Case Study: Dryad Digital Repository

Prepared for the ESA Workshop on *Creating and Implementing Sustainability Plans for Data Repositories*[1]
Nancy Maron, BlueSky to BluePrint

Background

The Dryad Digital Repository was created in 2009, when leaders of several societies and research journals in the field of evolutionary biology and ecology sought to create a community-based solution for archiving their research data. Dryad was launched with funding from the National Science Foundation, and the express mission of serving researchers in the sciences whose data sets – whether spreadsheets, video files, images, or other – had "no natural home" in other repositories.

In 2017, Dryad published 4,307 data packages, 10% more than the prior year, each containing on average of 3 files. It also supports researchers by making those data freely available; in 2016, there were approximately 475,000 file downloads of files in the over 20,212 data packages currently in Dryad.

Sustainability history

From the outset, funders and project leaders alike understood that sustainability was a goal of the project; by the end of the grant, the plan was for the project team to have developed a long-term support model.[2] To that end, in 2012, the team decided to spin off to become an independent 501c(3). Early governance came from an advisory board between 2009 and 2012, and then a formal Board of Directors with its own set of bylaws.

The sustainability strategy was to charge fees at the time of data publication (a Data Publication Charge or DPC) which would be fashioned after the APC, but be "much more affordable." According to Executive Director Meredith Morovati, "the DPC scales with use, is transparent and fair, and waivers are available for developing countries." Until 2013, Dryad had just brought in the data, without charging anything. The fee-based model began in the Fall of 2013, in partnership with the Ecological Society of America (ESA) and others, so that automatic workflows could facilitate data deposit, alongside publication of scholarly articles.

Charging for something when the competition is free

Meredith notes that it can be a challenge to have a fee-based service, when there is a popular competitor offering a similar service for what appears to be free to the user. She cites FigShare, which came to market around the same time that Dryad did, and more recently, Zenodo. Since both appeal to researchers and are free to end-users, Dryad's challenge is making its value proposition very clear to its target audience, who are faced with a range of options.

---

[1] Information in this article is based on interviews and email exchanges with Meredith Morovati, Executive Director, Dryad, as well as documentation provided by Morovati and available online.

[2] Dryad Annual Report 2016: http://datadryad.org/themes/Mirage/docs/DryadAnnualReport2016.pdf

Value Proposition

Leadership at Dryad see "data curation" as a key differentiator, one that they expect their target market of scholars and scholarly publishers to appreciate and be willing to pay for. While FigShare and Zenodo accept many content types, they lack a mechanism to confirm that it is academic in nature. Nor do they curate the data for improved quality and discoverability.

Indeed, in addition to being a not-for-profit organization, which can be appealing to scholarly societies, Dryad offers a level of curation to ensure that the data they host has been reviewed. First, they only accept data associated with scholarly, peer-reviewed or vetted content. Second, they employ a range of further activities, including (1) Simple checks, where staff open and review files to check that all permissions are in place for the content to be uploaded, since often authors will occasionally include items – such as work that is the Intellectual Property of someone else – in their materials; (2) Cleaning up metadata; and (3) Offering support via a help desk, to prompt researchers to maybe add something else, when needed. Reviewing for licensing issues is increasingly important, as is making sure that data do not contain human subject or rare species location.

Shifting Understanding of Audience

When Dryad started, their outreach efforts were primarily to the researchers themselves, those who would have the materials to deposit. Over time, they realized that this focus might need to shift. Among other things, marketing and outreach to individual users "is expensive and harder convert into a significant jump in users." But more fundamentally, while data curation seems to be a "nice to have" for researchers, curation services played a more immediately important role for publishers. Dryad staff, in vetting the data prior to ingesting it, review privacy issues and check to make sure links go to quality, academic data.

This has led to a strategic shift and a focus on pinpointing the publishers and institutions who are the core of the partnership strategy Dryad is pursuing, as a means to expand beyond life science, while doing it in a way that is targeted and efficient. Today, they have partnerships with several journal publishers, including Evolution, Molecular Biology, the Royal Society, some BMC journals, eLife. In 2013, an integration with PLOS journals helped them to expand beyond the discipline of ecology. Any organization may become members – and for an annual fee of $1,000 (for institutions with gross income of under $10 million) or $5,000 (for those with gross income above $10 million), they participate in governance of the organization and benefit from discounted submission rates. For a fee of $500, organizations may be members but without the benefit of discounted submission plans.

Dryad's understanding not just of users, but of potential paying customers has also developed over time. At first, according to Meredith, "we assumed that the fee was being paid by the grant, but often people were just paying for it." In 2016, they needed to increase the fee; the management team had found that the prices were not covering all of the costs of the work they were doing, and that prices had been kept low for three years. They were covering the majority of the direct expenses, but "not contributing back to running the organization the way we needed to." They raised prices in 2016, so that the highest per item rate on the schedule is not $120 (from $90). While they saw some fall off in individual users, "revenues stayed the same and slightly increased."

Marketing: Louder, more often, in more ways

She describes part of the challenge in Dryad's model as owing to the need for "culture change" among researchers, publishers, and everyone in the scholarly communications lifecycle; the service Dryad provides is "not always something that researchers feel they need to do, or are mandated to do… so part of the work of Dryad is to convince them of the value of archiving their data."

This means that communications and outreach are a key strategic aspect of their work. For Dryad, this involves not just one-off messaging at a specific conference, but being a voice and active in the community, which they do, for example, through social media. With over 12,200 followers, the Dryad Twitter feed serves as a means to share community information, as well as promoting the organization.

Meredith underlines just how important it is to get the word out about the service: "It's not just showing up to a conference once a year… We are seeking to encourage a culture change" which takes a lot of work and time. Press releases, blog posts, ongoing communications, are all part of the ongoing campaign to have relevance. She notes, "you always need to say it louder, more often, in more ways" to get the point across.

Sustainability Model Today

As of FY 2016, the operating costs of Dryad were approximately $750,000. This includes 9 staffers, of which 3 are part time. Of total annual revenues, 60% come from earned income and 40% from ongoing grant support. Some projects are still running and NSF does offer some support for operations.

Membership support yields approximately $180,000, and grants bring in around $300,000. Submission Fees in 2016 were $267,249, as documented by Dryad below.

Revenue, by payment plan, 2016 (in USD)

	Q1	Q2	Q3	Q4	Total
Subscription Plan	$27,576	$17,500	$13,228	$2,173	$60477
Voucher Plan	7,135	2,625	28,900	5,670	44330
Deferred Payment Plan	12,707	10,710	-	10,605	34022
Single vouchers	-	240	-	-	240
Individuals	30,035	35,825	32,130	30,190	128180
Total	**$77,453**	**$66,900**	**$74,258**	**$48,638**	**$267249**

Next Steps

Meredith sees the potential for further growth for Dryad is in getting more researchers to see the value in depositing data. She notes that despite funder mandates concerning DMPs and article deposit, there are still loopholes that researchers routinely use. For example, in a data management plan, it is possible to just check a box when talking about "supplementary materials." While answering the question, it permits scholars to side-step a direct response concerning where and how those materials are being deposited.

As they look ahead, the Dryad team continues to think about how much more data is out there? How to convince more researchers to deposit? How to reduce the friction in that decision?