Senior Curator Manual

Version 1.3

February 2014

Versions 1.0-1.3 by Sara Mannheimer, August 2013-February 2014
1. Get to know Dryad

• Browse Dryad content. Some good sample content is on the wiki here:
  http://wiki.datadryad.org/Sample_Dryad_Content
  – Please keep the sample content wiki page updated for outreach purposes.
  – If you happen to see something that is unusual – e.g. a good example of filetypes, types of publications accepted, differing authors for article/data, etc., add it!

• The user’s perspective: using dev.datadryad.org or staging.datadryad.org, submit test data packages to get a feel for the system. (Either testing site works. Dev can be unstable, but staging is not updated as often)
2. Curation practices

• Use the curation manual, found here: http://wiki.datadryad.org/Curation

• Not-so-frequently asked curation questions, the curation whiteboard, and other useful documents are available in the Dryad Google Drive
3. Releasing data from publication blackout

- Assistant curators regularly search for articles that have been published, then approve them out of pub-blackout.
- We also sign up for new article publication alerts from integrated journals.
- In the future, CrossRef will allow us to streamline this process. See Trello #20: https://trello.com/c/0pPUMt0W
3.1 Embargoes

• Files can have one of three embargoes: untilArticle Appears, oneyear, and custom. For “custom,” the length must be specified.
  – Usually, we’ll receive an email from the journal approving a certain embargo. Add the length in the file’s dryad.curatorNote field.
  – If the embargo is listed as “custom,” but the length isn’t specified, email the author (see Templates for Correspondence).

• Embargoes are always set at the file level. Edit the length of embargo in the file’s “item metadata” tab, in the “dc. type. embargo” field

• Once data is published, a curator manually updates the embargo date in the file’s “item embargo” tab.
  – Click “lift embargo now” if the embargo is “untilArticleAppears”
  – For longer embargoes, refer to the publication date, and enter the appropriate embargo end date in the “embargoed until date” field.
  – This process must be repeated for each file in the data package.
4. Review workflow

Some integrated journals allow authors to submit data to Dryad while the article is still in review. (see http://wiki.datadryad.org/Submission_Integration:Overview for more info) Dryad is technically not doing anything with this data. There is no data publication charge for submissions into review (see 13. Payment system).

- The curator@gmail.com account receives an alert email when items are submitted to the review workspace, titled “Data package received for review doi:xxxx”
- The process isn’t automated yet, so Ryan watches for a metadata email that says “accepted,” after which he pushes the data package through to curation (the “In Curation: Unclaimed” queue).
- Before assigning a submission previously in review to assistant curators, look at the acceptance email (in dryad.journal.submit@gmail.com) to make sure that the title and abstract are correct. Especially for MolEcol submissions, these may have changed since the data was deposited into review.
- Authors can only ADD files at the review stage. No other modification is possible. This is so that authors’ actions are completely transparent. We are working on adding status indicators, but for now, the author must email help@datadryad.org to tell the curators which files to keep or delete during curation. (See 7.3. and Templates for Correspondence, “Modifying a submission in Journal Review”)
4.1. Adding articles in review to the spreadsheet

When data is submitted associated with an article in review, we enter it into the spreadsheet. This can either be done by the senior curator, or delegated to an assistant curator.

- To add data in review to the spreadsheet:
- Refer to the email notification in the curator@gmail.com account (see 4. Review workflow) for title, author, doi, and review workspace URL.
- Search the spreadsheet for the title and manuscript number to make sure that it isn't a duplicate.
  - If it’s a duplicate, enter in red in the “Notes” column of each duplicate: “has dup associated with [earlier/later] submission.” Indicate which submission is most recent.
  - Duplicate submission warnings are in process of being created. Ideally, if someone has data in review and tries to re-submit it in the first-time submission area, they’ll be taken to their existing submission.
- Refer to the metadata email (dryad.journal.submit@gmail.com) to see if the submitter is the lead author. For submitter Smith and lead author Jones, write “Jones (submitted by coauthor Smith)”
- In the final field, paste the URL from the “received for review” email
- Make the row’s background color light gray.
5. Workflow Overview

Accessible from the Profile menu. From Workflow Overview, you can see all the data in Dryad, regardless of the workflow stage.

- When a submission is accessed via Workflow Overview, you can choose “change workflow step” from the profile menu to push an item from one workflow step to another (e.g. from review into curation). Ryan believes that this function may not always work perfectly, depending on the specific step/movement. Always keep a close eye on items whose workflow step you change to make sure nothing breaks.

- Via Workflow Overview, you can also click “return to submitter” to return an item to an author’s personal workspace. Type a reason, then click “Back to Submission.” It’s weird, but it works!
6. Fogbugz

- Emails sent to help@datadryad.org arrive in this system, which is shared by several Dryad staff members (although not accessible by assistant curators).
- Check these first thing to make sure they don’t apply to items in the “In Curation: Unclaimed” queue.
- It is the senior curator’s responsibility to respond to help tickets or assign them to other staff (most often Laura or Ryan).
  - When assigning to another staff member it often helps to:
    - change title to something that is more descriptive
    - set priority (levels 1-6)
    - add a note about what you know and why you're assigning it away.
- When responding to tickets, click “Send and Close.” We keep track of which help tickets need attention by which are still open.
  - If you don’t need to respond, click “Resolve,” and choose a reason.
  - Choose the “Spam” button when appropriate
- The status of data packages can almost always be determined by looking at fogbugz correspondence, metadata emails, and the spreadsheet. Always check these.
- Many of these tickets are commonly-asked questions (see 6.1.). Some tickets can require creative problem solving. If in doubt, search through old tickets to see if a curator has answered similar questions in the past.
- The wiki is also a helpful tool: wiki.datadryad.org
- If you forward a help ticket to dryadassistant@gmail, change the “from” field to cases@nescent.fogbugz.com -- their email filters messages from help@datadryad.org
6.1. Common helpdesk question topics

• Common help ticket topics are addressed on the “Templates for Correspondence” page of the wiki
  http://wiki.datadryad.org/Templates_for_Correspondence.

• Common questions tend to change as the database changes. Update, add to, and modify the templates as you see fit.
7. Modifying submissions

- Curators can often modify submissions as long as they are still in our workflow (not yet in the public archive)
- The versioning feature, currently in testing, (see 8. Versioning) is designed to allow modification of any data package, including archived packages.
7.1. Modification rules by workflow stage

• **Unfinished submissions**
  – These files are drafts and may be modified in any way by the author
  – Once the author clicks “Finalize submission,” the data package can no longer be edited by the author.

• **Data package in journal review**
  – Authors can add new files but cannot delete them. This restriction ensures that all reviewers see the files that were originally submitted. (See 4. and 7.3.)

• **Data package in publication blackout**
  – Curators can modify as necessary. Authors may not modify their submissions unless the senior curator returns the package to the author’s personal workspace (see 5. Workflow Overview).

• **Data package published but files still embargoed**
  – Since the data is still not public, it’s probably okay to change the files, although this isn’t best practice – provenance metadata will be incomplete.

• **Data package archived and public**
  – Adding files requires the versioning feature.
  – If the data was very recently published, we might agree to switch bitstreams (1-to-1). This is an exception to our rules – definitely not best practice.
7.2. Switching bitstreams

• Under your login menu, go to “Edit This Item.” Then click the “Item Bitstreams” tab.
  – Check the box for the bitstream you want to delete, then click “delete” and confirm delete.
  – Click “Upload new bitstream.”
    • Browse to the correct version of the file
    • Under “description:”
      – for a data file, enter: dataset-file.
      – For a ReadMe file, enter: dc_readme.
7.3. Adding or deleting files

- Under the profile menu, go to “My Tasks,” and search for the data package you want to modify.
- Click on the title of the package, then scroll to the bottom, where you can choose from green buttons. Choose “Edit item(s)”
- Then on the next screen, click “add new data file.” Add the file, then click “save file and continue.”
- To delete files, simply click the “delete” button next to the undesired data file. But be careful, because there’s no confirmation screen!
- **For data packages still in review:** tell the author to add corrected files to their submission and identify to you which files should be deleted. It’s easiest if a curator deletes the unwanted files after the article has been accepted and moved to the Unclaimed Queue, just prior to curation. Document this in the spreadsheet!
8. Versioning

The forthcoming versioning feature will simplify many of the complexities of data package modification (see 7. Modifying submissions).

- In the Profile menu, there is a listing “Add to or modify these data.” This is the versioning feature, but it doesn’t work perfectly yet.

- Versioning has a lot to do with dois:
  - new versions will have .1, .2, etc
  - the original doi will point to the latest version

- Indexing: search will pull up only latest version

- A warning heading is added to older versions to say they are old.

- Design considerations: If a submitter were to click “add to or modify these data,” they should get a screen that says:
  - You're about to add a new version of all of your data. This tool is for adding, removing, or modifying your data files. If you simply want to change file titles, descriptions, or other metadata, please contact the Dryad curatorial team at help@datadryad.org.
  - Once you have changed the files, a Dryad curator will approve your changes. Once approved, there will be a new version of your package. Your doi will remain the same and will resolve to the most recent version of your data. Here are the reasons you might want to do this: x, y, z.
  - On this screen, there should also be a button that says “YES, create a new version” or “Cancel and return to data package”

- Elaborate testing notes for the versioning feature are on Trello (https://trello.com/c/omM5p5Aw)

- Help tickets relating to versioning are nested in Fogbugz under case 4468.
8.1. Versioning: what seems to work now

Sara believes the following use-cases will create versioned packages without problems:

- **Adding a file:** the simple DOI for the new file must be manually registered. For doi:10.5061/dryad.63q27, she added a fifth file, with doi:10.5061/dryad.63q27.2/5. However, the simplest version of the DOI (doi:10.5061/dryad.63q27/5) was not registered and she had to do so manually (see https://docs.google.com/document/d/1eXCLIS65Zj84MzXGV8kbeSO25S1XlRAIsC_91Cb93hl/edit in Google drive for instructions on how to register a DOI)
  - Another bug: 1st version files should have doi:10.5061/dryad.63q27.1/1 direct to the new version instead (doi:10.5061/dryad.63q27.2/1). These files are the same in each version, though, so it doesn’t really matter.

- **Replacing an old file with a new version:** DO NOT delete the old file, then add a new file. You must edit the old file, deleting the old bitstream and uploading a new one. (You can either do this through the submitter interface, clicking “edit” beside the old file, then removing the old bitstream and browsing to upload a new one; OR you can click the file open in tabs as you would during curation, then click the “Item Bitstreams” tab, deleting the old bitstream and uploading a new one.)
  - If you delete the old file/add a new one, the new file will be registered with doi:10.5061/dryad.xxxx.2/2 instead of doi:10.5061/dryad.xxxx.2/1, which means that Google search will pull up the old version of the file (e.g. doi:10.5061/dryad.xxxx/1)
9. Assistant curator work assignments

- Every morning, after you’ve gone through the Fogbugz tickets, send emails to assistant curators assigning them new submissions to process. These emails go to dryadassistant@gmail.com, with “For [assistant curator name]” in the subject line.
- Choose one entry at a time from the “In Curation: Unclaimed” queue.
- First, check the shopping cart to make sure the Data Publication Charge is correct.
- Check the journal name – if it’s a journal we don’t yet have data for, see 10.6.
  - Let assistant curator know if it’s a new journal – they’ll make the name pink in the spreadsheet to facilitate Friday reports.
- Click “Show full item record.” Look at all the metadata to make sure that everything looks okay. (See Curation Manual. See also: 10.) If you notice something awry, it doesn’t hurt to make a note in the assignment email.
  - Integrated journal submissions should show journal contacts in the metadata. If these contact emails aren’t there, it means the author submitted outside of the normal workflow. Check the metadata emails – if no email, See 10.3.
  - To access the full data package without claiming the submission, create a URL like this: datadryad.org/resource/doi:10.5061/dryad.NNNN

- For integrated submissions formerly in review, check the metadata email and the spreadsheet to see if any modifications need to be made to the data package before it’s curated.
- Assignment emails should contain:
  - Last name of the submitter (not necessarily the lead author)
  - int/nonint, pub-blackout/no pub blackout, previously in review?
  - Any additional notes
10. Common submission issues

10.1. Simple errors

• E.g. Html tags? (we don’t support these); name cases? (our standard form is Smith, John T.); title in sentence case? (i.e. Host-associated genomic differentiation in congeneric butterflies: now you see it, now you don’t); scientific names in proper field? (dwc.scientificName)

10.2. Journal name not in authoritative form

• Non-integrated journals only. If an integrated journal name is not in its correct form, it means the author didn’t use the integrated workflow. Look the spreadsheet or browse the repository by journal name to find Dryad-accepted names. For new journals, see next slide.

10.3. Author didn’t use integrated workflow

- You can tell because there will be no MS# or journal contacts.
- First check for a metadata email to make sure article has been accepted.
- If there is no metadata email, reject using appropriate correspondence template ([wiki](#)).
- If we received a metadata email:
  - If the journal uses review workflow and the submission should be in review: reject the submission using appropriate correspondence template.
  - If journal doesn’t use review workflow and the associated MS has been accepted: it’s simplest to add the missing info yourself. The MS# will be in the metadata email (dc.identifier.manuscriptNumber), and the contacts are listed in the config file (workflow.archive.mailUsers). Be sure to document these actions in the “notes” section of the spreadsheet.

10.4. Published documents by other authors

- (Articles from JFWM often submit these as supplementary data).
- Publications by authors other than the Dryad submitter are often subject to copyright. Government docs are generally public domain, but docs by private authors may not be appropriate for CC0.

10.5. Human subject data

- Personally identifiable information is absolutely not accepted in Dryad. Because we operate under a Creative Commons Zero license, everything in the database must be appropriate for full public availability. See this article for more info: http://www.bmj.com/content/340/bmj.c181
- Take a careful look at all human subject data, and tell the assistant curators to contact the author if they have the slightest doubt. (template)
- Students should email authors from dryad.journal.submit@gmail, where they can send from help@datadryad.org.
- If, after consulting the author, data still seems questionable, you can consult with Dryad colleagues (Todd, Laura, Peggy, Ryan), and journal editors, but DO NOT send files via email.
  - We’re trying to figure out a more secure way for files to be viewed by Dryad staff.
  - For some journals, editors can access the Dryad submission with a reviewer code.
- We’re not legally required to ensure properly anonymized data. But we still want to be very cautious. If necessary, Dryad can decline to publish human subject data that has not been anonymized to our satisfaction (see wiki template for correspondence).

10.6. New journal title (not yet in Dryad)

- Search for title in WorldCat, UNC catalog, Ulrichs, or Journalseek.net – we want the authoritative title.
- Alert the processing curator that this is a new journal (in your assignment email). They will make the name pink the spreadsheet.
11. Scheduling assistant curators

- Assistant curator work hours tend to change each semester
- Ask curators for preferred hours, then schedule so that assistant curator work time is spread relatively evenly throughout the week.
12. Trello

- Every other week there is a meeting to discuss topics on the Trello board.
- Different boards address different perspectives in the project.
  - The senior curator manages the “Curators Perspective” board.
  - The main board is called “Dryad Development.”
- Advocate for curation issues each week. Move cards into place according to priority, both on the Curators Perspective board and the main board.
- Codes on Trello boards: FEAT (feature) BUG (bug), DOC (documentation related), or SKIP (something so important that it needs to skip the line)
- Tag other staff on Trello cards by typing @theirname (e.g. UX staff can tag Senior Curator on something in design that needs feedback.
- Labels: the color indicates what board a card came from
- You can attach files (e.g. extensive testing notes) that don't fit into the card itself
- Each card has a persistent link—if you end up moving the card (which you usually do) this link will always work.
13. Payment plan

• In September 2013, we began charging a fee for data publication. For details, see http://datadryad.org/pages/pricing and http://wiki.datadryad.org/Cost_Recovery_System.

• Consult with Laura for payment issues

• Vouchers: Create a voucher to zero out a shopping cart by going to the profile menu and clicking “shopping cart” → manage vouchers. Creator should be Dryad. Type a brief explanation of the voucher. The new voucher should appear first on the list. You can also search for your explanation.
14. New integrated journals

- Add new integrated journals to the spreadsheet’s “Notes” and “Weekly Template” tabs
- Go to Dryad dropbox ➔ partnerJournals ➔ PIQs to find the pre-integration questionnaire for the new journal. This will tell you whether they plan to use review workflow and pub-blackout.
- Create a new tab for the integrated journal, and move all of the entries from the “Other journals” tab into this new tab. In the “manuscript number” column, enter “pre-integration.”
- Check with communications coordinator to confirm contacts. In the config file, update the journal contacts for notifyWeekly and notifyOnArchive.
15. Friday reports

Distinctions in vocabulary:
• Submissions are submitted
• Deposits are approved
15. Friday reports, cont.

1. Record this week’s counts in the weekly report template spreadsheet (https://docs.google.com/spreadsheet/ccc?key=0Anksgfd-i0bDdF82YmtZQlhuUDdfdmgyQWNSSk1Xcnc&usp=sharing). You are counting from last Friday at 12:00am until yesterday (Thursday) at 11:59pm.

2. For review or accept notifications, look in the dryad.journal.submit@gmail, under each journal’s tag. Remember to expand all messages to get the full count (e.g. pbiology and Systbio)!
15. Friday reports, cont.

3. For reviewed or archived **deposits** (never in blackout), go to the spreadsheet. Anything from last Friday until yesterday (Thursday) should be counted. For deposits in review, still count duplicates and other weird things, as long as they're in the review workspace.

- Remember to sort spreadsheet by date just in case something came out of pub blackout and the date changed – it wouldn't necessarily be at the bottom of the spreadsheet in that case.
- A lot of journals have us plugged into their systems, so you may get “rejected,” or “published” emails that are irrelevant to you.
- Pbiology is high volume, but notices are nested in unique subject lines, so turn on “conversation view” when counting these notices.
- MolEcol and MolEcolRes request that we list articles in review as well as articles accepted, but we don’t do this for any other journals.
15. Friday reports, cont.

4. Update the wiki: [http://wiki.datadryad.org/Curation_Summaries](http://wiki.datadryad.org/Curation_Summaries)
   - Review: number accepted into review
   - Integrated: number accepted to pub-blackout plus number approved (columns e and f). **Note how many are in pub blackout** in the notes section of the wiki spreadsheet. Note how many from integrated journals got taken out of pub blackout (column g)
   - Go to “other Journals” in the spreadsheet. **Sort by date.**
   - Count a few different things:
     - data new to Dryad and in pub blackout
     - data simply accepted (never in pub blackout b/c they were submitted to us after article publication)
     - data that have been approved from the pub blackout queue
       - (in notes, these say “first accepted and queued on x/xx/xxxx)
     - “previously queued submissions entered main repository:” add those from non-integrated journals, plus those from integrated journals.
     - For new non-integrated submissions, how many are still in the queue (pub blackout)?
15. Friday reports, cont.

5. Go to the dryad.journal.submit@gmail.com
   • Template for weekly reports is in the “Drafts” folder. You can also refer to previous weeks’ reports, in the “Sent Mail” folder
   • Remember to send each message from curator@datadryad.org. There are groups set up for each journal, so the “to:” field should be automatically populated.
   • MolEcol and MolEcolRes request that we list the dois of data packages that went into review each week. We don’t do this for any other journals.
15. Friday reports, cont.

6. Send one last email to dryad-dev@googlegroups.com
   • Create a template in the “Drafts” folder of your email account
   • Featured picks for the week
     – Open formats, Readmes, and other good data practice
     – Interesting research
     – Well-known researchers
   • Go to “other journals” and copy/paste the journals one at a time into a spreadsheet (I use the extra tabs in the weekly report template).
     – Put them either under “had new submissions” and (how many) or “previously queued” (and write integrated if integrated) and (how many) or “already published”
     – Pink journal titles are new to Dryad, or will be new once the data is out of pub blackout.
   • Curator time tracking:
     – How many hours did curators spend directly on curation this week?
     – What percent of the total work hours was spent on curation? (for sr. and asst. curators)
     – We’ve been using Toggl (http://new.toggl.com) to track our hours – we track time spent on curation and time spent on side projects -- makes it easier to measure and more accurate.
Glossary

• **Data file**: The bitstream containing actual data. Can also include a ReadMe file.

• **Data package**: The entire set of data files associated with one publication plus the metadata describing the combined set.

• **Embargo**: Dryad displays the descriptions of the data files, but the actual data will not be public for a certain amount of time. We have three embargo options: untilArticlePublication, oneYear, and custom.

• **Metadata email**: The email sent to journal.submit@gmail.com by an integrated journal, alerting us to the status of an article.

• **Publication blackout (pub blackout)**: Some journals choose to keep the entire data package hidden until the associated article is published. For curators, this means we keep the data package claimed in the Dryad Queue profile. When the article is published (asst curators check periodically for published articles), these packages can be approved.

• **“The spreadsheet”**: The Journal Tracking Spreadsheet, found [here](#).
Web resources

• Config file: https://github.com/datadryad/dryad-repo/blob/dryad-master/dspace/config/DryadJournalSubmission.properties

• Curator office whiteboard: https://docs.google.com/document/d/1C4go10DYWF8856Obf8fwYwzFyp_808iYX3G4mfEHG8/edit?usp=sharing

• Dryad blog: http://blog.datadryad.org/

• Dryad development Google group: https://groups.google.com/forum/?hl=en#!forum/dryad-dev

• Encyclopedia of Life: http://eol.org/

• EZID blog: http://ezidstatus.wordpress.com/

• Journal submission tracking spreadsheet: https://docs.google.com/spreadsheet/ccc?key=0ApcFERZ3FxdbdGJtTU5vOVM2YkZBODZUQ201Mkp0bEE&usp=sharing

• Dryad Wiki: http://wiki.datadryad.org/

• All-Hands Meeting agendas: http://wiki.datadryad.org/WG:Team_Meetings

• Reports that curator would like to be able to generate: http://wiki.datadryad.org/Curator_Reports

• Dryad Calendar: http://wiki.datadryad.org/Calendar

• Simple website edits: http://wiki.datadryad.org/Editing_Website_Text
Senior curator accounts and permissions

- Fogbugz – ask Hilmar
- Wiki permissions – contact itadmin@nescent.org
- Send/receive from curator@datadryad.org – Hilmar
- Shared Dropbox folder: https://www.dropbox.com/sh/u7kfin3dh0x4a2m/b7-ygUJBxw - can be shared by any folder member
- Trello – create account
- Dryad-dev google group – join
- Google Drive documents – get access
- Toggl – create account and link to team
- GitHub – create account, ask Hilmar for permissions