A Brief Tour of the EDI Dashboard

Mark Servilla, Duane Costa, and James Brunt
Environmental Data Initiative, University of New Mexico

Introduction

As a user of the EDI Data Repository have you ever tried to upload a data package or access data through either the LTER or the EDI Data Portal and received an error message indicating that PASTA+ or another subsystem is not responding? Have you ever uploaded a data package and wanted to see if it is still being processed by PASTA+ or determine when it completed processing and was registered by PASTA+ as a published and archived data package, or for that matter, if it was successfully synchronized and indexed by DataONE? Were you ever interested in simply seeing how many new or updated data packages had been added to the EDI Data Repository over the last 24 hours, week, or month? If you responded “yes” to any of these questions, there is some help out there to answer them in the form of the “EDI Dashboard” website at https://dashboard.edirepository.org/dashboard. The EDI Dashboard was initially created as an internal tool for us to monitor the state-of-health of PASTA+ and related systems that keep the EDI Data Repository running smoothly. It has since blossomed into an EDI “Swiss Army knife” for reporting on and managing information about EDI, PASTA+, and the data packages that are under our care. The following article is a brief tour of what the EDI Dashboard provides to us, as EDI administrators, and you, as users of the EDI Data Repository.

The EDI Dashboard is currently partitioned into four major sections that are accessible from the website banner: Health, Reports, PASTA, and User Management, along with the typical website About page and a section of convenience links to the different EDI Data Portal environments (i.e., production, staging, and development). You may also notice a user login link on the right-side of the banner section. For the most part, the EDI Dashboard is accessible to the public. There are some actions within sections, however, that do require an administrative login for privacy reasons (e.g., under User Management or Reports), but we’ll still describe them just so you are aware of their purpose. Before we jump off into this tour, we do want to emphasize that this site is provided with no expectations and it is continually changing, mostly with new tools and features. Oh, and if you are the inquisitive sort, you may find an Easter egg or two sprinkled about.

Health at a Glance

The first major section of the EDI Dashboard is Health. In fact, you’ll notice that the default display you see when you reach the EDI Dashboard home page is the “Health at a glance” page, the grand view of all critical systems under EDI management. It is divided into six subsections that cover the different areas of EDI cyber-infrastructure: PASTA+ Server Infrastructure, EDI Portals, LTER Portals, EDI GMN, LTER GMN, and Related Services. Each subsection is defined by a hierarchy that is classified by one or more of the deployment environments we manage or by a particular service. At each level, the dashboard will display the state of that environment or service by indicating whether it is “ok” (in green) or “down” (in red). Drilling down further into one of the environment links (e.g., the Related Services subsection) shows you individual services, also with “ok” or “down” status indicators.

These status indicators will let you know immediately if there is a problem with a particular service, but the real details are found yet one layer down when you select the individual service link. At this lowest
layer, the EDI Dashboard breaks down the “state-of-health” into component level states that comprise the
duty of the service. For example, PASTA+’s Data Package Service is composed of three hierarchical
components: from the highest level component to lowest are the Ubuntu Operating System, Apache
Tomcat, and the Data Package Java application (which operates under Apache Tomcat). Each of these
components must be functioning correctly for the overall “state-of-health” to be “ok”. At this point, the
“state-of-health” check begins with an evaluation of the highest level component first, and only then
proceeds to the next lower component if the higher level is healthy (i.e., it doesn’t make good sense to
check if Apache Tomcat is running when the Operating System is not responding). You’ll notice that the
“state-of-health” indicators in this view have changed from “ok” and “down” to assertions, such as
“SERVER_DOWN” or “TOMCAT_DOWN”, followed by either “True” or “False”. For system administrators,
these assertions are more meaningful because they indicate a specific state condition that tells us where
to begin looking for a problem if one exists on a particular PASTA+ or related service.

The “state-of-health” process checks on all of our critical infrastructure once every 5 minutes. If you
watch any of the “state-of-health” web pages, they too update every 5 minutes so that you may see, at a
glance, the health of our systems. The sub-system performing the 5 minute health check also sends an
email to us whenever there is a change in status to any server we monitor. This capability complements
our use of Nagios, which will eventually be phased out. Like Nagios, the “state-of-health” service and the
EDI Dashboard web application run on a server in Amazon’s EC2 cloud so that they may continue to
function if our local infrastructure or network become compromised.

Reports

The Reports section of the EDI Dashboard is somewhat of a “catch-all” section for displaying various
information about the EDI Data Repository, PASTA+, and data packages. The first two reports are
accessible only by EDI administrators since they may expose what some may consider to be sensitive
information. The “No Public Access” report lists data packages in the EDI Data Repository that contain
access control elements that do not allow public access to one or more data entities or the entire data
package itself. It is important for us to keep track of data packages being submitted without public access since we strongly believe that all data should be open and accessible unless circumstances require privacy.

Similarly, the “Offline Data” report shows us the data packages that are using the “offline” attribute in the EML metadata. Offline data may be used in some cases where the data are too large for online access or the data is so very sensitive that it must be protected at an offsite location. Both reports are refreshed on a weekly basis. Thankfully, the number of records in either the “No Public Access” and the “Offline Data” reports is fairly low.

The next report, which is open for all to access, is the “Package Tracker”. This report takes a PASTA+ package identifier in the form of “scope.identifier.revision” as input and returns state information about that data package, including when it was uploaded and registered in PASTA+, if and when it was uploaded to the DataONE Generic Member Node (either the LTER or EDI GMN, respectively), if and when it was synchronized to the DataONE Coordinating Node, and if it had been indexed by DataONE’s Solr search engine. This report may be helpful to those users who would like to know more information about their data packages beyond an acknowledgement that it has been published into the EDI Data Repository. This particular report is both new and evolving, so the information displayed today may be very different from the information displayed tomorrow—new information may include the date and time...
when the data package was copied to Amazon’s Glacier storage and the date and time of the last checksum verification of any disk stored resource (metadata, data, and report) of the data package. Stay tuned for updates.

The last set of reports that you may view in this section are the “Recent Uploads” reports. These reports are divided into queries for the past 24 hours, week, and month, and display a time-series plot of the upload frequency for the time period specified, as well as a list of the recently uploaded data packages and the date and time they were uploaded. We find this report most helpful to quickly see how active the EDI Data Repository has been in the recent past.
The third section of the EDI Dashboard contains two convenience functions: the first displays a list of data package identifiers (applicable only to the “edi” scope at this time) that have been reserved by an individual and the second shows any data package that is actively being processed by PASTA+ as a result of either an evaluation or upload. Because the “edi” scope is shared across so many individuals and organizations, we found it helpful that users could set aside and reserve package identifiers for which they could use with a future data package. Unfortunately, some users would reserve a set of package identifiers, but immediately forget what identifiers they had reserved. The “Reservations” function displays a list of all reserved package identifiers that are not associated with a data package in the EDI Data Repository. The list shows the scope and identifier value of the package identifier, the full principal identity of who made the reservation, and the date and time of when the reservation occurred. This list is divided into sections designated for the production, staging, and development environments that we support.
The second function is the “Working On” table, which displays data packages that are actively being processed by PASTA+ as either an evaluation or an upload (labeled as a “create” operation in the table) and the date and time processing began. This table is also divided into sections for production, staging, and development environments. Large data tables can take extra time during processing to ensure its congruence quality. For anyone who has just started an evaluation or upload process through either the LTER or EDI Data Portal or PASTA+’s REST API, the "Working On" table is invaluable to see if your data package is still in the processing state. As EDI administrators, we often consult this table before we begin our Wednesday evening system patching or if we need to deploy software to fix a critical bug. You may find that your data package lingers in the “Working On” table if it requires extra time during the quality checking phase of processing, especially if it contains many or large data tables.

User Management

The last major section of the EDI Dashboard is “User Management”. At present, functions under “User Management” pertain only to users registered in the EDI LDAP user directory. Since EDI has broadened its scope to include communities outside of the LTER Network, and because we do not explicitly manage the LTER LDAP, we had to deploy an LDAP system that allowed us to register non-LTER users. To simplify the management of these users, we have developed a user management functions that gives EDI administrators the ability to create and delete users, but also allows individual
users the ability to modify their account information. The first three functions under “User Management” are restricted to EDI administrators: “Create User”, “Delete User”, and “List Users”. User account information is limited to login identifier, given name, surname, and email. Once a user account is initially created with the “Create User” function it is seeded with a random password, and a one-time password reset request is sent to the user’s email address. As EDI administrators, we do not manage the user’s password. The “Delete User” function does what it says, it deletes a user’s account permanently, and the “List Users” function simply provides a list of user login identifiers in the form LDAP distinguished names.

Users, on the other hand, can modify their account information using the functions “Update User”, “Change Password”, and “Reset Password”. The “Update User” function allows an authenticated user to change only their given name, surname, or email address. The “Change Password” allows a user to change a current password to a new password. Both the “Update User” and “Change Password” functions, as expected, require a current password to be successfully processed. The “Reset Password”, like the “Create User”, sends a one-time password reset request to the email address currently registered with the user’s account.

Summary

In summary, the EDI Dashboard provides a collage made up of vignettes into EDI cyberinfrastructure that is helpful to both EDI users and administrators. The website has evolved (and continues to evolve) to accommodate new tools and services necessary to perform our jobs. We do want to set an expectation that this website should be viewed as an ongoing development that may change without notice. With that in mind, we are also eager for new ideas to incorporate into the EDI Dashboard that will help improve the overall curation of environmental and ecological data in the EDI Data Repository. Just drop us a line.