http://datadryad.org

Notes from the DryadUK Workshop

The British Library
April 1st 2011

Brian Hole
DryadUK Project Manager
Brian.hole@bl.uk

Contents

Introduction

The DryadUK project held a workshop on sustainability at the British Library on April 1st. The aim of the workshop was to gather stakeholder feedback on appropriate funding models for Dryad as an international organisation, and examined mixed funding models involving subscriptions, submission fees, hosting of services, and grants.

We would like to thank the participants of the meeting for generously providing their time and expert input on these issues, as well as Kevin Ashley from the Digital Curation Centre who did an excellent job of facilitating.

The workshop began with an overview of the costs of running repositories by Neil Beagrie, followed by a presentation of the current Dryad fee structure by Todd Vision. The workshop then split into four breakout groups based on stakeholder type (publishers, journals, funding bodies and researchers), who discussed the following three questions:

1. What is the level of demand for Dryad in your community?
2. What should the funding model be?
3. What role should funding bodies play?

Following a break the groups then reconvened and discussed and debated the outcomes of the breakouts as one body. Participants were also encouraged to leave anonymous written feedback. The notes from the breakouts, plenary session and written feedback follow. As the workshop was held under the Chatham House Rules, all speakers have been anonymised, other than indicating which stakeholder group they belonged to. All participants are welcome to quote this document, as long as they continue to respect this rule.

Summary

The group and plenary discussions were lively and represented a range of positions and opinions between stakeholder groups of publishers, journals, funders and researchers. The following is a summary of the main themes that emerged.

Value to the community

There was consensus across the groups that Dryad offers increased discoverability, citation and tracking of data, along with peer review, all of which are highly valuable. There was recurrent discussion of whether Dryad should also take raw data, based on the fact that this is currently not being addressed well elsewhere. Several larger publishers expressed the view that Dryad needs to move to providing this kind of added value (also in terms of more curation) as otherwise they have the resources to modify their existing systems to achieve the same end. Smaller publishers and journals felt that Dryad would improve their status. There was some concern addressed by various stakeholders that the governance model does not sufficiently incorporate the interests of researchers, who nevertheless themselves were enthusiastic about Dryad's ability to make data available and trusted, while helping them to demonstrate impact.

The funding model

There was general consensus that the current funding model does addresses the situation of the most publishers and journals well, although large publishers felt that they should receive a discount for providing a greater volume of submissions, and that for this to be paid in advance it would need to be based on a realistic estimate of the number of submissions likely. There was some concern expressed that authors might be put off depositing via plan C, or might choose an A or B subscribing journal instead. Funders on the other hand were worried that plan C might result in a significantly increased burden for them if enough authors went to non-subscribing journals.

The role of funding bodies

Publishers, journals and researchers all voiced the opinion that funders do need to provide some support for Dryad. Suggested methods were for funders to ring-fence funds for data deposition, and to increase and enforce mandates. It was also mentioned that funders should directly support Dryad because they would save money through reuse and more efficient research in the long run. The funding bodies themselves were not inclined to fund Dryad directly for data deposition, but that they are open to encouraging inclusion of reasonable deposition costs in grant applications. They would be more open to funding research and innovation, and potentially to providing bridge funding as a consortium to help Dryad become self-sustaining. They also see publication data as only one part of a greater problem, and might be more compelled to fund the preservation of raw data as well.

Breakout session notes

Society publishers and researchers

Value to the community

Demand for Dryad was seen to be somewhat difficult to ascertain for research institutions. Any initiative that could increase access to the amount of data stranded in research labs is seen as highly positive, but here was concern that researchers could be charged a second time if also expected to deposit with an institutional repository, although most institutions seem more focused on big data at the moment. The fact that Dryad is working in partnership with national institutions such as the British Library was seen as a counterweight to this, as HE institutions may be more willing to see their data stored with such partners as an alternative.

The funding model

Society publishers see great value in being able to monitor the contributions of society members, authors and journals, and especially in being able to distinguish the impact of data as well as of articles. Dryad will also be a useful resource for supporting systematic reviews across the sciences, something that was seen as being of great, in addition to increased discoverability and accessibility, particularly for interdisciplinary work. For all societies but for the smaller ones in particular, there could also be the added benefit of raising a journal's impact factor and profile as a result of an increase in citations.

The role of funding bodies

Regarding the funding model, overall it was felt that the value proposition of Dryad was more compelling for smaller publishers, as they do not have the resources to look after data themselves, and may not trust big commercial interests. On the other hand, big publishers might be better able to run a data repository themselves, and to leverage this for commercial benefit. Dryad could therefore be key in getting to the long tail of small publishers and societies, and also offers a way for them to achieve author data deposition without paying for it themselves. It was thus felt likely that models B and C would become more popular as more smaller players became involved. The idea of larger societies absorbing the cost as a membership benefit was also discussed and found to have merit.

Non-society publishers

Value to the community

The views of this group were somewhat mixed, reflecting the fact that publishers are not homogenous in how they work. This was reflected in discussion that perhaps the Dryad proposition had been designed to this point around a relatively small sample of journals, and it would benefit from the input of larger publishers in particular. Some publishers saw great benefit in Dryad as a way to lessen the problem of dealing with supplementary material, and there was discussion about the degree to which this would happen, or whether Dryad could be persuaded to take all supplementary material and not only data. This was one example of the kind of value it was felt Dryad could add, and that it needed to articulate this better in terms of making life easier for publishers.

The funding model

The focus on added value was carried over into the discussion about the funding model. It was agreed that the costs need to be directly related to the value added, and that this needed to be transparent. Scalability was seen as a problem with the current model for bigger publishers in particular, both in terms of potentially continually mounting fees, and also accumulated indirect costs. It was felt that the former needed to be addressed by changing the model, while the latter needed to be better quantified to enable decision-making. Another topic was the need to bring the research community on-side, with concern noted that the option C fees were very high, and that it would be good to look at adding a waiver option to the model for disadvantaged authors.

The role of funding bodies

It was agreed that funders need to give credit for the deposition of data, and it was proposed that they ring fence a portion of funding for this activity. As part of this, Dryad could be seen as playing an integral role in supporting mandates.

Funding bodies

Value to the community

The funders group were generally in agreement that Dryad helps to tackle problems that are of importance to them. One research council reported that their data policy now includes a specific section on the importance of access to data underpinning research publications, and sees Dryad as being of help in this area. Another noted that this was a concrete way of solving the 'data behind the graph' problem. There was discussion about the actual value of the data, and it was noted that even if only 10-20% of it is of real value, it only takes one major reuse case for the cost of the repository for a year to be justified. It was pointed out that funders are concerned about all data (i.e. raw data as well as processed and publication data), and that while Dryad does partially address their priorities, they would like to see it consider taking extra material as well. Another concern was that, as a journal-driven system, Dryad might not be as representative of researchers' interests as funders would like, and thus also not as capable of inducing step-changing behaviour, so this should be kept in mind.

The funding model

With regard to the current funding model, it was felt that charges needed to be significantly lower than typical APCs, set at around 10-20%. There was a worry that a high charge for plan C would simply result in funders being asked to pay more when a journal was not part of Dryad, and it was proposed that it be a simple increase such as A: £20, B: £30, C: £40.

The role of funding bodies

There was consensus that funding bodies should not fund Dryad's operating expenses directly, but they do see a role for themselves in funding it indirectly, through payments to researchers and research institutions to deposit data. They would be willing to fund innovation should Dryad apply for this in conjunction with an eligible institution. There was however also agreement that a funders consortium might be possible to provide short-term bridge funding as the repository became established, along the lines of that provided to UKPMC.

Ways of lowering the cost of operating the repository were also suggested, including sponsored storage by companies such as Amazon, and the idea that Dryad could operate as the front end, accepting the data, curating the metadata and making it discoverable, but behind the scenes passing the data itself on to the research councils to store.

It was also felt that besides the funding bodies other institutions such as the British Library would be needed to support the research community in this. At the same time the publishers are seen as having to contribute something as well, with a good chance of recovering it through the added value to their subscribers of data being made available.

Journals

Value to the community
The journals group were in agreement that Dryad was of benefit to their community in a range of ways. These included the extra traffic generated to the journals from the datasets, with a possible increase in citation and impact factor, along with the perception of authors and readers that the journal and its data were widely accessible. Curation of the data was also seen as an important assumption of responsibility on the part of Dryad, and a relief for the journals. Making data available for review was also thought to be beneficial, as while it might not always be heavily reviewed, it could be very useful in resolving situations such as plagiarism. Even though difficult to quantify exactly, once again the prospect of Dryad taking care of all supplementary material and making it citable was seen as potentially high added value.

The funding model
At the journal level there was little discussion of the benefits of one model over the other – it was generally accepted that the proposed funding model was workable. Some members of the group were already using Dryad, and reported that they were trialling absorbing the Dryad deposit costs into their current APC. If this was not possible then rather than have an additional charge for data, they would like to look at recovering the cost from funders directly. For subscription-based journals the option of rolling the costs into subscription charges was also discussed, and it was felt that publishers should probably take a mixed approach, depending on what was sustainable for each journal individually, depending on the overall sales and income of each. In order to counter scalability problems, it was suggested that publishers should receive a discount on the cost per deposition if they contributed a large number of journals. For plan C, extending current APC discounts and waivers with developing countries was also thought to be necessary.

The role of funding bodies
It was hoped that funders would support Dryad and publishers as they would also benefit from the system, for example through less repetition of experiments, and the potential development of standards for data types.

Plenary discussion notes

Each of the breakout groups presented a summary of their sessions, and the issues arising were then openly discussed.

Value to the community

It was acknowledged that the strength of Dryad lies in its tackling of the discoverability and preservation of long tail of publication data including the 'data behind the graph', as well as a wide range of data types. A range of publishers expressed concern that Dryad only accounts for the 15% of data associated with publication, and does not address the most pressing issue of the remaining 85% of research data. This was countered by the fact that Dryad can in fact address this demand if it is what its members want. At the moment it is focusing on journal requirements for enough data to be made available to verify the statistics in a paper, according to the Joint Data Archiving Policy (JDAP). If there were demand from journals to expand this to include all data it could be done, either as a whole or on a journal-specific basis.

The value to researchers of Dryad was expressed as greater trust in the authority of participating journals, as published research and data would have been validated to a much higher standard. Several participants expressed the view that researchers should have more input to Dryad's policies, and that this could be achieved by some kind of a role for publishing societies.

The funding model

In regard to the additional costs that could be experienced by publishers, some division in expectations was expressed. Smaller publishers did not expect significant increases, and did not feel that making it available for peer review would be costly, but that it would in fact make this process easier. One journal that is already integrated with Dryad reported that the process was indeed simple and inexpensive. Larger publishers however continued to express concern that despite economies of scale, the accumulated costs would be more significant for them, and that they would require more customisation work. There was also discussion of ways to bring down the costs of storage, and agreement that through an economy of scale Dryad would be of most benefit to smaller journals and publishers.

The pricing plans were discussed, with a general consensus that they needed to be tailored to fit both large and small publishers. While some larger publishers stated that even a deposit fee of $25 could be too high when dealing with large numbers of journals, others stated that they would be willing to pay more if Dryad were able to demonstrate greater added value, principally by taking supplemental and raw data as well, and with additional levels of curation. This was also an issue for funders – if they are expected to pick up the costs of plan C submissions, the additional value associated with the higher cost needs to be made clear. It was also suggested that plan C could be priced at different levels depending on discipline.

Larger publishers also expressed a wish that advance payments be based on the actual number of submissions likely, rather than on a flat per journal basis, which would for example be more equitable for journals where submission was likely for only 5% of articles.

The role of funding bodies

There was initial debate about the nature and degree to which funders should support Dryad, based on a perceived gap between publisher, journal and researcher expectations and funder willingness. It was suggested that besides financial support, funders could also do more in terms of encouraging and even mandating data deposition, for example by only agreeing to new grants to researchers who have deposited data from previous funded work. Overall the funders were supportive, and the point was made that their incentive to fund Dryad would be even greater if it took a wider range of materials, especially raw data.

Written feedback

In addition to the breakout and plenary discussions, participants were also offered the chance to give anonymous written feedback. An additional question was also asked about the appropriate article charges for the proposed plans B and C.

Value to the community

Researchers described Dryad as addressing a situation where the majority of data is currently unobtainable, helping to demonstrate impact, increasing chances for collaboration, and providing a level of trust for shared data. Some publishers expressed a preference for hosting such data themselves, but several others were happy that it would save them from having to do so, with Dryad providing technical and subject expertise that they do not have. These publishers also felt that having data in Dryad would add to the value of the paper and thus also the journal. Journal representatives tended to list the most benefits, and especially to stress the advantages of open data in terms of increased availability, discoverability, citation, and reputation for the journal itself.

The funding model

Overall publishers commented that the proposed funding model was adaptable and affordable, especially for smaller publishers, and there was again some concern from the larger publishers that a per-article fee would be too expensive for them (if not all articles contributed data). Several also commented that a mix of funding sources would be a good thing, indicating wider acceptance and providing greater independence. Concern was once again voiced by some publishers and journals that plan C might be too expensive for many authors.

The role of funding bodies

Direct support for data publishers, and ring-fencing of funds within grants, especially if are mandating deposition. There was a general consensus among publishers, journals and researchers that funding bodies should commit funds if they see this as a priority issue, and that it would be most effective if they did produce clear data publishing policies, requirements, mandates and advocacy.

For their part, funders wrote that they would be the ultimate payers, and that this would be best achieved mostly through providing support through a percentage of grants, as well as some direct support for further development.

Appropriate article charges for plans B and C

This was an interesting question, as during the open discussion these plans had been described as both too high and too low by some participants. The average level suggested for plan B was around $25-$50, with not too much deviation. The responses to plan C were divided however, with half proposing a price of $50-$100, and half believing it should be over $125. It was pointed out that this amount needs to be fair, reasonable and transparent, while at the same time it should be higher than the cost per paper of plans A and B due to more curation work being required than for a non-integrated journal.

Appendix – The proposed Dryad fee structure presented at the workshop

	A *Full*	B *Associate*	C *Author pays*
Joining fee (waived for charter members)	$1000 / £624	$1,000 / £624	NA
Annual fee from journal	Prospective	Retrospective	0
a. all peer-reviewed articles in prior yr b. articles with data deposited to Dryad	$25/article / £16/article	$100/article / £62/article	
Author charge at deposit	0	0	$200 / £125
Length of contract	3 or 5 yrs	3 or 5 yrs	n/a
Legacy data deposits free?	Y	N	N
Can move between plans A & B	Y	Y	n/a
Representative on Consortium Board	Y	Y	N
Can vote on board and serve on executive committee	Y	N	N
Coordinated data deposit	Y	Y	Y/N
Data DOI in published article	Y	Y	Y/N
Branding of journal content	Y	Y	N