-- DRAFT – DRAFT --
Data Management Plan for
Development and analysis of a database of Landsat Thematic Mapper imagery to support cross-site research
2012 LTER Working group
Names?

Summary

With 2012 post-doc salary from the Network, Kyle Cavanaugh generated a collection of Landsat TM imagery (scenes and JPG 'quicklooks') that cover each of the 26 current sites in the LTER Network from 1985 to the present. The work is expected to generate approximately 10,000 data packages, each composed of one zipped file of scene(s) and one quicklook, and a total of 2Tb data. Packages will be made accessible through the Network catalog. In a second phase (Phase II, to be proposed for 2013) we will create higher-level cross-site data products including atmospherically corrected surface reflectance data. This document describes the data management needs for the 2012 work, but may be expanded for Phase II.

Data packages from this project are expected to be structurally uniform, i.e., each package will be composed of a zipped file of one or more Landsat TM scenes, and one 'quicklook' (JPG). The metadata are already available with each scene in text format. This means that necessary software can be lightweight and the process uncomplicated, which will help assure efficiency and portability. We plan to process metadata with XML transforms and Linux (bash) scripts. Any issues that arise will most likely be due to the size of the collection.

Goals

1. Data entities delivered to LTER Network Office, housed in an archive with long-term Internet accessibility
2. EML metadata for scenes and JPG quicklooks
3. Packages (EML metadata + data files) uploaded to LTER data catalog
4. Initial design for customized web interface for browsing and delivery (if necessary)
5. Maintenance, re-use in Phase II if indicated

Progress to date

Goal 1: Tasks were described in the post-doc award and were completed by Kyle Cavanaugh and John Vande Castle.

Goal 2: Tasks 2.1, 2.2, 2.3 were completed in 2012. Apart from Cavanaugh (supported by the post-doc award), these people contributed time: Margaret O'Brien (35 hours), Theresa Valentine (8 hours) and M. Gastil-Buhl (4 hours). Task 2.4 does not have responsibility assigned, and has not been started.

Tasks for other data management goals have not been completed (nor have responsible individuals been assigned). The Network catalog programmers became aware of this project in late 2012, and some consultation has begun.

Tasks and personnel

It was decided that the simplest route to EML metadata would be via XML transform of a template, populating specific fields with external metadata. Time estimates (below) assume that the individual already possesses a high degree of understanding of appropriate software and LTER practices, e.g., metadata requirements, the PASTA Quality Engine (QE), and XML transformations.

Goal 1: Data entities delivered to LTER Network office and housed in archive with long-term Internet accessibility

Task	Task Description	Responsible
1.1	Collect Scenes and quicklooks	Cavanaugh
1.2	House scenes at LNO	Vande Castle

Goal 2: EML metadata for scenes and JPG quicklooks

Task	Task Description	Responsible (consulting)
2.1	Develop a flat representation, in XML, of the text metadata already available for each scene	Kyle Cavanaugh (O'Brien)
2.2	EML metadata template	O'Brien (Cavanaugh, Gastil-Buhl)
2.3	XSL stylesheet to populate the template (see task 2.2) from flat XML (see task 2.1). Evaluate a minimal number of resulting EML documents for structural integrity and content (PASTA QE).	O'Brien (Cavanaugh, Valentine, Gastil-Buhl)
2.4	Generate EML metadata for 10,000 data packages a) Write a script to generate transforms with parameters, loop through XML files (task 2.1), run XSL transform with template (task 2.2). Scripting language TBD, consider reuse in Phase II. b) Organize inputs, run script for all scenes, assure success, and rerun as necessary. NOTE: Task2.4 should be tightly coupled to Task 3.1, since outcome of package evaluation (3.1) may require adaptation of the script. Time estimate: 25 hours	TBD

Goal 3: Packages (EML metadata + data files) uploaded to LTER data catalog

Task	Task Description	Responsible
3.1	Individually evaluate a subset of data packages (e.g. 5%, randomly chosen) against the PASTA QE. Implementation is TBD by whomever is responsible, but review of QE reports would be most easily completed with a script. Time estimate: 20 hours	TBD
3.2	Consult with PASTA developers. Confirm readiness of cataloging system for bulk upload of this scale Time estimate: 2-4 hours	TBD
3.3	Monitor upload. Confirm success of all packages Adapt script from task 3.1 for this purpose. Time estimate: 20 hours	TBD

Goal 4: Initial design for customized web interface for browsing and delivery

Task	Task Description	Responsible
4.1	Outline requirements, consider other catalog-interface activities already underway Time estimate:	TBD
4.2	Evaluate need for a custom interface, considering the size of this collection and its potential audience. Time estimate:	TBD

Goal 5: Maintenance, re-use in Phase II

Task	Task Description	Responsible
5.1	Assemble/write documentation. Organize software for potential use should these packages need to be revised (revisioned). Identify responsible Time estimate:	TBD
5.2	Assess usability of scripts and templates for Phase II, determine data management needs for that project. Time estimate:	TBD

Resources:

Original proposal (2012):
 http://intranet2.lternet.edu/content/development-and-analysis-database-landsat-thematic-mapper-imagery-support-cross-site-researc

Organization meeting at 2012 ASM (ad hoc working group):
http://asm2012.lternet.edu/working-groups/using-remote-sensing-tools-calculate-biomass-consistently-across-lter-sites

ASM working group which featured this project:
 http://asm2012.lternet.edu/working-groups/data-package-construction-landsat-imagery

Phase II proposal (2013):
http://intranet2.lternet.edu/content/atmospheric-correction-lter-landsat-catalog