

Dryad Curation Practices

August 2012

Important Curation Documents

- Curation manuals
 - <http://wiki.datadryad.org/Curation>
- dryadassistant@gmail.com
 - Notifications of new submissions, newly published articles, other assignments
 - Correspond with authors using this account, send as help@datadryad.org
- Curator office whiteboard
 - google doc shared with dryadassistant google account
- Submission tracking spreadsheet
 - google doc shared with dryadassistant google account
- Integrated journal metadata emails
 - access via dryad.journal.submit@gmail.com
- Templates for correspondence
 - http://wiki.datadryad.org/Templates_for_Correspondence

Integrated and Non-Integrated Journals

- Non-Integrated
 - No coordination between journal and Dryad (no metadata emails, journal contact addresses for reporting, etc.)
- Integrated
 - Metadata emails send info ahead of submission
 - May use review workflow or only archive data after manuscript acceptance
 - May require ‘blackout’ of Dryad submission until article publication
 - Journal contacts are notified upon submission for review (if using review workflow), acceptance to blackout queue, approval/archiving, and weekly summary

The original integrated workflow is represented to the right. Some journals now also use a review workflow with additional steps or require the Dryad data package to remain hidden until after article publication (what we call “blackout”).

Further integration details are available in the following presentation:
<http://wiki.datadryad.org/wg/dryad/images/c/c6/DryadIntegrationOverview.pdf>

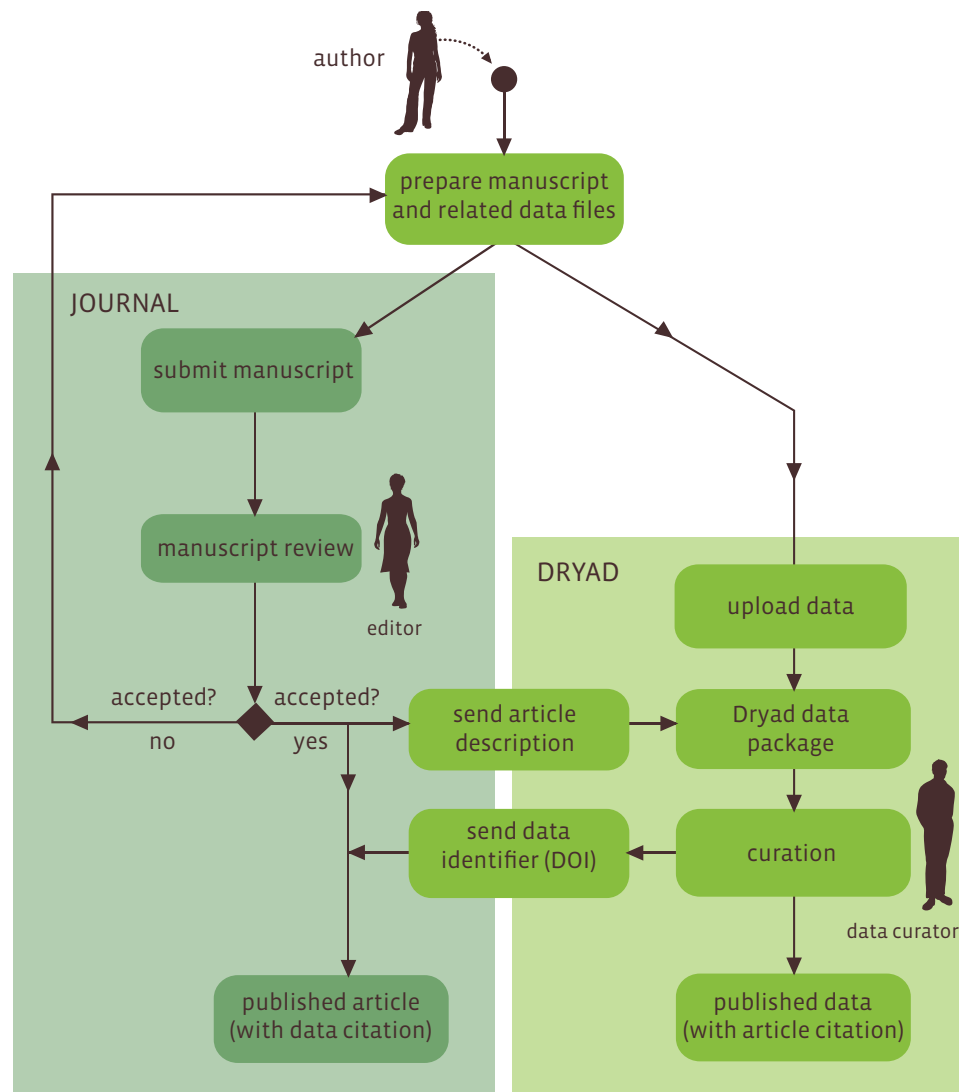
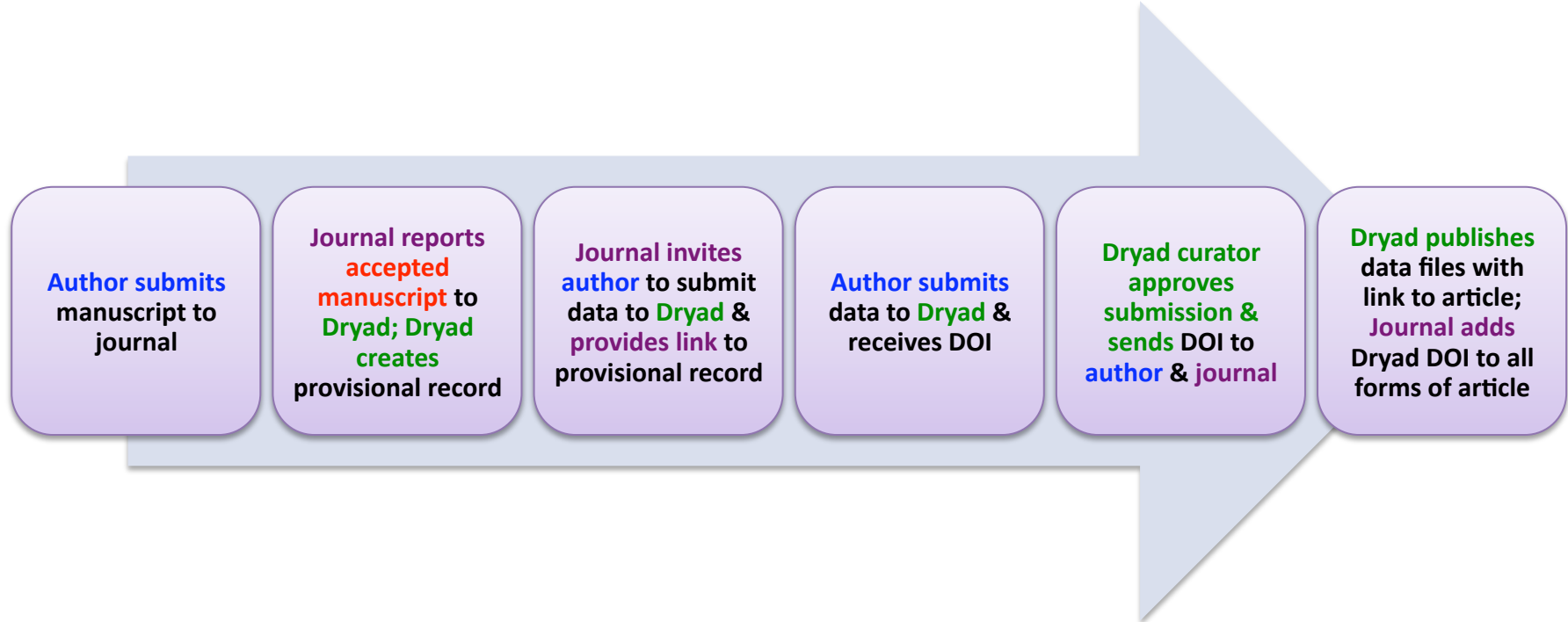


Image based on Lowry, R., E. Urban, and P. Pissierssens (2009), A New Approach to Data Publication in Ocean Sciences, Eos Trans. AGU, 90(50), doi:10.1029/2009EO500004.

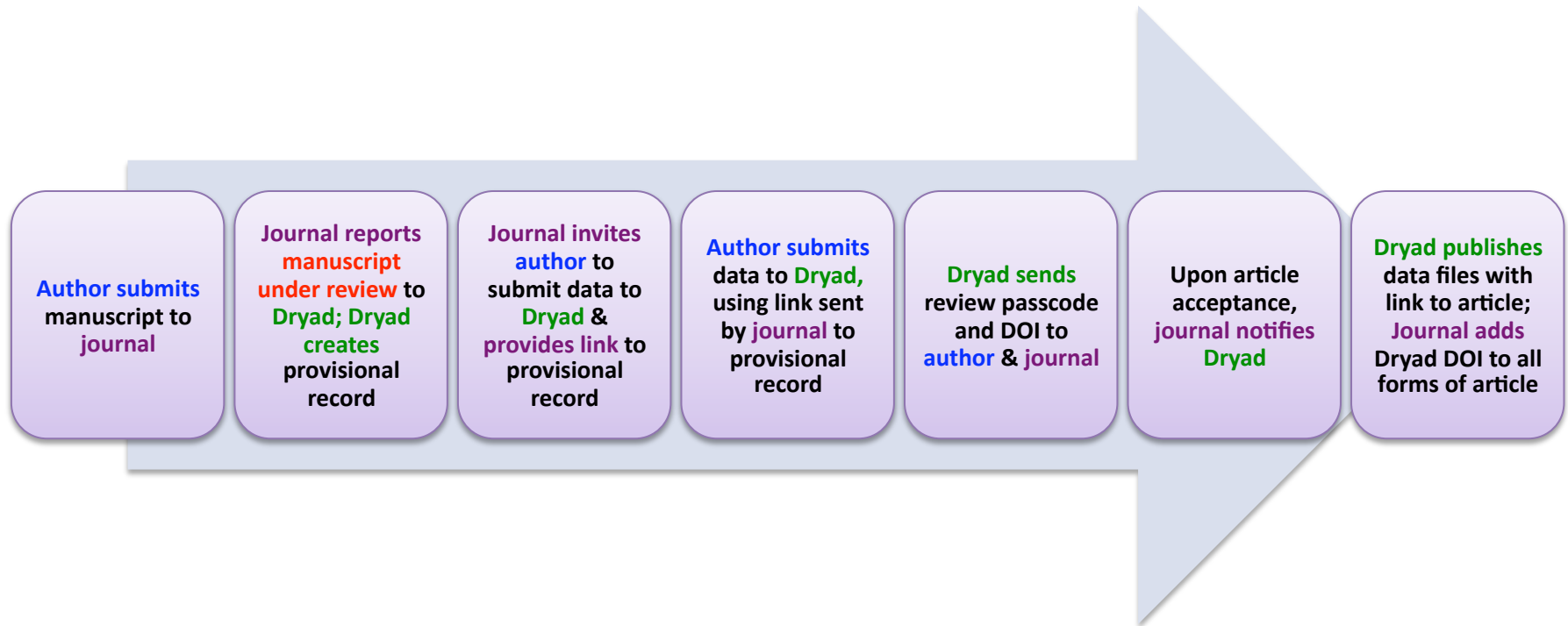
Basic Integrated Workflow (no review)



Review Workflow

- Journal sends manuscript information to Dryad before manuscript acceptance and invites authors to upload data.
- Dryad submission is routed to private review workspace, not main curation and publication queue. Passkey link is sent to journal for editor/reviewer access to Dryad submission.
- Author may continue to add files while submission is in review workspace.
- Journal sends second metadata email to Dryad with manuscript acceptance notification, triggering any associated submission to move from review to curation.
- Curator inspects and approves, queues, or rejects submission, as in basic workflow.

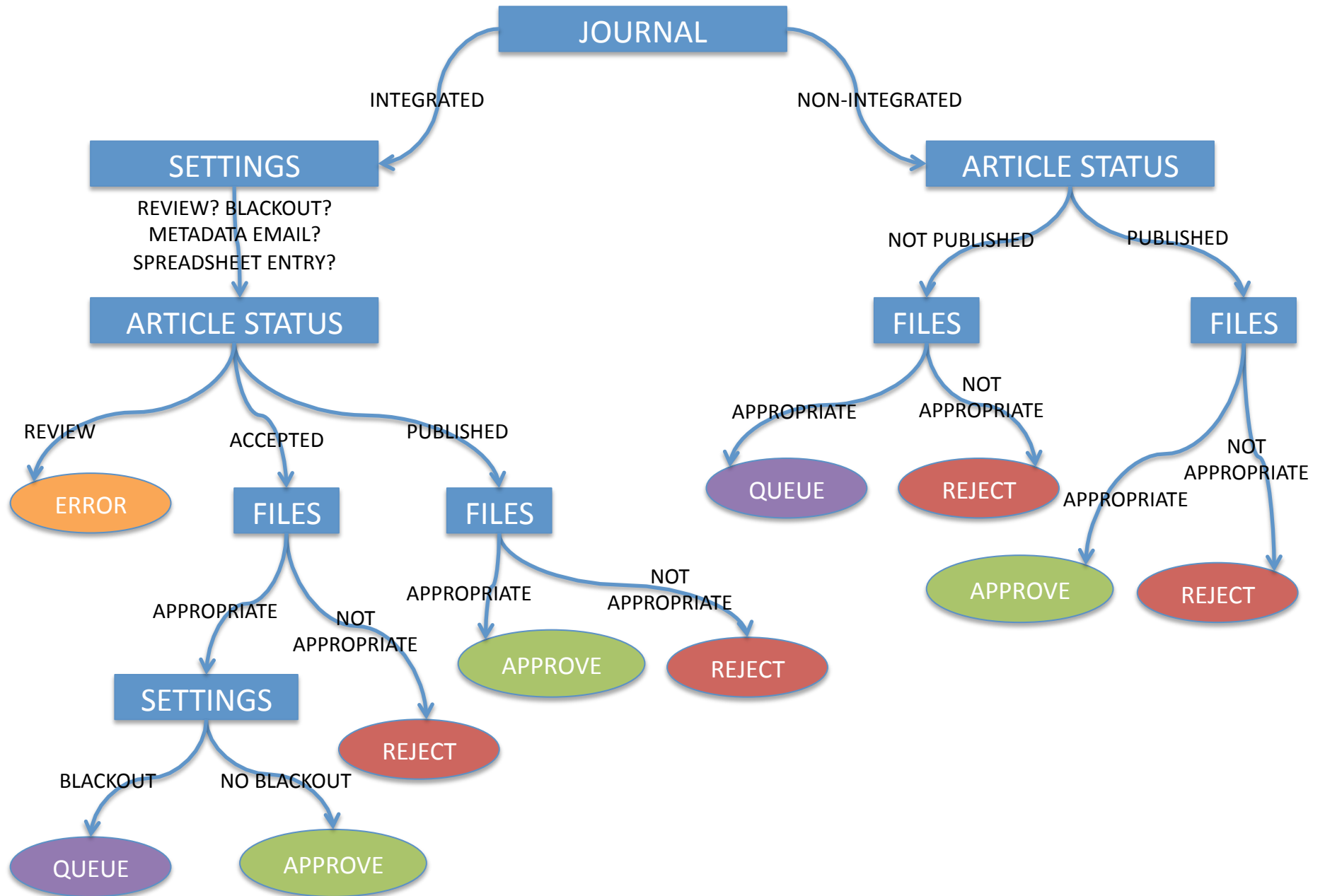
Review Workflow



Navigation

- Notifications of new tasks go to `dryadassistant@gmail.com`
- Log in to Dryad site with `dryadassistant@gmail.com` and `dryad.queue@gmail.com`, as appropriate
 - <http://datadryad.org/password-login>
- Dryad site left sidebar:
 - My Submissions are submissions you have created
 - My Tasks are submissions you can act on as a curator
 - Workflow Overview provides a way to search for items before or after archiving, and to force changes in their status that aren't always available in the interface

Overview of New Submission Processing



Journal Settings

- See JournalSubmissionTracking spreadsheet shared in google docs.
- First tab (“Notes”) lists each integrated journal and its review and blackout settings.
- Also search for duplicate submissions or notes in appropriate sheet.

Article Status

- Look for article DOI or volume information in the submitted metadata.
- Google search and/or visit publisher website.

Inspecting Files

- Check for technical problems, corrupt files, files that won't open in expected software, etc.
- Files should contain something that looks like data, with a very broad definition of data (supplementary figures, multimedia, etc., are ok, the manuscript itself is not).
- Look for copyright statements and licenses (not good).
- Look for identifiable human subject data (err on the side of caution).
- Look for duplicated files, data files uploaded in place of readme files, etc., and clean these up.

Editing Metadata

- REMEMBER: some metadata, such as author names, is repeated on the package and files and will need to be edited in both places.
- Scan over all metadata to see if it looks reasonable and identify problems.
- Author names should be formatted as
 - LastName, FirstName M. I.
 - remove any titles, such as “PhD”
- Data package title should be formatted as
 - Data from: Article title in sentence case
- Add values to specialized keywords (geographic, temporal, scientific name), move them from/to general subject keywords, as appropriate. Scientific names should be Latin (common names go in dc:subject instead) and should be recognized by <http://eol.org/>.
- Look for line breaks, especially in article abstract and file descriptions and edit these fields as needed for clarity when the content is displayed without line breaks.
- Check for inappropriate embargoes (custom when we have no info from journal, until article appears when article is out) and adjust as needed. If custom embargo, add embargo period (from journal) as dryad:curatorNote in file metadata.

Approving/Archiving

- Check for duplicates and notes in tracking spreadsheet, if you haven't already done so.
- Click Approve (will need to click twice if item is going through blackout first, based on settings). Email notification is sent automatically.
- Visit Dryad homepage and find item in Recently Published list (if not there, look for it on My Tasks page or track down any error).
- Check for duplicated package DOI and delete, if needed.
- Check that package DOI resolves correctly (may be a few minutes delay). Log in to EZID and check/fix, if needed.
- Update submission tracking spreadsheet.

Placing Submission in Publication Blackout Queue

- You should have already inspected files, edited metadata, and checked for duplicate submissions at this point.
- Register package DOI in EZID. Go to Create IDs -> Advanced in order to specify your DOI. Use <http://datadryad.org/publicationBlackout> as the location and leave all other description blank.
- Send acceptance email to submitter (and journal contacts, if integrated journal). Template is saved as a draft in dryadassistant gmail account.
- Add entry to submission tracking spreadsheet.

Updating Archived Items Once Article is Published

- Check author names, article title, and article abstract against published article and update if needed
- Edit package `dc:date.issued` to publication date of article (format as YYYY-MM-DD)
- Add article DOI to package `dc:relation.isreferencedby` (format as doi:####)
- Add article citation to package `dc:identifier.citation` (format as LastName F, LastName FM (YYYY) Article title in sentence case. Journal Name Vol(Num): page-page.)
- Lift embargoes or set embargo end dates, as appropriate

Approving Submission out of Publication Blackout

- Update metadata and embargoes as described in **Updating Archived Items Once Article is Published**.
- Click Approve. Find item in Publication Blackout list on My Tasks page. Click Approve again.
- Update package DOI in EZID (use Lookup function) and change location to point to public item URL

PACKAGE METADATA GUIDE

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

FILE METADATA GUIDE

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