

July 2010 Curation Checklist

Use RSS feed to monitor new Dryad submissions.

http://datadryad.org/repo/feed/rss_2.0/site

- For all records:
 - In all metadata fields, look for spelling and other errors. Generally check that the metadata looks reasonable.
 - In all metadata fields, look for special characters, punctuation marks, or formatting that isn't displaying well. Formatting tags need to be removed and special characters may need to be copied and pasted from notification email, unicode.org, or elsewhere.
 - Invert author names, if needed (Surname, Forename).
 - If time permits, search for existing content in Dryad from authors of the new deposit, and compare forms of name. Edit names for consistency across the repository, if needed, preferring the fullest form of the name. [Note: name authority control is a much larger project than can be accommodated in our current practices. Even at this level, research is often required to determine author identity and disambiguate names.]
- For data [file] records:
 - Download and open the file to see that it looks reasonable (contains data). [Note: virus scanning is now automated]
- For data package records:
 - If needed, add taxon names that are prominent in the title or abstract in `dwc:ScientificName` (will be `dwc:scientificName`) field(s). Likewise, possibly add geographic names or coordinates as `dc:coverage.spatial` (will be `dcterms:spatial`) and geologic timespans as `dc:coverage.temporal` (will be `dcterms:temporal`).
 - Add `dryad:status` with value "approved" after editing any other metadata
- Run the DOI minting script. Email depositor with their Dryad data package DOI.
 - [Note: there is not currently an easy way for the curator to know the identity of the depositor from within Dryad. The curator has access to a list of registered users, but these are not linked to deposits and submission without registration is possible through integrated journals. For content from integrated journals, author contact information from journal notification emails is used.]
 - [Note: the DOI minting script is a temporary work-around. Running it may flush out new deposits that have not yet appeared in the RSS feed, which only updates daily. After running the script, always check the "Recently Added" list for new deposits and curate them immediately.]
- Enter information about the data package and any submission errors in the "JournalSubmissionTracking" shared spreadsheet. Submission errors should also be recorded on the wiki:
http://www.nescent.org/wg_dryad/Submission_Integration_Issues
- Weekly, record on the wiki and report to managing editors at integrated journals the following:

- Number of new “notice of accepted publication” emails from that journal.
- Number of new Dryad deposits for that journal, with name of lead author and Dryad handle (for inclusion in publication)
- Any submission errors or issues encountered that week.

Use RSS feeds and/or email subscriptions to receive tables of contents for integrated journals. (currently American Naturalist, Evolution, Journal of Evolutionary Biology, Journal of Heredity, Molecular Ecology)

- As these notices arrive, search Dryad for data associated with each article.
- If a matching Dryad record is found, add the article DOI in a dc:relation.isReferencedBy (will be dcterms:isReferencedBy) field as “doi:#####” (not the URL). Add a citation to the article in a dc:identifier.citation field (if published in print and not just viewed online ahead of publication).
[Note: we are currently following PLoS guidelines for formatting article citations. We will be recording article citation metadata (volume, page numbers, etc.) in their own metadata fields once the Dryad application profile version 3.0 is implemented.]
- When an article has been published, update the dc:date.issued (will be dcterms:issued) and dc:date.embargoedUntil (will be dcterms:embargoedUntil) (if present) fields. In terms of legacy content, these fields are messy and need attention. The value in dc:date.issued/dcterms:issued should be the publication date of the article (use date first published online, if ahead of print). The value in dc:date.embargoedUntil/dcterms:embargoedUntil should be based on the publication date of the article and option chosen by the depositor (content embargoed for one year from publication defaults to one year from deposit, content embargoed until publication defaults to 9999-01-01, both need to be changed when article is published).