A Continuing Plan for the "FRBR-Inspired" Catalog 2.1? (Fall 2012)
Alison Babeu

Introduction

Work that began in 2005[1] to create an experimental "FRBRized" catalog for the Perseus Digital Library's (PDL) online classics collection has continued to the present day in the form of work on what I have labeled the "FRBR Inspired" catalog. The catalog and authority data created over the last seven odd years is in varying stages of completion[2] with differing levels of granularity and has seen the transition through two new releases of the MODS schema (currently at 3.4), the final release of the Resource Description and Access (RDA) cataloging code, the zombie apocalypse, and the explosion of library sources (as well as many other cultural heritage sources) as linked data. Consequently we have decided to release under a Creative Commons license both the catalog data (MODS[3] records) and authority data (MADS records) in its raw and frankly largely undocumented form to see if interesting uses beyond our rather limited initial "plan" (more on this below) might be realized. Fuller details on the cataloging workflow, the creation of different kinds of MODS (depending on the type of book being cataloged) and MADS records, and extensive sample records can be found in Babeu (2008).[4] The purpose of this largely informal document is to provide a very basic overview/explanation of the current state of the data, present some of its

---

[1] For more on the FRBR-Inspired catalog's auspicious beginnings, see David Mimno, Gregory Crane, Alison Jones. (2005). "Hierarchical Catalog Records: Implementing a FRBR Catalog." In *D-Lib Magazine*, 11 (10), 2005. http://www.dlib.org/dlib/october05/crane/10crane.html

[2] A number of MODS records for large component works, such as the fourth volume of the *Fragmenta Historicorum Graecorum* are unfinished and contain rather accusatory <LEFT OFF HERE> comments, with the goal of course being to return and finish cataloging them as time allows.

[3] MODS stands for Metadata Object Description Schema (http://www.loc.gov/standards/mods/), whereas MADS stands for Metadata Authority Description Schema (http://www.loc.gov/standards/mads/), both are standards that were created by the Library of Congress and are expressed in XML.

[4] Alison Babeu. (2008). "Building a "FRBR-Inspired" Catalog: The Perseus Digital Library Experience." http://www.perseus.tufts.edu/publications/PerseusFRBRExperiment.pdf

limitations/shortcomings, discuss some desired changes to be made to both the records and the workflow, and future plans.

So What Exactly is in the FRBR-Inspired Catalog Collection Anyway?

The Perseus FRBR Catalog collection currently includes MODS records for:

- About half of the current Perseus online collection of primary source and reference works available as fully TEI-compliant XML *(cataloging of the Perseus collection is ongoing).*
- A collection of 600 primary source editions as well as reference works in Greek and Latin that exist largely as image books with OCR (scanned in house at Perseus between 2006 and 2007) but with no TEI-XML transcriptions *(this cataloging is complete)*.
- A collection of about 250 classical editions that were scanned by the Open Content Alliance (OCA) for the PDL between 2007 and 2009 and are all fully available online (they also have accompanying OCR) through the Internet Archive *(cataloging of this test collection is mostly complete)*.
- A collection of about 3000 public domain editions of classical authors and reference works that were identified through the creation of a collection development spreadsheet based off of the standard editions used by four major classical reference works: Lewis and Short, the Oxford Latin Dictionary (OLD), the Liddle Scott Jones (LSJ) Lexicon, and the Thesaurus Linguae Graecae (TLG). This list was initially created as a digitization guide for the Mellon funded Cybereditions project to ensure that at least one public domain edition for every major surviving classical Greek and Latin author was available online[5] *(cataloging of this collection is ongoing)*.

This download also includes "enhanced" MADS records for:

- Hundreds of classical authors of primary source texts (including fragmentary authors)
- Hundreds of editors of public domain classical editions published in the major series (Teubner, Loeb, OCT, etc.)

One challenge with the data collection as it currently stands is its varying levels of completeness and granularity since the creation of both MODS and MADS records has been and largely continues as a perpetual work in progress. A number of the more prominent issues include:

- There are a number of large composite works that have only been partially cataloged (e.g. *Fragmenta Historicorum Graecorum (FHG)*, Vols 4 and 5).
- A number of large composite works have been cataloged but have not had individual expression records created (e.g. *Grammatici Romanae Fragmenta, Epistolographi Graeci*).[6]
- A large number of early MODS records (created in 2006 and 2007) do not contain as much data as later records.

---

[5] This list also provided the basis for a select number of these texts being made more fully available for download (full pages scans and OCR data, not simply PDF files), including a fully "open" list (that is volumes and data that are downloadable from outside the US) and a "limited distribution" list of volumes and data only downloadable from within the United States with a Google account.

[6] There is also the question of whether individual expression level records should continue to be created as this creates unnecessary duplication of data when searching the catalog in the eXtensible catalog interface as well as in the generation of the Atom Feeds. Some further discussion of this issue is found later in this document.

- Many MODS records do not contain appropriate links to online manifestations (such as in Google Books) because when the work was first cataloged there were no online editions available.
- A number of constituent records (<relatedItem type="constituent">) for individual works within larger MODS records (e.g. the *Comicorum Atticorum Fragmenta*) include records for fragmentary or rare authors that have not been linked to their online authority records (before the Virtual International Authority File (VIAF)[7], links to online authorities for many of these authors were not available)
- A number of authors do not have MADS authority records created for them due to time constraints or the lack of a downloadable authority record at the time their work was cataloged.
- A number of author have not had identifiers created for them as yet (part of an ongoing project).
- Many MADS authority records have not been enhanced (e.g. with work identifiers, URLs, etc.)
- New data sources have become available since works were cataloged (e.g. the VIAF as linked data), and MADS records are not up-to-date.
- Page level linking was not available for either the HathiTrust or the Open Content Alliance when many books were cataloged, and many MODS records do not contain this data.

In many cases adding this data would be prohibitively time consuming but it also means there is a great deal of inconsistency particularly in terms of data available in many of the records.

What's Contained in the Catalog Data (MODS records)

- Standard identifiers from library systems for author names and work titles (uniform names for authors, and uniform titles for works where possible), which have also been mapped to their corresponding names in classical canons.
- Unique work identifiers drawn from existing canons such as the TLG, PHI, the STOA registry of Latin literature, and Perseus ABOS (as well as CTS-URNS in MODS records for Perseus works).
- Analytical cataloging data (full lists of contained works and authors in structured XML, expanded manifestation level information such as addition of series information).
- Links to online bibliographic records (Worldcat.org, Library of Congress, Open Library)
- Links to online manifestations of books in the different mass digitization projects (including Google Books, the Open Content Alliance, Hathi Digital Trust), with page level linking to exact locations of works.

Figure 1 is a sample MODS "expression-level" record for a work by a Greek fragmentary historian, the image below does not include the complete MODS records, particularly the <relatedItem type="host"> section containing the full enhanced bibliographic information for the larger book in which this work is found.

```
?xml version="1.0" encoding="UTF-8"?>
<mods xmlns="http://www.loc.gov/mods/v3"
   xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
   xsi:schemaLocation="http://www.loc.gov/mods/v3 http://www.loc.gov/standards/mods/v3/mods-3-4.xsd">
  <titleInfo>
    <title>Fragmenta</title>
```

---

[7] http://viaf.org/

4

```xml
    </titleInfo>
    <titleInfo type="abbreviated" displayLabel="FHG">
      <title>FHG i p. 1</title>
    </titleInfo>
    <titleInfo type="abbreviated" displayLabel="FGrH">
      <title>FGrH #1</title>
    </titleInfo>
    <name authority="naf" type="personal" xmlns:ns2="http://www.w3.org/1999/xlink"
      ns2:href="http://errol.oclc.org/laf/n87-931007.html">
      <namePart>Hecataeus,</namePart>
      <namePart type="termsOfAddress">of Miletus</namePart>
      <role>
        <roleTerm authority="marcrelator" type="text">creator</roleTerm>
      </role>
    </name>
    <name authority="naf" type="personal" xmlns:ns2="http://www.w3.org/1999/xlink"
      ns2:href="http://errol.oclc.org/laf/nr95-8185.html">
      <namePart>Müller, Karl</namePart>
      <namePart type="termsOfAddress">of Paris</namePart>
      <role>
        <roleTerm authority="marcrelator" type="text">editor</roleTerm>
      </role>
    </name>
    <name authority="naf" type="personal" xmlns:ns2="http://www.w3.org/1999/xlink"
      ns2:href="http://errol.oclc.org/laf/nr95-8187.html">
      <namePart>Müller, Theodor,</namePart>
      <role>
        <roleTerm authority="marcrelator" type="text">editor</roleTerm>
      </role>
    </name>
    <name authority="naf" type="personal" xmlns:ns2="http://www.w3.org/1999/xlink"
      ns2:href="http://errol.oclc.org/laf/n84-135026.html">
      <namePart>Langlois, Victor</namePart>
      <namePart type="date">1829-1869</namePart>
      <role>
        <roleTerm authority="marcrelator" type="text">editor</roleTerm>
      </role>
    </name>
    <identifier type="tlg">0538.002</identifier>
    <language>
      <languageTerm authority="iso639-2b" type="code">lat</languageTerm>
    </language>
    <language>
      <languageTerm authority="iso639-2b" type="code">grc</languageTerm>
    </language>
    <part>
      <extent unit="pages">
        <start>1</start>
        <end>31</end>
      </extent>
    </part>
    <location>
      <url displayLabel="GoogleBooks"
        >http://books.google.com/books?id=M-oUAAAAQAAJ&amp;pg=PA1</url>
    </location>
    <location>
      <url displayLabel="Open Content Alliance"
        >http://www.archive.org/stream/fragmentahistori01mueluoft#page/xcii/mode/2up</url>
    </location>
```

Figure 1: "Expression" Level MODS Record

What's Found in the Enhanced Authority Records (MADS RECORDS)

- Lists of variant names with language encoded.

- Standard identifier for authority records (e.g. LCNAF number)
- Lists of work identifiers so that the authority records can be effectively linked to bibliographic records.
- Lists of <mads:fieldOfActivity> so that authors can be search for by genre/type of activity (e.g. historian, epic poet, geographer, etc.)
- Standard author abbreviated names from classical reference works.
- Links to important online reference sources for a author name and/or biography (e.g. Wikipedia, Worldcat Identities, the VIAF)

Figure 2 illustrates a sample MADS record of the fragmentary author listed in the MODS record, Figure 1.

```xml
<mads:mads version="2.0" xsi:schemaLocation="http://www.loc.gov/mads/v2 http://www.loc.gov/standards/mads/mads.xsd"
xmlns:mads="http://www.loc.gov/mads/v2" xmlns:xlink="http://www.w3.org/1999/xlink" xmlns:xsi="http://www.w3.org/2001/XMLSchema-
instance">
 <mads:authority geographicSubdivision="not applicable">
  <mads:name type="personal" authority="naf">
   <mads:namePart>Hecataeus</mads:namePart>
   <mads:namePart type="termsOfAddress">of Miletus</mads:namePart>
  </mads:name>
 </mads:authority>
 <mads:related type="equivalent" lang="grc">
  <mads:name type="personal" authority="Brill">
   <mads:namePart>Ἑκαταῖος</mads:namePart>
  </mads:name>
 </mads:related>
 <mads:variant type="other" lang="fre">
  <mads:name type="personal" authority="viaf">
   <mads:namePart>Hécatée</mads:namePart>
   <mads:namePart type="termsOfAddress">de Milet </mads:namePart>
  </mads:name>
 </mads:variant>
 <mads:variant type="other" lang="ger">
  <mads:name type="personal">
   <mads:namePart>Hekataios</mads:namePart>
   <mads:namePart type="termsOfAddress">von Miletos</mads:namePart>
  </mads:name>
 </mads:variant>
 <mads:variant type="other" lang="ger">
  <mads:name type="personal">
   <mads:namePart>Hekataios</mads:namePart>
   <mads:namePart type="termsOfAddress">von Milet</mads:namePart>
  </mads:name>
 </mads:variant>
 <mads:variant type="other" lang="ita">
  <mads:name type="personal">
   <mads:namePart>Ecateo</mads:namePart>
   <mads:namePart type="termsOfAddress">di Mileto</mads:namePart>
  </mads:name>
 </mads:variant>
 <mads:variant type="other" lang="pol">
  <mads:name type="personal">
   <mads:namePart>Hekatajos</mads:namePart>
   <mads:namePart type="termsOfAddress">z Miletu</mads:namePart>
  </mads:name>
 </mads:variant>
 <mads:variant type="other" lang="lat">
  <mads:name type="personal" authority="viaf">
   <mads:namePart>Hecataeus</mads:namePart>
   <mads:namePart type="termsOfAddress">Milesius</mads:namePart>
  </mads:name>
```

```xml
  </mads:variant>
  <mads:variant type="other">
   <mads:name type="personal" authority="lsj">
    <mads:namePart>Hecataeus Milesius</mads:namePart>
    <mads:namePart type="termsOfAddress">Historicus</mads:namePart>
   </mads:name>
  </mads:variant>
  <mads:variant type="abbreviation">
   <mads:name type="personal" authority="lsj">
    <mads:namePart>Hecat.</mads:namePart>
   </mads:name>
  </mads:variant>
  <mads:note type="source">Heidel, W. A. Hecataeus and the Egyptian priests in Herodotus, Book II, 1987: CIP t.p. (Hecataeus)</mads:note>
  <mads:note type="source">LC manual auth. cd. (hdg.: Hecataeus, of Miletus)</mads:note>
  <mads:note type="source">Collier's ency., c1973 (Hecataeus of Miletus; Grk. hist. &amp; geographer; b. at Miletus in 6th cent.,
B.C.)</mads:note>
  <mads:note type="source">Ecateo da Mileto ed i frammenti della Periēgēsis, 1896-1897</mads:note>
  <mads:note type="source">LCCN PA3998.H17 (Hecataeus, of Miletus, logographer, fl. ca. 500 B.C.)</mads:note>
  <mads:note type="source">Oxford comp. to class. lit. (Hecataeus (Hekataios))</mads:note>
  <mads:identifier type="lccn">n  87931007 </mads:identifier>
  <mads:classification>PA3998.H17InU</mads:classification>
  <mads:fieldOfActivity>Historian</mads:fieldOfActivity>
  <mads:fieldOfActivity>Geographer</mads:fieldOfActivity>
  <mads:fieldOfActivity>Logographer</mads:fieldOfActivity>
  <mads:note type="source">Smith's Dictionary of Greek and Roman Biography and Mythology, Vol , 1867, p. 1: </mads:note>
  <mads:note type="source">Brill's New Pauly:  "H. of Miletus Greek author, 6th cent. BC, on of Hegesander, c. 560-480 BC, first 'logographer'
(using the term - current since Fr. Creutzer, but somewhat incorrect - for the prose authors before Herodotus), from whom numerous fragments
exist (c. 370 in the collection of the FGrH 1). H. is of utmost significance for the development of Greek geography and historiography...." from
Meister, Klaus (Berlin). "Hecataeus." Brill's New Pauly. Antiquity volumes edited by: Hubert Cancik and ; Helmuth Schneider . Brill, 2011. Brill
Online. Tufts University Library. 22 September 2011</mads:note>
<mads:url displayLabel="Wikipedia">http://en.wikipedia.org/wiki/Hecataeus_of_Miletus</mads:url>
<mads:url displayLabel="Worldcat Identities">http://worldcat.org/identities/lccn-n87-931007</mads:url>
<mads:url displayLabel="Smith's Dictionary">http://www.perseus.tufts.edu/hopper/text?doc=Perseus:text:1999.04.0104:entry=hecataeus-bio-
2</mads:url>
<mads:url displayLabel="VIAF">http://viaf.org/viaf/60616393</mads:url>
  <mads:identifier type="tlg">0538</mads:identifier>
  <mads:extension>
   <mads:description>List of related work identifiers</mads:description>
   <identifier type="tlg">0538.002</identifier>
  </mads:extension>
  <mads:recordInfo>
   <mads:recordOrigin>Converted from MARCXML to MADS version 2.0 using MARC21slim2MODS-2-0.xsl (Revision
2.07)</mads:recordOrigin>
   <mads:recordContentSource authority="marcorg">DLC</mads:recordContentSource>
   <mads:recordChangeDate encoding="iso8601">20040518052251.0</mads:recordChangeDate>
   <mads:recordIdentifier source="DLC">n87931007</mads:recordIdentifier>
   <mads:descriptionStandard>aacr2</mads:descriptionStandard>
  </mads:recordInfo>
</mads:mads>
```

Figure 2: MADS Record for Hecataeus of Miletus

The Catalog's Initial Goals and Changing Goals

In 2008 (Babeu 2008), I created a list of initial goals for the catalog data and what we hoped to enable by
creating these records, which I have slightly altered here:

- Supporting greater interoperability of traditional library and classical domain data through the
  use of standard author identifiers/names and uniform names for works.
- Allow "FRBRized" browsing/searching through various multi-author/multi-work volumes by
  using standard work identifiers to identify the works contained therein.
- Provide greater intellectual access to large digitized volumes of rare and standard classical
  works through analytical cataloging, which will support more granular identification and

searching of digital resources.

- Link catalog records to multiple mass digitization projects and investigate the idea of creating customized collections within million book digital libraries.
- Use existing library standards such as MODS and MADS to promote metadata interoperability.
- Repurpose and enhance already existing catalog and authority data to the largest extent possible and illustrate the importance of making both catalog and bibliographic data both human and machine-readable.
- Make all catalog and authority files available as XML to promote reuse of the data.

While all of these goals are still important, one additional goal is to not only make all catalog and authority files available as XML but to also explore the potential of this data as *linked data* within the large and growing world of digital classical sources online that have started to make their data available such as Pleiades[8], Nomisma.org[9] and PELAGIOS.[10]

In addition, another major goal has been to develop an online environment for browsing and searching this data. To this end, an initial prototype browsing environment for the Perseus FRBR Catalog[11] has been developed by Bridget Almas, and makes use of the open source eXtensible Catalog[12] software, particularly its Metadata Services and Drupal Toolkit, which as described on the website "enables the XC user interface to present FRBRized, faceted navigation across a range of library resources." One significant barrier in working with the eXtensible catalog is that its Metadata Services Toolkit works only with MARC and Dublin Core metadata, but since all of the code is open source, Almas was able to work with the code and converted the MODS records within the Perseus FRBR Catalog into MARCXML for import into the eXtensible catalog environment.[13] Research is ongoing into the use of other potential open source tools and database environments (e.g. native XML databases such as Exist) in order to determine whether the eXtensible catalog presents the best long-term searching and browsing environment for the MODS records. Furthermore, as the current catalog implementation does not make use of the MADS authority records, research is also ongoing to determine how best to utilize both these sources together to provide access to the Perseus FRBR catalog data.

Current Status of Cataloging Work

As of September 2012, the current catalog includes over 1300 authors and over three thousand works. For many major authors and works (e.g. Cicero, Thucydides, Homer) there are MODS records for multiple

---

8 http://pleaides.stoa.org

9 http://nomisma.org

10 http://pelagios-project.blogspot.com/p/about-pelagios.html

11 http://catalog.perseus.tufts.edu/perseus.org/

12 http://www.extensiblecatalog.org/

13 Further documentation on this process is currently being written (Any additional information you can provide on the whole import process would be most helpful Bridget!)

editions and translations. The ultimate goal is for the Perseus FRBR catalog to include MODS records for all works included in the Perseus Digital Library and to include at least one cataloged representation of every surviving Latin and Greek author from the classical era, utilizing the collection development list created from the standard editions lists of LS, LSJ and the OLD.

One challenge moving forward, and one that may largely be due to a "mistaken" decision of mine in terms of analytical cataloging and through my desire to quickly navigate author/work files within Oxygen, has been the large creation of "duplicate" expression level records for large composite works. I have always created individual *expression* level records for the *individual* works found within large composite works, largely because it was a way for me to keep track of all the individual authors I had cataloged (especially for adding to the spreadsheets for collection management) and a quick way to gather up all the individual records for an author within an individual author folder. In retrospect this was probably not the best solution (this method is particularly problematic for automatically generated text inventories). I had originally tried to explain this problem in my earlier paper:

Since the analytical cataloging process is not automated and is very time consuming, some difficult decisions have to be made as regards to at what level to stop cataloging or creating records. Currently, we have created single XML manifestation level MODS records for large volumes such as the *Greek Anthology* and *Historicorum Romanorum* (that contain all of the component records for the individual works) and then created separate expression level records for each individual work that is linked to its manifestation. Similarly, for those manifestations that contain works by two or more *authors*, we have also created individual linked expression level records. Without these expression level records we would not know the full range of authors available within the collection. Nonetheless, the collection currently contains many volumes with *multiple works* by an *individual author* (such as Cicero) where *only the manifestation level* XML MODS records exists, but where component XML records for each individual work are nested within the one XML tree. Each of these component records, however, does have a *unique work identifier*, which will perhaps enable expression level records to be created algorithmically in the future. A related issue is that currently the component catalog records within some of these manifestation records (for example, manifestation records for the Loeb editions) list both the English and Latin or English and Greek expressions of a work in a single component record. In other words, these volumes do not yet have separate *expression level records* for each *language* ….. The current plan is to use work identifiers as way of pulling expressions level records out of all of these volumes.

Bridget Almas has worked on helping solve this problem algorithmically within the current eXtensible Catalog implementation, but this may create a potential problem of a large number of duplicate records. Duplicate expression records for the same *single expression* may be found for different authors and works. For example, someone searching/browsing the catalog may potentially find two records for each individual work (e.g. for an epigrammatist in the *Greek Anthology*) that is found within a larger component work MODS record. While the catalog currently only seems to find expression level records created as *separate* XML files within individual author name directories, all works found within larger component works may have records created in the long term if the system "indexes" larger component works in order to provide expression level access to all the individual work constituent records found within them. This will be required for many component works as described above that have not had individual expression level records created for the works within them, both volumes that only include works for individual authors as well as a number of large component volumes that have been cataloged but have not yet had individual expression level records created for the constituent works.[14]

A random thought:
One overriding desire I have is to delete the FRBR catalog structure as it currently exists in the CVS and create a new "clean" directory that is easier to work with where I can perhaps just move all the important MODS records (English Bible, Greek Library, Latin Library, MARCXML-MADS authority records, Reference Materials), perhaps keeping the schemas, spreadsheets, tasklists, templates, and XSLT sheets in

---

[14] As of this writing it does not appear that the eXtensible Catalog has created duplicate records due to this strategy, but in terms of the possible automatic creation of constituent records for large works in the future to speed cataloging a solution will need to be determined.

a separate location.

Some Desired Enhancements to the MODS and MADS records

While there are a number of ways in which the MODS and MADS records could be greatly improved, the following list reflects the most pressing improvements:

1) Convert all records to MODS 3.4 Schema in order to support new linked data features of this latest MODS schema.

2) After converting all records to MODS 3.4, explore the linked data opportunities of this latest version of the MODS schema:

- Convert author and editor names within each MODS record into a linked data representation from their current links to either unstable NAF HTML splash pages available through an OCLC web service or to non-permanent URLs within the VIAF.

  Current:
  ```
  <name authority="naf" type="personal" xmlns:ns2="http://www.w3.org/1999/xlink"
  ns2:href="http://errol.oclc.org/laf/n79-32166.html">
  <namePart>Cicero, Marcus Tullius.</namePart>
  <role>
     <roleTerm authority="marcrelator" type="text">creator</roleTerm>
  </role>
  </name>
  ```

  Desired:
  ```
  <name type="personal" authorityURI="http://id.loc.gov/authorities/names/" valueURI="http://id.loc.gov/authorities/names/n79032166">
  <namePart>Cicero, Marcus Tullius.</namePart>
  <role>
     <roleTerm authority="marcrelator" type="text">creator</roleTerm>
  </role>
   </name>
  ```

- Explore the possibilities of converting place names within MODS records into linked data
  There are two place names typically encoded within each MODS record, both under <originInfo> and are used to indicate where a manifestation was published. The potential subelement that could be used is <place> and its subelement <placeTerm>. There are two ways of encoding place within MODS, either to use the marccountry code or to use a generic textual place term. Most MODS records include both:

  Current:
  ```
  <place>
          <placeTerm type="code" authority="marccountry">enk</placeTerm>
       </place>
       <place>
         <placeTerm type="text">London</placeTerm>
       </place>
  ```

  Desired:
  ```
  <place>
          <placeTerm authority="marccountry"  type="code" authorityURI="http://id.loc.gov/vocabulary/countries"
            valueURI="http://id.loc.gov/vocabulary/countries/enk">enk</placeTerm>
       </place>
       <place>
  ```

```
            <placeTerm authorityURI="http://dbpedia.org" valueURI="http://dbpedia.org/page/London"
         type="text">London</placeTerm>
       </place>
```

(For the textual place term, it is an open question what source to use for URIs such as DBPedia or Geonames?)

All of the Library of Congress vocabularies including subject headings (LCSH), MARC countries, etc. are now available as linked data and this may present some interesting opportunities for visualization, etc.

3) Enhancing MADS authority records
- Many MADS records haven't been "enhanced", they have not had work identifiers added, variant names entered, or appropriate links to online URIs added (e.g. VIAF, ID.gov (LCNAF),) or to other resources (Wikipedia)
- Linking MADS authority records to the new URIs within Perseus for Smith's Dictionary (any other possible places to link as well)
- Upgrading the URLS for VIAF (currently have links to non-permanent URLs)--

Cataloging Priorities/Desires

The ultimate goal of the last few years has been to have MODS records for every work/expression in Perseus as well as at least one cataloged expression (with links to online manifestations in OCA, Google Books, Hathi Trust, etc.) for every author (fragmentary or surviving) that is found within the Authors-Abbreviations-Edition spreadsheet (essentially the LSJ-LS-OLD-TLG editions checklist). The following list outlines my priorities for the next year ideally:

1) The most important task is to create individual MODS records for all of the different classical works within Perseus (about half way there).

2) I would like ideally to finish as well as start cataloging some important fragmentary author and large multi-author component series:
- *Fragmenta Historicorum Graecorum* (Vols 4 and 5 need to be finished)
- Other ones that are top priority are the CAG (*Commentaria in Aristotelem Graeca*) and the *Die Fragmente der Vorsokratiker*

3) Some large composite works have been cataloged but have no expression level component records (*Grammatici Romanae Fragmenta*, *Epistolographi Graeci*, *FHG* II and *FHG* III)

4) A large number of authority records have not yet been downloaded/created for various authors.

5) I would like to add the full list of PHI authors to our main collection spreadsheet (Authors-Abbreviations-Editions) since we have many but not all represented, as well as records for the PHI works to the catalog. All of the PHI editions are now freely available online[15] and MARC records are available for the works.[16]

---

15 http://latin.packhum.org/index

16 http://www.uflib.ufl.edu/catmet/documents/Classical-Latin-Texts-records.zip--While there are MARC records available for all the works for download as a batch, they are designed to be manipulated in UNIX to import into a library catalog and I was unable to determine how to get at individual MARC records that could then ideally be converted into MARCXML and then MODS.

Some Tasks To Automate/Partial Automation?

1) Automating creation of component/constituent records for the individual works found within larger books (such as for *FHG* II, *Grammatici Romanae*, etc.) that have already been cataloged. The larger question is whether creating individual expression level records at this point will create a large amount of redundancy in searching and browsing the catalog. I am uncertain to what level the current eXtensible catalog implementation is indexing and providing access to the individual constituent works (more on this below).

2) Automating creation of page links to the Open Content Alliance for constituent works. In many cases the composite MODS record has a base URL to the digitized book in the OCA, but a larger problem with automation is that their formatting for page level links isn't always consistent (especially when a book has extensive front matter) so I'm not sure if this is even possible.

3) Finding already cataloged composite books in the Hathi Trust and creating page level links. This is particularly desirable to me as they are a library digitization project, but again I'm not sure how realistic a goal this might be.

Current Issues with the FRBR Catalog eXtensible Catalog implementation

1) A large number of **authors records** need to be **merged**, due to slight variants in how the name appears (typically this is due to a missing period after an A.D. or a B.C., a missing comma, or some other slight typographical error/variant on my part (e.g. Aelian 3rd cent. and Aelian, 3rd. cent)

2) There seem to be a large number of random MODS **records missing**. In addition, some times the constituent records within larger works seems to have been indexed and other times they have not.
For example, there is a missing edition for Theophrastus *Characteres*
(http://catalog.perseus.tufts.edu/perseus.org/xc_browse/browse_works/17/metadata_type_s%3Aexpression%20dcterms__identifier_cts_t%3A%22urn%3Acts%3AgreekLang%3Atlg0093.tlg009%22?search_type=browse&hide_query=1)

Within the data for the catalog there is a Loeb edition by Edmonds (theophrastus.Loeb(Edmonds-1929).Characters.mods.xml), but this is not represented in the eXtensible Catalog.

3) The **language identification** for expressions seems to be partial and sometimes inaccurate.
Using the same example, above, for Theophrastus *Characteres*, the catalog lists one Greek edition (the Perseus edition) and incorrectly lists two Latin editions (perseus-latX1 and perseus-latX3). Another confusing issue is that both of these CTS-URN identifiers have been generated for the same OCT-Diels edition of Theophrastus Characters that is represented by one MODS record (theophrasti.Oxford.Characteres.xml).

Is this perhaps because the exTensible catalog can't manage two languages in a record (because it is using MARCXML representations? This MODS record has two languages listed:

```
<language objectPart="preface">
    <languageTerm authority="iso639-2b" type="code">lat</languageTerm>
  </language>
  <language objectPart="text">
    <languageTerm authority="iso639-2b" type="code">grc</languageTerm>
  </language>
```

Is there a way to modify the code to make sure that it only bases the language of an expression on a

language element with the attribute <objectPart="text">?

This issue with language identification for expressions seems to be a larger problem with the catalog in general, for example under Theophrastus, *Historia Plantarum* the catalog incorrectly lists three Latin editions based off one MODS record: (http://catalog.perseus.tufts.edu/perseus.org/xc_browse/browse_works/17/metadata_type_s%3Aexpression%20dcterms__identifier_cts_t%3A%22urn%3Acts%3AgreekLang%3Atlg0093.tlg001%22?search_type=browse&hide_query=1).

A related issue is that even though this work is found within one single Teubner volume and has only one MODS record (theophrastus.Teubner.Opera.VolI.xml) it has generated three CTS URNS. The language is also represented the same way in this MODS record as the example above. Another issue is that the list of editions for *Historia Plantarum* for Theophrastus is also missing the **Loeb** edition (theophrastus.Loeb(Hort-1916).EnquiryIntoPlantsAndMinorWorks.Vol1.mods.xml and theophrastus.Loeb(Hort-1916).EnquiryIntoPlantsAndMinorWorks.Vol2.mods.xml), although strangely it has indexed the other works in volume 2 (e.g. De Odoribus, though it again is only indicating that these works are available in **English** not both English and Greek because it is a Loeb edition) (http://catalog.perseus.tufts.edu/perseus.org/xc_browse/browse_works/17/metadata_type_s%3Aexpression%20dcterms__identifier_cts_t%3A%22urn%3Acts%3AgreekLang%3Atlg0093.tlg010x01%22?search_type=browse&hide_query=1)

4) There seems to be unexplained **duplication** (not the duplication issue explained above regarding double indexing). For example, why does the catalog list the OCT edition of Theophrastus twice? Or one single-volume (and single work) Teubner edition as three separate CTS URN works?
Are there lots of hidden duplicate MODS records perhaps hiding in the catalog that I didn't manage to delete?

For example, under Achilles Tatius, there is one Loeb edition, with one work (Leucippe and Clitophon) (and one MODS record) (achilles-tatius.loeb.gaselee.cml), yet the search http://catalog.perseus.tufts.edu/perseus.org/xc/search/xc__contributor_t%3A%28Gaselee%2C%20Stephen%20Sir%201882-1943%20translator%29?ql=1 lists four separate results for this edition:
Three results have CTS English identifiers and one has a Greek CTS URN, but the Greek record then doesn't list any online editions.
(http://catalog.perseus.tufts.edu/perseus.org/content/oaimstscaifedlorgmetadataservicestoolkitmarctoxctransformation36093?hit=2&caller=xc-search)
Why are four separate expressions generated for this one MODS record?

The catalog search for Achilles Tatius lists 10 editions:
http://catalog.perseus.tufts.edu/perseus.org/xc_browse/browse_works/17/metadata_type_s%3Aexpression%20dcterms__identifier_cts_t%3A%22urn%3Acts%3AgreekLang%3Atlg0532.tlg001%22?search_type=browse&hide_query=1

But at the same time the text inventory for Fedora lists only 4 (the Perseus edition plus 3 others), which is correct.

http://services.perseus.tufts.edu/fedora/objects/org.perseus:greekLang.tlg0532.tlg001/datastreams/CTS/content

Current Issues With FRBR Feeds in Dropbox
1) FRBR feeds are not fully capturing the identifier attributes—the displayLabel attribute is being used to define (for lack of a better method) whether the identifier text in the element is for the actual work itself, or whether it is for some type of related work (scholia, commentary, introduction, etc.)

For example within a MODS record for Euclid's *Catoptrica*, for the work itself, the identifier is encoded

as follows:
`<identifier type="tlg">1799.011</identifier>`

But for a scholia on this work by Euclid, that also has a separate TLG identifier itself, the component record contains:
`<identifier type="tlg">5022.005</identifier>`
`<identifier type="tlg" displayLabel="isScholiaTo">1799.011</identifier>`

Currently the FRBR feeds are aggregating by work identifier only.

2) The FRBR feeds as well as the eXtensible catalog also seem to be missing some works in their aggregation (such as **some** of the constituent works but not **all** found within the 3 volume Reimer-Bekker-Aristotle catalog records:
(aristotle.Reimer.Bekker(1831).Opera.Vol1.mods.xml,;
aristotle.Reimer.Bekker(1831).Opera.Vol2.mods.xml;
aristotle.Reimer.Bekker(1831).Opera.Vol3.mods.xml)

It seems like in many cases they are not indexing/including all of the **constituent** works that are only contained within a larger work. For example, the Perseus FRBR feed, for *Mechanica* (tlg0086.tlg023) includes the constituent work from the Bekker edition Volume 2 (but the feeds for two other works that are also include in this same volume Rhetoric (tlg0086.tlg038) (https://dl-web.dropbox.com/get/FRBR.feeds.all.20120605/tlg0086.tlg038.xml?w=7924a5d3
) and Poetics (tlg0086.tlg034) (https://dl-web.dropbox.com/get/FRBR.feeds.all.20120605/tlg0086.tlg034.xml?w=404188ed) do not.

Why is this? Does this have something to do with these works also having Perseus versions? What drives the creation of these feeds?