

# AN OVERVIEW OF THE EDI DATA REPOSITORY AND DATA PORTAL (AND HOW TO USE IT FOR PUBLISHING DATA)

by Mark Servilla  
Environmental Data Initiative



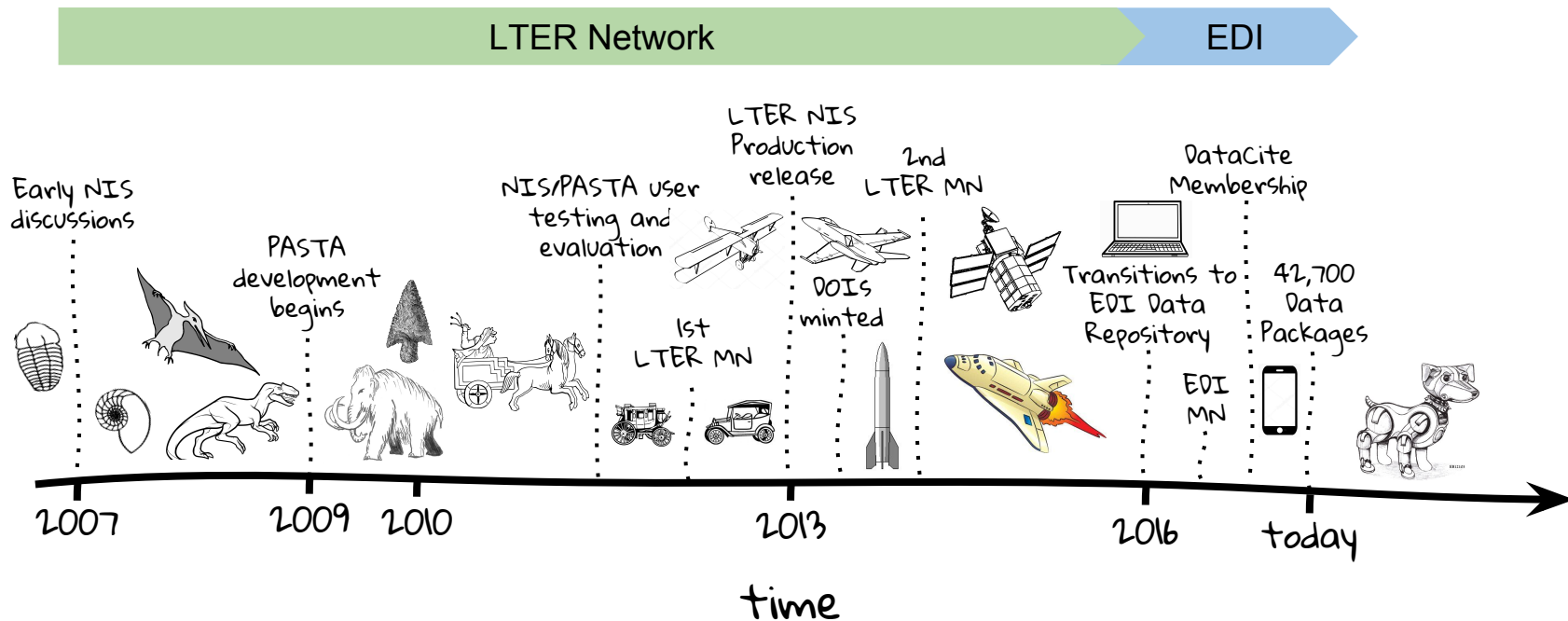
# TOPICS

- Overview of the
  - EDI Data Repository
  - PASTA software
- Use of the EDI Data Portal to
  - evaluate data packages
  - upload and publish data packages

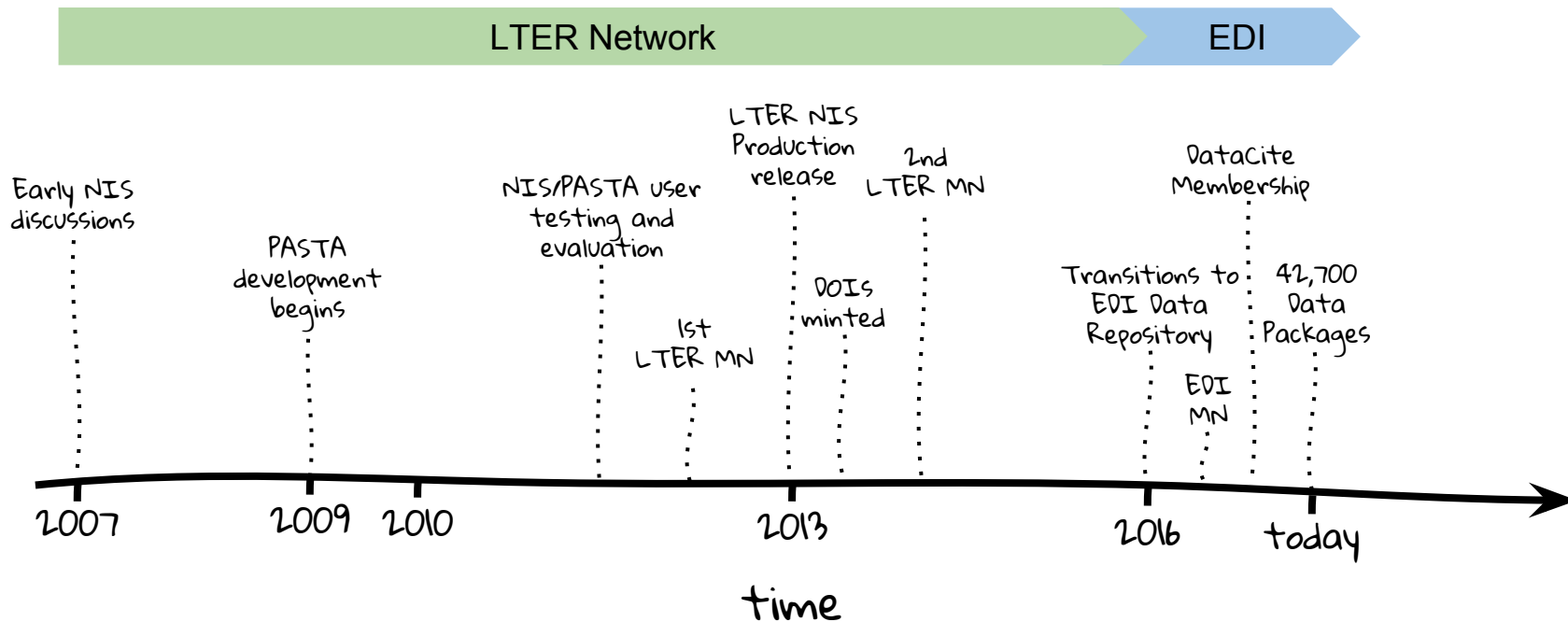
# PREMISES

- You are familiar with the Ecological Metadata Language (EML)
- You know how to describe a data package with EML

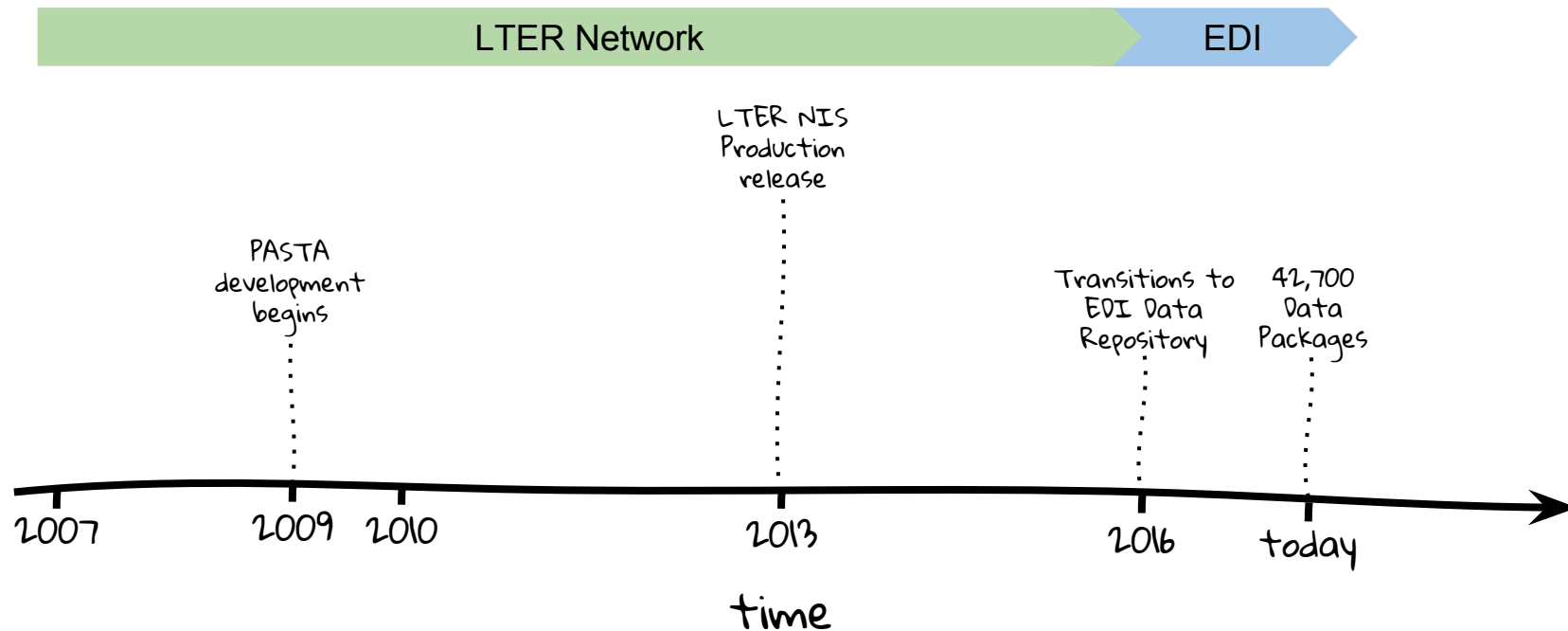
# HISTORY (ANNOTATED)



# HISTORY



# HISTORY (HIGHLIGHTS)



# WHAT IS THE EDI DATA REPOSITORY?

- An Internet accessible "open" data repository
- Uses the PASTA data repository software
- Generates Digital Object Identifiers for all public data packages
- Is a DataONE member node operator (two)
- Contains about 42,700 data packages
- Stores about 8.5TB of data
- Recognized by Springer Nature, Scientific Data, ESA, r3data, PloS, ...

# WHAT IS THE EDI DATA REPOSITORY?

- An Internet accessible "open" data repository
- Uses the PASTA data repository software
- Generates Digital Object Identifiers for all public data packages
- Is a DataONE member node operator (two)
- Contains about 42,700 data packages
- Stores about 8.5TB of data
- Recognized by Springer Nature, Scientific Data, ESA, r3data, PloS, ...



# WHAT IS THE EDI DATA REPOSITORY?

- An Internet accessible "open" data repository
  - Long Term Ecological Research (LTER) Network
  - Long Term Research in Environmental Biology (LTREB)
  - Macro-systems Biology (MSB)
  - Organization of Biological Field Stations (OBFS)

...but open to any environmental/ecological (size friendly) data

# WHAT IS THE EDI DATA REPOSITORY?

- An Internet accessible "open" data repository
- Uses the PASTA data repository software
- Generates Digital Object Identifiers for all public data packages
- Is a DataONE member node operator (two)
- Contains about 42,700 data packages
- Stores about 8.5TB of data
- Recognized by Springer Nature, Scientific Data, ESA, r3data, PloS, ...

# WHAT IS THE EDI DATA REPOSITORY?

- An Internet accessible "open" data repository
- Uses the PASTA data repository software
- Generates Digital Object Identifiers for all public data packages
- Is a DataONE member node operator (two)
- Contains about 42,700 data packages
- Stores about 8.5TB of data
- Recognized by Springer Nature, Scientific Data, ESA, r3data, PloS, ...

# WHAT IS THE EDI DATA REPOSITORY?

- An Internet accessible "open" data repository
- Uses the PASTA data repository software
- Generates Digital Object Identifiers for all public data packages
- Is a DataONE member node operator (two)
- Contains about 42,700 data packages
- Stores about 8.5TB of data
- Recognized by Springer Nature, Scientific Data, ESA, r3data, PloS, ...

# WHAT IS THE EDI DATA REPOSITORY?

- An Internet accessible "open" data repository
- Uses the PASTA data repository software
- Generates Digital Object Identifiers for all public data packages
- Is a DataONE member node operator (two)
- Contains about 42,700 data packages
- Stores about 8.5TB of data
- Recognized by Springer Nature, Scientific Data, ESA, r3data, PloS, ...

# WHAT IS THE EDI DATA REPOSITORY?

- An Internet accessible "open" data repository
- Uses the PASTA data repository software
- Generates Digital Object Identifiers for all public data packages
- Is a DataONE member node operator (two)
- Contains about 42,700 data packages
  - 6,300 researcher contributed
  - 21,000 Landsat images
  - 15,400 Ecotrends project

# WHAT IS THE EDI DATA REPOSITORY?

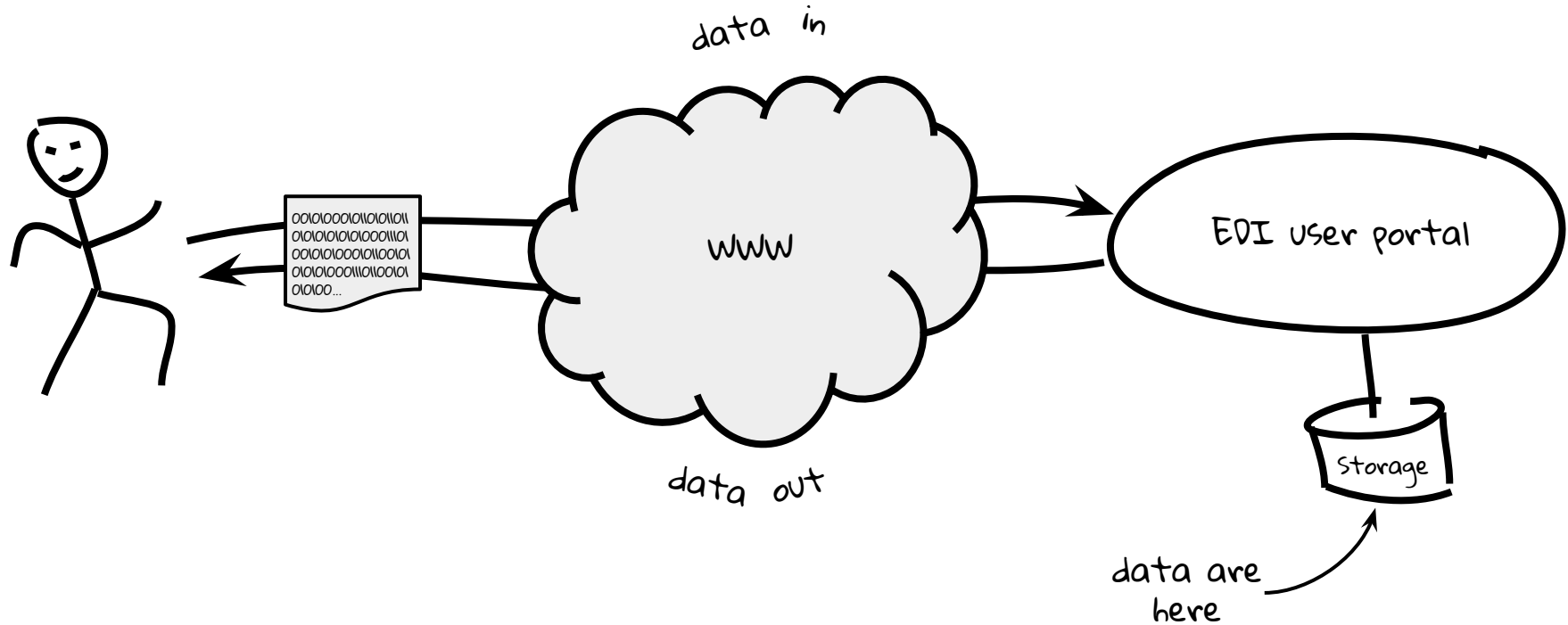
- An Internet accessible "open" data repository
- Uses the PASTA data repository software
- Generates Digital Object Identifiers for all public data packages
- Is a DataONE member node operator (two)
- Contains about 42,700 data packages
- Stores about 8.5TB of data
- Recognized by Springer Nature, Scientific Data, ESA, r3data, PloS, ...

# WHAT IS THE EDI DATA REPOSITORY?

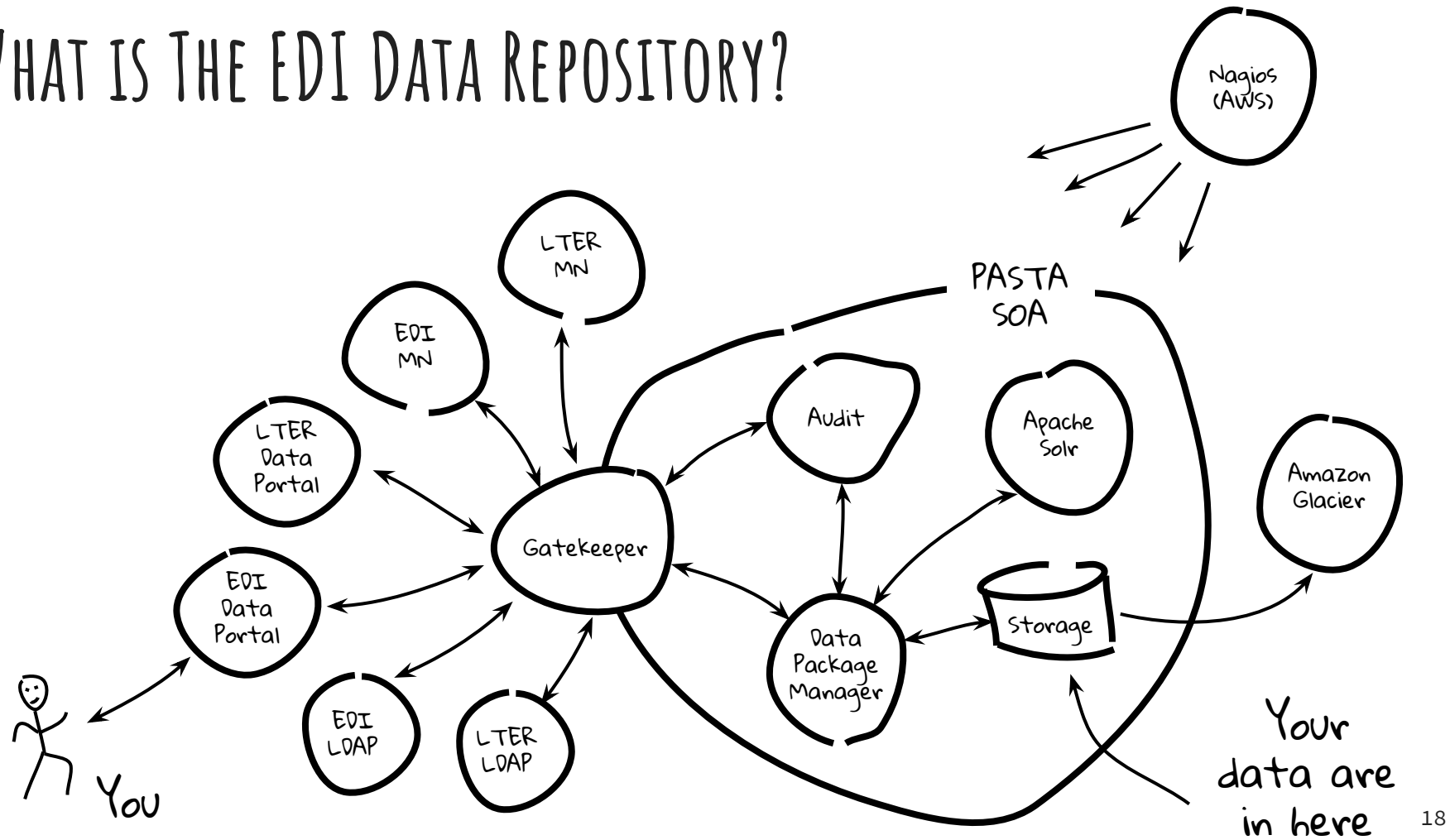
- An Internet accessible "open" data repository
- Uses the PASTA data repository software
- Generates Digital Object Identifiers for all public data packages
- Is a DataONE member node operator (two)
- Contains about 42,700 data packages
- Stores about 8.5TB of data
- Recognized by Springer Nature, Scientific Data, ESA, r3data, Plos, ...



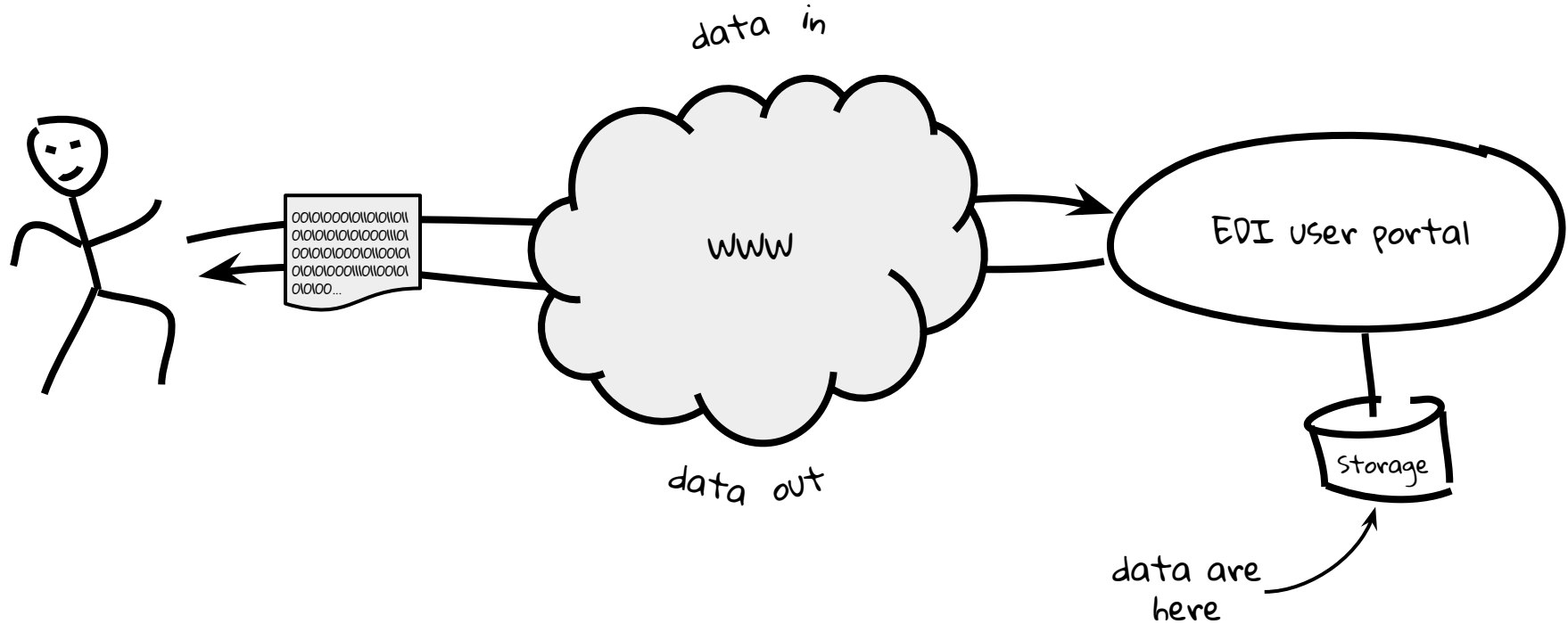
# WHAT IS THE EDI DATA REPOSITORY?



# WHAT IS THE EDI DATA REPOSITORY?



# WHAT IS THE EDI DATA REPOSITORY?



# WHAT IS PASTA?

- A metadata-driven data repository software system that supports:

# WHAT IS PASTA?

- A metadata-driven data repository software system that supports:
  - a "data package" model

# WHAT IS PASTA?

- A metadata-driven data repository software system that supports:
  - a "data package" model
  - data package quality evaluation

# WHAT IS PASTA?

- A metadata-driven data repository software system that supports:
  - a "data package" model
  - data package quality evaluation
  - provenance tracking

# WHAT IS PASTA?

- A metadata-driven data repository software system that supports:
  - a "data package" model
  - data package quality evaluation
  - provenance tracking
  - "strong" versioning



# WHAT IS PASTA?

- A metadata-driven data repository software system that supports:
  - a "data package" model
  - data package quality evaluation
  - provenance tracking
  - "strong" versioning
  - DOI assignment at the data package level

# WHAT IS PASTA?

- A metadata-driven data repository software system that supports:
  - a "data package" model
  - data package quality evaluation
  - provenance tracking
  - "strong" versioning
  - DOI assignment at the data package level
  - access control at the package and data level

# WHAT IS PASTA?

- A metadata-driven data repository software system that supports:
  - a "data package" model
  - data package quality evaluation
  - provenance tracking
  - "strong" versioning
  - DOI assignment at the data package level
  - access control at the package and data level
- Based on a Service Oriented Architecture design pattern

# WHAT IS PASTA?

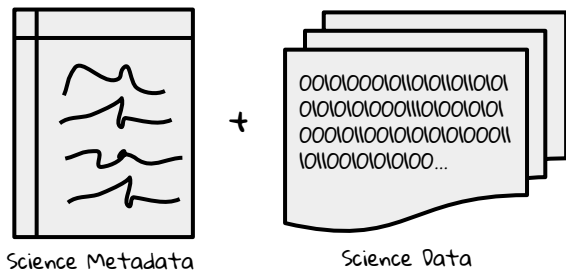
- A metadata-driven data repository software system that supports:
  - a "data package" model
  - data package quality evaluation
  - provenance tracking
  - "strong" versioning
  - DOI assignment at the data package level
  - access control at the package and data level
- Based on a Service Oriented Architecture design pattern
- Supports a REST web-service interface for CRUD operations

# WHAT IS PASTA?

- A metadata-driven data repository software system that supports:
  - a "data package" model
  - data package quality evaluation
  - provenance tracking
  - "strong" versioning
  - DOI assignment at the data package level
  - access control at the package and data level
- Based on a Service Oriented Architecture design pattern
- Supports a REST web-service interface for CRUD operations

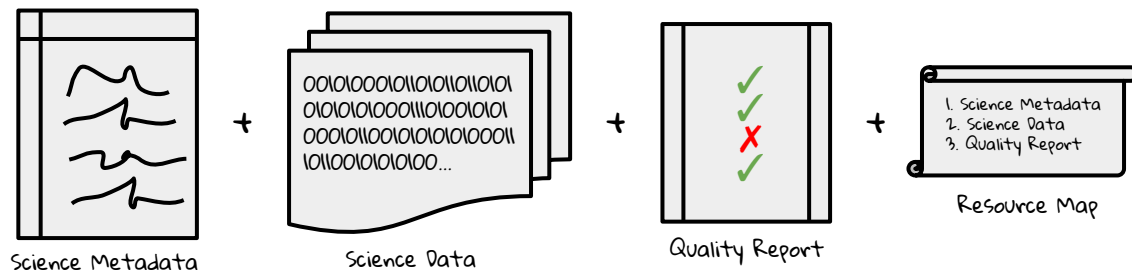
# WHAT'S A DATA PACKAGE?

**Data Package** (noun): an assemblage of science metadata (e.g., EML) and one or more science data objects; PASTA data packages include a quality report object and are described by package metadata called a "resource map" (i.e., manifest)



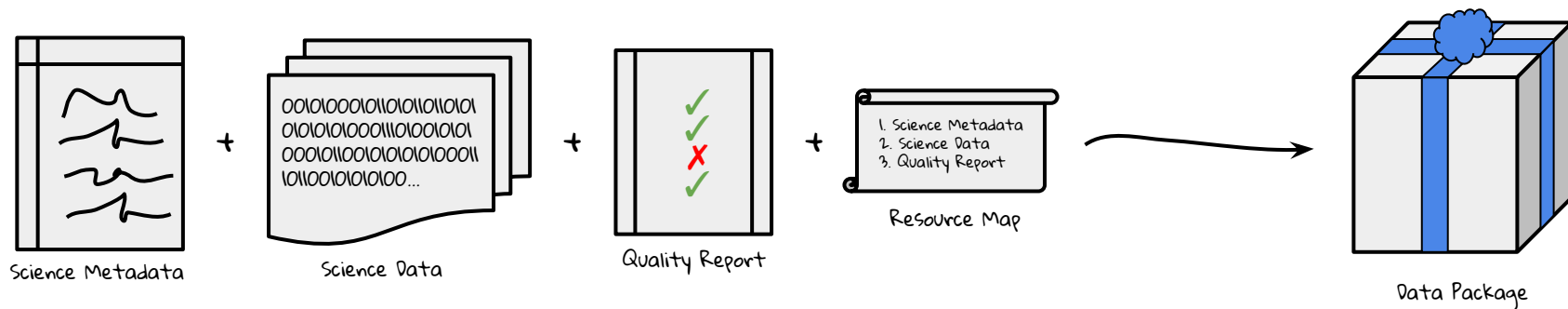
# WHAT'S A DATA PACKAGE?

**Data Package** (noun): an assemblage of science metadata (e.g., EML) and one or more science data objects; PASTA data packages include a quality report object and are described by package metadata called a "resource map" (i.e., manifest)



# WHAT'S A DATA PACKAGE?

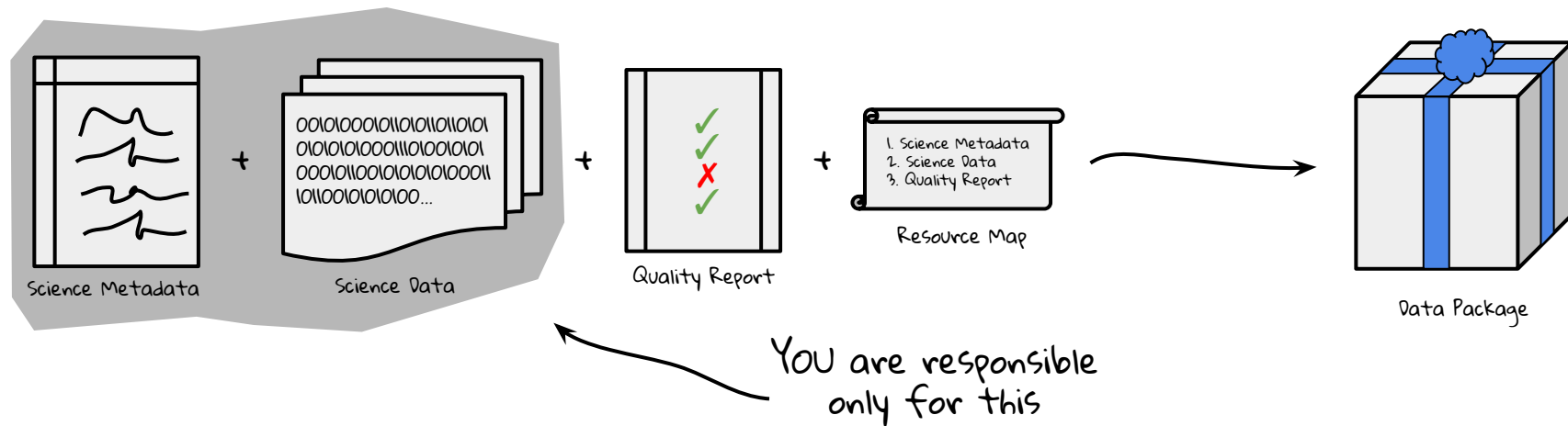
**Data Package** (noun): an assemblage of science metadata (e.g., EML) and one or more science data objects; PASTA data packages include a quality report object and are described by package metadata called a "resource map" (i.e., manifest)





# WHAT'S A DATA PACKAGE?

**Data Package** (noun): an assemblage of science metadata (e.g., EML) and one or more science data objects; PASTA data packages include a quality report object and are described by package metadata called a "resource map" (i.e., manifest)



# PASTA PACKAGE IDENTIFIERS

PASTA data package unique identifiers:

`edi.10.1`

# PASTA PACKAGE IDENTIFIERS

PASTA data package unique identifiers:

`edi.10.1`

`scope:identifier:revision`

# PASTA PACKAGE IDENTIFIERS

PASTA data package unique identifiers:

```
edi.10.1
```

```
scope:identifier:revision
```

String value that identifies the organization, project, or theme of the data package

# PASTA PACKAGE IDENTIFIERS

PASTA data package unique identifiers:

edi.10.1

scope:identifier:revision

Integer value that uniquely identifies the data package in the namespace of the scope

# PASTA PACKAGE IDENTIFIERS

PASTA data package unique identifiers:

edi.10.1

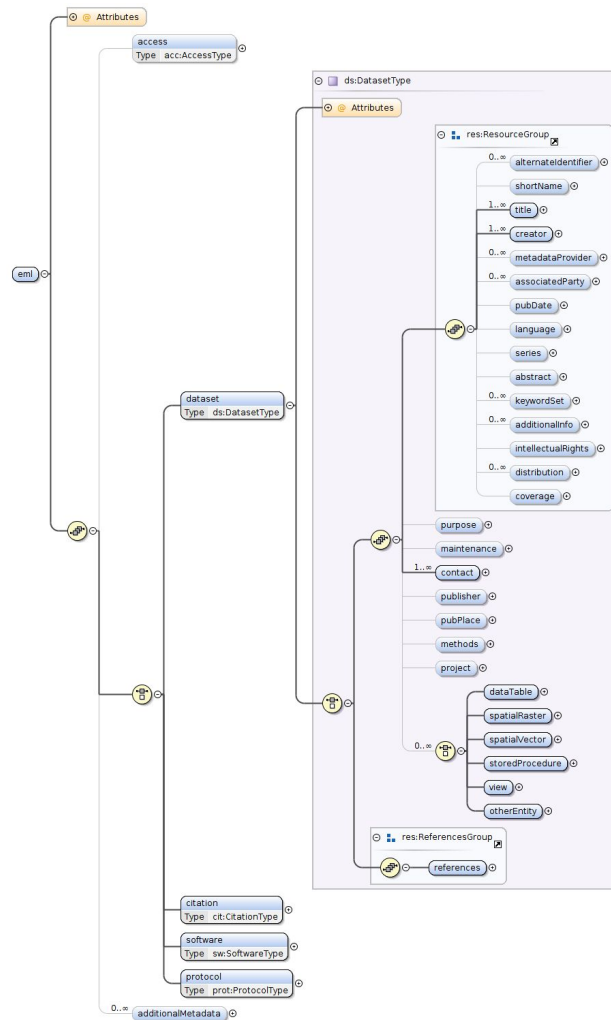
scope:identifier:revision

Integer value in increasing order that identifies the version of the data package

# METADATA-DRIVEN...

- Ecological Metadata Language (EML)
  - XML Schema
  - Current version 2.1.1  
(Under active development - 2.2.0 coming soon)
  - Describes the full data package  
(access, who, where, when, how, and what)

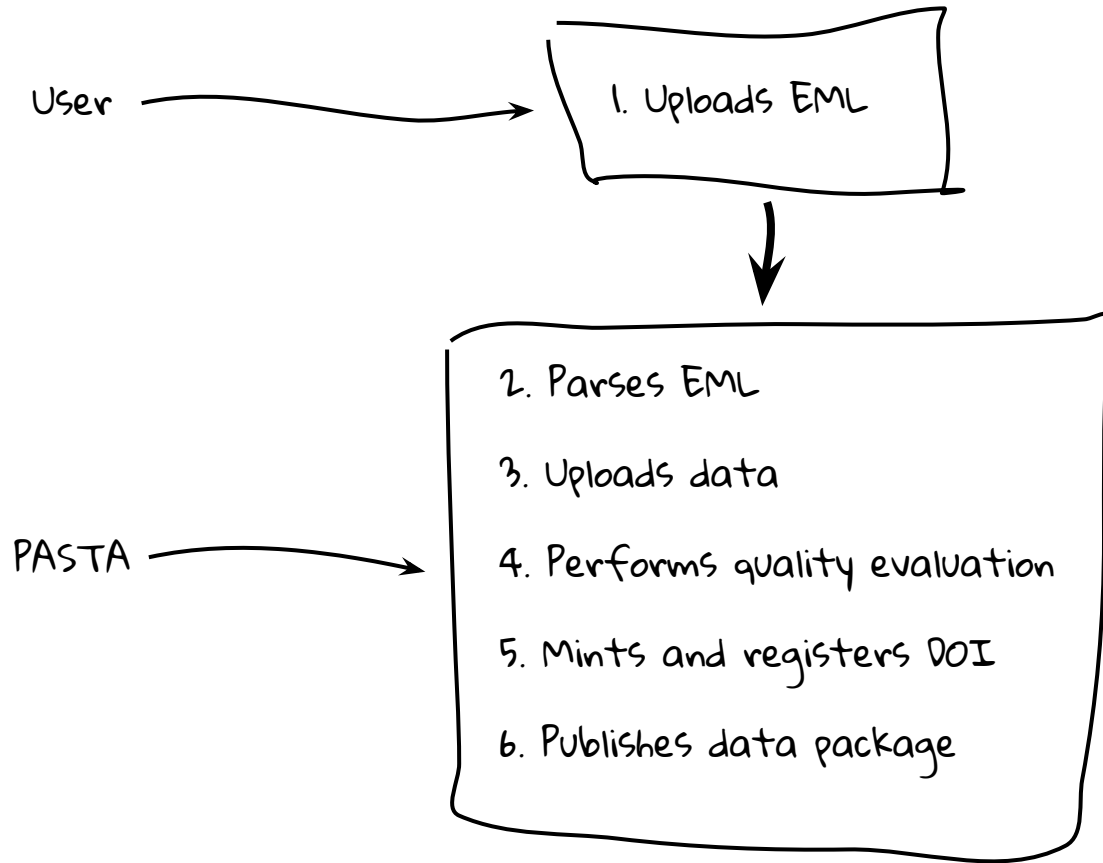
<https://knb.ecoinformatics.org/#external//emlparser/docs/index.html>

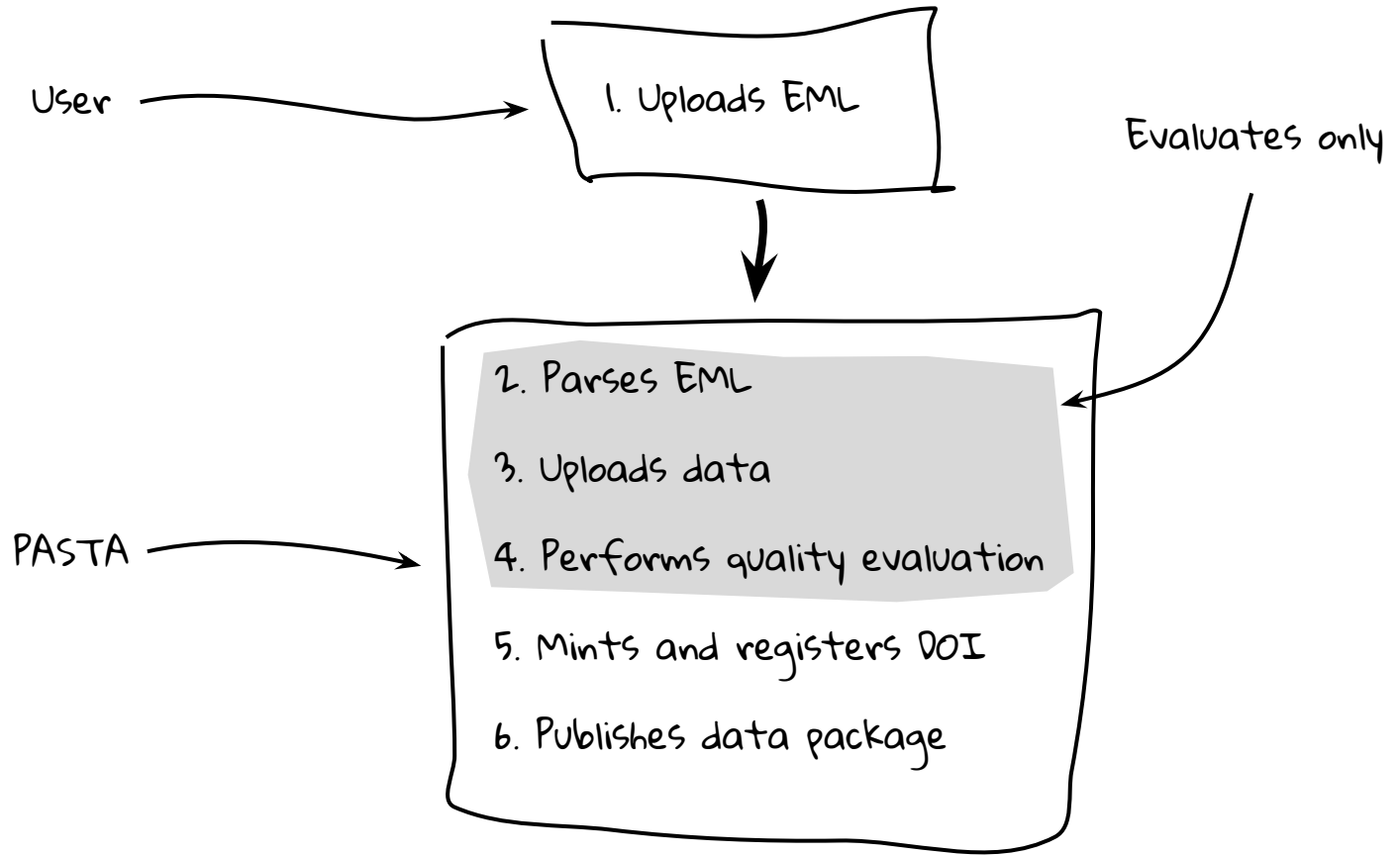


# METADATA-DRIVEN...

- User provides the EML document, PASTA does the rest
  - EML describes access control of both the full data package or individual data objects
  - EML describes who was involved in the research (individual, organization, role)
  - EML describes where the research occurred
  - EML describes when the research occurred
  - EML describes how the research was conducted
  - EML describes what was researched --- the data (type, physical, and attributes)
    - type: tabular, spatial, database, other
    - physical: size, checksum, source location
    - attributes: detailed information (e.g., column information for tabular)







# WHAT IS DATA PACKAGE QUALITY EVALUATION?

# WHAT IS DATA PACKAGE QUALITY EVALUATION?

A series of quality checks for...

# WHAT IS DATA PACKAGE QUALITY EVALUATION?

A series of quality checks for...

- Metadata validation
  - Well formed and schema valid
  - Content validation (does content match best practices?)

# WHAT IS DATA PACKAGE QUALITY EVALUATION?

A series of quality checks for...

- Metadata validation
  - Well formed and schema valid
  - Content validation (does content match best practices?)
- Data validation
  - Accessible (can data be downloaded?)

# WHAT IS DATA PACKAGE QUALITY EVALUATION?

A series of quality checks for...

- Metadata validation
  - Well formed and schema valid
  - Content validation (does content match best practices?)
- Data validation
  - Accessible (can data be downloaded?)
- Congruence validation
  - Metadata description of data matches physical structure of data (e.g., correct number of columns, rows, datatype, delimiters)

# con·gru·ence

/kən'gruːəns/

*noun*

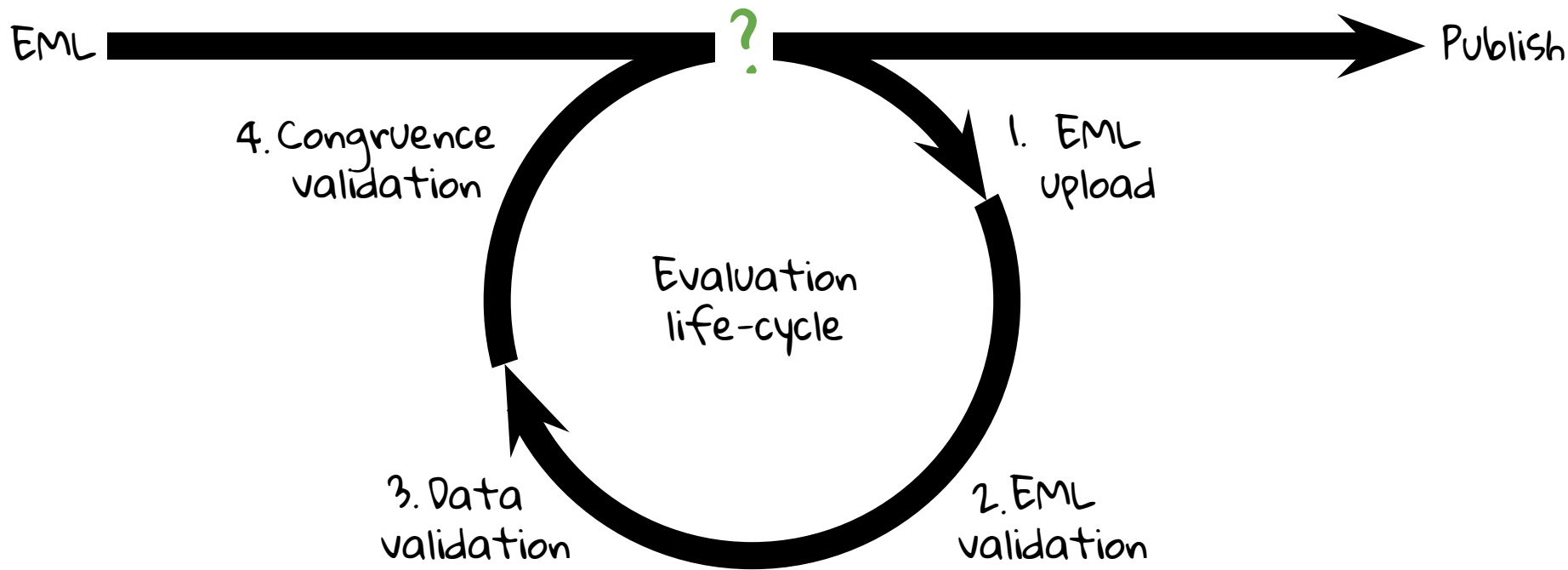
agreement or harmony; compatibility.

"the results show quite good **congruence with** recent studies"

*synonyms:* compatibility, consistency, conformity, match, balance, consonance, congruity; [More](#)



# THE QUALITY EVALUATION LIFE-CYCLE



# QUALITY EVALUATION REPORT

**Valid** - quality check meets criteria

**Warn** - quality check does not meet criteria, but does not fail upload

**Error** - quality check does not meet criteria, results in failed upload

**Info** - quality check only provides information

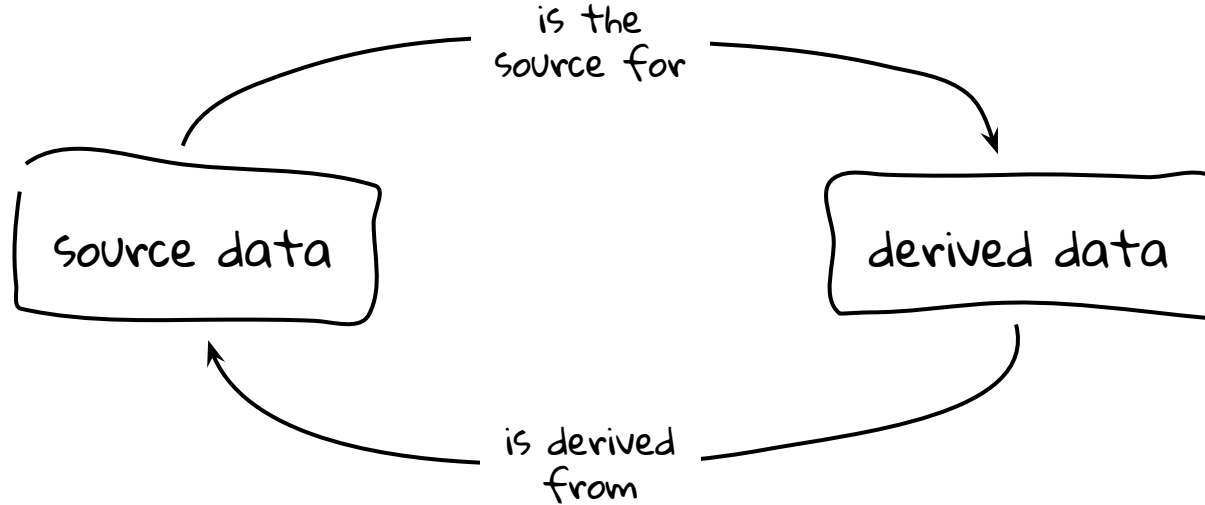
PackageId: knb-lter-nwk.1424.50

Report Date/Time: 2017-11-22T20:38:36

## Dataset Report

#	Identifier	Status	Quality Check	Name	Description	Expected	Found
1	packageIdPattern	valid	Type: metadata System: lter On Failure: error	packageId pattern matches "scope.identifier.revision"	Check against LTER requirements for scope.identifier.revision	'scope.n.m', where 'n' and 'm' are integers and 'scope' is one of an allowed set of values	knb-lter-nwk.1424.50
2	emlVersion	valid	Type: metadata System: lter On Failure: error	EML version 2.1.0 or greater	Check the EML version	2.1.0	eml://ecoinformatics.org/eml-2.1.1/namespaces
6	keywordPresent	warn	Type: metadata System: lter On Failure: warn	keyword element is present	Checks to see if at least one keyword is present	Presence of one or more keyword elements	0 'keyword' element(s) found
7	methodsElementPresent	valid	Type: metadata System: lter On Failure: warn	A 'methods' element is present	All datasets should contain a 'methods' element, at a minimum a link to a separate methods doc.	presence of 'methods' at one or more xpaths.	2 'methods' element(s) found
8	coveragePresent	warn	Type: metadata System: lter On Failure: warn	coverage element is present	At least one coverage element should be present in a dataset.	At least one of geographicCoverage, taxonomicCoverage, or temporalCoverage is present in the EML.	0 'coverage' element(s) found
9	geographicCoveragePresent	info	Type: metadata System: lter	geographicCoverage is present	Check that geographicCoverage exists in EML at the dataset level.	geographicCoverage at least at the dataset level.	0 'geographicCoverage' element(s) found
12	onlineURLs	valid	Type: congruency System: knb On Failure: error	Online URLs exist	Online URLs return something		
13	integrityChecksum	error	Type: congruency System: lter On Failure: error	Compare the metadata checksum for an entity to the checksum of the downloaded entity	Two possible responses: valid if checksums match; error if checksums do not match.	915a52bd06ef5730ca5ef33dd359380a59c86ef5	815a52bd06ef5730ca5ef33dd359380a59c86ef4

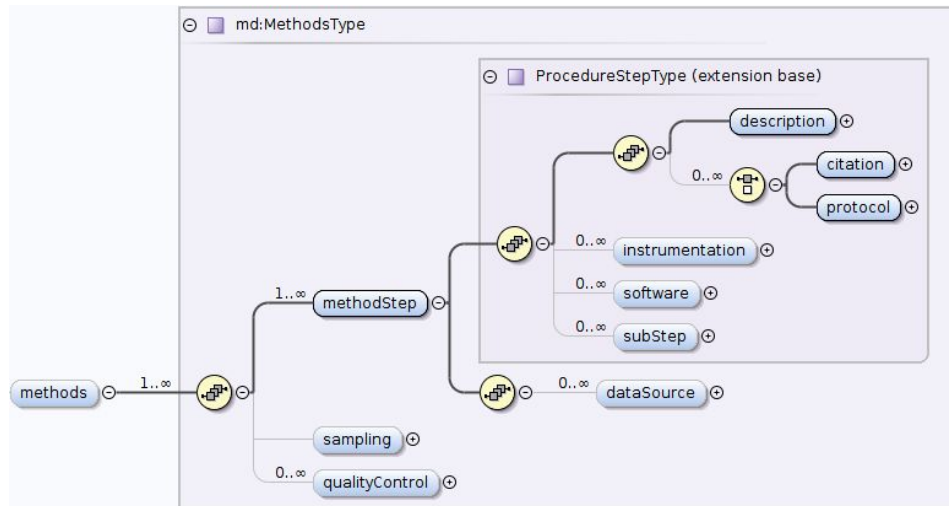
# PROVENANCE TRACKING THROUGH METADATA




# PROVENANCE TRACKING THROUGH METADATA

## EML

- Description method step
- Data Source
  - Title
  - Creator
  - Distribution
    - Online description
    - Online URL
  - Contact information






N: 44.4077808 S: 44.0384061 E: -68.1933176 W: -68.61187

**Package ID:** edi.20.1

**Resources:** Metadata  
Report  
Data  
1. ME\_ANP\_post\_2006.csv (38.5K)

is the  
source  
for



N: 49.3477 S: 34.61761 E: -66.99892 W: -97.90363

**Package ID:** edi.101.1  
next revision  
all revisions

**Resources:** Metadata  
Report  
Data  
1. LAGOSNE\_epinutr10871.csv (100M)  
2. LAGOSNE\_lakeslimno10871.csv (3.8M)  
3. LAGOSNE\_secchi10871.csv (97.1M)  
4. LAGOSNE\_sourceprogram10871.csv (22.4M)

**PASTA Identifier:** <https://pasta.lternet.edu/package/eml/edi/20/1>

**Provenance:** This data package is a source for the following data packages:

1. LAGOS-NE-LIMNO v1.087.1: A module for LAGOS-NE, a multi-scaled geospatial and temporal database of lake ecological context and water quality for thousands of U.S. Lakes: 1925-2013
2. LAGOS-NE-LIMNO v1.087.1: A module for LAGOS-NE, a multi-scaled geospatial and temporal database of lake ecological context and water quality for thousands of U.S. Lakes: 1925-2013

Generate provenance metadata for use within your derived data package

**Journal Citations:** [Add Journal Citation](#)

is  
derived  
from

**PASTA Identifier:** <https://pasta.lternet.edu/package/eml/edi/101/1>

**Provenance:** This data package is derived from the following sources:

1. Acadia National Park, U.S. National Park Service using Lakes and Stream Monitoring Protocol for National Parks in the Northeast Temperate Network, Version 1.1 (2006-2011)
2. Adirondack Lakes Long Term Monitoring (NY), 1992-2011
3. Adirondack Lakes Survey (NY): summer surface chemistry data, 1984-1987
4. Biocomplexity at North Temperate Lakes LTER; Coordinated Field Studies: Chemical Limnology 2001 - 2004
5. Biocomplexity at North Temperate Lakes LTER; Coordinated Field Studies: Color 2001 - 2004
6. Biocomplexity at North Temperate Lakes LTER; Coordinated Field Studies: Lakes 2001 - 2004

# STRONG VERSIONING THROUGH IDENTIFIERS

edi.10.1

time



# STRONG VERSIONING THROUGH IDENTIFIERS

knb-lter-nin.2.5

edi.10.1

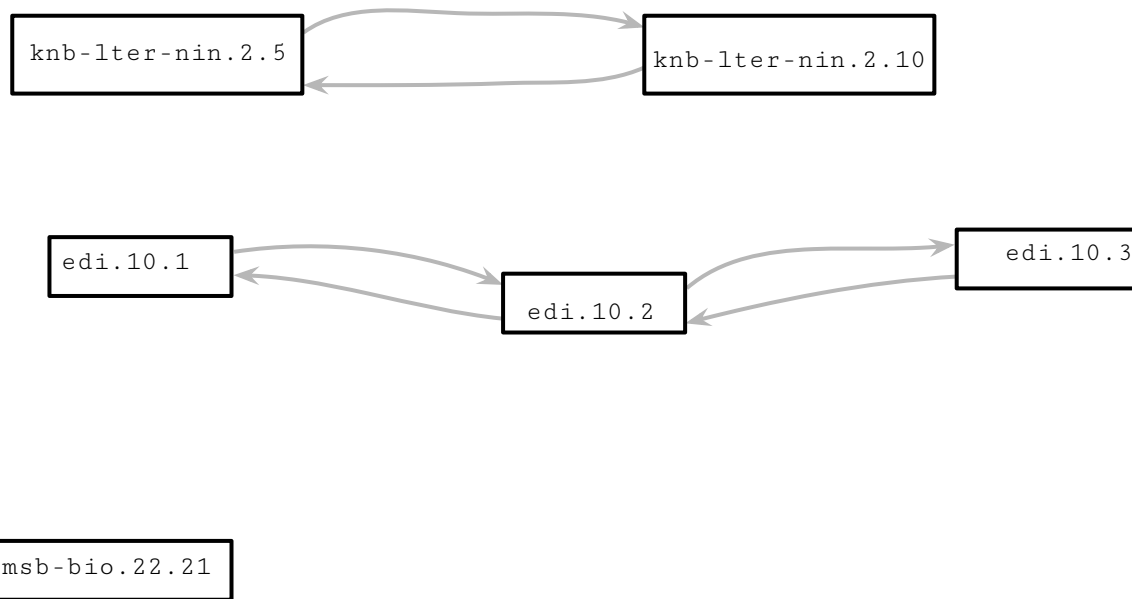
edi.10.2



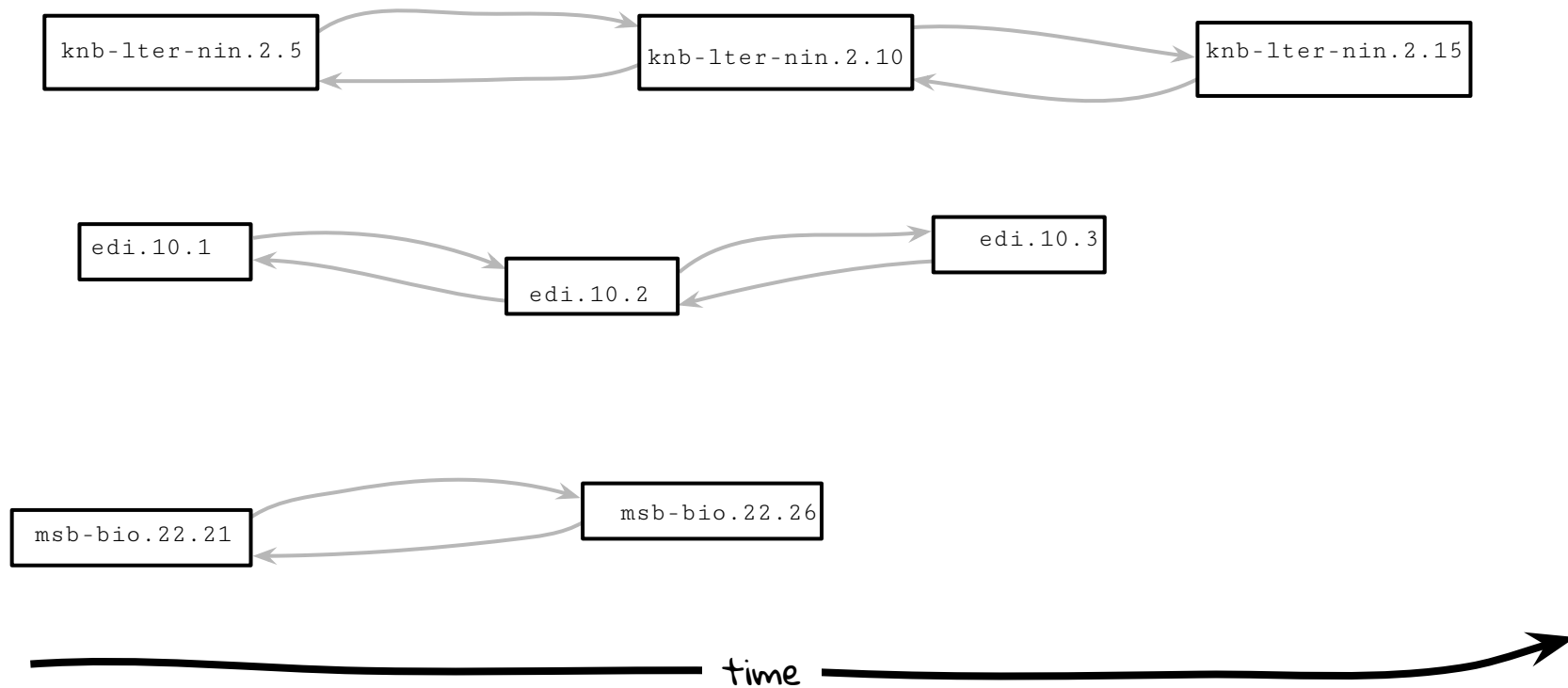
time



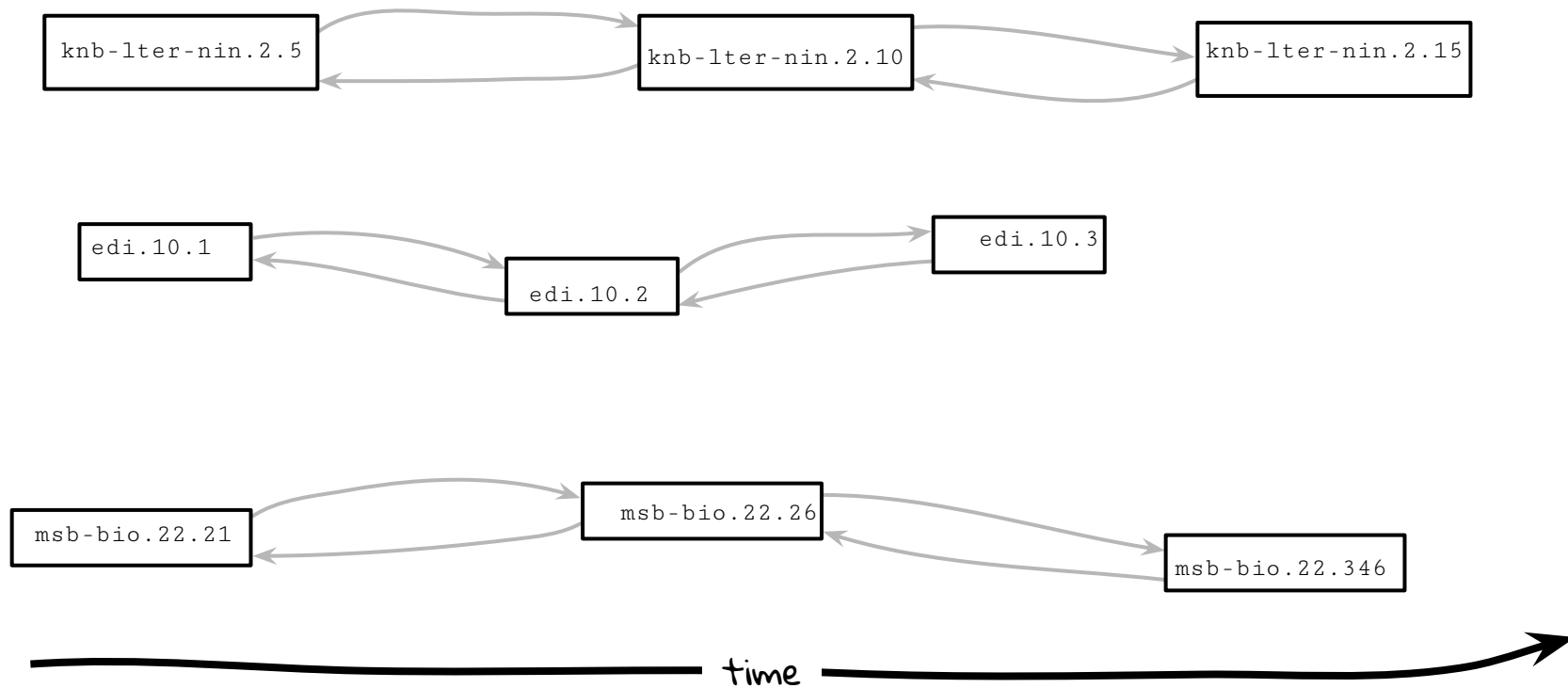
# STRONG VERSIONING THROUGH IDENTIFIERS



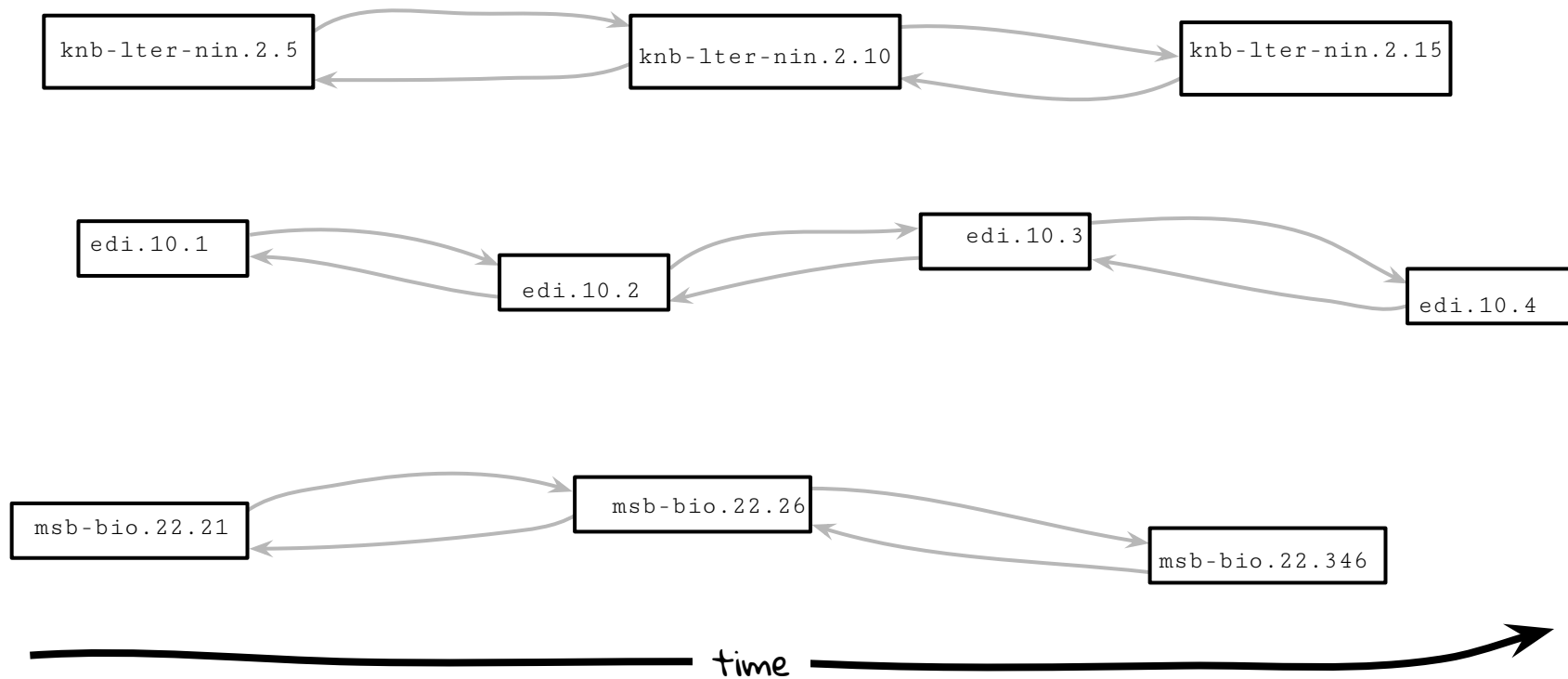
# STRONG VERSIONING THROUGH IDENTIFIERS



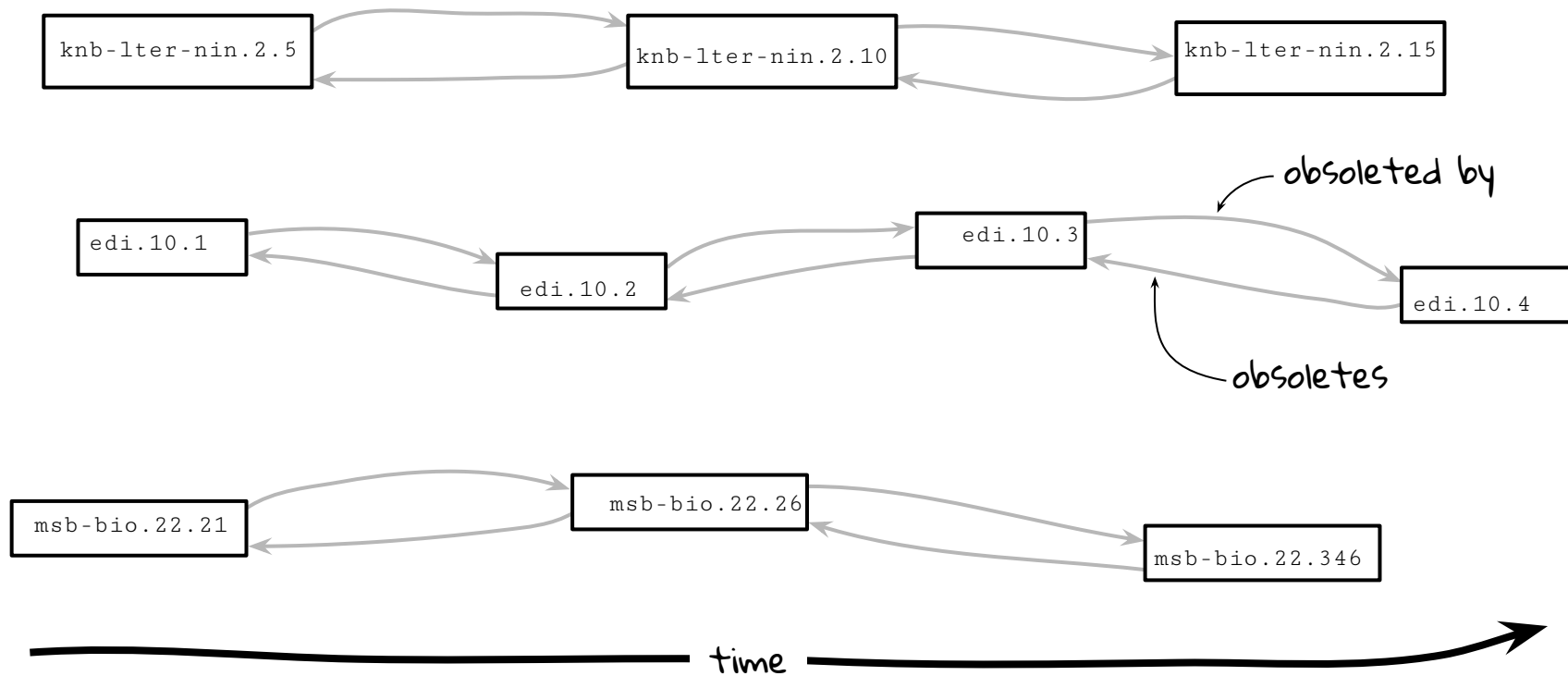
# STRONG VERSIONING THROUGH IDENTIFIERS



# STRONG VERSIONING THROUGH IDENTIFIERS



# STRONG VERSIONING THROUGH IDENTIFIERS



# STRONG VERSIONING THROUGH IDENTIFIERS

## Browse Data by Package Identifier

Select a data package *scope.identifier.revision* value to view a specific revision of the data package lineage.

1. edi.99.3  
2. edi.99.4  
3. edi.99.5

## Data Package Summary [View Full Metadata](#)

**Title:** LAGOS-NE-GEO v1.05: A module for LAGOS-NE, a multi-scaled geospatial and temporal database of lake ecological context and water quality for thousands of U.S. Lakes: 1925-2013

**Creators:** Soranno, Patricia A; Michigan State University  
Cheruvelli, Kendra S; Michigan State University

**Publication Date:** 2017-05-03

**Citation:** Soranno P., K. Cheruvelli. 2017. LAGOS-NE-GEO v1.05: A module for LAGOS-NE, a multi-scaled geospatial and temporal database of lake ecological context and water quality for thousands of U.S. Lakes: 1925-2013. Environmental Data Initiative. <http://dx.doi.org/10.6073/pasta/83350a5b1cae9def7d4465a82d12dd12>. Dataset accessed 1/21/2018.

**Abstract:** This data package, LAGOS-NE-GEO v1.05, is 1 of 5 data packages associated with the LAGOS-NE database-- the LAKE multi-scaled GeoSpatial and temporal database. Three of the data packages each contain different types of da...  
[Show more >](#)

**Spatial Coverage:**



N: 49.3477 S: 34.61761 E: -66.99892 W: -97.90363

**Package ID:** edi.99.4  
[previous revision](#)  
[next revision](#)  
[all revisions](#)

**Resources:** Metadata  
Report  
Data  
1. LAGOSNE\_buffer100m\_105.csv (1.2M)  
2. LAGOSNE\_buffer100m\_lulc105.csv (26.6M)  
3. LAGOSNE\_buffer500m\_105.csv (1.3M)  
4. LAGOSNE\_buffer500m\_conn105.csv (3.1M)

# DIGITAL OBJECT IDENTIFIER (DOI) ASSIGNMENT

Internal: `edi.10.1`

External: `https://pasta.lternet.edu/package/eml/edi/10/1`

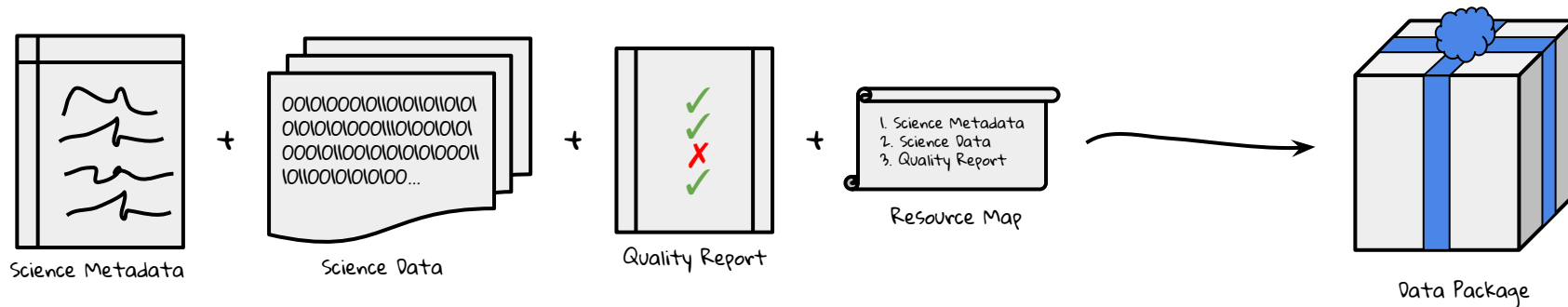
DOI: `doi:10.6073/pasta/8996dd613fd0227e8d041822bbd3071d`

# DIGITAL OBJECT IDENTIFIER (DOI) ASSIGNMENT

Internal: `edi.10.1`

External: `https://pasta.lternet.edu/package/eml/edi/10/1`

DOI: `doi:10.6073/pasta/8996dd613fd0227e8d041822bbd3071d`





# DIGITAL OBJECT IDENTIFIER (DOI) ASSIGNMENT

**Internal:** `edi.10.1`

**External:** `https://pasta.lternet.edu/package/eml/edi/10/1`

**DOI:** `doi:10.6073/pasta/8996dd613fd0227e8d041822bbd3071d`

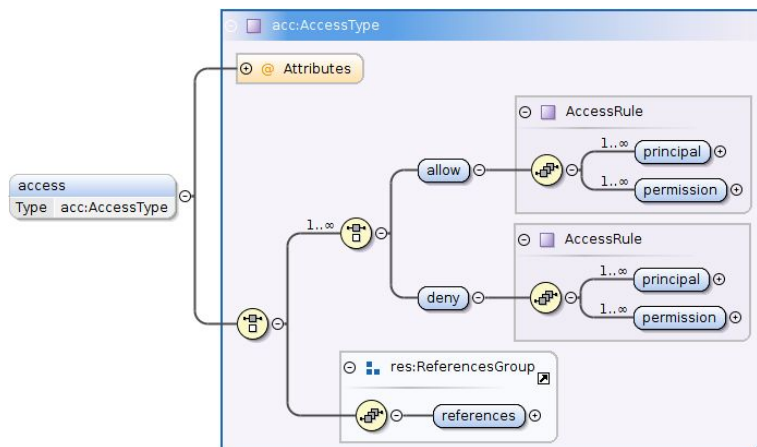


# DIGITAL OBJECT IDENTIFIER (DOI) ASSIGNMENT

EZID California Digital Library, University of California: 2013 - 2016

DataCite DOI allocating agent: 2017...

# DATA PACKAGE ACCESS CONTROL



read, write, or all (change-permission)

```
<access authSystem="https://pasta.edirepository.org/authentication"
  order="allowFirst" scope="document" system="https://pasta.edirepository.org">
  <allow>
    <principal>uid=EDI,o=LTER,dc=ecoinformatics,dc=org</principal>
    <permission>all</permission>
  </allow>
  <allow>
    <principal>uid=SBC,o=LTER,dc=ecoinformatics,dc=org</principal>
    <permission>all</permission>
  </allow>
  <allow>
    <principal>uid=lkui,dc=ecoinformatics,dc=org</principal>
    <permission>all</permission>
  </allow>
  <deny>
    <principal>public</principal>
    <permission>read</permission>
  </deny>
</access>
```





# Environmental Data Initiative

- Create . Package . Archive . Discover . Reuse -

[WELCOME](#) [ABOUT](#) [DATA](#) [NEWS](#) [EVENTS](#) [RESOURCES](#) [EDI-BLOGS](#)

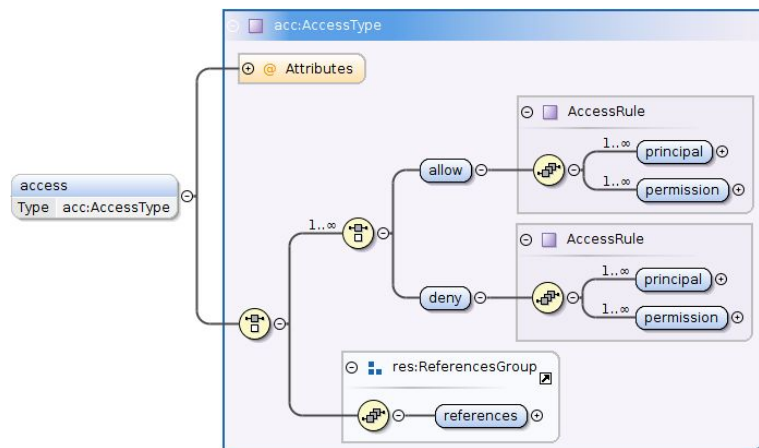
## EDI Data Policy

The Environmental Data Initiative (EDI) strives to make environmental research data open and accessible to the general public without undue restrictions or barriers. Although EDI strongly recommends making all data publicly available, we recognize that some data may require limited access while it is under review during manuscript preparation. In these cases, the EDI Data Repository supports access control to data when justified by the data provider, thereby limiting exposure of the data resource to only users with appropriate permission\*. Such access control must be identified in the data package metadata and requires a

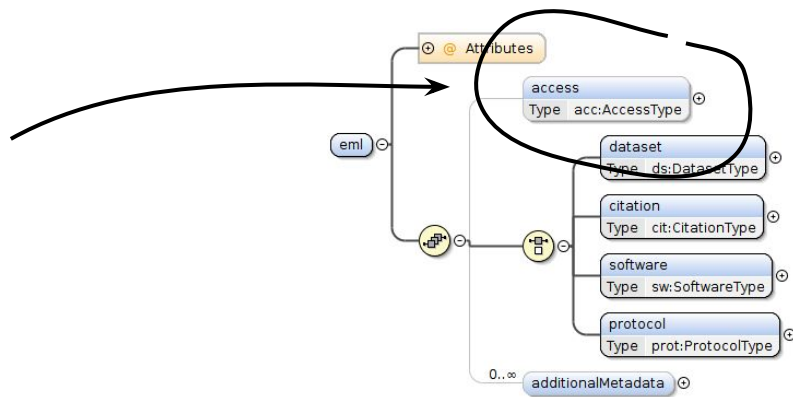
EDI makes every effort to ensure that all data are curated with intellectual rights defined by the data provider as found in the data package metadata. Although EDI advocates for open and unfettered access to data packages without use restrictions, we do not forbid data providers from declaring more restrictive licensing agreements for use of their data packages. Data providers should include a statement of Intellectual Rights in the metadata of their submissions. If they do not, EDI reserves the right to add a default declaration of intellectual rights to the data package metadata. The default declaration of intellectual rights used by EDI is based on the Creative Commons CC0 “No Rights Reserved” waiver. See below for the full default statement:

“ This data package is released to the “public domain” under Creative Commons CC0 1.0 “No Rights Reserved” (see: <https://creativecommons.org/publicdomain/zero/1.0/>). It is considered professional etiquette to provide attribution of the original work if this data package is shared in whole or by individual components. A generic citation is provided for this data package on the website <https://portal.edirepository.org> (herein “website”) in the summary metadata page. Communication (and collaboration) with the creators of this data package is recommended to prevent duplicate research or publication. This data package (and its components) is made available “as is” and with no warranty of accuracy or fitness for use. The creators of this data package and the website shall not be liable for any damages resulting from misinterpretation or misuse of the data package or its components. Periodic updates of this data package may be available from the website. Thank you.

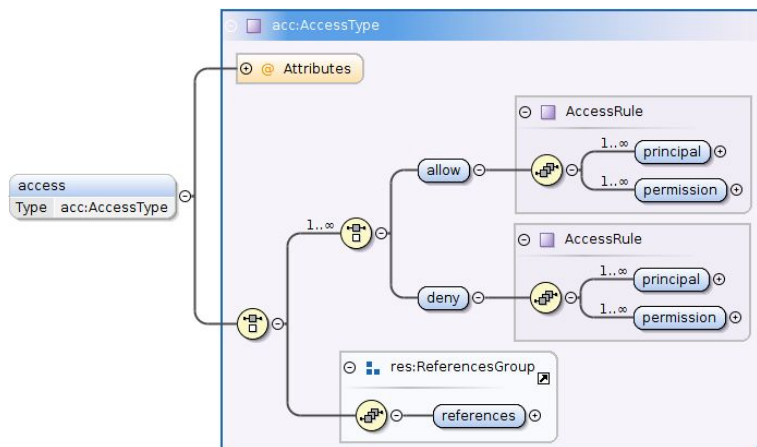
# DATA PACKAGE ACCESS CONTROL



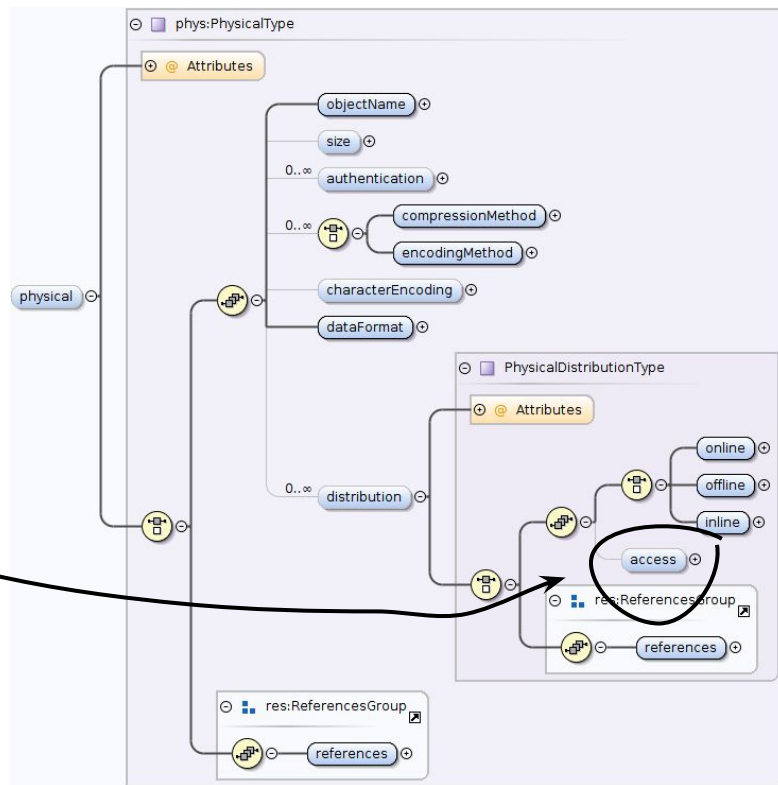
data package level



# DATA PACKAGE ACCESS CONTROL



data object level



# DATA PACKAGE ACCESS CONTROL

An error has occurred in the Data Portal:

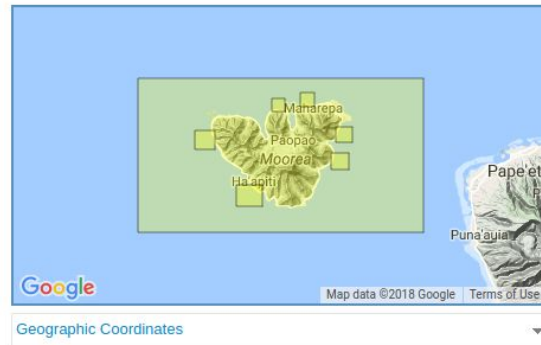
`edu.lternet.pasta.common.security.access.UnauthorizedException:`  
PASTA returned status code 401: User public does not have permission to read this metadata document: <https://pasta-d.lternet.edu/package/metadata/eml/edi/121/1> Logging into the LTER Data Portal *may* let you read this resource.

For further assistance, please contact the Environmental Data Initiative. Please copy the error message shown above into your email message, along with any other information that might help us to assist you more promptly.

data package level

data object level

Spatial Coverage:



Package ID: knb-lter-mcr.1035.8

Resources: Metadata  
Report  
Data

1. Data object [\[more info\]](#)
2. Data object [\[more info\]](#)
3. Data object [\[more info\]](#)
4. Data object [\[more info\]](#)
5. Data object [\[more info\]](#)
6. Data object [\[more info\]](#)
7. Data object [\[more info\]](#)

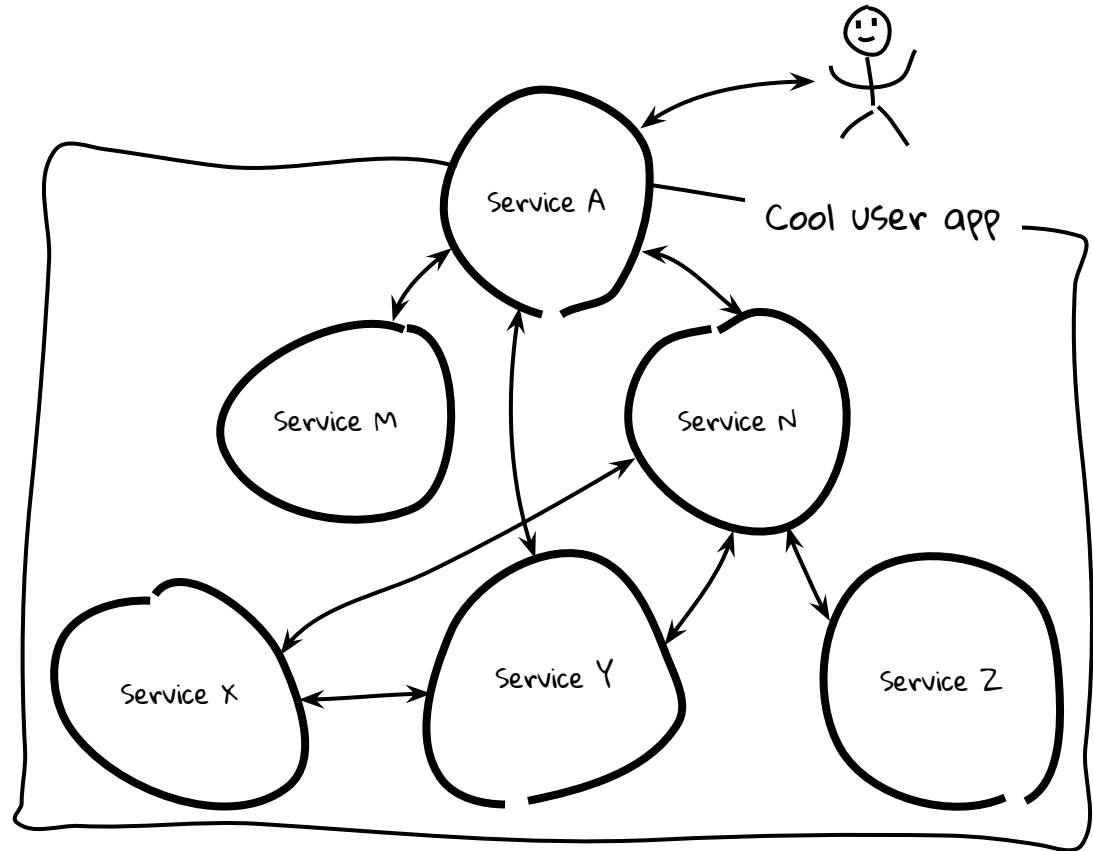
Download Zip Archive

You may need to log in before you can access this data object.



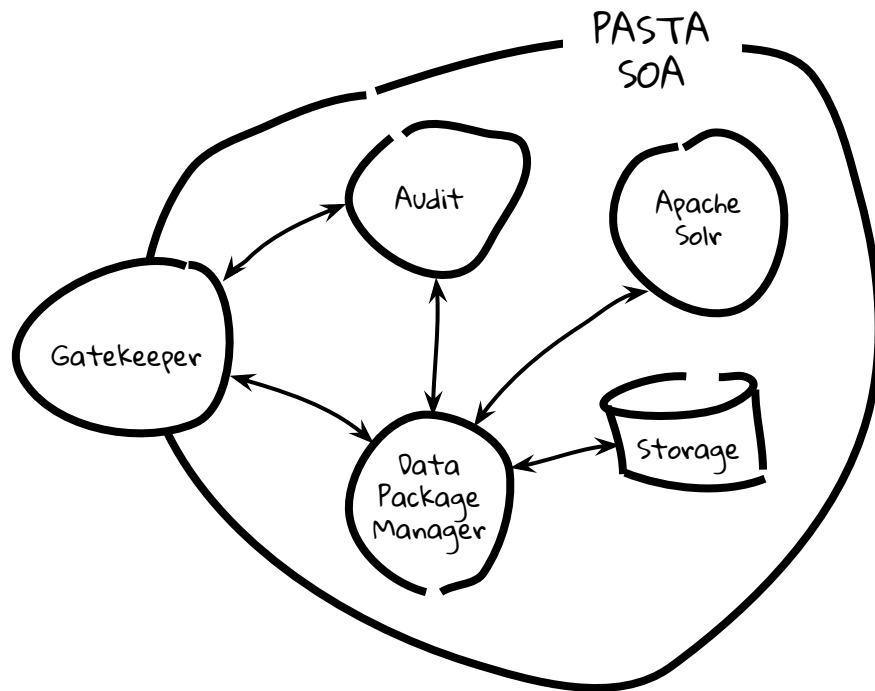
# WHAT IS A SERVICE ORIENTED ARCHITECTURE?

"a collection of interoperable, network-accessible services that, when working together, form a large, singular application"

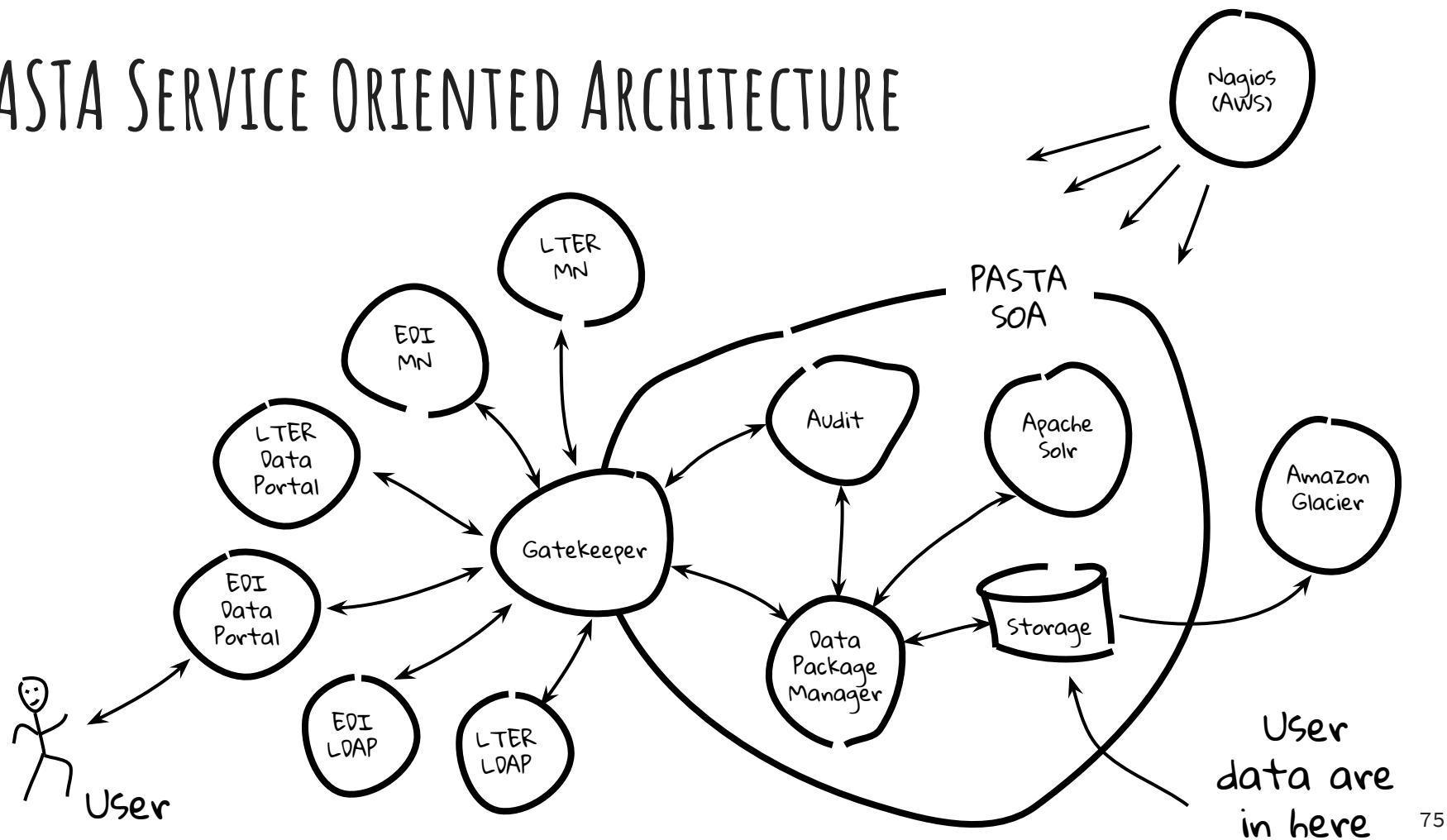


# PASTA SERVICE ORIENTED ARCHITECTURE

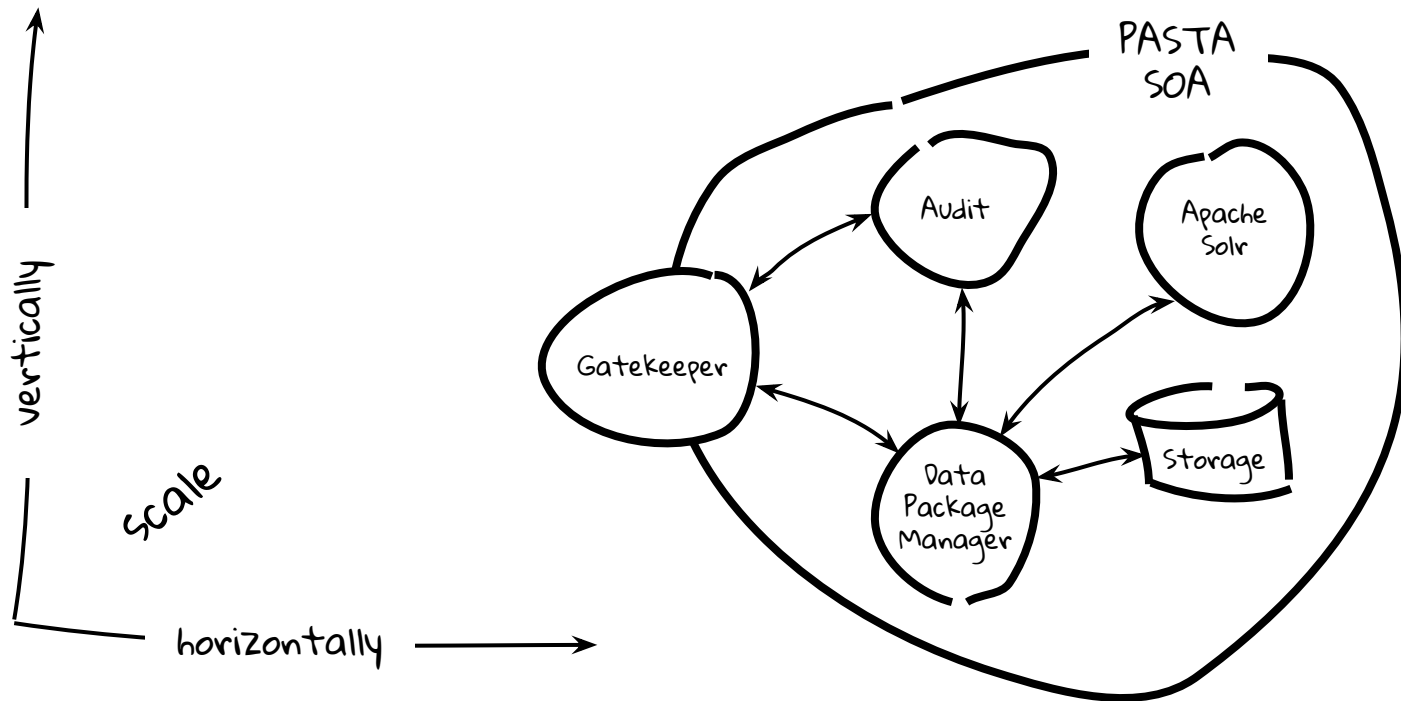
- Gatekeeper
  - reverse-proxy
  - user identity management
- Data Package Manager
  - packaging
  - quality evaluation
  - provenance
  - versioning
  - DOI creation
  - access control
- Audit
  - logging service
- Apache Solr
  - metadata indexing
  - search



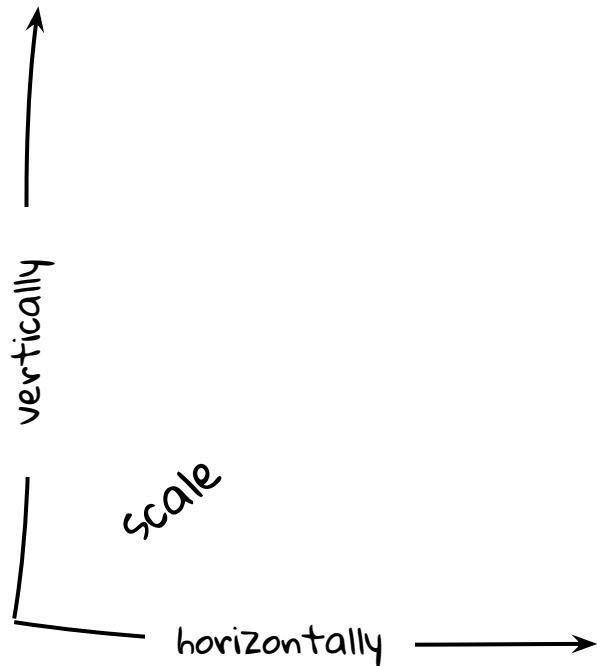
# PASTA SERVICE ORIENTED ARCHITECTURE



# PASTA SERVICE ORIENTED ARCHITECTURE



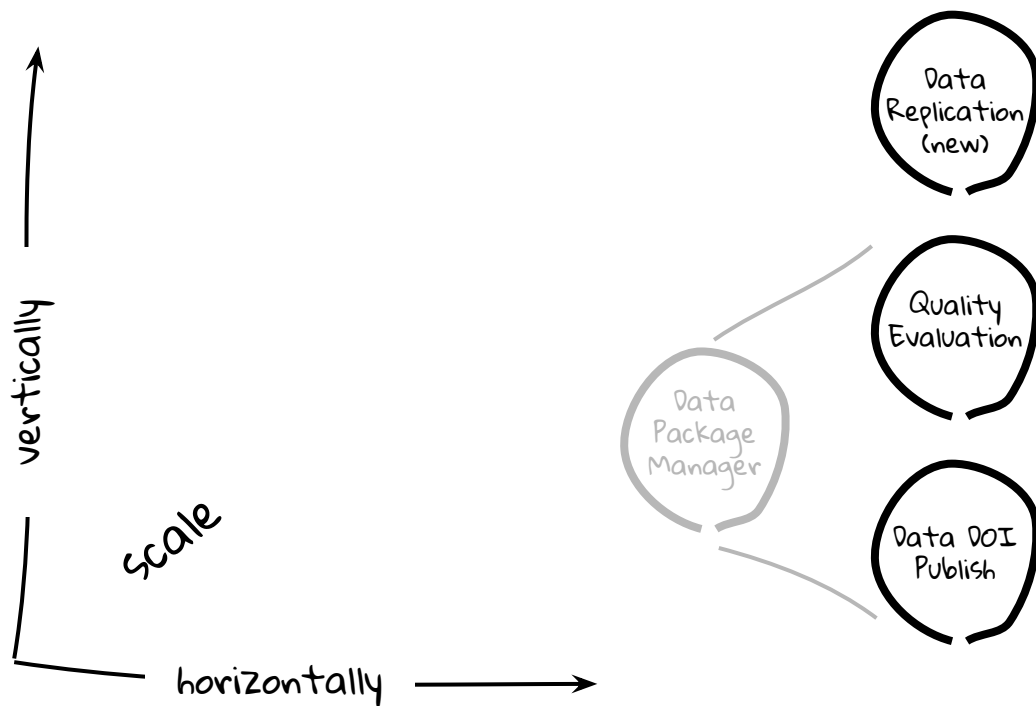
# PASTA SERVICE ORIENTED ARCHITECTURE



Scaling horizontally to  
improve performance and  
availability

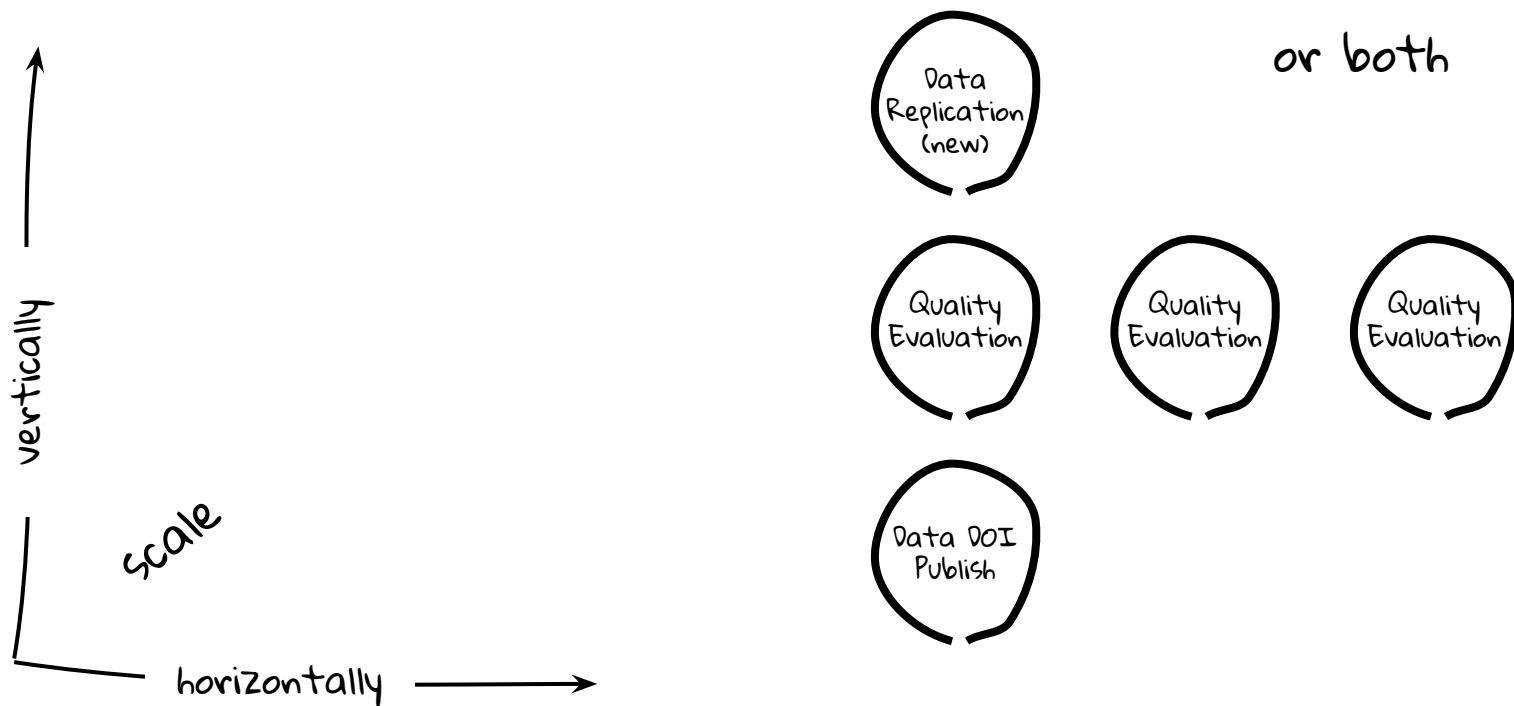


# PASTA SERVICE ORIENTED ARCHITECTURE



Scaling  
vertically to  
add new  
features or  
create  
microservices

# PASTA SERVICE ORIENTED ARCHITECTURE



# PASTA REST WEB SERVICE API

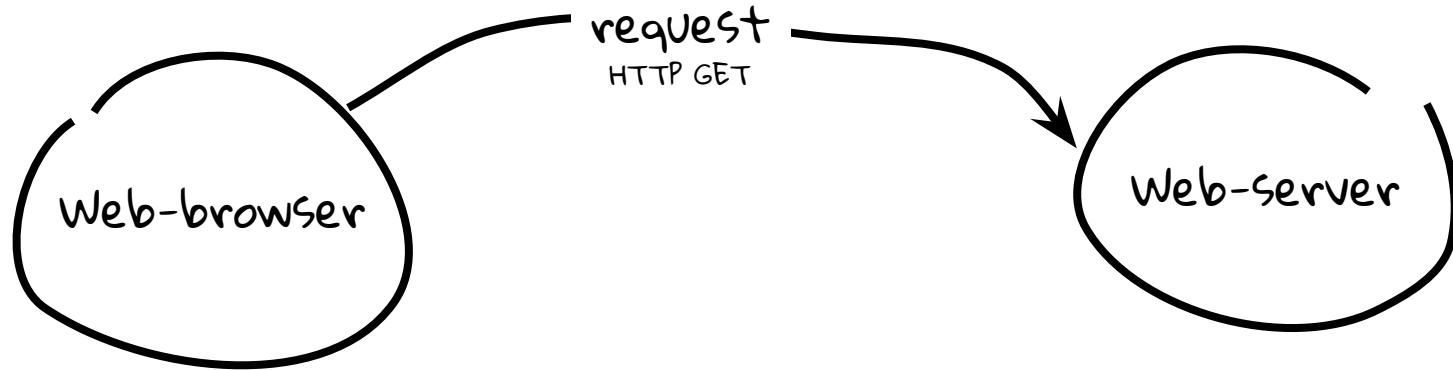
REST = Representational State Transfer

Web services allow "platform-independent" computing between servers that are connected to the Internet (WWW).

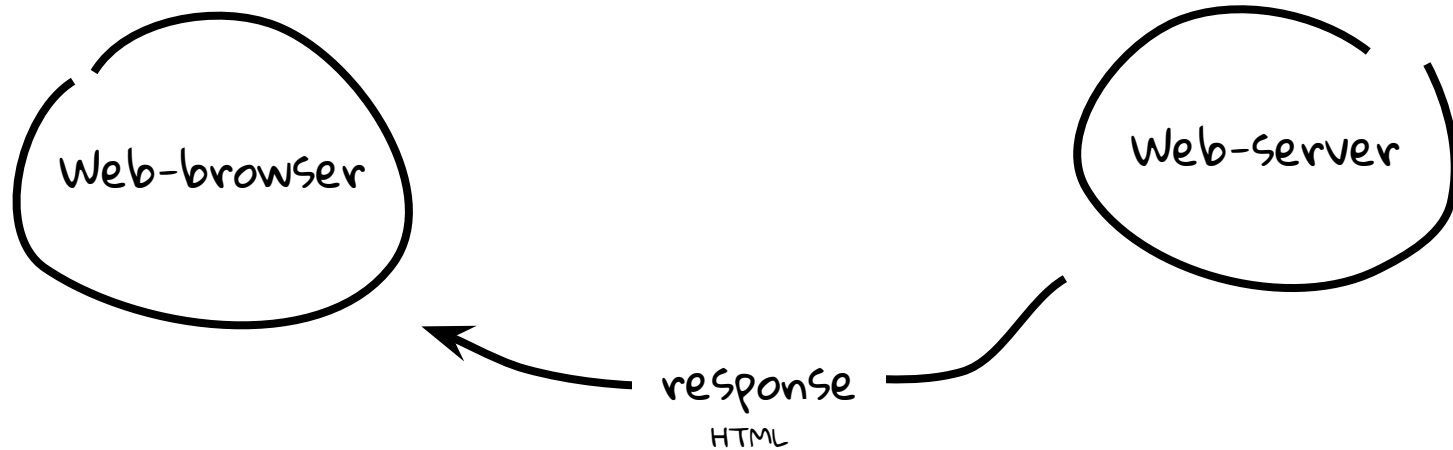
REST is a design pattern for application-to-application processing of web resources based on the HTTP protocol.



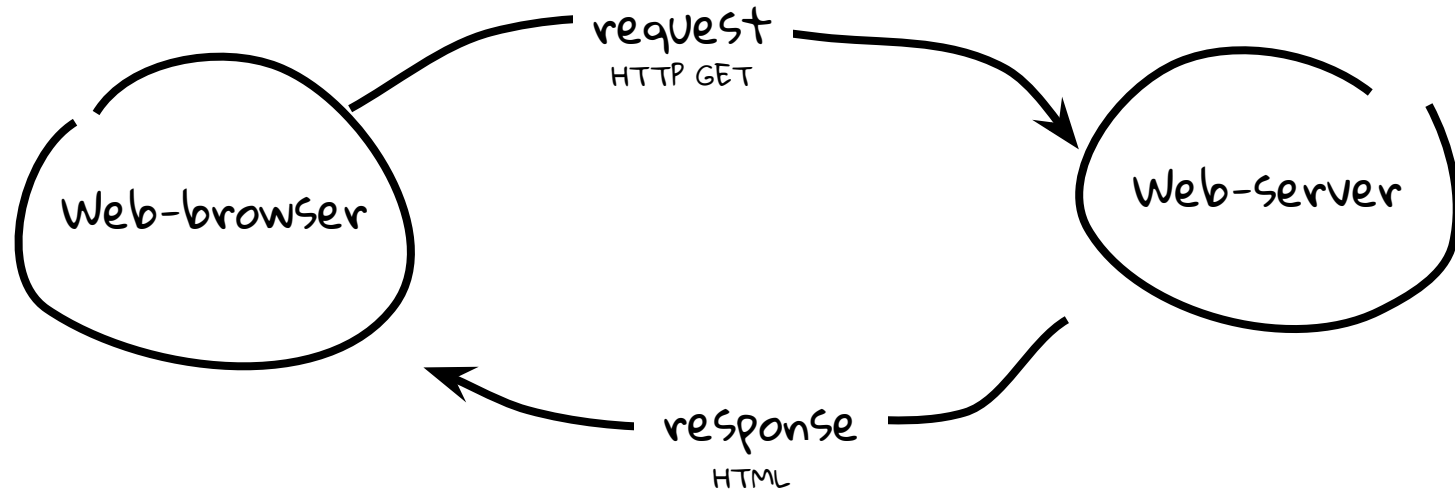
# PASTA REST WEB SERVICE API



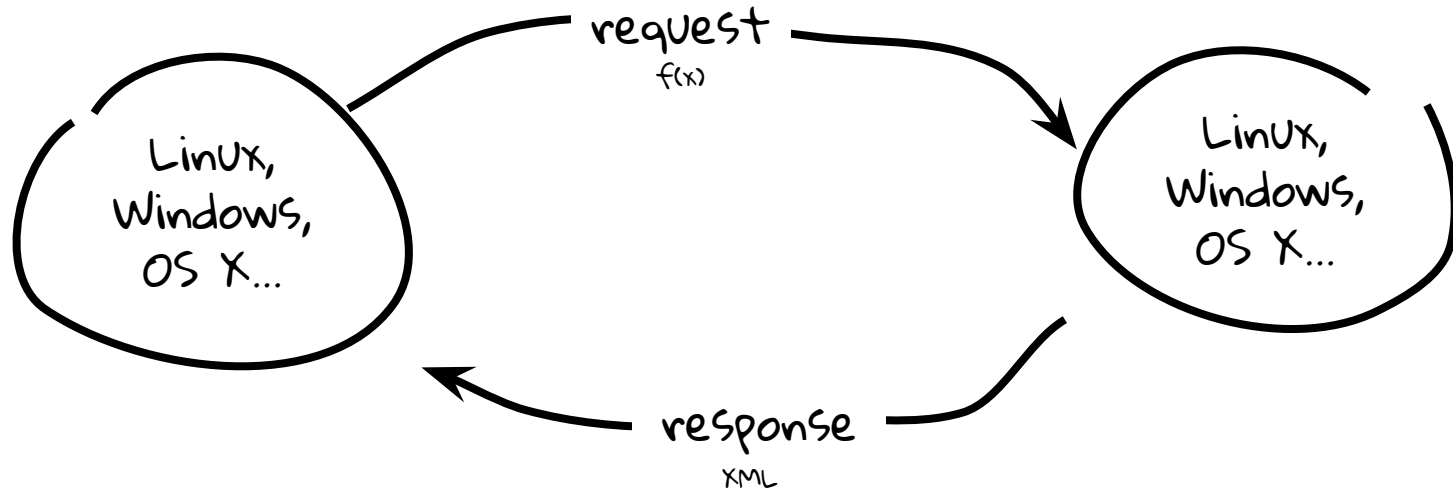
# PASTA REST WEB SERVICE API



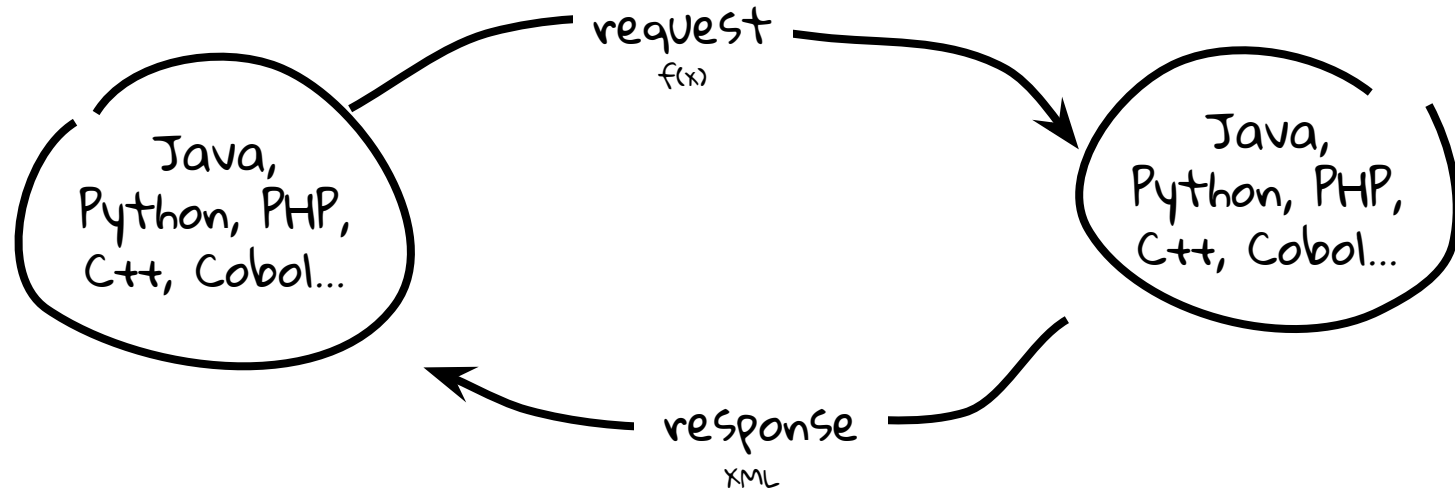
# PASTA REST WEB SERVICE API

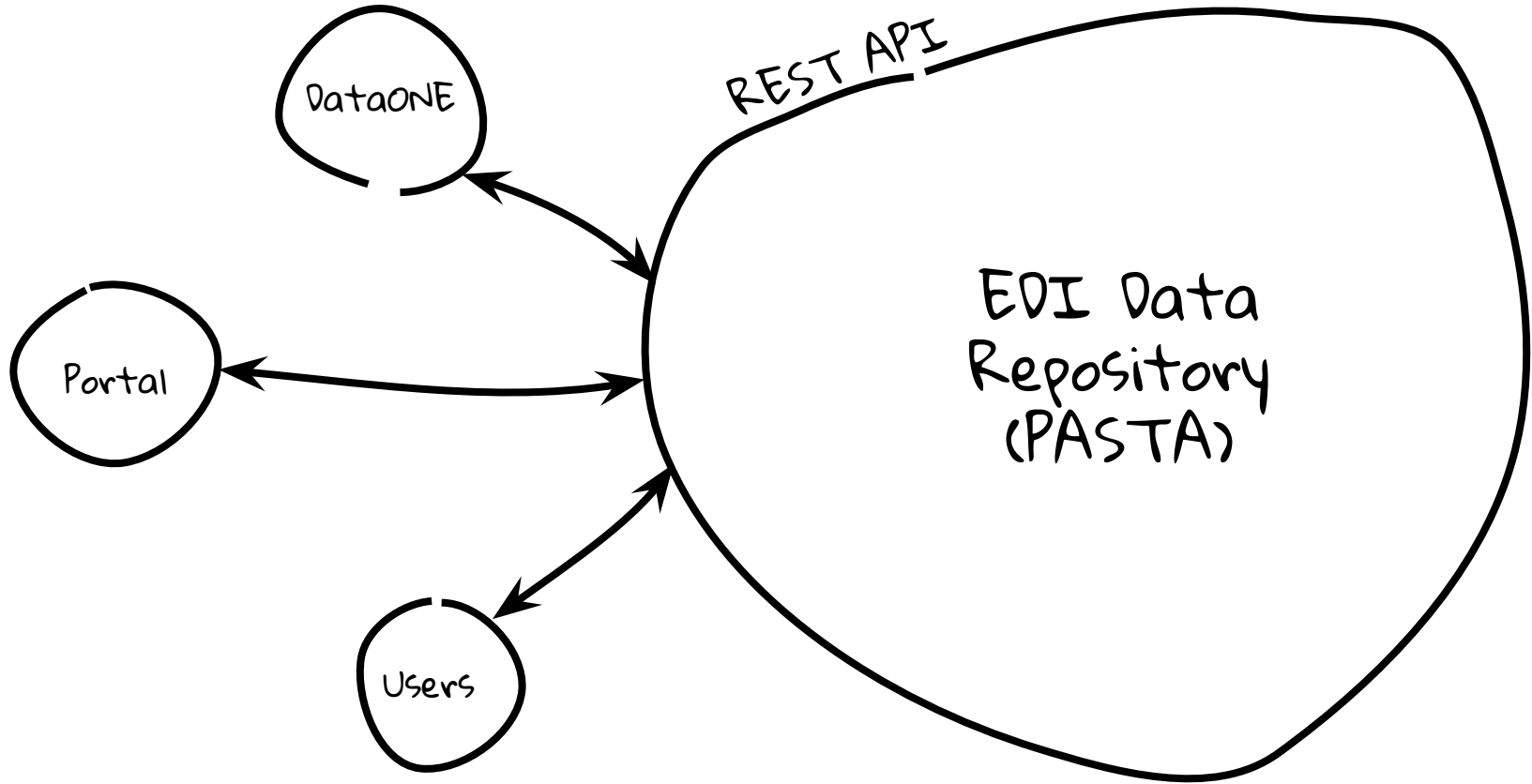


# PASTA REST WEB SERVICE API



# PASTA REST WEB SERVICE API





# PAST REST WEB SERVICE API

"CRUD"

- Create (a resource): HTTP POST
- Read (a resource): HTTP GET
- Update (a resource): HTTP PUT
- Delete (a resource): HTTP DELETE

# PAST REST WEB SERVICE API

- List functions (read)
  - List scopes for data packages in PASTA
  - List scope.identifiers for data packages in PASTA
  - List scope.identifier.revisions for data packages in PASTA
- Get resource functions (read)
  - Get metadata
  - Get data
  - Get quality report
- Post/Put/Delete resource functions (create, update, delete)
  - Post (aka upload) data packages
  - Put (aka revise) data packages
  - Post event subscriptions
  - Delete (aka archive) data packages



# PAST REST WEB SERVICE API

66 total REST web service methods  
calls available

<https://pasta.lternet.edu/audit/docs/api>

<https://pasta.lternet.edu/package/docs/api>

- List functions (read)
  - List scopes for data packages in PASTA
  - List scope.identifiers for data packages in PASTA
  - List scope.identifier.revisions for data packages in PASTA
- Get resource functions (read)
  - Get metadata
  - Get data
  - Get quality report
- Post/Put/Delete resource functions (create, update, delete)
  - Post (aka upload) data packages
  - Put (aka revise) data packages
  - Post event subscriptions
  - Delete (aka archive) data packages

# PAST REST WEB SERVICE API

```
pasta.lternet.edu/package/metadata/ x +
https://pasta.lternet.edu/package/metadata/eml/edi/19/1
This XML file does not appear to have any style information associated with it. The document tree is shown below.
<?xml version="1.0" encoding="UTF-8"?>
<eml:eml packageId="edi.19.1" system="https://pasta.edirepository.org"
  xsi:schemaLocation="eml://ecoinformatics.org/eml-2.1.1 http://mis.lternet.edu/schemas/EML/eml-2.1.1/eml.xsd">
  <access authSystem="https://pasta.edirepository.org/authentication" order="allowFirst" scope="document"
    system="https://pasta.edirepository.org">
    <allow>
      <principal>uid=csmith,o=LTER,dc=ecoinformatics,dc=org</principal>
      <permission>all</permission>
    </allow>
    <allow>
      <principal>public</principal>
      <permission>read</permission>
    </allow>
  </access>
  <dataset>
    <alternateIdentifier system="https://doi.org">doi:10.6073/pasta/7280dedf74ffa54706c894025e7ccd</alternateIdentifier>
  </title>
  Massachusetts Department of Conservation and Recreation (DCR): Quabbin Reservoir and Ware River watershed Secchi depth, 2005-2008
  </title>
  <creator>
    <individualName>
      <givenName>Paul</givenName>
      <surName>Reyes</surName>
    </individualName>
    <organizationName>
      Massachusetts Department of Conservation and Recreation Office of Watershed Management
    </organizationName>
    <positionName>Environmental Engineer</positionName>
  </creator>
  <address>
    <deliveryPoint>485 Ware Road</deliveryPoint>
    <city>Belchertown</city>
    <administrativeArea>MA</administrativeArea>
    <postalCode>01007</postalCode>
    <country>USA</country>
  </address>
  <phone phonetype="voice">413-323-6921, ext. 307</phone>
  <electronicMailAddress>paul.reyes@state.ma.us</electronicMailAddress>
  </creator>
  <pubDate>2017-04-06</pubDate>
  <abstract>
    <para>
      The Department of Conservation and Recreation, Division of Water Supply Protection is the state agency charged with the responsibility of managing Quabbin Reservoir and its surrounding natural resources in order to protect, preserve and enhance the environment of the Commonwealth and to assure the availability of pure water to future generations. As part of this effort, the Environmental Quality Section maintains a comprehensive water quality monitoring program to ensure that Quabbin Reservoir and its tributaries meet state water quality
    </para>
  </abstract>
  <keywordSet>
    <keyword>Massachusetts</keyword>
    <keyword>Quabbin Reservoir</keyword>
    <keyword>Secchi</keyword>
    <keywordSet>
    <keyword>
      aquatic ecosystems</keyword>
    </keywordSet>
  </keywordSet>
  </dataset>
</eml>
```

```
File Edit View Search Terminal Help
servilla@gaucha ~$ curl -s -X GET https://pasta.lternet.edu/package/metadata/eml/edi/19/1
<?xml version="1.0" encoding="UTF-8"?>
<eml:eml packageId="edi.19.1" system="https://pasta.edirepository.org"
  xmlns:eml="eml://ecoinformatics.org/eml-2.1.1"
  xmlns:stml="http://www.w3.org/2001/XMLSchema-instance"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema"
  xsi:schemaLocation="eml://ecoinformatics.org/eml-2.1.1 http://mis.lternet.edu/schemas/EML/eml-2.1.1/eml.xsd">
  <access authSystem="https://pasta.edirepository.org/authentication"
    order="allowFirst" scope="document" system="https://pasta.edirepository.org">
    <allow>
      <principal>uid=csmith,o=LTER,dc=ecoinformatics,dc=org</principal>
      <permission>all</permission>
    </allow>
    <allow>
      <principal>public</principal>
      <permission>read</permission>
    </allow>
  </access>
  <dataset>
    <alternateIdentifier system="https://doi.org">doi:10.6073/pasta/7280dedf74ffa54706c894025e7ccd</alternateIdentifier>
    <title>Massachusetts Department of Conservation and Recreation (DCR): Quabbin Reservoir and Ware River watershed Secchi depth, 2005-2008</title>
    <creator>
      <individualName>
        <givenName>Paul</givenName>
        <surName>Reyes</surName>
      </individualName>
      <organizationName>Massachusetts Department of Conservation and Recreation Office of Watershed Management</organizationName>
      <positionName>Environmental Engineer</positionName>
      <address>
        <deliveryPoint>485 Ware Road</deliveryPoint>
        <city>Belchertown</city>
        <administrativeArea>MA</administrativeArea>
        <postalCode>01007</postalCode>
        <country>USA</country>
      </address>
      <phone phonetype="voice">413-323-6921, ext. 307</phone>
      <electronicMailAddress>paul.reyes@state.ma.us</electronicMailAddress>
    </creator>
    <pubDate>2017-04-06</pubDate>
    <abstract>
      <para>The Department of Conservation and Recreation, Division of Water Supply Protection is the state agency charged with the responsibility of managing Quabbin Reservoir and its surrounding natural resources in order to protect, preserve and enhance the environment of the Commonwealth and to assure the availability of pure water to future generations. As part of this effort, the Environmental Quality Section maintains a comprehensive water quality monitoring program to ensure that Quabbin Reservoir and its tributaries meet state water quality standards. As part of this task, the Environmental Quality Section performs the necessary field work, interprets water quality data and prepares reports of findings. This annual summary is intended to meet the needs of the decision makers, the concerned public and others whose decisions must reflect water quality considerations</para>
    </abstract>
    <keywordSet>
      <keyword>Massachusetts</keyword>
      <keyword>Quabbin Reservoir</keyword>
      <keyword>Secchi</keyword>
    </keywordSet>
    <keywordSet>
      <keyword>
        aquatic ecosystems</keyword>
      </keywordSet>
    </keywordSet>
  </dataset>
</eml>
```

# THE FOUR STEPS TO EDI DATA HAPPINESS

1. Login using your user id
2. Reserve your data package identifier (to avoid namespace collision)
3. Evaluate your data package
4. Upload your data package

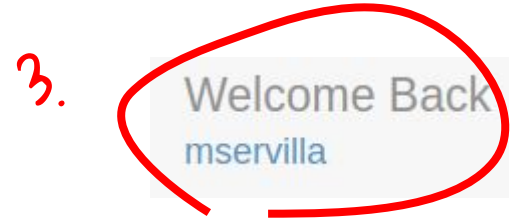
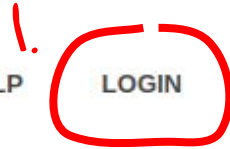
# LOGIN USING YOUR USER ID



DATA

TOOLS

HELP



## Login

2.

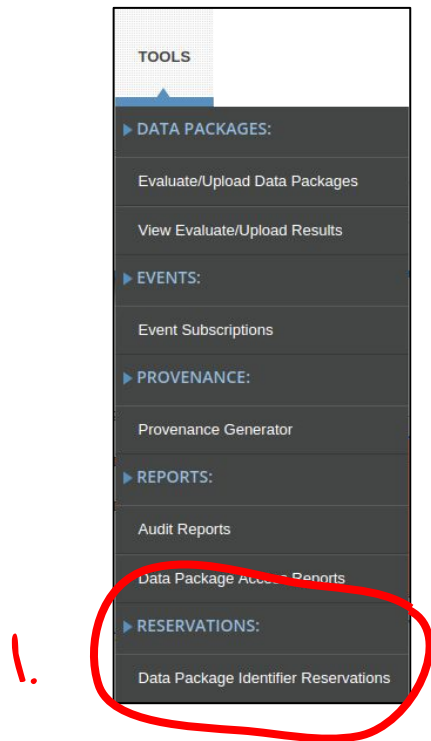
User Name: mservilla

Password: .....

Affiliation: EDI ▼

Login Clear

# RESERVE A DATA PACKAGE IDENTIFIER



## Data Package Identifier Reservations

### Reserve Identifiers

Reserve the next available identifier values for new data packages that you intend to **insert** into the EDI Repository (currently restricted to the **edi** scope):

2.

How many identifiers in the edi scope would you like to reserve?:

Reserve Next Available Identifier(s)

## Data Package Identifier Reservations

### Reserve Identifiers

Reserve the next available identifier values for new data packages that you intend to **insert** into the EDI Repository (currently restricted to the **edi** scope):

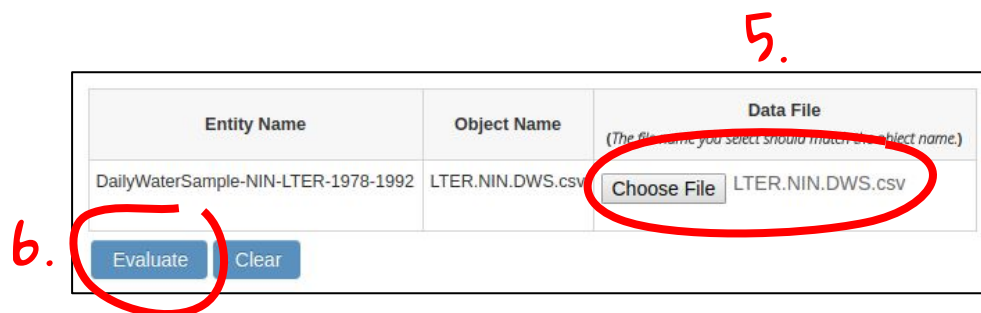
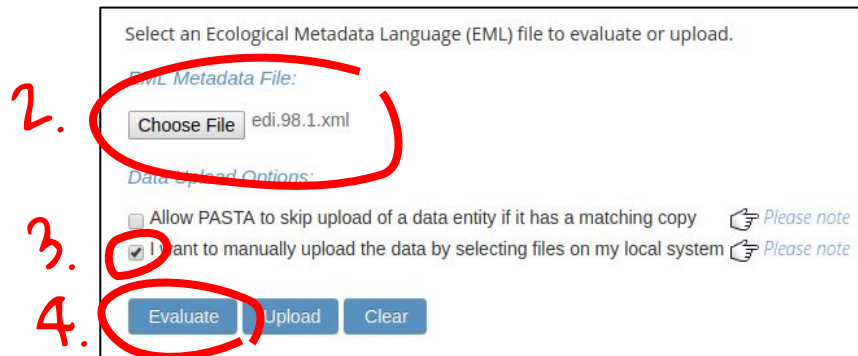
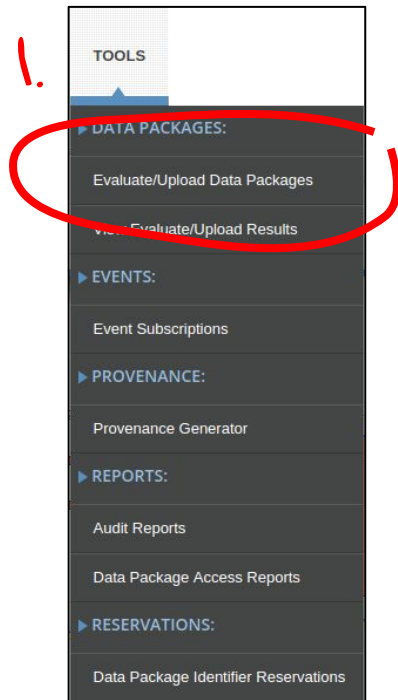
How many identifiers in the edi scope would you like to reserve?:

Reserve Next Available Identifier(s)

3.

The following identifier has been reserved for user uid=mservilla,o=EDI,dc=edirepository,dc=org: **edi.98**

# EVALUATE YOUR DATA PACKAGE



# EVALUATE YOUR DATA PACKAGE

## View Evaluate/Upload Results

Select the **Evaluate** or **Upload** results to view (results are stored for 180 days):

Wed Nov 15 13:03:35 MST 2017 (evaluate) ▼

Package Id	Was evaluated	Report	Total Quality Checks	Valid	Info	Warn	Error	System Message
edi.98.1	Yes	<a href="#">view</a>	35	25	7	1	2	

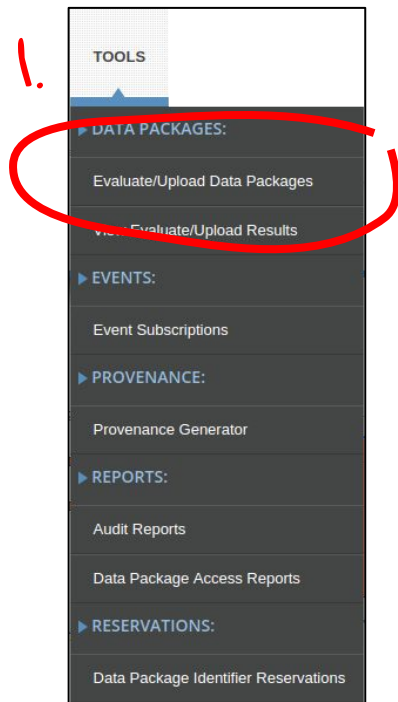
## View Evaluate/Upload Results

Select the **Evaluate** or **Upload** results to view (results are stored for 180 days):

Wed Nov 15 13:07:56 MST 2017 (evaluate) ▼

Package Id	Was evaluated	Report	Total Quality Checks	Valid	Info	Warn	Error	System Message
edi.98.1	Yes	<a href="#">view</a>	37	30	7	0	0	

# UPLOAD YOUR DATA PACKAGE



2.

3.

Select an Ecological Metadata Language (EML) file to evaluate or upload.

*EML Metadata File:*

edi.98.1.xml

*Data Upload Options:*

☐ Allow PASTA to skip upload of a data entity if it has a matching copy [Please note](#)

☒ I want to manually upload the data by selecting files on my local system [Please note](#)

4.

5.

6.

Entity Name	Object Name	Data File <small>(The file name you select should match the object name.)</small>
DailyWaterSample-NIN-LTER-1978-1992	LTER.NIN.DWS.csv	<input type="button" value="Choose File"/> LTER.NIN.DWS.csv



# UPLOAD YOUR DATA PACKAGE

### View Evaluate/Upload Results

Select the **Evaluate** or **Upload** results to view (results are stored for 180 days):

Wed Nov 15 13:21:06 MST 2017 (upload) ▼

Package Id	Was uploaded	Report	Total Quality Checks	Valid	Info	Warn	Error	System Message
<a href="#">edi.98.1</a>	Yes	<a href="#">View</a>	37	30	7	0	0	

# UPLOAD YOUR DATA PACKAGE

**Data Package Summary** [View Full Metadata](#)

---

**Title:** Daily Water Sample Nutrient Data for North Inlet Estuary, South Carolina, from 1978 to 1992, North Inlet LTER


**Creators:** Gaucho, Chase; Conquerors of the Useless

**Publication Date:** 2006


**Citation:** Gaucho C. 2006. Daily Water Sample Nutrient Data for North Inlet Estuary, South Carolina, from 1978 to 1992, North Inlet LTER. Environmental Data Initiative. <http://dx.doi.org/10.5072/FK2/777109ef542e04740af0b4521ebfae08>. Dataset accessed 11/15/2017.

**Abstract:** This data package consists of Daily Water Sample Nutrient Data for North Inlet Estuary, South Carolina, from 1978 to 1992, North Inlet LTER. Its purpose is to establish a long term data base on the nutrient dynamics of a...  
[Show more >](#)

**Spatial Coverage:**



S: 33.1925 E: -79.1002 W: -79.2936

**Package ID:** [edi.98.1](#) 

**Resources:** [Metadata](#) [Report](#) [Data](#)

1. DailyWaterSample-NIN-LTER-1978-1992 (902K)

[Download Zip Archive](#)

**Intellectual Rights:** LTER Network Data Access Requirements The access to all LTER data is subject to requirements set forth by this policy document to enable data providers to track usage, evaluate its impact in the community, and confirm us...  
[Show more >](#)

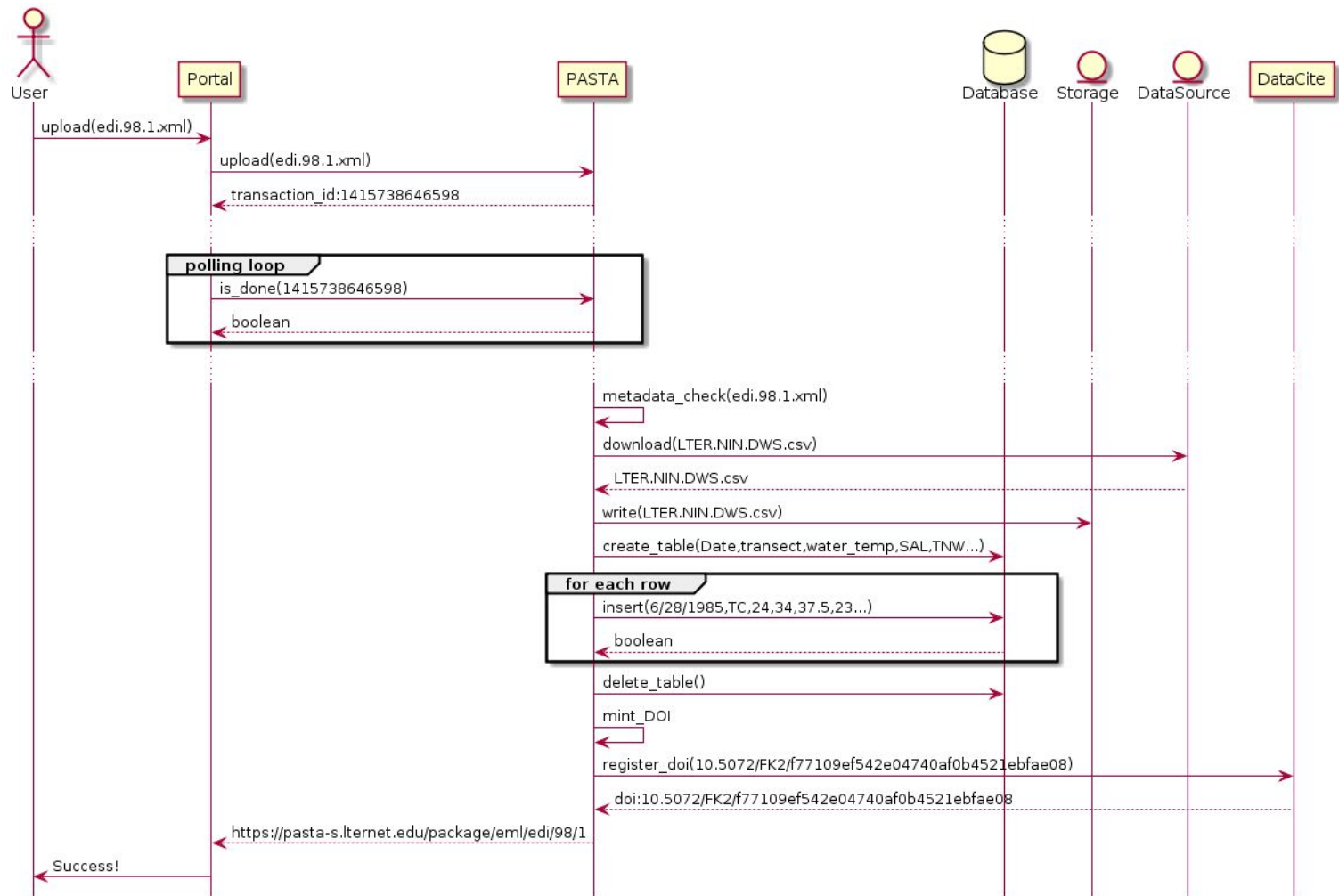
**Digital Object Identifier:** [doi:10.5072/FK2/777109ef542e04740af0b4521ebfae08](http://dx.doi.org/10.5072/FK2/777109ef542e04740af0b4521ebfae08)

**PASTA Identifier:** <https://pasta-s.internet.edu/package/eml/edi/98/1>

**Provenance:** Generate [provenance metadata](#) for use within your derived data package

**Code Generation:** Analyze this data package using [Matlab](#), [R](#), [SAS](#), [SPSS](#)

Success!



# RESOURCES

Read-the-docs - <http://pastaplus-core.readthedocs.io>

Package API - <https://pasta.lternet.edu/package/docs/api>

Audit API - <https://pasta.lternet.edu/audit/docs/api>

PASTA+ on GitHub - <https://github.com/PASTAplus/PASTA>

Examples - <https://github.com/EDIorg/tutorials>

Questions - [info@environmentaldatainitiative.org](mailto:info@environmentaldatainitiative.org)

THANK YOU

