DHCuration

a community resource guide to data curation in the humanities

Contents

Welcome

Data curation is an emerging problem for the humanities as both data and analytical practices become increasingly digital. Research groups working with cultural content as well as libraries, museums, archives, and other institutions are all in need of new expertise.

This guide is the first stop on your way to mastering the essentials of data curation for the humanities. It offers concise, expert introductions to key topics plus annotated links to the most important standards, documentation, articles, and projects in the field.

Intended to help students and those new to the field, the Community Resource Guide also provides a quick reference for teachers, administrators, and anyone seeking an orientation in the issues and practicalities of data curation.

As indicated by the name, the Community Resource Guide is intended to be a living, participatory document. Readers are encouraged to review and comment on every part of this guide and to suggest additional resources. Contributions from readers are incorporated at intervals to keep the Guide at the cutting edge.

Browse, comment, contribute! The table of contents provides a road map to the Guide's current topics and those to be added soon.

GRADUATE SCHOOL OF **LIBRARY AND INFORMATION SCIENCE** The iSchool at Illinois

INSTITUTE of **Museum** and **Library** SERVICES

CIRSS

DHCuration Guide (beta) main page

Classics, "Digital Classics" and Issues for Data Curation

Alison Babeu, Perseus Project, Tufts University

Introduction

View and add comments »

¶ 1

View and add comments »

The field of classics encompasses a large number of related disciplines such as archaeology, epigraphy, manuscript studies, numismatics, philology and papyrology, each with their own unique set of research methods, objects of study (including various types of artifacts, ancient monuments, ancient documentary and literary texts, coins, inscriptions, papyri, etc.), and various ways of sharing and preserving data. Classical studies is thus an inherently interdisciplinary field that has also long made use of advanced digital technologies (e.g. advanced imaging and document recognition for classical languages such as Sumerian, Ancient Greek and Latin, the 3D reconstruction and visualization of ancient monuments, and the use of TEI-XML to create digital editions of classical texts). The relatively advanced digital nature of many classical disciplines has in turn shaped both the research practices of the field and increased the need for data curation strategies that address the complex needs both of specific disciplinary research methods and specific types of digital data that are created as part of the research process. While the field of classics has long faced the issue of preserving fragile physical artifacts such as damaged manuscripts and centuries old fragments of papyri, it now increasingly faces the challenges of preserving digital objects created to represent these artifacts as well.

¶ 2

View and add comments »

Complicating matters is the fact that for many of the disciplines of classics, preserving the interpretative stages and individual decisions involved in creating a "final" scholarly argument can be as important as preserving the final result of such scholarship, such as a virtual reconstruction or digital edition of an inscription. For example, the digital reconstructions of archaeological monuments typically involve a large amount of uncertainty and individual scholarly interpretation, yet many visualizations are often viewed by students as complete and accurate representations of "reality." Similarly, in creating a digital edition of a classical text (e.g. a play of Aeschylus with many manuscript sources), many individual scholarly decisions are made in terms of what text variants to include or what manuscript witnesses are considered more reliable, yet many digital editions lack the "apparatus criticus" that contains such decisions and can give the illusion of one text. These issues and projects/solutions that have been created to address them will receive further attention below.

¶ 3

View and add comments »

For data curators, the key question to consider is how the research practices of "digital classics" are creating new challenges for data curation, and indeed a number of significant projects across the disciplines are currently working to address some of these challenges. Despite often seemingly huge differences between various disciplines in the field of classics, there are a number of common themes that will require further research and collaboration between classical scholars and those working in data curation.

¶ 4

View and add comments »

To begin with, the difficulties of defining both the complex semantics and structure of classical data that needs to be preserved must be addressed. Data in classics is extensively multi-lingual and multi-script (with many different languages such as Ancient Greek, Latin, Sumerian, Sanskrit, etc.) The same data (e.g. data about the same classical place, person or other named entity, descriptions of the same archaeological object, multiple images of the same inscription) found across different projects is also often described using very different vocabularies. Similarly, more research will need to focus on how meaningful data integration might be used to create larger digital classical resources that could then possibly be more effectively curated. A variety of issues complicate this process, including the fact that multiple digital facsimiles of objects exist in various digital data collections (often with greatly varying levels of metadata). To solve this issue, many projects have chosen to create virtual data centers or utilized a federated approach allowing data to remain distributed and independent. In fact, complete interoperability or full data integration may be impossible to attain and is also not necessarily an ideal solution according to many practitioners.

¶ 5

View and add comments »

As indicated by the projects that will be covered here, the process of data curation and ensuring data sustainability has many components. Many consider the technical components of sustainability to be the easiest task for the long run and stress that long-term financial planning for the organization/s that will host/curate the data is far more essential. Similarly, political considerations of both the needs of data

XML

```
1    <?xml version="1.0" encoding="UTF-8"?>
2
3    <?xml-model href="../bpg_article.rng" type="application/xml" schematypens="http://relaxng.org/ns/structure/1.0"?>
4    <article status="draft" xml:lang="en">
5        <header>
6            <authorInfo xml:id="alisonbabeu">
7                <author_name>
8                    <family>Babeu</family>
9                    <given>Alison</given>
10               </author_name>
11               <affiliation>Perseus Project, Tufts University</affiliation>
12               <email/>
13           </authorInfo>
14           <article_title>Classics, "Digital Classics" and Issues for Data Curation</article_title>
15           <keywords>
16               <keyword/>
17           </keywords>
18           <changes>
19               <change when="2011-06-29" who="PMC"/>
20           </changes>
21       </header>
22       <body>
23           <introduction>
24               <para xml:id="p01">The field of classics encompasses a large number of related
25                   disciplines such as archaeology, epigraphy, manuscript studies, numismatics,
26                   philology and papyrology, each with their own unique set of research methods,
27                   objects of study (including various types of artifacts, ancient monuments, ancient
28                   documentary and literary texts, coins, inscriptions, papyri, etc.), and various ways
29                   of sharing and preserving data. Classical studies is thus an inherently
30                   interdisciplinary field that has also long made use of advanced digital technologies
31                   (e.g. advanced imaging and document recognition for classical languages such as
32                   Sumerian, Ancient Greek and Latin, the 3D reconstruction and visualization of
33                   ancient monuments, and the use of <term ref="glossary.xml#TEI">TEI-XML</term> to
```
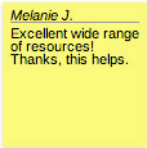
HTML

Classics, "Digital Classics" and Issues for Data Curation

Alison Babeu, Perseus Project, Tufts University
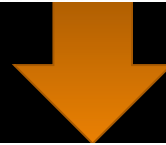
Melanie J.
Excellent wide range of resources! Thanks, this helps.

All article comments »

Introduction

View and add comments »

The field of classics encompasses a large number of related disciplines such as archaeology, epigraphy, manuscript studies, numismatics, philology and papyrology, each with their own unique set of research methods, objects of study (including various types of artifacts, ancient monuments, ancient

¶ 1

View and add comments »

Mr. Phileas Fogg lived, in 1872, at No. 7, Saville Row, Burlington Gardens, the house in which Sheridan died in 1814. He was one of the most noticeable members of the Reform Club, though he seemed always to avoid attracting attention; an enigmatical personage, about whom little was known, except that he was a polished man of the world. People said that he resembled Byron—at least that his head was Byronic; but he was a bearded, tranquil Byron, who might live on a thousand years without growing old.

Plain HTML

MR. PHILEAS FOGG LIVED, IN 1872, AT NO. 7, SAVILLE ROW, BURLINGTON GARDENS, THE HOUSE IN WHICH SHERIDAN DIED IN 1814. HE WAS ONE OF THE MOST NOTICEABLE MEMBERS OF THE REFORM CLUB, THOUGH HE SEEMED ALWAYS TO AVOID ATTRACTING ATTENTION; AN ENIGMATI-CAL PERSONAGE, ABOUT WHOM LITTLE WAS KNOWN, EXCEPT THAT HE WAS A POLISHED MAN OF THE WORLD. PEOPLE SAID THAT HE RESEMBLED BYRON—AT LEAST THAT HIS HEAD WAS BYRONIC; BUT HE WAS A BEARDED, TRAN-QUIL BYRON, WHO MIGHT LIVE ON A THOUSAND YEARS WITHOUT GROWING OLD.

HTML with CSS

XSLT

- eXtensible Stylesheet Language Transformations
- List of rules to transform contents of an XML document into...
  - Another XML document
  - PDF
  - HTML
  - Etc...

XML

<?xml version="1.0" encoding="UTF-8"?>

<?xml-model href="./bpg_article.rng" type="application/xml" schematypens="http://relaxng.org/ns/structure/1.0"?>
<article status="draft"

<?xml version="1.0" encoding="UTF-8"?>

<?xml-model href="./bpg_article.rng" type="application/xml" schematypens="http://relaxng.org/ns/structure/1.0"?>
<article status="draft"

<?xml version="1.0" encoding="UTF-8"?>

<?xml-model href="./bpg_article.rng" type="application/xml" schematypens="http://relaxng.org/ns/structure/1.0"?>
<article status="draft"

<?xml version="1.0" encoding="UTF-8"?>

<?xml-model href="./bpg_article.rng" type="application/xml" schematypens="http://relaxng.org/ns/structure/1.0"?>
<article status="draft"

<?xml version="1.0" encoding="UTF-8"?>

<?xml-model href="./bpg_article.rng" type="application/xml" schematypens="http://relaxng.org/ns/structure/1.0"?>
<article status="draft"

XSLT

<?xml version="1.0" encoding="UTF-8"?>
<xsl:stylesheet xmlns:xsl="http://www.w3.org/1999/XSL/Transform"

xmlns="http://www.w3.org/1999/xhtml"
xmlns:xsd="http://www.

HTML

<!DOCTYPE html><html xmlns="http://www.w3.org/2001/XMLSchema" xmlns="http://www.w3.org/1999/xhtml">
<head>
<meta http-equiv="Content-

<!DOCTYPE html><html xmlns="http://www.w3.org/2001/XMLSchema" xmlns="http://www.w3.org/1999/xhtml">
<head>
<meta http-equiv="Content-

<!DOCTYPE html><html xmlns="http://www.w3.org/2001/XMLSchema" xmlns="http://www.w3.org/1999/xhtml">
<head>
<meta http-equiv="Content-

<!DOCTYPE html><html xmlns="http://www.w3.org/2001/XMLSchema" xmlns="http://www.w3.org/1999/xhtml">
<head>
<meta http-equiv="Content-

<!DOCTYPE html><html xmlns="http://www.w3.org/2001/XMLSchema" xmlns="http://www.w3.org/1999/xhtml">
<head>
<meta http-equiv="Content-

XML

```
1    <?xml version="1.0" encoding="UTF-8"?>
2
3    <?xml-model href="../bpg_article.rng" type="application/xml" schematypens="http://relaxng.org/ns/structure/1.0"?>
4    <article status="draft" xml:lang="en">
5        <header>
6            <authorInfo xml:id="alisonbabeu">
7                <author_name>
8                    <family>Babeu</family>
9                    <given>Alison</given>
10               </author_name>
11               <affiliation>Perseus Project, Tufts University</affiliation>
12               <email/>
13           </authorInfo>
14           <article_title>Classics, "Digital Classics" and Issues for Data Curation</article_title>
15           <keywords>
16
17
18
19
20
21   </h
22   <bo
23
24
25
26
27
28
29
30
31
32
33
```

XSLT

```
295
296
297    <xsl:template name="headerTitleAboveArticle_section">
298        <div class="sectionContainer" id="pageTitle">
299            <div class="sectionContentContainer">
300                <span id="siteTitleForPrint"><strong>DHCuration: </strong>a community resource guide to data curation in the humanities</span>
301                <div class="paraSidebarContainer">
302                    <div class="paraSidebar articleSidebar" id="articleCommentsAboveText">
303                        <ul class="sectionComments" id="ac_top">
304                            <li class="addComment">
305                                <a href="#articleComments" id=" ">all article comments &#187;</a>
306                            </li>
307                        </ul>
308                    </div>
309                </div>
310                <!-- /paraSidebarContainer articleSidebar-->
311                <div class="paraContainer" id="articleInfoAboveTextContainer">
312                    <div class="paraTextContainer" id="articleInfoAboveText">
313                        <span id="articleTitleAboveText">
314                            <xsl:value-of select="article/header/article_title"/>
315                        </span>
316                        <xsl:for-each select="/article/header/authorInfo">
317                            <span class="authorNameAboveText">
318                                <a href="#" title="Contributor page">
319                                    <xsl:value-of select="author_name/given"/>
320                                    <xsl:text> </xsl:text>
321                                    <xsl:value-of select="author_name/family"/>
322                                    <xsl:text>, </xsl:text>
323                                    <xsl:value-of select="affiliation"/>
324                                </a>
325                            </span>
326                            <br/>
327
328    </di
329    <!--
330
331            </div>
332            <!-- /pa
333        </div>
334        <!-- /secti
335    </div>
336    </xsl:template>
```

HTML

```
195    <div id="contentWrapper">
196        <div id="article">
197            <div class="sectionContainer" id="pageTitle">
198                <div class="sectionContentContainer">
199                    <span id="siteTitleForPrint">
200                        <strong>DHCuration: </strong>a community resource guide to data curation in the humanities</span>
201                    <div class="paraSidebarContainer">
202                        <div class="paraSidebar articleSidebar" id="articleCommentsAboveText">
203                            <ul class="sectionComments" id="ac_top">
204                                <li class="comment highlighted"><a href="javascript:;" onClick="window.open('comments.html','myWin','width=400,scrol
205                                    <span class="commentAuthor">Melanie J.</span> Excellent wide range of resources! Thanks, this helps.</a></li>
206                                <li>
207                                    <a href="#articleComments" class="addComment">All article comments &#187;</a>
208                                </li>
209                            </ul>
210                        </div>
211                    </div>
212                    <div class="paraContainer" id="articleInfoAboveTextContainer">
213                        <div class="paraTextContainer" id="articleInfoAboveText">
214                            <span id="articleTitleAboveText">Classics, "Digital Classics" and Issues for Data Curation</span>
215                            <span class="authorNameAboveText">
216                                Alison Babeu, Perseus Project, Tufts University
217                            </span>
218                            <br/>
219                        </div>
220                    </div>
221
```

XML

```
<authorInfo xml:id="alisonbabeu">
    <author_name>
        <family>Babeu</family>
        <given>Alison</given>
    </author_name>
    <affiliation>Perseus Project, Tufts University</affiliation>
    <email/>
</authorInfo>
<article_title>Classics, "Digital Classics" and Issues for Data Curation</article_title>
```

XSLT

```
<span id="articleTitleAboveText">
    <xsl:value-of select="article/header/article_title"/>
</span>
<xsl:for-each select="/article/header/authorInfo">
    <span class="authorNameAboveText">
        <a href="#" title="Contributor page">
            <xsl:value-of select="author_name/given"/>
            <xsl:text> </xsl:text>
            <xsl:value-of select="author_name/family"/>
            <xsl:text>, </xsl:text>
            <xsl:value-of select="affiliation"/>
        </a>
    </span>
```

HTML

```
<div class="paraTextContainer" id="articleInfoAboveText">
    <span id="articleTitleAboveText">Classics, "Digital Classics" and Issues for Data Curation</span>
    <span class="authorNameAboveText">
    Alison Babeu, Perseus Project, Tufts University
    </span>
```

Looks like...

cirssweb2.lis.illinois.edu/hdcbpg/research-practices/classics/

Classics, "Digital Classics" and Issues for Data Curation

Alison Babeu, Perseus Project, Tufts University

Introduction

The field of classics encompasses a large number of related disciplines such as archaeology, epigraphy, manuscript studies,

¶ 1

View and add comments »

An Introduction to Humanities Data Curation

Julia Flanders, Brown University

Trevor Muñoz, University of Maryland

these must be in order on the page

§ 1
What is data curation?

#p001

unique paragraph identifiers

¶ 1

At present, there are a number of competing terms used to describe the activity of managing digital materials for research: digital curation, digital stewardship, data curation, digital archiving. There is overlap among these definitions or visions. The variation that does exist is due to more than confusion or carelessness. Each of these terms has significant connotations and attempts to align the relatively new activity of caring for digital materials with an older tradition, discipline, or profession.

these comments must stick with...

¶ 2

#p002

The term and the idea of "curation" is experiencing a moment of cultural cachet that simultaneously threatens to empty it of meaning. Nonetheless, curation is a strong and suitably-encompassing term for the practices we want to describe. When we speak of "curation", what set of activities do we have in mind? Etymological guidance from the word's root meaning and early usage gives us the idea of curation as healing, and of the curator as one who "has the care or charge of a person

this text!

"guardian" (OED). As the word has emerged in the contexts of museums and also rare book and manuscript librarianship

Susan Bridle
I think this is indeed a distinct risk and I wonder whether even the recourse to etymology here is sufficient to pin down the word...

Jack Yi
The problem here is surely the "dual emphasis" this term carries. Isn't there a fundamental tension between "protection"...

P.M. Monroy
I have to dissent slightly from the previous commenter: it's useful (practically as well as strategically)

DHCuration Guide modeling choices

- XML > XSLT > HTML+CSS for articles
  - One XML schema and one XSLT stylesheet to rule them all (consistency, convenience)

- Just HTML+CSS for back matter (FAQ, Glossary)