointed by the societies that oversee the respective journals. The membership initially is comprised of research journals in ecology, evolution and/or organismal biology that have an international readership and are taxonomically broad in scope. Expansion of the membership is expected, and one of the roles of the DCMB is to set the rules for this expansion.

### 3. Executive summary of outcomes

- *Interim Executive Committee* appointed.
  - Membership: Mike Whitlock (chair), William Michener, Allen Moore, Marcy Uyenoyama, Tim Vines, Todd Vision (ex officio).
  - The initial charge for the committee includes presenting governance documents for approval at the next DCMB meeting, articulating "community norms" for data citations, deciding whether to purchase CrossRef DOIs for data records, and developing a phased strategy for what needs to be in place in order to make data deposition a requirement.
- *Top priorities* for Dryad staff: integration with Partner journals on an accelerated timescale, scaling up curation capacity, internationalization, and sustainability planning.
- *Joint Data Archiving Policy* wording revised.
- *Intellectual property*: adopted Creative Commons Zero as the sole license and public domain dedication, and decided to supplement this with "community norm" guidelines for data citation.
- *Embargoes*: Partner journals have the option of allowing authors to request a no-questions-asked 1 yr embargo, and granting longer embargoes for sensitive data on at the discretion of the journal editor.
- *Limits on data package size*: As an initial policy, a limit of 1 GB for an individual file, and of 10 GB for all files associated with a single article, was adopted.
- *Next meeting*: The DCMB will convene again in early December, at a place to be determined. The agenda will include adoption of the governance documents, and consideration of alternative budget and funding options.

### 4. Governance

One goal of the meeting was to discuss the future operation of the DCMB, including the rights and obligations of members, the means by which decisions will be reached, and how the DCMB will be governed and members replaced. An interim voting process was adopted whereby each representative had one vote and a majority was required to pass a resolution. A Constitution and Bylaws, previously drafted by Dryad staff, were raised for

consideration. These documents define the Dryad Consortium; define the levels of partnership, specify that the unit of partnership and voting is to be the individual journal, and set out the financial responsibility of the Partner journals for repository sustainability.

The distinction between Partners and Associate Partners is currently defined as follows:

### **Partners**

- have adopted the Joint Data Archiving Policy or have a substantially equivalent policy of required data submission
- are committed to providing for the long-term sustainability of the repository (although there is no immediate financial commitment since the repository is sustained by the NSF award through 2012); we will be talking about the sustainability later - by the time the commitment comes, it will be very explicit
- designate a voting representative (one per journal) to the DCMB
- will enjoy special services, such as customized handshaking and branding of journal content in the repository

### **Associate Partners**

- have not yet committed to the full suite of Partner responsibilities, but demonstrate a desire to do so
- are invited to have a nonvoting representative on the DCMB

The Constitution and By-laws were drafted with the intent that they could be used if and when the Dryad Consortium chooses to apply for legal status as a nonprofit organization. An interim executive committee was appointed, consisting of Mike Whitlock (chair), William Michener, Allen Moore, Marcy Uyenoyama, Tim Vines, and Todd Vision (ex officio), and one of the initial charges of this committee was to review the draft governance documents and present these for a vote at the next meeting of the DCMB.

## **5. Joint Data Archiving Policy**

The Joint Data Archiving Policy (JDAP) is distinct from Dryad; the policy can be adopted without reliance on Dryad, through archiving in supplemental materials, for example. However, it was recognized that Dryad is designed in order to make the JDAP easier, and without JDAP there would likely be limited adoption of Dryad; thus the two efforts are mutually reinforcing. Representatives described the current status of their journal or society with respect to the JDAP, and the details of its implementation were discussed. After discussion of various concerns about the current wording, the following suggested revised JDAP was agreed upon by the Board.

**"<<Journal >>requires, as a condition for publication, that data supporting the results in the paper should be archived in an appropriate public archive, such as <<list of approved archives here>>. Data are important products of the scientific enterprise, and they should be preserved and usable for decades in the future. Authors may elect to have the data publicly available at time of publication, or, if the technology of the archive allows, may opt to embargo access to the data for a period up to a year after publication. Exceptions may be granted at the discretion of the editor, especially for sensitive information such as human subject data or the location of endangered species."**

The exact wording may vary from one journal to the next. The important features are that the data which support the results in the paper be archived in a publicly accessible form, and that any default embargo period not exceed a year.

The aim is for the policy of required deposition to be adopted in a coordinated fashion by journals (*i.e.* required deposition is not immediately required). One charge of the interim executive committee is to bring to the full board a proposal for the steps in the process of coordinated adoption, including any requirements for the functionality of repositories such as Dryad. Partners that have not yet formally adopted the JDAP are expected to bring it to the consideration of their governing bodies at the earliest opportunity.

There was considerable discussion of the embargo terms. It was decided that journals would have the right to have a more restrictive policy, in which embargoes for non-sensitive data are disallowed. Furthermore, Dryad would need to present the appropriate options (1: no embargo; 2: no embargo or 1 year embargo) based on the identity of the partner journal.

## **6. Intellectual Property**

The DCMB considered a data use policy that would be clear to authors and users and would promote the intended purpose of the repository to advance science through the reuse of data. James Boyle, professor of law at Duke and one of the founders of Science Commons, presented a summary of intellectual property law and how it pertains to scientific data. Boyle described Creative Commons Zero (CC0), a license and public domain dedication that is based on the principles of the Science Commons Open Access Data Protocol, and is applicable to scientific facts and creative output (which are not universally subject to copyright). Advantages of CC0 are that it is valid as a legal document across international jurisdictions, can be presented in a form that is easily understood by both data depositors and users, and also is machine readable. The DCMB adopted the policy of requiring authors to deposit data to Dryad under the terms of CC0. This is a change from Dryad's initial policy, which asked authors to deposit their under the terms of a Creative Commons 3.0 license. It should be made explicit that the CC0 terms apply to the metadata as well as the data. Extraction of metadata from the publisher's record may require a special arrangement that does not violate the terms of use for that publication.

The CC0 makes no requirements regarding conventions for attribution. This is generally articulated by "community norms" specific to the domain of use. There was considerable discussion of the various options for attribution and citation of Dryad data, but no specific convention was agreed upon. It was determined that the interim executive committee would be charged with the task of proposing a citation convention at the next DCMB meeting.

## 7. Sustainability

The initial costs of repository development were borne by NESCent and current development is funded primarily by a 4-year grant from the National Science Foundation (expiring 2012), as well as funds from the Institute for Museum and Library Services. In the long term, some combination of page charge receipts, society and institutional subscriptions, grants and other sources will be needed to ensure that the repository is not dependent on ephemeral grant funding, and one of the major tasks of the DCMB will be to reach agreement among the journals on a sustainability plan and help implement it. Lorraine Eakin (UNC School of Information and Library Science, and a contributor to the Blue Ribbon Task Force on Sustainable Digital Preservation and Access <http://brtf.sdsc.edu/>) gave a presentation that emphasized the need to consider sustainability early, as well as providing some models for assessing value and determining costs. She discussed how this issue has been approached at other repositories, ranging from the Protein Data Bank(<http://www.rcsb.org/pdb/>) to ICPSR (<http://www.icpsr.umich.edu/>). Participants offered the perspective of their journals, publishers and societies about opportunities and liabilities for different sustainability models. Peter McCartney of the National Science Foundation also called in to discuss NSF's perspective on sustainability; NSF is very open to a diversity of models for database and repository sustainability, and does not disallow cost recovery for NSF products. NSF can provide some continuing funding through traditional grants as long as the community that benefits is willing and able to shoulder the majority of the operating costs.

The DCMB charged the Dryad project staff with developing a business model would identify the value-added benefit of the repository to different parties (depositors, users of archived data, journals, publishers, societies, funding organizations, etc), the costs of operation, and a mix of potential sources of revenue. Costs would need to include such things as physical infrastructure, system administration, data curation, business/management and communications. Sources of revenue could include:

- an archiving charge (similar to a page charge)
- pay-per-use, or individual subscriptions, for access to repository contents
- institutional subscriptions (possibly for higher service levels)
- subscriptions from societies and journals (possibly in return for full partnership benefits)
- fees from publishers
- recovery of cost from archiving of large data packages

- grants from government funding agencies across the globe as well as private foundations
- angel donors

The DCMB would like to be in a position to discuss the business plan in considerable depth at the next meeting, and charged Dryad staff with making this a top priority.

One sentiment that resonated with many board members was that the cost per paper would need to be on the order of \$20 or less in order for the repository to be supported by scientists and societies; this figure will be useful for determining the appropriate level of service the repository can provide. There were a wide range of viewpoints as to the feasibility and desirability of different revenue streams, and much discussion was focused on the obstacles to obtaining archiving charges from authors. It was pointed out that the collection of receipts from individual users could impose considerable overhead costs. Furthermore, a number of partner journals do not have a culture of page charges (e.g. *The Biological Journal of the Linnean Society*, *Integrative & Comparative Biology*, *Molecular Ecology*); some have page charges waived for members, grant exemptions for most articles, or only require them on a voluntary basis (e.g. *The American Naturalist*, *Heredity*, *Journal of Evolutionary Biology*, *Conservation*, *Paleobiology*); some have regular page charges (e.g. *Molecular Biology and Evolution*); and some charge an additional fee for archival of large data sets (Ecological Society of America journals). If journals or publishers are to be charged, models that recognize the limited resources of many small journals and societies, will need to be considered.

The board had a number of suggestions for making the business case to stakeholders. If it is anticipated that societies, journals and publishers would be investing resources in individualized data archives in the future in the absence of a trusted, shared repository, then the true costs of hosting supplemental data now, and under that scenario, need to be understood. The value-added benefit of economy of scale, and of professional curation, should be carefully considered. The board also discussed ways to align incentives with the business model. For example, if there is a charge to authors in non-Partner journals, while Partner journals pay a reduced per-paper fee in bulk, this would recoup the added curation expenses from non-Partner submissions, while offering an incentive for new Partners to join and contribute. The business model may also consider phased revenue streams. There was general agreement that the value of deposition needs to be made very clear to the scientists who will be expected to deposit data and possibly contribute financially for doing so. A study of the citation impact of posting supplemental data in ecology and evolution would be extremely useful, if difficult to do in a controlled fashion. The case may also be made for the long-term preservational value of data stewardship. It was also pointed out that the incentives and value of data archiving and sharing to authors in academic fields may differ from those in more applied disciplines, such as conservation (e.g. citation vs. regulatory compliance). A vigorous informational campaign should be implemented once the business case is in hand, assuming the board itself is convinced of its merits.

Page 6 of 15

*Dryad Consortium MB report Spring FINAL.doc*

by Todd Vision & Peggy Schaeffer, June-August, 2009

[https://www.nescent.org/wg\\_dryad/Management\\_Board\\_Meeting\\_Spring\\_2009](https://www.nescent.org/wg_dryad/Management_Board_Meeting_Spring_2009)

In the interim, it was suggested that Dryad staff and executive board members approach potential international partners and funding sources to raise, and gauge, interest. It was generally agreed that international buy-in would greatly assist with long-term sustainability.

## **8. Journal integration**

Dryad staff reported the current status of data submissions and journal integration activities. Voluntary data submissions are currently being accepted through the website. Legacy data from *Systematic Biology* (specifically, from the SystBio archives) is being deposited by Dryad staff. Most significantly, Dryad staff have worked with *The American Naturalist* to integrate data submission into the article acceptance process. Currently the basic publication metadata from all accepted articles (going back to mid-2008) can be pulled up by entering the manuscript number; this saves the author the burden of entering the manuscript information, and makes it possible for the journal to send the authors an article-specific link to Dryad's submission system. Authors are now encouraged by *American Naturalist* to deposit their data in Dryad (or another archive) as part of the acceptance letter. The journal now sends us publication metadata (including manuscript number) at the same time the acceptance letter with the customized link to Dryad deposition interface is sent to the author. When the author deposits their data, the identifiers are sent back to the journals so that the publisher can include those links in the proofs and/or on the website. This basic workflow can be adapted for additional journals as needed.

The discussion centered on how to make sure data was linked from the paper and vice versa. Some members felt that incorporating links to the data on the publisher website might be difficult, but that including it the paper itself should be less so, since it is within the editor's purview to implement. Another point of discussion concerned when in the process authors should be asked to deposit data, and it was determined that this may differ among journals. It was generally agreed that the integration with the journals would need to be engineered in such a way that we could ensure that at least the single data package identifier was included in each published paper. The board discussed the merits of instituting a convention among journals for where to include this identifier, with one suggestion being a special section entitled Supporting Data. In addition, the board was in agreement that it is important for the Dryad data record to contain a valid and resolvable link to the original article, which suggests that data should not be published on Dryad until the DOI has been assigned to the original paper.

## **9. Repository Development and Policies**

### ***9.1 Development priorities***

Dryad staff presented a thumbnail sketch of planned repository development needs, including integration with journals and handshaking with specialized repositories, so that the DCMB could provide their perspective on the priorities and timeline of milestones.

Work is currently underway to harvest metadata about data holdings in related repositories and data networks. An OAI-PMH gateway for harvesting of data from the Knowledge Network of Biocomplexity, and other Metacat providers, will be completed soon. A similar mechanism for harvesting the metadata from TreeBASE is scheduled for the coming year. Next year will also see the repository physically transferred to the Digital Library Program at North Carolina State University, who will eventually oversee replication of the contents across multiple institutions to ensure long-term preservation and security. The NSF grant also specifies that Dryad will experiment with "handshaking", or serving as a portal for data deposition to other repositories and data networks. This is conceived as a way to lower data deposition burden for authors that have data types for which different specialized repositories already exist, and to ensure that all the components of an article's data package remain linked via machine-stored identifiers. Additional areas of major development that the management board was asked to prioritize include improvements to the curation interface, the submission interface, and the search interface; enrichment of select data packages targeted at undergraduate and graduate educators; and publication of Dryad contents using semantic web standards.

Many board members felt that integration with Partner journals should take precedence over many of these other developments. One suggested target was that all Partner journals should have the level of integration now achieved with *American Naturalist* by the end of 2009. This reflected a concern that delaying integration would in turn delay adoption of the JDAP, and the sense that it is critically important to sustain the momentum for the JDAP. A minority of members expressed the view that rushing integration would expose the repository interface before it was bug-free and full-featured; the value of the repository would be more evident after such features as handshaking were already in place. The final conclusion was that Dryad staff should investigate how to revise the timeline to accelerate journal integration while providing sufficient time and effort for testing and incorporating the feedback of early users. The first journals in the queue should be the large number that use the Manuscript Central system (Tim Vines offered to help facilitate this process). Only a few Partner journals (e.g. *American Naturalist*, *Paleobiology*, and those published by Elsevier) use other systems. Other priorities should include international journals, such as *Journal of Evolutionary Biology*, and *Systematic Biology*, which already requires data deposition, has a very large readership, and which is a testbed for handshaking with TreeBASE and GenBank.

The discussion then turned to handshaking. Opinions regarding the value of handshaking at deposit varied widely, and concerns were raised about the maintenance burden of this functionality. It was concluded that handshaking would need to be evaluated in light of the overall cost model, and that more information would be needed about its value to Dryad's depositors and end-users. The NSF grant commits Dryad to undertake handshaking with two external repositories, TreeBASE and GenBank. The board discussed other repositories and data networks that could be candidates in the future, should handshaking turn out to be both feasible and a valued service. These candidates included:

Page 8 of 15

Dryad Consortium MB report Spring FINAL.doc

by Todd Vision & Peggy Schaeffer, June-August, 2009

[https://www.nescent.org/wg\\_dryad/Management\\_Board\\_Meeting\\_Spring\\_2009](https://www.nescent.org/wg_dryad/Management_Board_Meeting_Spring_2009)



- The Paleobiology Database
- Functional genomic data repositories such as Gene Expression Omnibus (U.S.A.), ArrayExpress (Europe), Center for Information Biology Gene Expression database (Japan), and the Stanford Microarray Database
- OBIS: Ocean Biogeographic Information System
- Image databases such as Morphbank and MorphoBank

## **9.2 Repository policies**

Dryad staff presented a number of proposals regarding repository policy for considering of the management board.

### **9.2.1 Storage limits**

Storage of a few very large datasets could considerably elevate the costs of the repository, may often be more “raw” than required to satisfy the JDAP, and at present cannot even be efficiently transferred from the user to the repository. The storage cost for the average Dryad record is currently ~\$0.0045/yr, as compared to ~\$1,000 for hypothetical 1 TB dataset, or data package. Given the budget for hardware in the current NSF grant, we could store the data for nearly 7 million papers of the current average size, while we could store the data for only 41 of the hypothetical genomics papers. Thus, Dryad staff recommended to provisionally limit individual data files to 1 GB, and 10 GB for the sum of all data files, i.e. the data package, associated with one paper. (corresponding to a maximum storage cost of ~\$1/yr per file and ~\$10/yr per paper). This proposal was approved by a vote of 9 for, 0 against, 4 abstaining. Concern was expressed that requiring deposit, and then disallowing deposit for certain datasets, could send a mixed message, and Dryad staff were urged to explore a surcharge for data packages that are appropriate for archiving yet over the limit.

### **9.2.2. Data Submission and Modification**

A number of issues revolve around the issue of who has the authority to deposit and modify records. The board generally agreed with the proposal of Dryad staff to implement the following policies:

- One of the original authors must deposit the data and agree to the data usage terms (except in the case of legacy data obtained by curators, where permission from an author must still be obtained).
- For invited depositions, the corresponding author would be provided with a PIN, and anyone with a valid user account on Dryad, and the PIN for a publication, would be able to make modifications to that record.
- All new data packages, and modifications to existing packages (including deletions), need to be approved by a curator before going live.
- No data package should go live before the DOI (or equivalent) for the paper has been assigned and the final publication metadata has been obtained.
- Datasets receiving the most use will receive the most curator attention.

- All data packages will be versioned, and prior versions will be stored in a dark section of the repository.

### ***9.2.3. Conventions for data citations***

A community norm document to accompany the CC0 license would ideally provide guidance on how data found in Dryad should be cited. Dryad staff recommended that the board set a policy for all partner journals to follow in order to encourage consistent citation practice. There was heated discussion over what elements should be included in such a citation, and the relative emphasis that should be placed on the Dryad record relative to the original publication. A nonbinding vote was taken to determine the sense of the board as to whether the convention should be [1] to cite the original paper (assuming the data identifiers are available within the paper), [2] to cite the data record directly (assuming citations to the data identifier can be tracked), or [3] to cite both. Strong arguments were made on all sides. There were six votes for [1] and [3], and one vote for [2]; One of the points raised was whether the best solution depended on the number of data files being cited in one paper. It was decided that this issue would need to be discussed further among the executive committee, with input from the data citation study group within DataONE. The executive committee would in turn present a recommendation in six months time. In the meantime, the suggested format of data citations on the Dryad website will remain in effect.

### ***9.2.4. Identifier scheme***

Dryad currently assigns regular Handles to all individual data files and to the data package as a whole. The board was asked to consider the merits of assigning DOIs (a special type of Handle) instead. The pros would be that DOIs are well-known by scientists, facilitating their adoption, and that certain DOIs are supported by indexing services (e.g. CrossRef, ISI). The cons are that they cost somewhat more to assign (depending on the registry, anywhere from \$0.06 to \$0.40 per identifier) and carry a somewhat higher administrative burden. Dryad staff proposed that DOIs be used for the data package as a whole, while Handles continue to be used for individual data files. Some board members familiar with the technology urged that Dryad use CrossRef registered DOIs, to demonstrate seriousness of purpose and so that the full benefits of citation services would accrue. There was some discussion of how versioning would affect the identifier. Overall, the board felt that this could not be decided without further information, and that the decision rested in some measure on the decision regarding data citation. It was agreed that the executive committee would need to discuss the issues surrounding identifiers and provide a recommendation in six months time.

## **10. Metadata issues**

Jane Greenberg, professor at the School of Information and Library Science at UNC Chapel Hill, presented an overview of how metadata is being used within Dryad. The emphasis was on the competing goals of engineering a system that can take advantage of rich and high quality metadata while at the same time reducing the burden on depositors

*Page 10 of 15*

*Dryad Consortium MB report Spring FINAL.doc*

*by Todd Vision & Peggy Schaeffer, June-August, 2009*

*[https://www.nescent.org/wg\\_dryad/Management\\_Board\\_Meeting\\_Spring\\_2009](https://www.nescent.org/wg_dryad/Management_Board_Meeting_Spring_2009)*

of providing that metadata. Dryad is being designed with an expectation of minimal depositor-required metadata, some metadata enrichment and quality control using automation, and some professional metadata quality control (i.e. curation). The board discussed the metadata requirements for the submission interface. Management board members generally agreed on the following points:

- Data deposition should not take longer than about ten minutes. The system will need to be non-burdensome even for depositors with large numbers of files (e.g. many individual images).
- It will be important to employ automatic techniques to expedite the metadata generation process and help with quality control.
- Content metadata should be pre-populated, where possible, from article content (e.g., abstract and keywords, possibly also text and data captions). Experiments with metadata generation using full text should be undertaken to determine its utility.
- Provisions should be made allowing for the optional creation of metadata that is richer and more extensive than the minimal requirement.
- Standardization of metadata values is important (e.g., species/taxonomic names, geographic names and coordinates, temporal data) in order to ensure its usefulness. At the same time, interest was expressed in providing some "open" metadata fields, where depositors could optionally share additional (non-standardized) information on the data being deposited.
- Many Dryad depositors will have had some experience in assigning metadata from their use of other archives, but examples should be included to aid depositors during metadata generation, and having a trained curator to help maintain quality control will still be important. The board did not express concern about curators heavily editing metadata, as long as the original author-provided metadata could be recovered.
- There appears to be some debate about the copyright and terms of use for publication metadata (abstracts, keywords); this issue will need to be resolved in order to determine how much of these data can be used and redistributed within Dryad.
- It was also noted that many journals do not have keywords at the time of acceptance and so cannot provide them to Dryad at that time (e.g. *American Naturalist*); a separate process for acquiring those keywords will need to be developed.
- Some interest was expressed in visualization tools for aiding metadata creation (e.g. a graphical map for selecting coordinate boxes or geo-referenced place names), but, at the same time, the board expressed caution about the development effort that would be required to implement such interfaces.
- There is some interest in Web2.0 features, such as community ranking and tagging with folksonomies, to enrich the metadata after submission, but this is not perceived as critical for early adoption.

The results from a 2008 survey of >400 evolutionary biologists on data sharing attitudes and practices was discussed in light of what it suggested regarding system requirements. One surprising finding was the extent to which respondents valued data archiving as a way to allow repurposing of data. This suggests that the ultimate requirements for detailed metadata will be hard to predict in advance for any given study, and that there may be a limited return on investment for attempting to capture the metadata that the author thinks is most important at the time of submission.

The board also discussed how the mandate for accelerated integration with Partner journals (see 9.1 above) will necessitate simultaneously accelerating efforts to build up curation capacity, including developing a more automated curation interface, optimized curation procedures and a trained curation staff. The time required for curation needs to be the subject of careful optimization and experimentation, so that the costs of curation can be controlled in the overall budget.

## 11. Next steps

The board agreed upon the appointment of an interim executive committee to provide leadership in the interim between now and the next full management board meeting. The committee consists of Mike Whitlock (chair), William Michener, Allen Moore, Marcy Uyenoyama, Tim Vines, and Todd Vision (*ex officio*). The committee is empowered to include other members, as needed, for specific input. The executive committee will serve until a vote on the constitution and by-laws provides a more permanent governance structure. The charge of the executive committee is as follows:

- Propose governance documents (Constitution, Bylaws).
- Propose a "community norms" document that specifies the data citation policy.
- Propose an identifier scheme for data files and data packages.
- Developing a phased strategy for what needs to be in place in order to make data deposition a requirement.
- Coordinate a simultaneous editorial on data archiving in the Fall.

The board agreed to meet again in late fall/early winter (likely December), at a site to be decided by the executive committee. Dryad staff will investigate European meeting sites, which will be valuable in emphasizing the international nature of the initiative. Representatives from European funding bodies may be invited to participate at the meeting.

The issue of community engagement was discussed. It was agreed that stakeholders need to be kept better informed about the progress of the initiative. A number of upcoming opportunities were enumerated:

- An oral announcement about Dryad during a special graduate student session at the Evolution meetings in Idaho in June (Allen Moore)

- A briefing to the Joint Council of SSE/ASN/SSB (Joel Kingsolver), the AGA Council (Scott Baker), the ESEB council (Allen Moore), the ESA (William Michener), and the SMBE council (Marcy Uyenoyama), among others.
- Introductory slides and handouts will be posted on the Dryad wiki for people to use when presenting to stakeholder groups (Peggy Schaeffer)
- An editorial about the JDAP to be published in one or more Partner Journals this fall (Michael Whitlock)
- An announcement about the availability of Dryad to appear in journals, society newsletters (Peggy Schaeffer will coordinate with each Partner)
- A general press release and EvolDir/ECOLOG/NESCent newsletter announcement to be released once integration is accomplished with a critical mass of Partner journals (Peggy Schaeffer & Robin Smith, to be reviewed by executive committee)

The Dryad Consortium currently includes a large number of the top, discipline-specific journals in ecology and evolutionary biology that have pan-taxonomic scope. The board discussed when would be the best time to actively invite additional Partners, assuming that the board feels a larger consortium would be both beneficial and manageable. The general sense of the board was to hold off on expanding further invitations to the Consortium until the current Partner journals have been integrated and the open issues currently under discussion (especially concerning governance and sustainability) have been resolved.

## *Appendix*

### **Meeting Participants:**

#### Journal representatives:

- *The American Naturalist* - Michael Whitlock (University of British Columbia)
- *Biological Journal of the Linnean Society* - John Allen (University of Southampton)
- *Conservation Biology* - Erica Fleishman (University of California Santa Barbara)
- *Ecological Applications, Ecological Monographs, Ecology, Frontiers in Ecology and the Environment* - William Michener (University of New Mexico)
- *Evolution* - Mohamed Noor (Duke University)
- *Evolutionary Applications* - Michelle Tseng (University of British Columbia) [remote participant]
- *Integrative and Comparative Biology* - Harold Heatwole (North Carolina State University)
- *Journal of Heredity* - Scott Baker (Oregon State University)
- *Journal of Evolutionary Biology* - Allen Moore (University of Exeter)
- *Molecular Biology and Evolution* - Marcy Uyenoyama (Duke University)
- *Molecular Ecology* - Tim Vines (University of British Columbia)
- *Molecular Phylogenetics and Evolution* - Derek Wildman (Wayne State University)
- *Paleobiology* - Peter Wagner (Smithsonian Institution)
- *Systematic Biology* - Rod Page (University of Glasgow) [remote participant]

#### Other participants:

- *NESCent*: Hilmar Lapp, Peggy Schaeffer, Ryan Scherle, Robin Smith, Kathleen Smith, and Todd Vision
- *University of North Carolina School of Information and Library Science*
  - *Metadata Research Center*: Jane Greenberg, Sarah Carrier, and Hollie White
  - Lorraine Eakin
- *NC State University Digital Libraries*: Kristin Antelman
- *Science Commons/Duke University School of Law*: James Boyle

### **Tasks for Dryad staff in the next six months:**

- Integrate with Partner journals on an accelerated timescale; improve the process for collecting final publication metadata.
- Scale up curation capacity in anticipation of an elevated rate of deposition.
- Internationalization: increase visibility outside the U.S. and plan for a global reach in all activities (e.g. translations of the website into languages other than English).
- Continue public relations efforts to raise the visibility of Dryad among prospective data depositors and users, including coordination of the public announcements listed above.
- Sustainability planning: develop a business model that articulates benefits to stakeholders, costs of repository operation and projects potential revenue streams.
- Gather data about the impact of Supplementary Data Archives on article citation rates.
- Draft a document describing roles and responsibilities for Partners, and a technical guide to journal integration.
- Research the technical and financial details of DOIs with CrossRef, DataCite, and other potential registrars.
- Replace current licensing with CC0.
- Assist executive committee with tasks listed above.
- Arrange next DCMB meeting.
- Revise Yr2 repository development plan.