

Jornada Basin LTER Information Management 2020 Annual Report (DRAFT)

Gregory E. Maurer with contributions from the IM Team (John Anderson, Darren James, N

13 January 2021

Introduction

This document summarizes the accomplishments of the Jornada Basin LTER (JRN) Information Management (IM) team in 2020, and outlines activity planned for 2021. Most of the work described here has been, or will be, guided by what the IM team wrote in the Data Management Plan submitted with our 2020 LTER proposal. In the late summer of 2020 we received word that the LTER proposal was recommended for funding beginning in early 2021. The Data Management Plan received generally favorable feedback from the NSF review panel.

Data publishing progress

There were a number of notable accomplishments in packaging and publishing JRN data products in 2020. All of these involved significant additions to, or revision of, Jornada metadata.

- We finished updating “Core/Signature” data packages in early March 2020 (59, 2 remain) in time for submission of our LTER proposal.
- John and Geovany added 107 new data packages to EDI for Cross-Scale project meteorology data and have made monthly updates .
- 10 new non-meteorology data packages were added to EDI.
- We updated half the JRN “non-core” (older, completed) data packages on EDI.
- Two Jornada eddy covariance towers joined the AmeriFlux network and are, or will be, publishing 30 minute fluxes with the AmeriFlux data repository (Vivoni group’s Tromble tower, and Systems Ecology Lab Bajada site).
- Savannah lab UAV data (4 new packages) is staged, and we are working towards publishing several other complex “non-tabular” datasets to EDI.

Please see further statistics and figures on our activity at the end of this document.

New JRN IM features, infrastructure, and data products

We added several new elements to the Jornada Basin LTER Information Management system, including a new website and an IM-focused database. We also improved several existing features of the IM system and planned for further improvements to be made in 2021. These changes have improved the accessibility and integrity of JRN datasets and lay the groundwork for increases in the efficiency of publishing JRN data and greater use of this data locally and in the wider scientific community.

JRN Websites

- A new JRN website was built and began service in June 2020.
- A new partner data catalog was added to the website.
- A new interactive data viewer was made available on the website (Geovany), and further additions to this were discussed.
- An improved spatial data catalog was added to the new website, and further improvements were planned.
- A number of updates to the Jornada IM website were completed, including new documentation, improved metadata templates, IM diagnostics, and data access/analysis tools.

Metadata database

- JRN-metabase, a PostgreSQL database following the LTER-core-metabase schema, was created and made available for testing on the DASH server.
- Greg and Shelly (and formerly Haneen Omari) populated this database with metadata for all JRN “core” packages, and all recently updated “non-core” data packages.
- We successfully made valid EML (XML metadata documents) using the database as a backend.
- We expect JRN-metabase (and associated R code) to enter service for building metadata documents (EML) for the Jornada’s EDI data packages by March 2021.
- Alternative server locations for hosting this database are currently under consideration.

Harmonizing Jornada data

- Greg began a project to assemble and harmonize Jornada NPP and related data.
- Darren James and the Peters research group began working on new projects harmonizing Jornada precipitation data.
- The IM team identified several potential harmonized datasets to add to the interactive data viewer beginning in 2021.

Outreach, education, and LTER Network Activity

- The Non-tabular working group (LTER/EDI, Greg is a member) concluded a first draft of its “Best practices for non-tabular data archiving” document.
- Greg joined the LTER network’s IM Executive committee.
- Greg joined the LTER-core-metabase and Southwest IM data management/science training working groups.
- Language clarifying data publishing requirements for Jornada researchers (USDA and LTER) was agreed upon and added to the JER and JRN websites.
- A new Jornada Spatial Data Release Policy ([link](#)) and procedures for providing spatial data to researchers upon request, were agreed upon by LTAR and LTER.
- A Jornada IT and data management working group was established (B. Bestelmeyer, S. McCord, D. James, J. Ragosta, G. Maurer) and will begin coordinating upgrades to Jornada IT infrastructure, informatics systems, and data management practices in 2021. Preliminary ideas were presented at the 2020 Jornada Symposium ([see here](#)).
- At the 2020 Jornada Desert Ecology Short Course the IM team led or participated in 3 data-focused events - an introduction to data management, a metadata clinic, and a data jam. Details available [here](#).

Future initiatives

The JRN IM system has become more robust and effective over the past 12 months, approaching our goal that information management at JRN must meet or exceed all LTER Network standards and expectations. Planned activities and improvements during 2021 are below.

- Improvements to the efficiency of publishing Jornada datasets, i.e. lowering the time and effort required per dataset published, would be beneficial. New database initiatives (JRN metabase) and restructuring of project approval and metadata workflows (see “Challenges” below) are planned to help facilitate these improvements.
- We have applied to host an EDI fellow in the summer of 2021 ([details here](#)). If accepted Jornada’s fellow will develop workflows to publish and improve access to our non-tabular datasets, including UAV images, phenocam imagery, and derived spatial datasets.
- There is a need for better outreach and training for Jornada research personnel (students, investigators, and staff) on topics around data management, access, and publication. Several training initiatives and activities are under development:
 - Greg is considering being certified as a Carpentries instructor in 2021 with the goal of bringing data management and data science

workshops to the Jornada. His training is supported by the NM EpsCOR program.

- Data Bounty Hunters
 - Improved data-focused activities at the 2021 Ecology Short Course
 - Video tutorials on the JRN website (e.g. using the interactive data viewer)
 - New LTER data management & data science training materials and workshops are being developed in coordination with MCM, SEV, and NWT LTER IMs.
- We will add data and features to the interactive data viewer based on input gathered from numerous PIs this fall. Greg and Geovany will coordinate this.
 - There is an initiative in the LTER network to publish site meteorology data at CUAHSI HydroServer. Greg, Geovany, and John plan to develop a workflow for this change using a limited subset of our met data (GIBPE 15min data only at first).

Challenges to overcome

Below are a few challenges the IM team faces in 2021. Any ideas on how we might overcome some of them are welcome.

Project-level metadata and spatial data issues

There are some workflows for creating, organizing, and tracking projects and their location data at the Jornada. For instance, John Anderson’s “Research notification” system tracks approved research projects and should, in theory, initiate collection of metadata for each project. However, research projects commonly suffer from a lack of metadata collection and accessibility after the approval phase, and information gathered during the approval phase is not widely accessible to Jornada researchers or the public. Some things that would help address this problem:

1. A project metadata database (what experiments are happening or have happened, who is in charge, what are the questions, measurements, infrastructure) that is accessible to John and other project managers.
2. A “canonical” spatial database for the Jornada that links to project IDs. This spatial database would house plot locations, sensor locations, and other coordinates telling users where experiments happened over time.
3. A reference collection of ancillary spatial data products (roads, fence lines, vegetation, geomorphology). We have some of this, but its not very complete and there are multiple copies of everything.

Each of these resources would be most useful when shared among multiple partners at the Jornada (LTAR, LTER, etc.). The Jornada IT infrastructure and data management working group, which includes USDA, Landscape Data

Commons, and LTER staff, is beginning to address some of these issues and tasks in a coordinated way.

Outreach initiatives

- There is no shared communication platform for messaging among Jornada investigators, students, and staff. This tends to fragment Jornada communication and makes cohesive messaging for JRN community activities a challenge. Some of our future initiatives, such as Data Bounty Hunters, would greatly benefit from such a platform. A Mailman server, mailing list/group web services (MailChimp, Groups.io), or messaging platforms (Zulip, Slack) are possible candidates, but we need either infrastructure or budget to adopt these.
- It may be difficult to spend LTER budget for non-research activity, but we need some funds to make Jornada T-Shirts, hats, or stickers that could be used to encourage/reward participation in Jornada outreach initiatives (Data Bounty Hunters, for instance again).

Infrastructure

As described in above, the JRN IM Team needs server resources to host databases, web services, a communication platform, and potentially a future version of our website. Because JRN is partnered with the USDA Jornada Experimental Range (JER) and JER has existing server resources and sysadmins, appropriating some of these JER resources for JRN use (and contributing back to supporting the system) would be the most efficient & cost-effective solution for JRN to meet its IT needs. Currently, however, this option does not appear viable because the JER server infrastructure is out of date and undergoing a long transition to a more modern, stable, and supportable system. Thus, in 2020 the IM team was unable to provision a virtual server on this system. While we await progress on JER server systems, we are investigating ways we can meet our IT needs using NMSU or cloud providers (AWS, Google services), but these will require some LTER budget. Any suggestions regarding server resources available at Jornada or partner institutions are welcomed by the IM team.

Again, the Jornada IT infrastructure and data management working group, is beginning to address some of these issues in a coordinated way.

Tables and figures

Table 1: Data packaging updates for all, and non-meteorology JRN data packages in 2020.

	Total	Non-Met
New packages created	127	10
Package updates	2159	62
Unique package updates	270	47

JRN activity on EDI in 2020

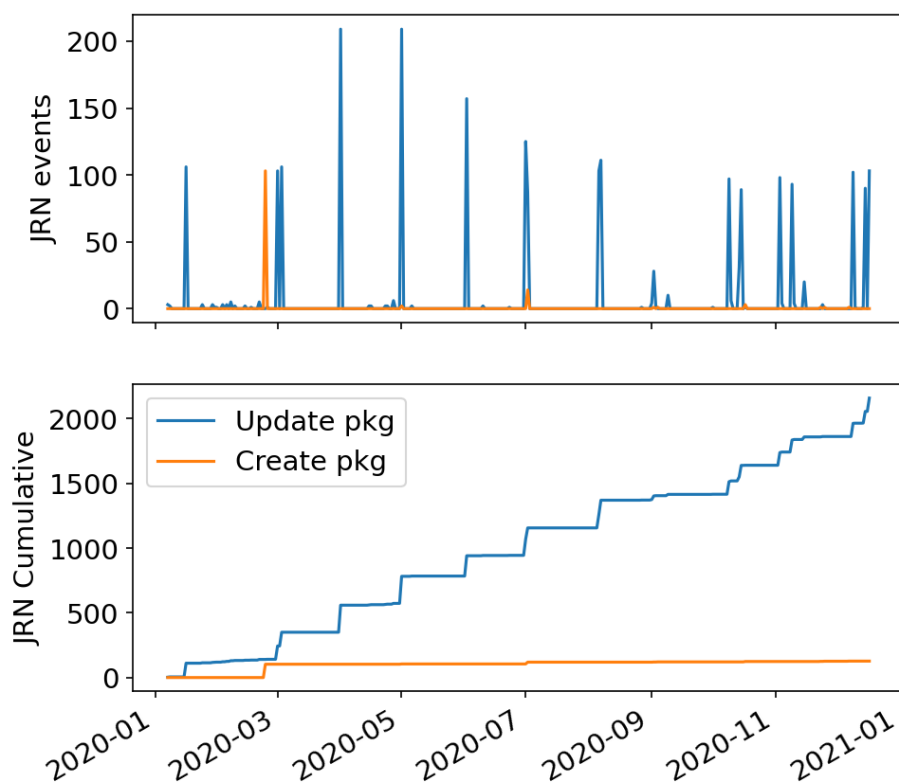


Figure 1: A plot of JRN’s PASTA+ database activity (knb-lter-jrn scope), including package creation and recurring updates, at the EDI repository in 2020. Top panel: Daily actions for all packages (spikes = meteorological data package creation/updates). Bottom panel: Cumulative actions for all data packages.

JRN activity on EDI (Excluding met)

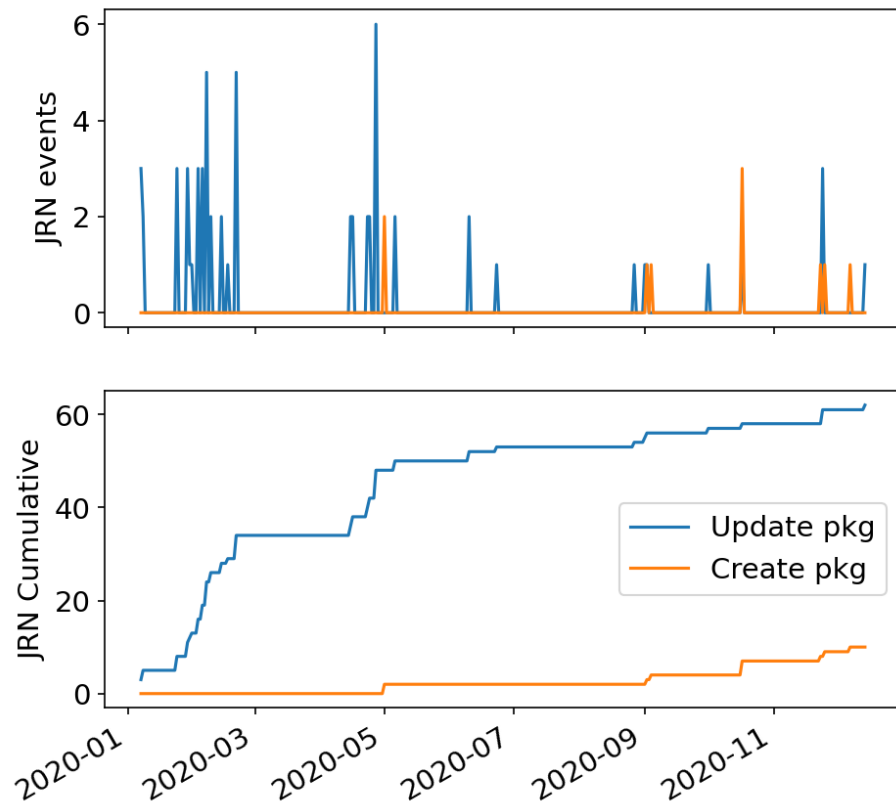


Figure 2: JRN's PASTA+ database activity, as in Fig 1, but excluding the meteorology data packages John and Geovany update. Top panel: Daily actions. Bottom panel: Cumulative actions.

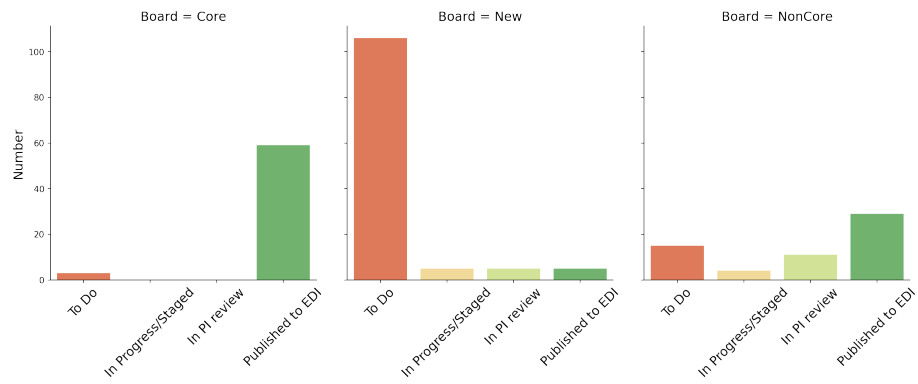


Figure 3: Number of data packages listed under the stages of progress on our "Core", "Non-core", and "New" JRN IM Trello boards.