

April 2017 Water Cooler: PASTA Quality Engine (aka “The Checker”)



Published on *LTER Information Management* (<http://im.lternet.edu>)

Home > April 2017 Water Cooler: PASTA Quality Engine (aka “The Checker”)

April 2017 Water Cooler: PASTA Quality Engine (aka “The Checker”)

Mon, 04/10/2017 - 11:07am — mgastilbuhl

The topic of the 10 April 2017 Virtual Water Cooler is

The PASTA Quality Engine

Noon pacific, 1pm mountain, 2pm central, 3pm eastern.

Connection info: <https://ucsb.zoom.us/j/322175707> [1]

Margaret will present about ten slides today. See pdf attached below.

There are also a few side-topics. See below.

The following message was Margaret sent out on April 3 for your brainstorm.
Hope to see you all. Thanks!

Hi all -

As you know, an LTER IMC watercooler is scheduled for next Monday, April 10 (3pm EDT). One of the topics will be advances planned for the ECC -

the system performing dataset checking for PASTA. This msg is a short description of what we would like to cover.

In case you have forgotten, back in 2012 an IMC working group finalized 72 checks, and ~25 of these were running when PASTA went into production in 2013. In the intervening time, additional checks were implemented depending on resources available, and other checks proposed. Fast forward to today: EDI is up and running, and we have resources budgeted to work on this more systematically.

The new checks to be implemented are related to specific feature requests. These involve data integrity, and they are important to review with you because failure will generate an 'error' and block upload of the dataset.

1. checksum (2 checks, details on request):

These will confirm entity integrity during upload. The checksum can be used later by PASTA to minimize entity duplication.

2. DOIs:

PASTA now adds package DOIs to L1 EML. This means that L0 EML should not contain a DOI (e.g., a DOI may have been inadvertently left behind if an EML doc was recycled). This check will prevent confusion due to possible conflicting ids.

During the watercooler, we will outline specifics about the checks. As with other PASTA improvements, checks will be developed and can be tested on portal-d (i.e., you can pre-evaluate your trial EML), and portal-s is reserved as the staging platform for production. For a summary of the checker, its behavior, and results from the first few years with LTER datasets, see this paper in the recent Ecoinformatics special issue, DOI: 10.1016/j.ecoinf.2016.08.001

Best,
EDI team

Side Topics:

* Today April 10 is deadline to submit an ESIP Session Proposal.

* Please indicate on the doodle [2] if you plan to (or hope to) attend the annual IMC meeting [3] at ESIP.

* If you have not already done the survey from the IM Training Group please do.

* Comments on the Data Access and Intel Rights document are due 15 April to Corinna.

(See pdf attached to an email of March 7th from Corinna for the doc titled 'DataAccessPolicyRequestForComments.pdf')

* Two seats will open on IM Exec this year.

* The May Virtual Water Cooler is (tentatively) about IM Training

Attachment	Size
2017-04-10_ECC_update.pdf [4]	409.21 KB

- Virtual Updates [5]

- Copyright © 2012 Long Term Ecological Research Network, Albuquerque, NM -
This material is based upon work supported by the National Science Foundation
under Cooperative Agreement #DEB-0236154. Any opinions, findings, conclu-
sions, 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.

Source URL: <http://im.lternet.edu/node/1300>

Links:

- [1] <https://ucsb.zoom.us/j/322175707>
- [2] <http://doodle.com/poll/da2dn384v7pdb9gw>
- [3] <http://im.lternet.edu/news/meetings/2017>
- [4] http://im.lternet.edu/sites/im.lternet.edu/files/2017-04-10_ECC_update.pdf
- [5] <http://im.lternet.edu/taxonomy/term/169>