

Practical Innovations of the NIS and PASTA

Duane Costa

NIS Lead Programmer

2012 Mid-term Review



**LTER NETWORK
OFFICE**

Four Key Practical Innovations

INTEROPERABILITY

PROVENANCE

ACCESS

QUALITY



Why “practical innovations”?

- ▶ What’s *practical* about them?
 - All four can be demonstrated in the NIS prototype
 - Add to the flexibility, usefulness, responsiveness, and data accessibility for individual LTER sites and users outside of the LTER community
- ▶ What’s *innovative* about them?
 - All four capabilities are either deficient or absent from the legacy LTER Data Catalog
 - Innovative even in the context of the greater ecoinformatics community, in particular:
 - Provenance tracking
 - Quality reporting



Legacy LTER Data Catalog

<http://metacat.lternet.edu>

LTRE Data Portal - Mozilla Firefox

File Edit View History Bookmarks Tools Help

metacat.lternet.edu/das/lter/advancedsearch.jsp

Google Calendar AT-del AT-gra AT-lav AT-tep AT-tro DI-624 DP-del DP-gra DP-lav DP-tep DP-tro EVO Google Java SE 6 >>

LTRE Home | Intranet | LNO LTER Site Home Pages Go

LTER The US Long Term Ecological Research Network
A founding member of the International Long Term Ecological Research Network

Login

Search

Browse

LTER Data Portal Advanced Search

LTER Sites

Andrews LTER
Arctic LTER
Baltimore Ecosystem Study
Bonanza Creek LTER

Subject

Keywords Only contains All Terms

☒ Add more specific terms ☐ Add related terms ☐ Add related terms and their more specific terms

Owners/Creators

Individual's Last Name: contains

Organization: contains

Spatial Criteria

Zoom in to the region you would like to search

North 90.0
West -180.0
South -90.0

☐ Dataset must be fully contained within the specified region

Geographic Place Name:

POWERED BY Google

Terms of Use

Taxonomic Criteria

LTRE Data Portal - Mozilla Firefox

File Edit View History Bookmarks Tools Help

metacat.lternet.edu/das/lter/browseforward.jsp

Google Calendar AT-del AT-gra AT-lav AT-tep AT-tro DI-624 DP-del DP-gra DP-lav DP-tep DP-tro EVO Google Java SE 6 >>

LTRE Home | Intranet | LNO LTER Site Home Pages Go Google Custom Search Search

LTER The US Long Term Ecological Research Network
A founding member of the International Long Term Ecological Research Network

Login

333 data packages found

[Total data packages](#) 333

[LTER data packages](#) 333

[CAP](#) 333

Search Results (click on title for more information)

- Package includes URL(s) that should link directly to data
- Package includes a URL that may link to information, metadata, or data

Show All Hide All

View	LTER Site	Data Package Title/Owners/Creators	Data
+/-	CAP LTER	"Arbuscular mycorrhizal fungal diversity and functioning in urban desert preserves and surrounding deserts" - Stutz Ontiveros	•
+/-	CAP LTER	"Hierarchical regulation of nitrogen export from urban catchments: Interactions of storms and landscapes." - Lewis Grimm	•
+/-	CAP LTER	"Phoenix Area Social Survey (PASS)" - Harlan	•
+/-	CAP LTER	"Lichen Resurvey with Heavy Metal Analysis: Distribution of Praseodymium concentration in lichen tissue in Maricopa County" - Gries Zschau	•
+/-	CAP LTER	"Residential completions in Maricopa County" - Walton	•
+/-	CAP LTER	"False Color Landsat Image of Greater Phoenix" - National Aeronautic and Space Administration, Geology Remote Sensing Lab, ASU	•
+/-	CAP LTER	"Environmental Risk and Justice: Facilities 2000 with Toxic Release Inventory data" - Bolin Atwood	•
+/-	CAP LTER	"Environmental Risk and Justice: Facilities 2000 with Toxic Release Inventory data in 1970 tracts" - Bolin Atwood	•
+/-	CAP LTER	"Environmental Risk and Justice: Facilities 1995 with Toxic Release Inventory data in 1970 tracts" - Bolin Atwood	•
+/-	CAP LTER	"Environmental Risk and Justice: Facilities 1995 with Toxic Release Inventory data" - Bolin Atwood	•
+/-	CAP LTER	"Environmental Risk and Justice: Facilities 1990 locations from Toxic Release Inventory" - Bolin Atwood	•

LTER Network Office

Legacy LTER Data Catalog: Does some things well

- ▶ 2004–Present
- ▶ 7000+ LTER metadata (EML) documents
- ▶ Good search capability
 - Simple and advanced search
 - Utilizes LTER Controlled Vocabulary
- ▶ Reasonable performance
- ▶ Recent improvements to UI
 - Search results presentation
 - Metadata presentation
 - Data access

<http://metacat.lternet.edu>



Legacy LTER Data Catalog: Where it needs improvement

- ▶ Performance could be better
- ▶ Mostly closed system:
 - Web application, not web services
 - Machine-to-machine interaction available via Metacat back-end, but not prominent
- ▶ Data access has improved but is still uneven
 - Historically, a large source of user frustration
- ▶ Lacks provenance tracking
 - Important to synthesis efforts
- ▶ Minimal quality control
 - Valid EML metadata is the only requirement for insertion

<http://metacat.lternet.edu>



Four Key Practical Innovations

INTEROPERABILITY

PROVENANCE

ACCESS

QUALITY



Interoperability

- ▶ Open system
 - Service Oriented Architecture, Web services API
- ▶ Flexibility to build the client application *you* want using the platform and programming language of *your* choice
- ▶ Could potentially be utilized by NEON, GLEON, CUASHI, etc.



PASTA Web Service API

Data Package Manager API - PASTA Design & Architecture - LTER NIS Confluence - Mozilla Firefox

File Edit View History Bookmarks Tools Help

Google Calendar AT-del AT-gra AT-lav AT-tep AT-tro DI-624 DP-del DP-gra DP-lav DP-tep DP-tro EVO Google Java SE 6 NIS NIS Confluence Rivendell

Dashboard > PASTA Design & Architecture > ... > Data Package Manager API

Browse Duane Costa Search Confluence

Added by [Danielle Stevens](#), last edited by [Danielle Stevens](#) on Apr 12, 2012 ([view change](#))

Data Package Manager web service API

Base URL - <https://package.lternet.edu/package>

The Data Package Manager Web Service provides a suite of operations to create, evaluate, read, update, delete, list, and search data package resources in the PASTA system. Data package resources include metadata documents, data entities, and quality reports.

Request Summary	
HTTP Verb : Relative URL	Brief description
POST : /eml	Create Data Package operation, specifying the EML document describing the data package to be created in the message body.
DELETE : /eml/{scope}/{identifier}	Delete Data Package operation, specifying the scope and identifier of the data package to be deleted in the URI.
POST : /evaluate/eml	Evaluate Data Package operation, specifying the EML document describing the data package to be evaluated in the message body.
GET : /data/eml/{scope}/{identifier}/{revision}	List Data Entities operation, specifying the scope, identifier, and revision values to match in the URI.
GET : /eml/{scope}/{identifier}	List Data Package Revisions operation, specifying the scope and identifier values to match in the URI.
GET : /eml/{scope}	List Data Package Identifiers operation, specifying the scope value to match in the URI.
GET : /eml	List Data Package Scopes operation, returning all scope values extant in the data package registry.
GET : /eml/deleted	List Deleted Data Packages operation, returning all document identifiers (excluding revision values) that have

Powered by Atlassian Confluence 4.0, the Enterprise Wiki | [Report a bug](#) | [Atlassian News](#)

Copyright © 2009-2011 [Long Term Ecological Research Network](#). This material is based upon work supported by the [National Science Foundation](#) under Cooperative Agreements [#DEB-0832652](#) and [#DEB-0936498](#). Any opinions, findings, conclusions, or recommendations expressed in the material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation. Please [contact us](#) with questions, comments, or for technical assistance regarding this web site or the LTER Network.

Find: Network Next Previous Highlight all Match case Reached end of page, continued from top

Interoperability: Examples

- ▶ Example 1: EML Congruency Checker (Summer, 2011)
 - Margaret O'Brien, SBC Information Manager
 - Generated LTER-wide quality reports using an early implementation of the Quality Engine
 - Perl scripts and shell scripts
- ▶ Example 2: EML Pre-flight Checker (Winter, 2012)
 - Sven Bohm, KBS Information Manager
 - Updated version of the ECC using a newer PASTA API
 - Ruby on Rails
- ▶ Example 3: NIS Data Portal (Winter-Spring, 2012)
 - Serves as a reference implementation of a PASTA client application
- ▶ Example 4: Audit report web application for a particular LTER site
 - Example of a potential client application



“The PASTA web services were very powerful in their assessment of the data, but also simple to access using only Linux command line tools.”

Margaret O'Brien

Information Manager, Santa Barbara Coastal LTER

Used with permission



Four Key Practical Innovations

INTEROPERABILITY

PROVENANCE

ACCESS

QUALITY

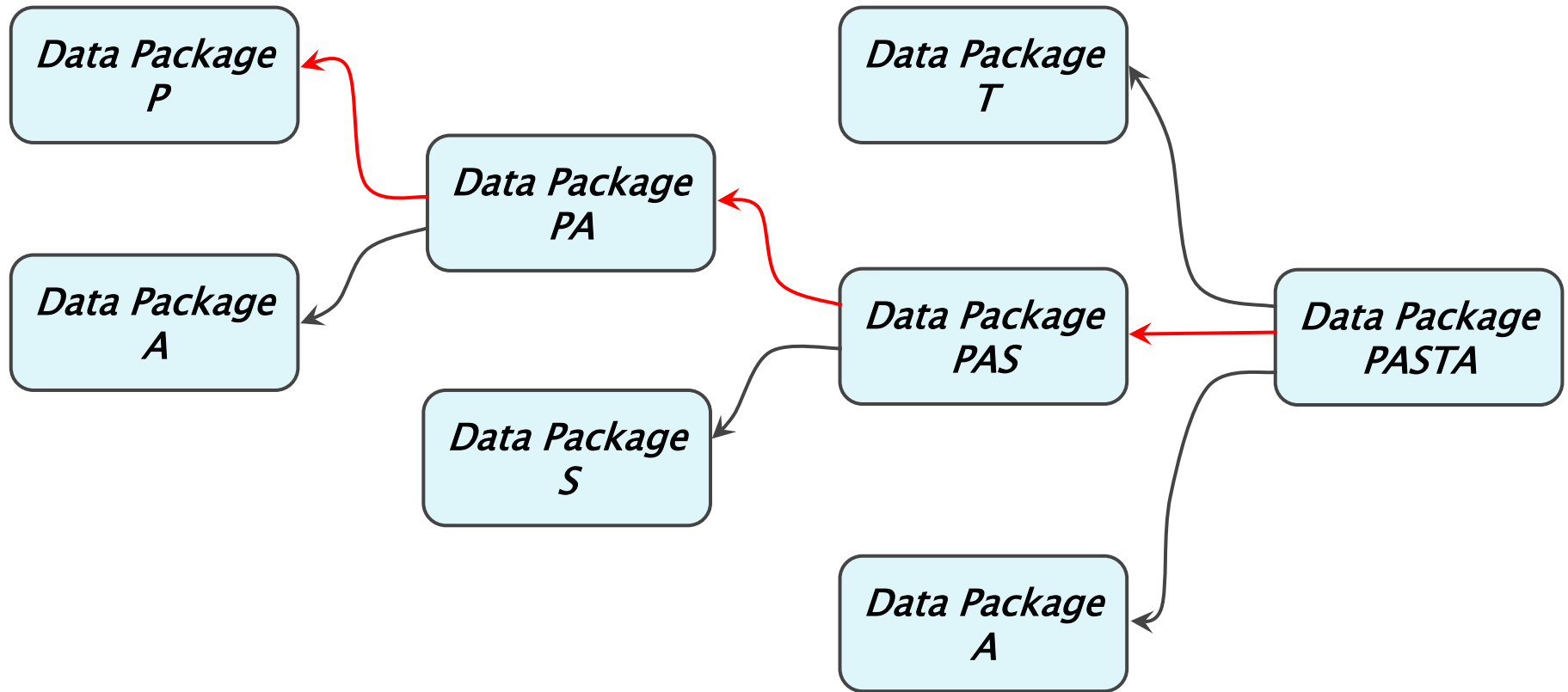


Provenance

- ▶ It's the "P" in "PASTA"
- ▶ Critical for documentation and understanding of synthesis/derived products
- ▶ Innovative use of the "methods" section of EML to document provenance
- ▶ Provenance Factory generates provenance block in EML metadata

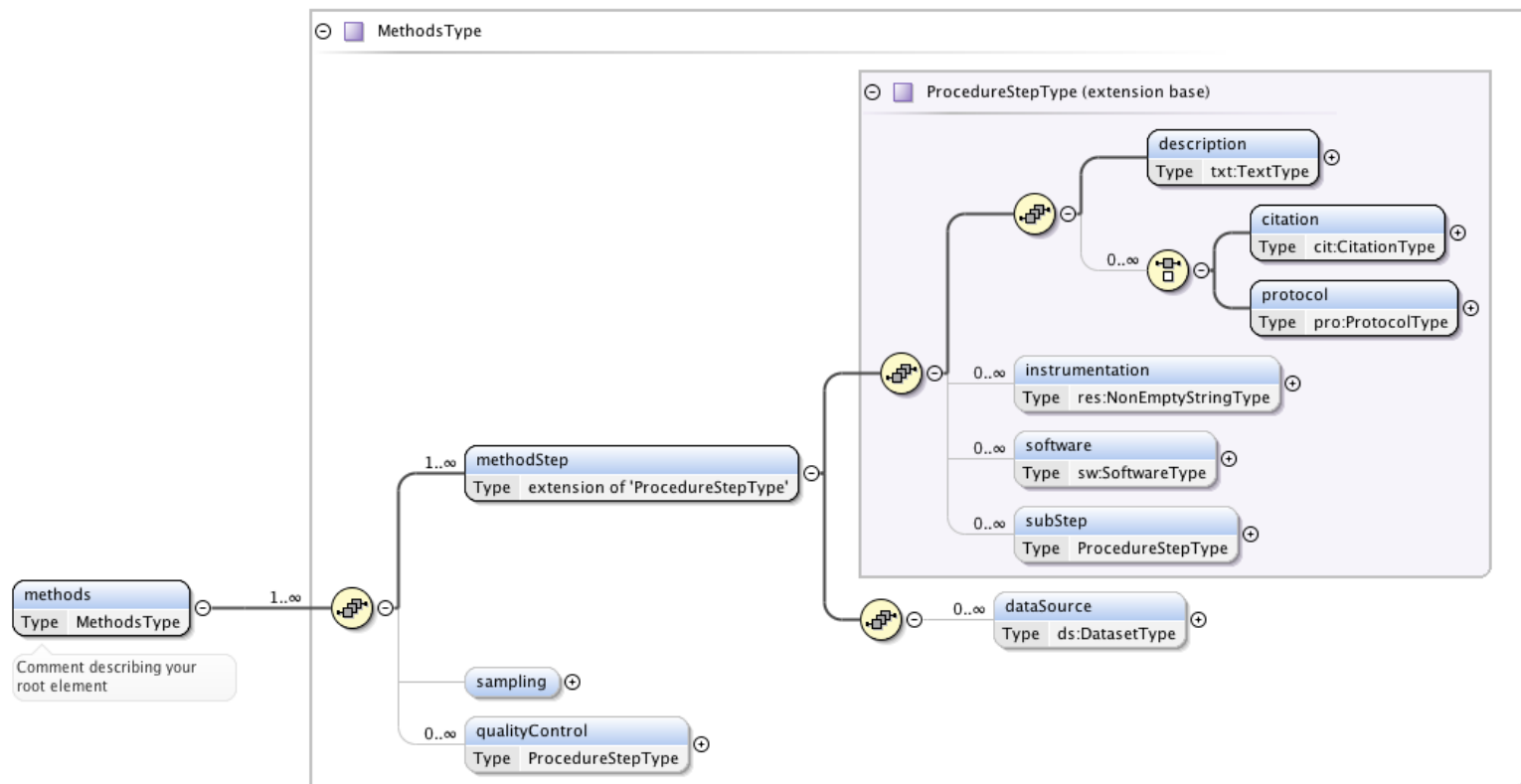


Provenance Chaining



Provenance Factory...

Generates provenance metadata for use in PASTA dependent derived products



Provenance in PASTA

Geographic Coverage	
Geographic Description:	North Inlet encompasses about 2,630 hectares of tidal marshes and wetlands near Georgetown, South Carolina, USA. North Inlet-Winyah Bay features high quality, ocean-dominated waters and salt marshes in North Inlet, contrasting with the brackish waters and marshes of Winyah Bay. The bay estuary is dominated by riverine discharges from a watershed impacted by agricultural, municipal and industrial development. Former rice fields and canals provide another system for study within the Reserve. The Debidue site is located at the confluence of Town Creek and Debidue Creek. The Bread and Butter site is located along the western shoreline of Town Creek adjacent to the mouth of Clambank Creek. Oyster Landing in Crab Haul creek within the NORTH INLET ESTUARY SYSTEM Georgetown, South Carolina. 33,20 lat. 79,11 long.
Bounding Coordinates:	-79.2936 W, -79.1042 E, 33.357 N, 33.2125 S
Geographic Coverage	
Geographic Description:	Oyster Landing in Crab Haul Creek 33.21'2" Lat., 79.11'27" Long.
Bounding Coordinates:	-79.1175 W, -79.1175 E, 33.2106 N, 33.2106 S
Methods	
Method Step 1 :	<p>Computational methods</p> <p>The dew point values are computed based on the daily mean relative humidity and the current temperature using the following equation:</p> $T_d = \frac{b \cdot \gamma(T, RH)}{a - \gamma(T, RH)}, \text{ where } \gamma(T, RH) = \frac{a \cdot T}{(b + T) + \ln(RH/100)} \text{ and } a = 17.271 \text{ and } b = 237.7 \text{ degC}$
Method Step 2 :	<p>The following data package was used in the creation of this product:</p> <p>National Weather Service data for North Inlet Estuary, South Carolina, from 1986 to 1992, North Inlet LTER (Click here to view metadata)</p>



Four Key Practical Innovations

INTEROPERABILITY

PROVENANCE

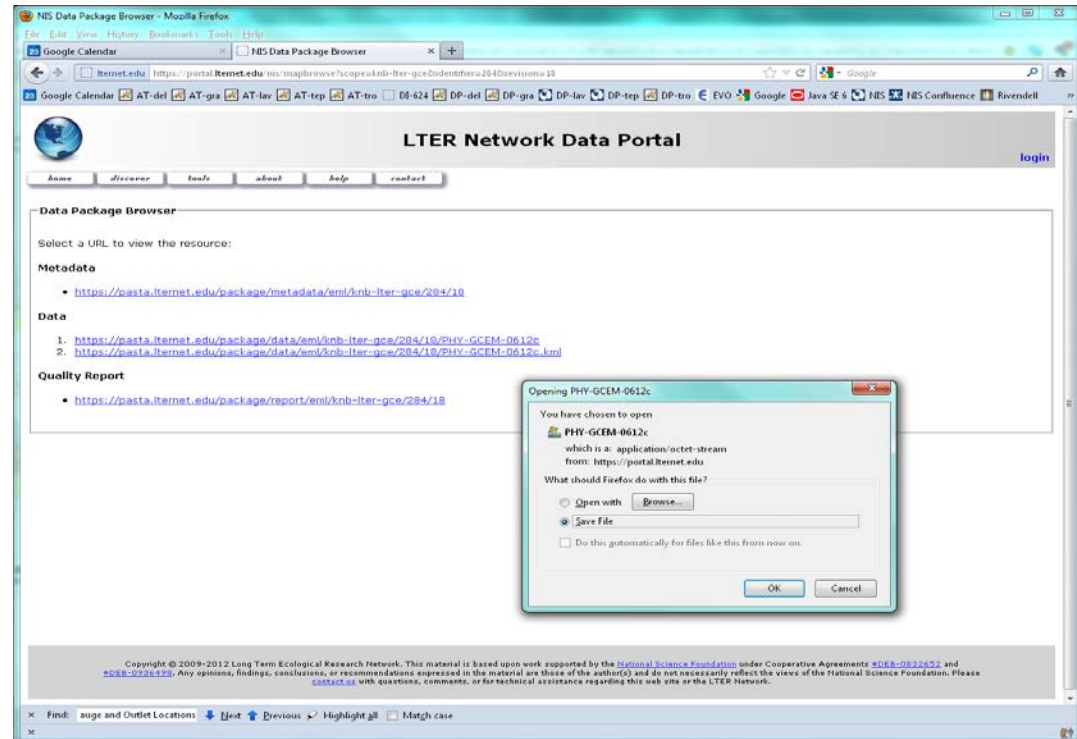
ACCESS

QUALITY



Access

- ▶ All NIS data packages link to data
 - A *quality check* guarantees this (more on this later)
- ▶ Local data storage
 - Snapshot of data entities at the time the data package was submitted by the LTER site
 - Consistent and reliable storage and retrieval



Four Key Practical Innovations

INTEROPERABILITY

PROVENANCE

ACCESS

QUALITY



Quality

- ▶ Could deliver a talk titled "Quality, Quality, Quality, and More Quality"
- ▶ LTER EML Metrics Working Group
 - ASM 2009 to present
 - Very active and productive group; includes DataONE participants
- ▶ NIS Tiger Teams
 - Data Manager, Data Package Manager, Metadata Quality



Quality Engine

- ▶ A subcomponent of the Data Package Manager
- ▶ Generates a quality report for each data package
- ▶ A quality report contains a set of quality checks
- ▶ Stored as XML but rendered in HTML for human readability
- ▶ 19 quality checks implemented in the NIS prototype
- ▶ 50+ quality checks documented by EML Metrics Working Group and Metadata Quality Tiger Team
- ▶ Quality Engine is available to the greater ecoinformatics community via the Data Manager Library (ecoinformatics.org)



What's a Quality Check?

- ▶ An individual metric or a best practice
- ▶ It may involve looking at:
 - metadata (independent of data), *or*
 - data (independent of metadata), *or*
 - congruency between metadata and data
- ▶ Can result in one of four statuses
 - **valid**
 - **info**
 - **warn**
 - **error**




How is the Quality Engine used in the NIS?

- ▶ Used as a litmus test
 - Any **error** status reported by a quality check blocks insertion of the data package into PASTA
- ▶ Users can evaluate data packages before inserting them into PASTA
 - Key idea contributed by the Data Manager Tiger Team
 - Helps site Information Managers prepare their data packages for insertion
- ▶ Quality report is a resource of the data package
 - Persists and can be accessed alongside metadata and data resources



Data Package Resource Map and Quality Report


LTER Network Data Portal
NIN
logout

[home](#)
[discover](#)
[tools](#)
[about](#)
[help](#)
[contact](#)

Data Package Browser

Select a URL to view the resource:

Metadata

- <https://pasta.lternet.edu/package/metadata/eml/knb-lter-nin/1/2>

Data

- <https://pasta.lternet.edu/package/data/eml/knb-lter-nin/1/2/DailyWaterSample-NIN-LTER-1978-1992>

Quality Report

- <https://pasta.lternet.edu/package/report/eml/knb-lter-nin/1/2>

Copyright © 2009-2011
Cooperative Agreements
of the author(s) and do

7	databaseTableCreated	valid	<ul style="list-style-type: none"> Type: metadata System: knb On Failure: error 	Database table created	Status of creating a database table	A database table is expected to be generated from the EML attributes.	A database table was generated from the description
8	examineRecordDelimiter	valid	<ul style="list-style-type: none"> Type: congruency System: knb On Failure: warn 	Data are examined and possible record delimiters are displayed	If no record delimiter was specified, we assume that '\r\n' is the delimiter. Search the first row for other record delimiters and see if other delimiters are found.	No other potential record delimiters expected in the first row.	No other potential record delimiters were detected previously detected
9	tooManyFields	error	<ul style="list-style-type: none"> Type: congruency System: knb On Failure: error 	Data does not have more fields than metadata attributes	Compare number of fields specified in metadata to number of fields found in a data record	15 fields	17 fields
10	dataLoadStatus	warn	<ul style="list-style-type: none"> Type: congruency System: knb On Failure: 	Data can be loaded into the database	Status of loading the data table into a database	No errors expected during data loading or data loading was not attempted for this data entity	One or more errors during data loading

Four Key Practical Innovations

LTER NIS is now poised to utilize these four key practical innovations as enabled by PASTA

INTEROPERABILITY

PROVENANCE

ACCESS

QUALITY

