"Readme" for Draft v0.1 EML Congruence Checker reports, September 2011

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

Home > IM Working Groups > "Readme" for Draft v0.1 EML Congruence Checker reports, September 2011

"Readme" for Draft v0.1 EML Congruence Checker reports, September 2011

Fri, 09/23/2011 - 4:20pm — mobrien

=== README_ECC_v0dot1_Report_Draft.txt ===
Fri Aug 26 12:13:05 PDT 2011

=== Contents ===
1. This readme file

2. Copy of the EML Congruence Checker 2011 Plan (HTML)
Also can be found at http://im.lternet.edu [1], and was distributed in email in June 2011.

3. xxx_docid.txt
list of the docids in metacat that are included in this report

4. xxx_report.xml
concatenated reports from all datasets in the docid list, and the input for the transformed HTML file

5. xxx_report.html
transformed version with some report responses summarized at the top

=== Introduction ===
This report is the result of running the v0.1 NIS DataManager web services. All datasets from all sites (contributions as of early August) were processed. The main benefit to be gained from running the code over all the datasets is that we've given the IMC's EML Metrics working group and the NIS developers a look at the range of dataset construction styles and a starting place for planning how to proceed. Since this exercise used site-data, the sites should also be kept aware of the results, and that's why you are receiving this report for your site. We encourage you to share the report and other material about datasets and PASTA with your local PIs.

Please keep in mind that at this time, the report is for information only; there are no failures. Also keep in mind that this is the first iteration of the checker. The list of requested checks is very long, and only a few have been implemented in v0.1. Each check must be examined in great detail as it is implemented. In V 0.1, only tabular data can be handled, but in the future, PASTA (and the checker) will handle other types of data as well.

=== Checker behavior ===
V 0.1 of the ECC has one goal: to load a data table described by EML into a relational database table (in this case, Postgres). The code responses from the individual steps are translated and give a status: "Valid", "Info", "Warn", "Error". The conditions for each status depend on the check being implemented. Determining each of these checks, conditions, responses and status has been the purview of the EML Metrics WG, the DataManager Tiger Team, and the NIS developers.

=== Report HTML View ===
The HTML view was transformed from the collected XML reports. It attempts to summarize the code responses. If the report cuts off information that is important to you, see the xxx_report.xml file. The report has 3 parts:

== 1. Summary ==
The summary tallies up differnet kinds of responses, for both the entire dataset and for indivdual data tables (remember, only data tables are addressed in V0.1)

Some datasets return server errors instead of reports, and here, these are summarized as "Currently uncheckable". These are not broken datasets; they represent code-errors that need to be fixed, and future checks to implement. For some of these, the dataset features that caused the code-error was descernable, and these were tallied.

== 2. Links ==
The next section contains a list of data entities generating report "error", links to their detailed reports, and links to the original dataset. Other sections may be added for "warn" and "info, valid".

== 3. Report Detail ==

Each data entity (table) that the checker attempted to load has an individual report with the response for each of the current (5) checks. Remember that as additional checks are implemented, the entity-level report detail could become quite long.

=== What IMs should do ===

1. You are the most knowledgeable about your site's datasets. Please examine the all the parts, and see if the responses are what you'd expect.

2. Share the information with your PIs. In some cases, PIs will not be interested in the detail of the report, but they should be aware of the growing capabilities of the NIS, and that these tools are being developed for us.

3. Respond to Margaret with comments on the report content, especially what features would be most useful to see aggregated across all sites.

4. Think about how you want to use this tool at your own site, and also how the network should it.

=== Upcoming events (2011) ===

1. at the IMC meeting, there will be a breakout session to focus on dataset reports and their use. See the agenda on http://im.lternet.edu/meetings/2011 [2] for developments, and write to imexec with comments on the breakout agendas.

2. at the EIMC, there will be a BoF session on the functional requirements of the checker tool. The session is planned to include a short demonstration of the web services and reports, a summary of the project status, and the developing rationale behind the classification of the list of checks. The bulk of the session is to be a discussion of current planned requirements and reporting scheme.

- EML Metrics and Congruency Checker [3]

Source URL: http://im.lternet.edu/node/923

Links:
[1] http://im.lternet.edu
[2] http://im.lternet.edu/meetings/2011
[3] http://im.lternet.edu/taxonomy/term/211