

# November 2011 Curation Checklist

## New submissions

1. Integrated or non-integrated submission
  - Integrated journal
    - Look up metadata email. If no metadata email, send back to depositor.
    - Follow journal settings for publication blackout.
      - If publication blackout is in place and article has not yet been published (and everything else about submission is appropriate), you will place item in Dryad Queue's claimed tasks, manually generate acceptance/DOI email to depositor and journal contacts, and manually register the package DOI to point to the publication blackout page.
      - If no publication blackout is in place (and everything else about submission is appropriate), you will approve the submission. The notification email will be automatically generated (make sure email addresses are present in metadata).
  - Non-integrated journal
    - For all non-integrated submissions, control the journal name. Check the journal browse list for previous submissions from this journal and use the established form of the journal name, if present. If no previous submissions from this journal, use library catalogs (UNC, LoC, Worldcat, as needed) to determine the uniform title.
    - If article has been published, edit as for any published item and approve the submission.
    - If article hasn't been published yet, follow publication blackout procedures: claim item in Dryad Queue account, check files and check/edit metadata, manually register package DOI to point to publication blackout page, and manually generate email to depositor.
2. Article/manuscript status
  - Published
    - If everything else checks out, the item can enter the public repository and files may be exposed (or not) according to the depositor's embargo selections.
    - Cues that an article has been published:
      - Check metadata for dc:relation.ispartofseries and dc:relation.isreferencedby.
      - Google search based on title and/or other metadata, or search at journal website.

- Special metadata editing:
          - Add PLoS-style article citation to data package metadata as dc:identifier.citation.
          - Add article DOI to data package metadata as dc:relation.isreferencedby with format “doi:####”.
          - Adjust file embargoes as appropriate.
          - After approval, edit dc:date.issued to article publication date.
        - Create a complete, final entry in the submission tracking spreadsheet.
      - Accepted (integrated journal) / Ahead of publication (non-integrated journal)
        - Non-integrated submissions ahead of article publication will be in publication blackout until associated article is published.
        - Submissions associated with accepted manuscripts in integrated journals can be approved unless the journal has specified a publication blackout.
        - Create an entry in the submission tracking spreadsheet – a dark grey entry with red note if item is queued, a regular entry with yellow highlighted article citation field if the item is approved.
      - In review
        - Depositors will only be able to access the review workflow if they have submitted to an integrated journal that is using this feature. Curation tasks such as metadata editing and examination of files do not occur until or unless the manuscript is accepted for publication and the item enters the regular curation queue.
        - Create a grey entry in the submission tracking spreadsheet for each data package in the review workspace. There is no normal handle/URL at this time, so make sure to include the reviewer key link in the notes field.
3. Appropriate content: If the files are not appropriate, as described below, either send the item back to the depositor (“Reject” it) with an explanation, or email the depositor directly to inquire.
- Files should contain something like data, script, sequences, etc. Watch out for uploaded manuscripts and other items that don’t fit even a very broad definition of data.
  - Rights concerns: The uploaded files should not contain material that is clearly copyrighted and should not contain copyright statements or non-CC0 license statements (look for these in the file description field and any readme files in particular).
  - Human subject concerns: Any files containing human subject data

(be alert when curating content associated with biomedical journals) should be anonymized. Anything questionable should be held back from becoming public while inquiring with the author and journal editor.

4. Editing new submissions
  - Look over all metadata for stray formatting tags, mangled characters, and spelling errors (allow international spelling variation, as in neighbour/neighbor). These are most common in the article abstract and author names. For integrated submissions in particular, look for errors like manuscript details tacked on to the end of the abstract.
  - Check all author names in both package and file records and make sure that they appear to be correctly formatted as “LastName, FirstName M. I.” Initials should be followed by periods and multiple initials should also be separated by spaces.
  - Add taxon names to dwc:ScientificName, geographic locations to dc:coverage.spatial, and geologic timespans to dc:coverage.temporal, as appropriate. This may include moving them from dc:subject.
  - Remove line breaks and multiple spaces in all metadata fields. Look for these especially in article abstract and file descriptions. Replace the line breaks, as appropriate, with other symbols to maintain clarity.
  - For submissions associated with articles that have already been published, add article metadata to the package (article citation in dc:identifier.citation, article DOI in dc:relation.isreferencedby, article publication date in dc:date.issued, check author names against article) and modify file embargoes if needed (setting end dates, removing if already past).
5. Enter all new submissions in submission tracking spreadsheet immediately upon approval, sending to queue, or viewing in the review workspace. Do not record items that have been returned to the depositor (rejected).

## Reporting

Each Friday, report emails summarizing the week’s activity are generated by the curator. See the Templates for Correspondence wiki page.

1. Contacts from each integrated journal receive the following information:
  - Number of notifications (metadata emails) received from this journal. Break out review and accepted status, if journal uses review workflow.
  - Number of deposits associated with this journal. Break out items in review workspace and main repository, if journal uses review workflow.

- List of archived deposits as: LeadAuthorLastName (manuscript number) data DOI.
- 2. Send an email to the dryad-dev Google group with the following information:
  - Counts of integrated deposits, non-integrated deposits, and deposits to the review workspace. For each, specify how many are queued and how many are in the main repository. Give a separate count of previously queued items that entered the main repository during the reporting period.
  - Names of journals associated with non-integrated deposits.
  - Any new journals in the repository or other interesting events or problems from the week.

## Publication tracking and updating

1. Tracking
  - Integrated journals: dryadassistant@gmail.com receives email alerts for new articles and issues.
  - Non-integrated journals: not systematically tracked at this time.
2. Searching
  - Integrated journals: As emails arrive, search the journal submission tracking spreadsheet for data packages associated with each new item. (Searching the repository rather than the spreadsheet works for some journals but will miss queued items for journals using publication blackout.)
  - Nonintegrated journals: Most of these items will appear in the Dryad Queue account's claimed tasks, but search the "Other journals" tab of the submission tracking spreadsheet for complete results.
3. Editing metadata once a match is made
  - Compare author names and other metadata (title and abstract) that may have changed between working manuscript and published article, and update these in Dryad. Make sure to edit author names for each file record, in addition to the data package record, and don't forget the corresponding author at dc:contributor.correspondingauthor in the package metadata.
  - Add PLoS-style article citation to data package metadata as dc:identifier.citation.
  - Add article DOI to data package metadata as dc:relation.isreferencedby with format "doi:####".
  - Edit dc:date.issued in package metadata to article publication date. (Queued items: this can only be performed after approval.)
4. Editing embargoes once a match is made (all embargo metadata is stored in file records)
  - Items in the repository:
    - If dc:type.embargo is set to "none", no action is required.

- If dc:type.embargo is set to “untilArticleAppears”, go to the “Item Embargo” pane when editing the item and click to disable the embargo.
  - If dc:type.embargo is set to “oneyear”, set dc.date.embargoeduntil to one year after the earliest article publication date, formatted as YYYY-MM-DD.
  - If dc:type.embargo is set to “custom”, look for dryad:curatornote for an explanation and set the date as appropriate.
- Queued items:
  - If dc:type.embargo is set to “none”, no action is required.
  - If dc:type.embargo is set to “untilArticleAppears”, change to “none” before approving.
  - If dc:type.embargo is set to “oneyear”, set dc.date.embargoeduntil to one year after the earliest article publication date, formatted as YYYY-MM-DD (this can only be performed after approval).
  - If dc:type.embargo is set to “custom”, look for dryad:curatornote for an explanation and set the date as appropriate (this can only be performed after approval).
- 5. Approving queued submissions: If a data package is queued until article publication, click Approve after making all edits possible before approval. Please note: you will need to manually edit the registration for the package DOI so that it points to the approved package rather than the publication blackout page. After approving, make any additional edits, make sure it’s displaying correctly, and give the metadata a once-over (look for buggy duplicated identifiers, etc.)
- 6. Update journal submission tracking spreadsheet to record any status changes.
  - If a queued item was approved, remove grey background, change record date to approval date and record previous submit date in notes column, remove red “needs to be finalized” warning, etc
  - If an item in the repository was updated with article citation, change that value to “online” (in yellow) if online in advance of print or “Y” (no color) if the article citation is in its final form (with volume and page numbers, or if the journal is online-only).

## Notes and reminders

- Journal setting information:  
<http://code.google.com/p/dryad/source/browse/trunk/dryad/dspace/config/DryadJournalSubmission.properties>
- Templates for correspondence and reporting:  
[https://www.nescent.org/wg\\_dryad/Templates\\_for\\_Correspondence](https://www.nescent.org/wg_dryad/Templates_for_Correspondence)
- Authors appear on both the package and the data file records. They may

- differ (which is rare), but the same form of name must be used in both places. For example, if you add a middle initial to an author name in the data package metadata, don't forget to do this for each file as well.
- Embargoes are set for each data file, and they may differ between the files in a given data package. Be careful lifting/editing embargoes, since embargo until publication and one year embargo look the same on the "Item Embargo" pane when editing an item.
  - Dates (article publication date, embargoed until date, etc) can only be edited after an item is approved.
  - Don't forget to manually update the DOI registration for previously queued packages upon their approval into the archive.
  - There are some defunct metadata fields that are no longer in use (i.e., dc:publisher) and some fields whose use has changed over time (i.e., dc:type). The use of many metadata fields will change dramatically when the 3.0 application profile is implemented.

PLoS style (modified for Dryad) citation examples:

- Smith PJ, Feinstein EM, Greenberg JA (2011) Dryad metadata is really fascinating. *Journal of Digital Repositories*, online in advance of print.
- Smith PJ, Feinstein EM, Greenberg JA (2011) Dryad metadata is really fascinating. *Journal of Digital Repositories* 2(6): 112-117.

Data package metadata				
Authors	dc:contributor.author	repeatable	required	LastName, FirstName M.
Corresponding author	dc:contributor.correspondingAuthor	not repeatable	required	LastName, FirstName M.
Spatial coverage	dc:coverage.spatial	repeatable	optional	place names, geographic coordinates, etc
Temporal coverage	dc:coverage.temporal	repeatable	optional	intended for geologic timespans, but years and other values are accepted
Approval timestamp	dc:date.accessioned	not repeatable	required	system-generated upon submission approval
Approval timestamp	dc:date.available	not repeatable	required	system-generated upon submission approval
Article publication date	dc:date.issued	not repeatable	required	system-generated to match approval date, later edited by curator to article publication
Data package DOI	dc:identifier	not repeatable	required	doi:10.5061/dryad.####
Article citation	dc:identifier.citation	not repeatable	optional	modified PLoS citation style
Journal's manuscript ID	dc:identifier.manuscriptNumber	not repeatable	optional	only for integrated submissions
Data package handle	dc:identifier.uri	not repeatable	required	http://hdl.handle.net/10255/dryad.####, system-generated upon submission approval
Abstract	dc:description	not repeatable	optional	article abstract
Component data file DOIs	dc:relation.haspart	repeatable	required	doi:10.5061/dryad.####/1, doi:10.5061/dryad.####/2, etc
Article volume, issue, year	dc:relation.ispartofseries	not repeatable	optional	only present if entered by depositor during submission
Article DOI	dc:relation.isreferencedby	not repeatable	optional	doi:####
Keywords	dc:subject	repeatable	optional	
Data package title	dc:title	not repeatable	required	Data from: Article title
Record type	dc:type	not repeatable	required	system-generated, now set to "Article"
Curator note	dryad.curatorNote	repeatable	optional	rarely used
Scientific names	dwc:ScientificName	repeatable	optional	Latin taxon names
Journal name	prism:publicationName	not repeatable	required	use authorized form of name only

Data file metadata				
Authors	dc:contributor.author	repeatable	required	LastName, FirstName M.
Spatial coverage	dc:coverage.spatial	repeatable	optional	place names, geographic coordinates, etc
Temporal coverage	dc:coverage.temporal	repeatable	optional	intended for geologic timespans, but years and other values are accepted
Approval timestamp	dc:date.accessioned	not repeatable	required	system-generated upon submission approval
Bitstream availability timestamp	dc:date.available	not repeatable	required	system-generated upon availability of bitstreams for download (will not appear if file is embargoed)
Approval date	dc:date.issued	not repeatable	required	system-generated upon submission approval
Data file DOI	dc:identifier	not repeatable	required	doi:10.5061/dryad.####/#
Data file handle	dc:identifier.uri	not repeatable	required	http://hdl.handle.net/10255/dryad.####, system-generated upon submission approval
File description	dc:description	not repeatable	optional	brief file description entered by depositor
Associated data package DOI	dc:relation.ispartof	not repeatable	required	doi:10.5061/dryad.####
Rights information	dc:rights.uri	not repeatable	required	CC0 URI for all items except a few legacy items under Original License
Keywords	dc:subject	repeatable	optional	
Data file title	dc:title	not repeatable	required	
Record type	dc:type	not repeatable	required	system-generated, now set to "Dataset"
Curator note	dryad.curatorNote	repeatable	optional	rarely used, mostly to specify custom embargo dates
Scientific names	dwc:ScientificName	repeatable	optional	Latin taxon names