

Sharing Between Data Repositories

Kevin S. Clarke
ksclarke@nescent.org

Thanks to the Dryad Data Repository contributors and funders:

Ryan Scherle, Todd J. Vision, Hilmar Lapp (NESCent)
Ben Bosman, Mark Diggory, Kevin Van de Velde (@mire, Inc.)



The Bio-Reposphere



DRYAD

(Generic Subject Repository)



FlyBase

(Subject Specific Repository)



MorphBank



TreeBASE

(General Scholarly Repository)



NCBI
GenBank



Generic vs. Specific Repos



- ✓ Easy submission
- ✓ Simple metadata
- ✓ Data is a “black box”
- ✓ No “orphaned” data

- ✓ Complex submission
- ✓ More useful metadata
- ✓ Well structured data
- ✓ Specific type of data

A Dryad Data Package



Profile: Kevin Clarke | Logout

Search Data

[Submit Data Now!](#)

[See how to submit](#)

Data from: Phylogenetic inference of reciprocal effects between geographic range evolution and diversification

When using this data, please cite the original article:

Citation is not yet available for this article from Systematic Biology. It will become available shortly after the article is published.

Additionally, please cite the Dryad data package:

Goldberg EE, Lancaster LT, Ree RH (2011) Data from: Phylogenetic inference of reciprocal effects between geographic range evolution and diversification. Dryad Digital Repository. doi:10.5061/dryad.8343

[Cite](#) | [Share](#)

Dryad Package Identifier	doi:10.5061/dryad.8343	61 views
Individual Data Files	http://datadryad.org/handle/10255/dryad.8355	26 views 5 downloads
	http://datadryad.org/handle/10255/dryad.8356	23 views 0 downloads
	http://datadryad.org/handle/10255/dryad.8357	23 views 0 downloads
	http://datadryad.org/handle/10255/dryad.8358	24 views 0 downloads
	http://datadryad.org/handle/10255/dryad.8359	24 views 0 downloads
	http://datadryad.org/handle/10255/dryad.8360	24 views 0 downloads
	http://datadryad.org/handle/10255/dryad.8362	23 views 0 downloads
	http://datadryad.org/handle/10255/dryad.8363	24 views 0 downloads
	http://datadryad.org/handle/10255/dryad.8364	23 views 0 downloads
	http://datadryad.org/handle/10255/dryad.8365	23 views 0 downloads
	http://datadryad.org/handle/10255/dryad.8366	24 views 0 downloads

Abstract Geographic characters—traits relating to the spatial distribution of a species—may both affect and be affected by processes associated with lineage birth and death. This is potentially confounding to comparative analyses of species distributions because current models do not allow reciprocal interactions between the evolution of

My Account

[My Submissions](#)
[Logout](#)
[Profile](#)

Context

[Edit this item](#)
[Export Item](#)
[Export Metadata](#)

Administrative

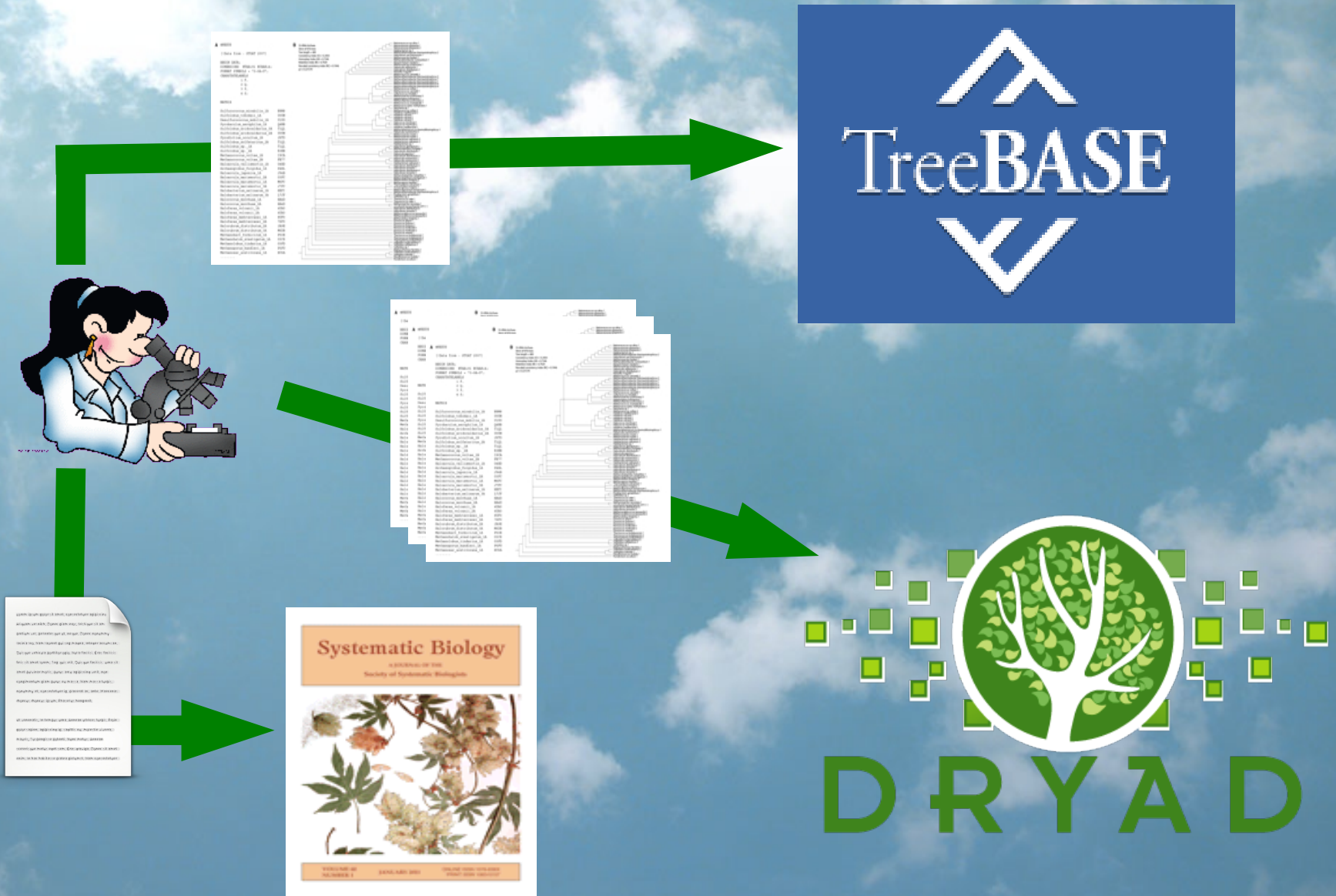
Access Control

[People](#)
[Groups](#)
[Authorizations](#)

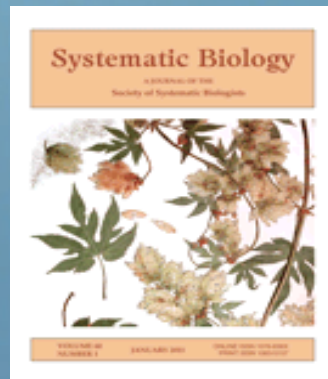
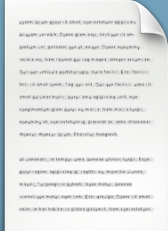
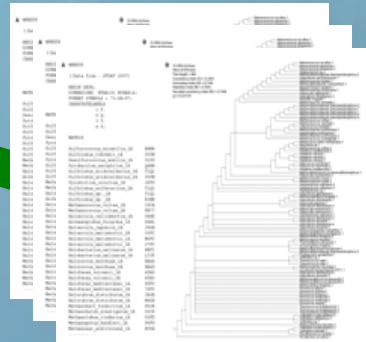
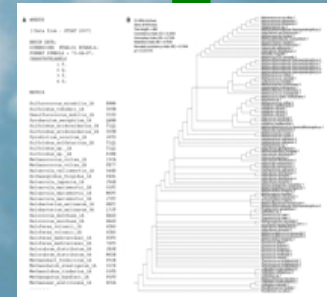
Registries

[Metadata](#)
[Format](#)

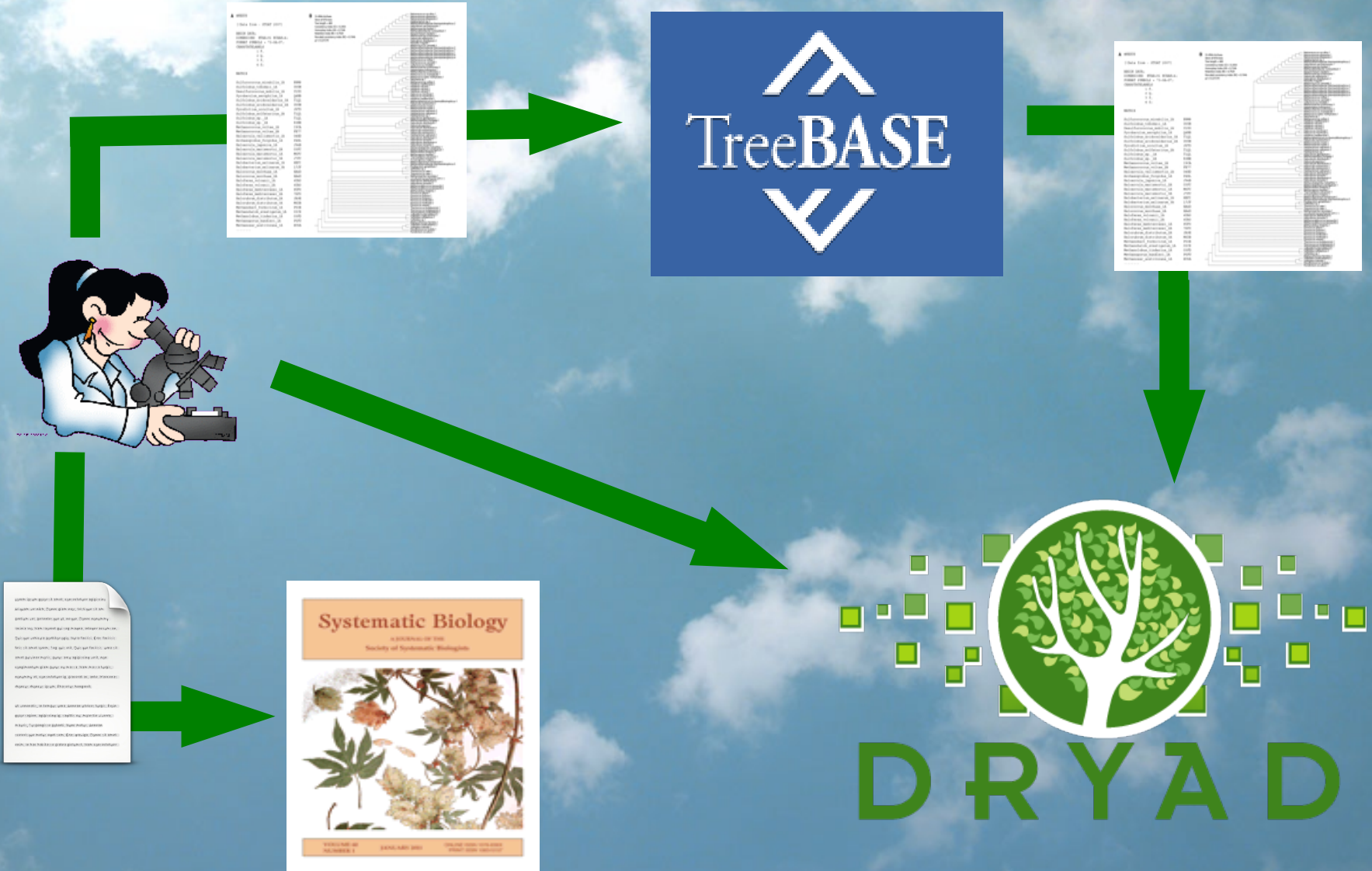
One Possible Workflow



“Save the Time of the User” #1



“Save the Time of the User” #2



Three Simple Steps



Profile: Kevin Clarke | Logout

Search Data ?

[Submit Data Now!](#)

[See how to submit](#)

Describe publication

Submitting data to Dryad consists of three simple steps:

1. **Describe your publication**
2. Upload and describe your data files
3. Approve data for publication



Please describe your publication in as much detail as possible. Providing a detailed description will make it easier for others to find your data in Dryad. Please describe the **publication only**. Do not enter information specific to your data files on this page.

Fields marked with an asterisk (*) are required. For more information on expected contents for a field, hold your mouse over the field in question.

Publication metadata

Title*:

Authors*:
Last name, e.g. Smith First name + initial, e.g. Donald F.

Journal name*:

Abstract:

DOI:

My Account

[My Submissions](#)
[Logout](#)
[Profile](#)

Context

[Edit Collection](#)
[Item Mapper](#)
[Export Collection](#)
[Export Metadata](#)

Administrative

Access Control
[People](#)
[Groups](#)
[Authorizations](#)

Registries

Case 1: TreeBASE Data Import

Publication

Data from: Title of my article, Clarke, Kevin S.

Data file *

Please upload your data file or provide the identifier of a file located in another repository.

No file chosen



Data file description

Title*:

Description:

ReadMe file: No file chosen

Author:
Last name, e.g. Smith First name + initial, e.g. Donald F.
 Clarke, Kevin S.

Harvesting and Web Services

OAI-PMH



PhyloWS

Case 2: Data Uploaded to Dryad

Publication

Data from: Title of my article, Clarke, Kevin S.

Data file *

Please upload your data file or provide the identifier of a file located in another repository.

- No file chosen 
-

Data file description

Title*:

Description:

ReadMe file: No file chosen

Author:
Last name, e.g. Smith First name + initial, e.g. Donald F.
 Clarke, Kevin S.

Partner Repository Upload

Submission overview

Submitting data to Dryad consists of three simple steps:

1. Describe your publication
2. Upload and describe your data files
3. **Approve data for publication**

Press a button to edit the description of your publication, add or modify all attached data files, or finalize the submission.

Describe publication:

Data from: Title of my article

Add and describe data files:

phylotree

Upload data files to partner repositories (optional):

phylotree.nex

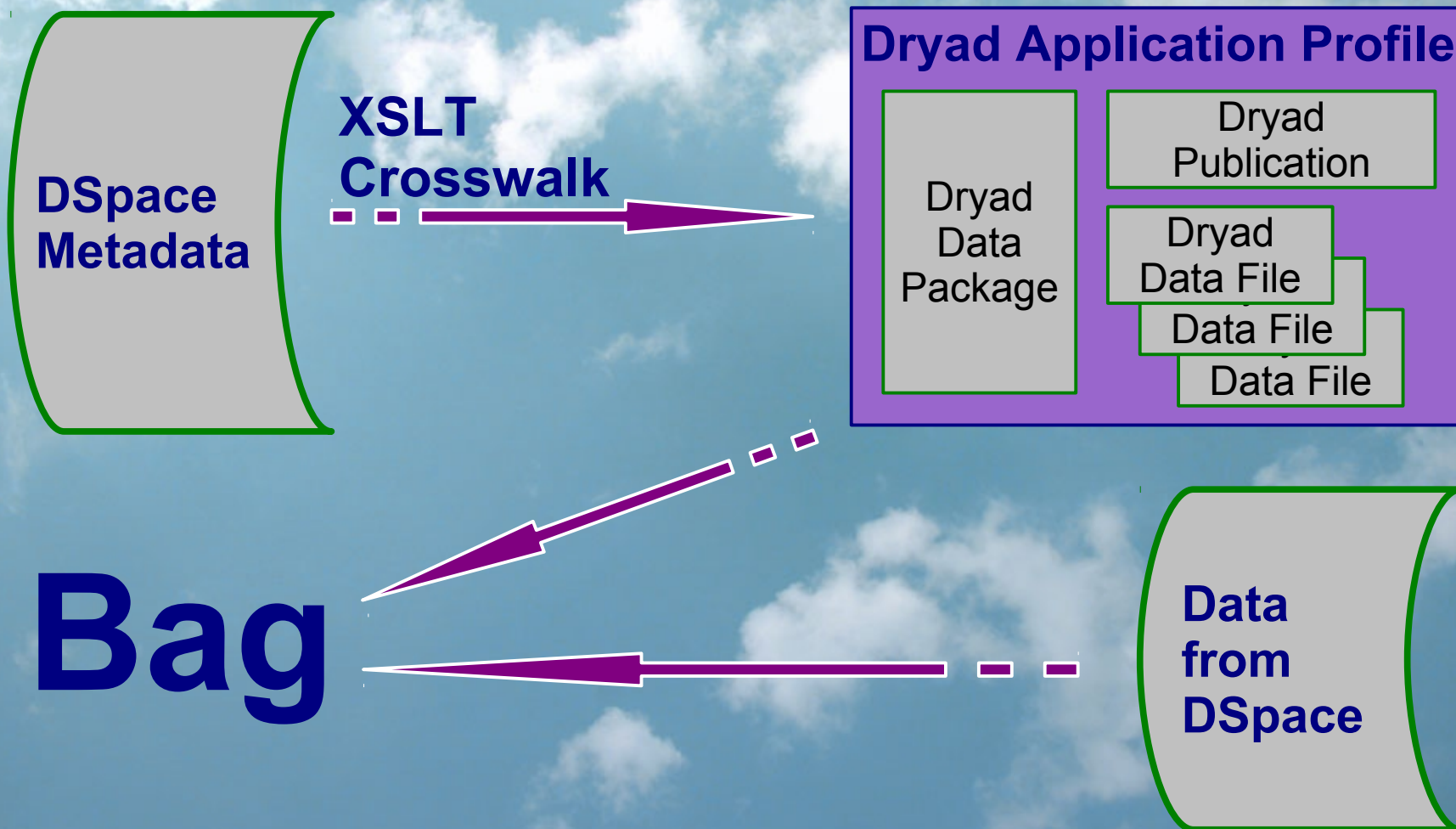


Finalize submission:

When you have added all data files for this publication, press the button below to finish the submission. A Dryad curator will review the submission to ensure the descriptions are complete and the files are not corrupted. After this brief review, your data will be archived in Dryad.

BagIt Disseminator

(implements DSpace PackageDisseminator)



A BagIt Bag

bag-info.txt

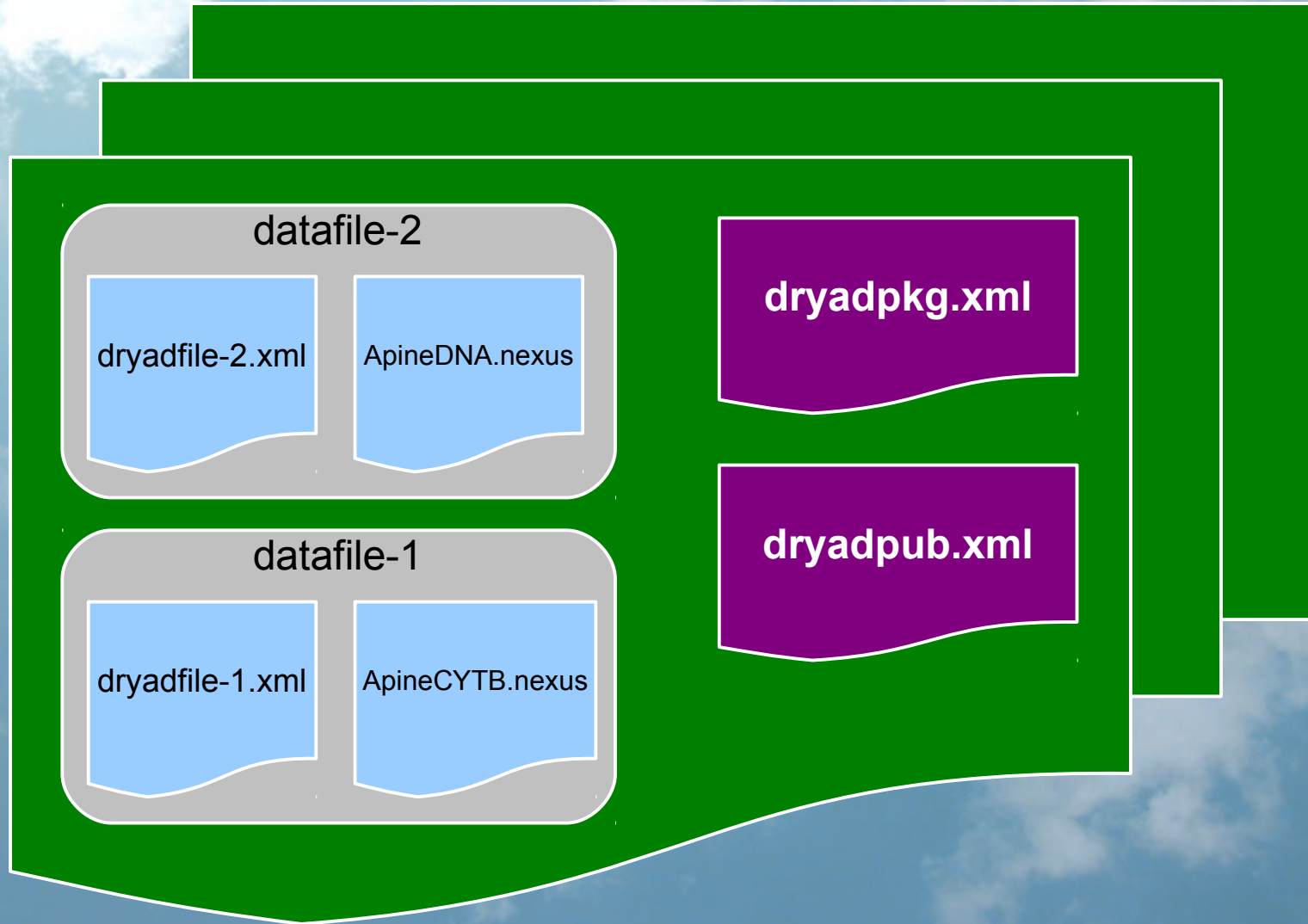
bagit.txt

manifest-md5.txt

data

tagmanifest-md5.txt

Dryad Data in the Bag



HTTP PUT Handshake



Lessons Learned

- ✓ Just enough to get the job done and no more
- ✓ Less local conventions and more “standards”
- ✓ There will always be custom solutions
- ✓ Options are developing quickly in this space

Future Directions

Less reliance on local conventions

- ✓ Plan to use OAI-ORE and Pairtree(s) within BagIt

OAI-ORE: Because it's Linked Data

Pairtree Filesystem

- ✓ So we can dereference URIs in ORE Resource Maps
<http://dx.doi.org/10.5061/dryad.8343>

URI prefix: <http://dx.doi.org/10.5061/dryad>.

Path: 83/43

83/43/Arctostaphylos.nex

Other Interesting Developments

DataONE

- ✓ Distributing data files and metadata
- ✓ May support packages in the future

“Dropbox of Bags”

or Bag replication network (BagNet?)

METS in Bags (in contrast to ORE)

The background of the slide is a bright blue sky filled with numerous small, white, fluffy clouds. The clouds are scattered across the entire frame, with some appearing larger and more distinct than others. The overall tone is light and airy.

The End

The cake was a lie

References

Dryad Code

<http://dryad.googlecode.com>

Dryad Data Repository

<http://datadryad.org>

BagIt

<http://en.wikipedia.org/wiki/BagIt>

OAI-ORE Primer

<http://www.openarchives.org/ore/1.0/primer>

OAI-ORE in BagIt

http://groups.google.com/group/oai-ore/browse_thread/thread/3ebfa7fcb4588048

ADMIRAL Data Packages (Planning ORE in BagIt)

http://imageweb.zoo.ox.ac.uk/wiki/index.php/ADMIRAL_data_packages

DSpace Packagers

<https://wiki.duraspace.org/display/DSPACE/PackagerPlugins>