

# PAL: Toward a Recommendation System for Manuscripts

Scott Ziegler and  
Richard Shrake

---

## ABSTRACT

*Book-recommendation systems are increasingly common, from Amazon to public library interfaces. However, for archives and special collections, such automated assistance has been rare. This is partly due to the complexity of descriptions (finding aids describing whole collections) and partly due to the complexity of the collections themselves (what is this collection about and how is it related to another collection?).*

*The American Philosophical Society Library is using circulation data collected through the collection-management software package, Aeon, to automate recommendations. In our system, which we're calling PAL (People Also Liked), recommendations are offered in two ways: based on interests ("You're interested in X, other people interested in X looked at these collections") and on specific requests ("You've looked at Y, other people who looked at Y also looked that these collections").*

*This article will discuss the development of PAL and plans for the system. We will also discuss ongoing concerns and issues, how patron privacy is protected, and the possibility of generalizing beyond any specific software solution.*

## INTRODUCTION

The American Philosophical Society Library (APS) is an independent research library in Philadelphia. Founded in 1743, the library houses a wide variety of material in early American history, history of science, and Native American linguistics. The majority of the library's holdings are manuscripts, with a large amount of audio material, maps, and graphics, nearly all of which are described in finding aids created using Encoded Archival Description (EAD) standards.

Like similar institutions, the APS has long struggled to find new ways to help library users discover material relevant to their research. In addition to traditional in-person, email, and phone reference, the APS has spent years creating search and browse interfaces, subject guides, and web exhibitions to promote the collections.<sup>1</sup>

As part of these ongoing efforts to connect users with collections, the APS is working on an automated recommendation system to reuse circulation data gathered through Aeon. Developed by Atlas Systems, Aeon is a "request and workflow management software specifically designed for special collections libraries and archives," and it enables the APS to gather statistics on both the use of our manuscript collections and on aspects of the library's users.<sup>2</sup> The automated recommendation system, which we're calling PAL, for "People Also Liked," is an ongoing effort. This article presents a snapshot of current work.

---

**Scott Ziegler** (sziegler1@lsu.edu) is Head of Digital Programs and Services, Louisiana State University Libraries. Prior to this position, Ziegler was the Head of Digital Scholarship and Technology, American Philosophical Society. **Richard Shrake** (shraker13@gmail.com) is a Library Technology Consultant based in Burlington, Vermont.



---

## LITERATURE REVIEW

The benefits of recommendations in library OPACs has long been recognized. Writing in 2008 about the library recommendation system BibTip, itself started in the early 2000s, Mönnich and Spiering observe that “library services are well suited for the adoption of recommendation systems, especially services that support the user in search of literature in the catalog.” By 2011 OCLC Research and the Information School at the University of Sheffield began exploring a recommendation system for OCLC’s Worldcat.<sup>3</sup>

Recommendations for library OPACs commonly fall into one of two categories, content-based or collaborative filtering. Content-based recommendations pair specific users to library items based on the metadata of the item and what is known about the user. For example, if a user indicates in some way that they enjoy mystery novels, items identified as mystery novels might be recommended to them. Collaborative filtering combines users in some way and creates recommendations for one user based on the preferences of another user.

There can be a dark side to recommendations. The algorithms that determine which users are similar and thus which recommendations to make are not often understood. Writing about algorithms in library discovery systems broadly, Reidsma points out that “in librarianship over the past few decades, the profession has had to grapple with the perception that computers are better at finding relevant information than people.”<sup>4</sup> The algorithms that are doing the finding, however, often carry the same hidden biases that their programmers have. Reidsma encourages a broader understanding of algorithms in general and deeper understanding of recommendation algorithms in particular.

The history of recommendation systems in libraries has informed the ongoing development of PAL. We use both the content-based and the collaborative filtering approach to offering recommendations to users. For the purposes of communicating them to nontechnical patrons, we refer to them as “interest-based” and “request-based,” respectively. Furthermore, we are cautious about the role algorithms play in determining which recommendations users see. Our help text reinforces the continued importance of working directly with in-house experts, and we promote PAL as one tool among the many offered by the library.

We are not aware of any literature on the development of recommendation tools for archives or special-collections libraries. The nature of the material held in these institutions presents special challenges. For example, unlike book collections, many manuscript and archival collections are described in aggregate: one description might refer to many letters. These issues are discussed in detail below.

### **PUTTING DATA TO USE: RECOMMENDATIONS BASED ON INTERESTS AND REQUESTS**

The use of Aeon allows the APS to gather and store data, including both data that users supply through the registration form and data concerning which collections are requested. PAL use both types of data to create recommendations.

#### ***Interest-Based Recommendations***

The first type of recommendation uses self-identified research interest data that researchers supply when creating an Aeon account. When registering, a user has the option to select from a list of sixty-four topics grouped into seven broad categories (figure 1). The APS selected these

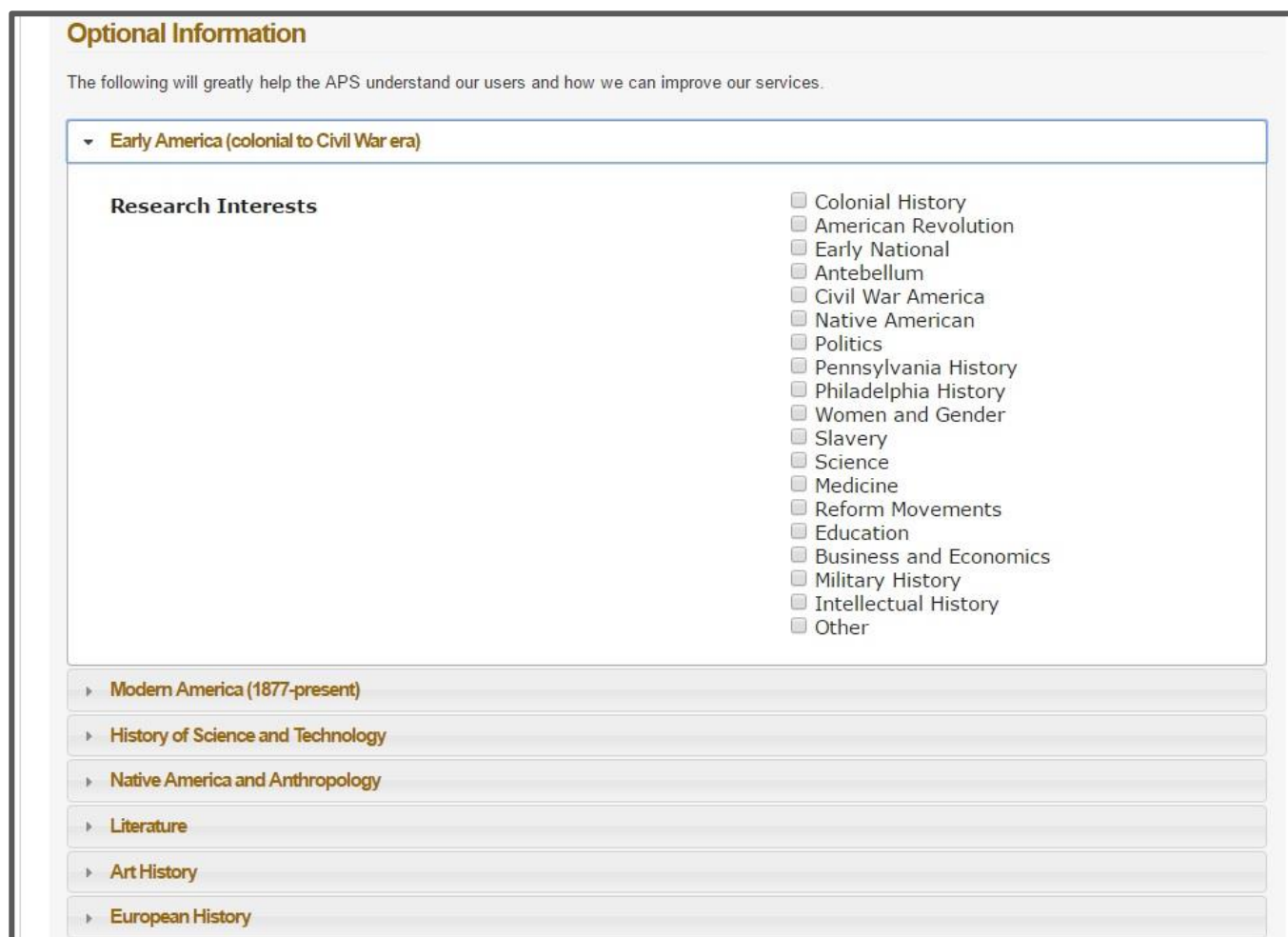
---

interests based on suggestions from researchers as well as categories common in the field of academic history.

Upon signing in, a registered user sees a list of links (figure 2); each link leads to a full-page view of collection recommendations (figure 3). These recommendations follow the model, “You’re interested in X, other people interested in X looked at these collections.”

### ***Request-Based Recommendations***

Using the circulation data that Aeon collects, we are able to automate recommendations in PAL based on request information. Upon clicking a request link in a finding aid, the user is presented with a list of recommendations on the sidebar in Aeon (figure 4). Each link opens the finding aid for the collection listed.



**Figure 1.** List of interests a user sees when registering for the first time. A user can also revisit this list to modify their choices at any point by following links through the Aeon interface. The selected interests generate recommendations.

**Aeon**

Choose an option from the choices below.

Active All Search

Outstanding Requests

TN	Title	Author/Creator	Volume/Box	Status	Order Status
73825	No. 2179 George Washington to Nathanael Greene			Awaiting Future Request Processing	

© 2015 Atlas Systems, Inc. All Rights Reserved.

What is this?

- View collections requested by people who expressed interest in Modern America-Philadelphia History
- View collections requested by people who expressed interest in History of Science and Technology-20th Century
- View collections requested by people who expressed interest in Native America and Anthropology-History of Anthropology
- View collections requested by people who expressed interest in Native America and Anthropology-Indian Policy
- View collections requested by people who expressed interest in Art History

**Figure 2.** List of links appearing on the right-hand sidebar, based on interests that users select.

**Aeon | APS Request Management**

Active All Search

What is this?

The collections most frequently requested from researchers who expressed an interest in Modern America-Philadelphia History are listed below with links to each collection's finding aid and the number of times each collection has been requested.

COLLECTION	CALL NUMBER	# OF REQUESTS	# OF USERS
<a href="#">The Sol Feinstone Collection of the American Revolution, ca. 1760s-1850s</a>	Mss B.F327	13	1
<a href="#">Peale-Sellers Family Collection, 1686-1963: Series 1: Correspondence</a>	Mss B.P31	5	1
<a href="#">Emma Weigley Collection of Sarah Tyson Rorer Papers</a>	Mss.Ms.Coll.136	5	1
<a href="#">Emma Weigley Collection of Sarah Tyson Rorer Papers: Publications</a>	Mss.Ms.Coll.136	3	1
<a href="#">Peale-Sellers Family Collection, 1686-1963</a>	Mss B.P31	1	1
<a href="#">Frans M. Olbrechts papers: Series I: General Haudenosaunee</a>	Mss.497.3.OL2	1	1
<a href="#">John Frederick Lewis papers, 1878-1932</a>	Mss B.L585	1	1

**Figure 3.** Recommended collections, based on interest, showing collection name (with a link to finding aid), call number, number of requests, and number of users who have requested from the collections. The user sees this list after clicking on option from sidebar, as shown in figure 2.

**Aeon | APS Request Management**

Active All Search

**New Manuscript/Archives Request!** \* Indicates required field

Enter information below and press the Submit Request button to send.

**Describe the item you want**

\*Title: ALS to John Bayard, 1738-1807  
Please do not abbreviate unless your citation is abbreviated

Author: Paine, Thomas, 1737-1809

Date(s): 1777 October 30

Collection Name/Accession No.: Colonel Richard Gimbel Collection of Thomas Paine Papers: Series I. Paine Material

Call Number: Mss. B. P165

Box(es)/Folder(s): Box 1

Request for: Shrake, Ph (prshrake)

Notes: 2p.  
Enter any additional information to help us identify the specific item(s) you are requesting.

End of Request?  
To facilitate efficient paging, select "Finished" when you submit the last item in your request.

Unfinished  
 Finished

Keep In Review   
Keep request in a review state before submitting to processing.

**Recommendations**  
Researchers who requested items from the collection Mss.B.P165 also requested materials from the following collections:

- Mss.Ms.Coll.200 - Miscellaneous Manuscripts Collection, 1668-1983
- APS.Archives - American Philosophical Society Archives
- Mss.B.P31 - Peale-Sellers Family Collection, 1686-1963
- Mss.B.V46p - Benjamin Vaughan Papers
- Mss.B.B284d - Violetta Delafield-Benjamin Smith Barton Collection
- Mss.Ms.Coll.52 - Wyck Association Collection
- Mss.B.V462 - John Vaughan papers, 1768 - Circa 1936
- APS.Archives.IIa - American Philosophical Society Archives. Record Group IIa
- Mss.B.F327 - The Sol Feinstone Collection of the American Revolution, ca. 1760s-1850s
- Mss.B.F85.misc - Miscellaneous Benjamin Franklin Collections
- Mss.B.F85inventory12 - Benjamin Franklin Papers Part 12 -- Letters from Franklin
- Mss.B.F85inventory01 - Benjamin Franklin Papers Part 1 -- Letters to Franklin

**Figure 4.** Request-based recommendation links appearing on the right-hand sidebar after a patron requests an item from a finding aid.

## THE PROCESS

Currently, the data that drives these two functions is obtained from a semidynamic process via daily, automated SQL query exports. Usernames are employed to tie together requests and interests but are subsequently purged from the data before the results are presented to users and staff. This section explains the process in detail and presents code snippets where available. All code is available on GitHub.<sup>5</sup>

### *Interest-Based Recommendations*

For interest-based recommendations, we employ two queries. The first query pulls every collection requested by a user for each topic for which that user has expressed an interest. The second aggregates the data for every user in the system. The following queries get data from the Microsoft SQL database, via a Microsoft Access intermediary, that Aeon uses to store data. Because of the number of interest options in the registration form, and the character length of some of them (“Early America - Colonial History,” for example) we encode the interests in shortened form. “Early America - Colonial History” becomes “EA-ColHist” so as not to run into character limits in the database. This section explores each of these queries in more detail and provides example code.



The first query gathers research topics for all users who are not staff (user status is 'Researcher'), and where at least one research topic is chosen ('ResearchTopics' is not null). The data is exported into an XML file that we call "aeonMssReg."

```
SELECT AeonData.dbo.Users.ResearchTopics, AeonData.dbo.Transactions.CallNumber,
AeonData.dbo.Transactions.Location

FROM AeonData.dbo.Transactions INNER JOIN AeonData.dbo.Users ON (AeonData.dbo.Users.UserName =
AeonData.dbo.Transactions.Username) AND (AeonData.dbo.Transactions.Username =
AeonData.dbo.Users.UserName)

WHERE (((AeonData.dbo.Users.ResearchTopics) Is Not Null) AND
((AeonData.dbo.Transactions.CallNumber) Like 'mss%' Or (AeonData.dbo.Transactions.CallNumber)
Like 'aps.%') AND ((AeonData.dbo.Users.Status)='Researcher'))

FOR XML RAW ('aeonMssReq'), ROOT ('dataroot'), ELEMENTS;
```

The second query combines all data for all users and exports an XML file 'aeonMssUsers.'

```
SELECT DISTINCT AeonData.dbo.Users.ResearchTopics, AeonData.dbo.Transactions.CallNumber,
AeonData.dbo.Transactions.Location, AeonData.dbo.Transactions.Username

FROM AeonData.dbo.Transactions INNER JOIN AeonData.dbo.Users ON (AeonData.dbo.Users.UserName =
AeonData.dbo.Transactions.Username) AND (AeonData.dbo.Transactions.Username =
AeonData.dbo.Users.UserName)

WHERE (((AeonData.dbo.Users.ResearchTopics) Is Not Null) AND
((AeonData.dbo.Transactions.CallNumber) Like 'mss%' Or (AeonData.dbo.Transactions.CallNumber)
Like 'aps.%') AND ((AeonData.dbo.Users.Status)='Researcher'))

FOR XML RAW ('aeonMssUsers'), ROOT ('dataroot'), ELEMENTS;
```

Each query produces an XML file. These files are parsed using XSL stylesheets into subsets for each research interest. The stylesheets also generate counts of users requesting a collection and number of total requests for a collection by users sharing an interest.

An example is the following stylesheet for the topic "Early America - Colonial History," which pulls from the XML file "aeonMssReg":

```
<xsl:template match="dataroot">
  <xsl:result-document method="xml" href="aeonMssReqParsed.xml">
    <collections>

      <interest name='Early America-Colonial History' code='EA-ColHis'>
        <xsl:for-each select="aeonMssReq">
          <xsl:if test="contains(ResearchTopics, 'EA-ColHis')">
            <xsl:if test="Location != ''">
              <item>
                <callnumber>
                  <xsl:value-of select="CallNumber" />
                </callnumber>
                <location>
                  <xsl:value-of select="Location" />
                </location>
              </item>
            </xsl:if>
          </xsl:if>
        </xsl:for-each>
      </interest>
    </collections>
  </result-document>
</template>
```

---

This process is repeated for each interest.

The data from the query that we modify with XSLT is presented as HTML that we insert into Aeon templates. This HTML includes the collection name (linked to finding aid), call number, number of requests, and number of users in a table. See figure 3 for how this appears to the user. The following shows how XSL is wrapped in HTML.

```
<p>The collections most frequently requested from researchers who expressed an interest in
<b><xsl:value-of select="@name"/></b> are listed below with links to each collection's finding
aid and the number of times each collection has been requested.</p>
<table class="rec-table">
  <tr>
    <th>Collection</th>
    <th>Call Number</th>
    <th># of Requests</th>
    <th># of Users</th>
  </tr>
  <xsl:for-each select="record">
    <tr>
      <td>
        <a href="http://www.amphilsoc.org/collections/view?docId=ead/{callnumber}-ead.xml"
          target="new">
          <xsl:value-of select="collection"/>
        </a>
      </td>
      <td>
        <xsl:value-of select="callnumber"/>
      </td>
      <td>
        <xsl:value-of select="requests"/>
      </td>
      <td>
        <xsl:value-of select="users"/>
      </td>
    </tr>
  </xsl:for-each>
</table>
```

To ensure a user only sees the links that match the interests they have selected, we use JavaScript to determine the expressed interests of the current user and display the corresponding links to the HTML pages in a sidebar. This approach works well, but we must account for two quirks. The first is that many interests in the database do not conform to the current list of options because many users predate our current registration form and wrote in free-form interests. Secondly, Aeon stores the research information as an array rather than in a separate table, so we must account for the fact that the Aeon database contains an array of values that includes both controlled and uncontrolled vocabulary.

First, we set the array as a variable so we can look for a value that matches our controlled vocabulary and separate the array into individual values for manipulation:

```
// Use var message to check for presence of controlled list of topics
var message = "<#USER field='ResearchTopics'>";
// Use var values to separate topics that are collected in one string
var values = "<#USER field='ResearchTopics'>".split(",");
```



---

We also create variables to generate the HTML entries and links out when we have extracted our research topics:

```
var open = "<ul id='navrec'><li class='active'><a href= '?Action=10&Form=20&Value=topic";  
var middle = "'>Because You Are Interested In <strong>";  
var close = "</strong></a></li></ul>"
```

Next we set a conditional to determine if one of our controlled vocabulary terms appears in the array:

```
//Determine if user has an interest topic from the controlled list  
if ((message.indexOf("EA-ColHis") > -1) || (message.indexOf("EA-AmRev") > -1) ||  
(message.indexOf("EA-EarlyNat") > -1) || (message.indexOf("EA-Antebellum") > -1) ||
```

...

If the array contains a value from our controlled vocabulary, we generate a link and translate our internal code back into a human-friendly research topic (“EA-ColHist,” for example, becomes once again “Early American - Colonial History”):

```
for (var i = 0; i < values.length; ++i) {  
  if (values[i]=="EA-ColHis"){  
    document.getElementById("topic").innerHTML += (open + values[i] + middle + "Early  
    America-Colonial History" + close);}  
  else if (values[i]=="EA-AmRev"){  
    document.getElementById("topic").innerHTML += (open + values[i] + middle + "Early America-  
    American Revolution" + close);}  
  else if (values[i]=="EA-EarlyNat"){  
    document.getElementById("topic").innerHTML += (open + values[i] + middle + "Early America-  
    Early National" + close);} else if (values[i]=="EA-Antebellum"){  
    document.getElementById("topic").innerHTML += (open + values[i] + middle + "Early America-  
    Antebellum" + close);}  
}
```

...

See figure 2 for how this appears to the user. Users only see the links that correspond to their stated interest.

If the array does not contain a value from our controlled vocabulary, we display the research-topic interests associated with the user account, note that we don’t currently have a recommendation, and provide a link to update the research topics for the account.

```
Else  
{document.getElementById("notopic").innerHTML = "<p>You expressed interest in: <div  
class='declared-interests'><#USER field='ResearchTopics'></div></p><p>We are unable to provide  
a specific collection recommendation for you. Please visit <a  
href='https://aeon.amphilsoc.org/aeon.dll?Action=10&Form=81'> our User Profile</a> page to  
select from our list of research topics.</p>"  
}
```

### ***Request-Based Recommendations***

In addition to interest-based recommendations, PAL supplies recommendations based on past requests a user has made. This section details how these recommendations are generated.

Aeon allows users to request materials directly from a finding aid (see figure 6). To generate our request-based recommendations we employ a query depicting the call number and user of every request in the system and export the results to an XML file called “aeonLikeCollections.”



```

SELECT subquery.CallNumber, subquery.Username,
IIf(Right(subquery.trimLocation,1)='.',Left(subquery.trimLocation,Len(subquery.trimLocation)-
1),subquery.trimLocation) AS finallocation
FROM (
SELECT DISTINCT AeonData.dbo.Transactions.CallNumber, AeonData.dbo.Transactions.Username,
IIf(CHARINDEX(':',[Location])>0,Left([Location],CHARINDEX(':',[Location])-1),[Location]) AS
trimLocation
FROM AeonData.dbo.Transactions INNER JOIN AeonData.dbo.Users ON (AeonData.dbo.Users.UserName =
AeonData.dbo.Transactions.Username) AND (AeonData.dbo.Transactions.Username =
AeonData.dbo.Users.UserName)
WHERE (((AeonData.dbo.Transactions.CallNumber) Like 'mss%' Or
(AeonData.dbo.Transactions.CallNumber) Like 'aps.%') AND ((AeonData.dbo.Transactions.Location)
Is Not Null) AND ((AeonData.dbo.Users.Status)='Researcher')) subquery
ORDER BY subquery.CallNumber
FOR XML RAW ('aeonLikeCollections'), ROOT ('dataroot'), ELEMENTS;

```

We then process the “aeonLikeCollections” file through a series of XSLT stylesheets, creating lists of every other collection that every user of the current collection has requested.

First the stylesheets remove collections that have only been requested once. Then we count the number of times each collection has been requested:

```

<collections>
  <xsl:for-each select="aeonLikeCollections">
    <xsl:variable name="variable">
      <xsl:value-of select="CallNumber"/>
    </xsl:variable>
    <xsl:if test="((CallNumber = following-sibling::*[1]/CallNumber) or (CallNumber =
preceding-sibling::*[1]/CallNumber))">
      <aeonLikeCollections>
        <CallNumber>
          <xsl:value-of select="CallNumber"/>
        </CallNumber>
        <Username>
          <xsl:value-of select="Username"/>
        </Username>
        <finallocation>
          <xsl:value-of select="finallocation"/>
        </finallocation>
        <count>
          <xsl:value-of select="count(key('callnumbercount', $variable))"/>
        </count>
      </aeonLikeCollections>
    </xsl:if>
  </xsl:for-each>
</collections>

```

We sort on the collection name and username and then re-sort to combine groups of requested collections with users who have requested each collection.

```

<collections>
  <xsl:for-each-group select="record" group-by="original">
    <group name="{current-grouping-key()}">
      <xsl:for-each select="current-group()/callnumber">
        <xsl:copy-of select="./following-sibling::*[1]"/>
        <xsl:copy-of select="."/>
        <xsl:copy-of select="./following-sibling::*[2]"/>
      </xsl:for-each>
    </group>
  </xsl:for-each-group>
</collections>

```



```

    </xsl:for-each-group>
</collections>

```

We then create a new XML file that is organized by our collection groupings. The following snippet shows a populated XML file generated by the XSLT stylesheet above.

```

<?xml version="1.0" encoding="UTF-8"?>
<collections>
  <group name="APS.Archives">
    <url>Mss.497.3.B63c</url>
    <callnumber>Mss.497.3.B63c - American Council of Learned Societies ... </callnumber>
    <count>94</count>
    <url>Mss.Ms.Coll.200</url>
    <callnumber>Mss.Ms.Coll.200 - Miscellaneous Manuscripts Collection ... </callnumber>
    <count>92</count>
  </group>

```

We use JavaScript to determine the call number of the user's current request and display the list of other collections that users who have requested the current collection have also requested.

```

<script>
  $(document).ready(function() {
    $.ajax({
      type: 'GET',
      url: 'aeonLike_f_Final.xml',
      dataType: 'xml',
      success: function(xml) {
        $(xml).find('group[name="<#PARAM name='CallNumber'>"]').each(function() {
          $('<.recintro').replaceWith('Researchers who requested items from the collection
          <b><#PARAM name='CallNumber'></b> also requested materials from the following
          collections:');
          $('<.read-more').append('See more');
          $(this).find('url').each(function() {
            $('#collections').append(
              '<li><a href="http://amphilsoc.org/collections/view?docId=ead/' + $(this).text()
              + '-ead.xml" target="_blank">' + $(this).next().text() + '</a> '
            );
          });
        });
      });
    });
  });
</script>

```

See figure 4 for how these links appear to the user.

All of the exports and processing are handled automatically through a daily scheduled task. The only personally identifiable data that is contained in these processes are usernames, which are used for counting purposes, but they are removed from the final products through the XSLT processing on an internal administrative server, are never stored in the Aeon web directory, and are never available for other library users or staff to see.

## POTENTIAL PITFALLS AND WHAT TO DO ABOUT THEM

PAL allows us to see new things about our users, and we hope that our users are able to see new collections in the library. However, there are potential pitfalls to the way we've been working on this project. We're calling the two biggest pitfalls the "bias toward well-described collections" and the "problem of aboutness."

### The Bias toward Well-Described Collections

The bias toward well-described collections is best understood by examining how the APS integrates Aeon into our finding aids. We offer request links at every available level of description: collection, series, folder, and item.

If a patron spends all day in our reading room and looks at the entirety of an item-level collection, they could have made between twenty and one hundred individual requests from that collection. For our statistics, each request will be counted as that collection being used. Figure 6 shows a collection described at the item level; each item can be individually requested, giving the impression that this collection is very heavily used even if it is only one patron doing all the requesting.

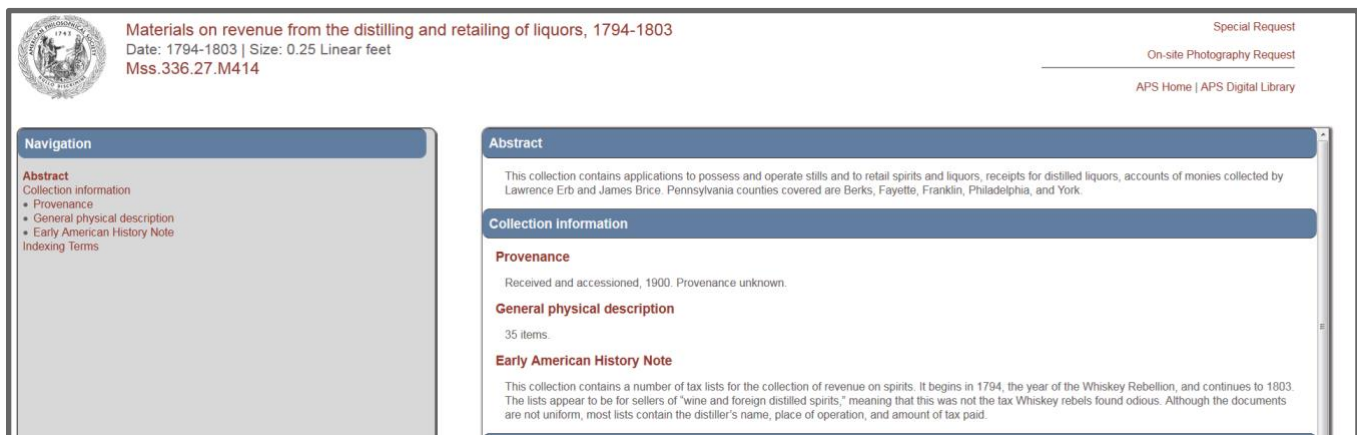
Detailed Inventory					
Series I. Paine Material		1775-1808		Request Series for Reading Room Access ?	
Request a Copy of this Series ?					
Contains correspondence to and from Thomas Paine. There are also manuscript copies of documents written by Paine.					
Paine, Thomas, 1737-1809. ALS to John Bayard, 1738-1807	1777 October 30	2p.	Box 1	Request Item for Reading Room Access ?	
Request a Copy of this Item ?					
Paine, Thomas, 1737-1809. ALS to Henry Laurens, 1724-1792	1778 April 11	3p.	Box 1	Request Item for Reading Room Access ?	
Request a Copy of this Item ?					
Paine, Thomas, 1737-1809. ALS to Henry Laurens, 1724-1792	1779 January 9	2p.	Box 1	Request Item for Reading Room Access ?	
Request a Copy of this Item ?					
Paine, Thomas, 1737-1809. ALS to Pennsylvania. Supreme Executive Council	1779 October 11	2p.	Box 1	Request Item for Reading Room Access ?	
Request a Copy of this Item ?					
Paine, Thomas, 1737-1809. ADS to David Rittenhouse, 1733-1796	1779 November 27	1p.	Box 1	Request Item for Reading Room Access ?	

**Figure 6.** Finding aid of collection described at the item level. A patron making their way through this collection could make as many as one hundred individual requests.

For collections described at the collection level, however, the patron has only one link to click to see the entire collection. For PAL, however, it looks like that collection was only used once, as shown in figure 7.

A patron sitting all day in our reading room looking at a collection with little description might use the collection more heavily than a patron clicking select items in a well-described collection. However, when we review the numbers, all we see is that the well-described collections get more clicks.





**Figure 7.** Screenshot of finding aid with only collection-level description. This collection has only one request link, the “Special Request” link at the top right. A patron looking through the entirety of this collection will only log a single request from the point of view of our statistics.

### ***The Problem of Aboutness***

When we speak of the problem of aboutness, we draw attention to the fact that manuscript collections can be about many different things. One researcher might come to a collection for one reason, another researcher for another reason. A good example at the APS Library is the William Parker Foulke Papers.<sup>6</sup> This collection contains approximately three thousand items and represents a wide variety of the interests of the eponymous Mr. Foulke. He discovered the first full dinosaur skeleton, promoted prison reform, worked toward abolition, and championed arctic exploration. A patron looking at this collection could be interested in any of these topics, or others.

PAL, however, isn’t able to account for these nuances. If a researcher interested in prison reform requests items from the Foulke Papers, they’ll see the same suggestion as a researcher who came to the collection for arctic exploration.

### ***What to Do about This***

Identifying these pitfalls is a good first step to avoiding them, but it’s only a first step. There are technical solutions, and we’ll continue to explore them. For example, the bias toward well-described collections is mitigated by showing both the number of requests and the number of users who have requested from a collection (see figure 3). We hope that by presenting both numbers, we move a little toward overcoming this bias.

However, we’re also interested in the nontechnical approaches to these issues. As mentioned in the introduction, the APS relies heavily on traditional reference service, both remote and in-house. Nontechnical solutions acknowledge the shortcomings of any constructed solution and injects a healthy amount of humility into our work. Additionally, the subject guides, search tools, and web exhibitions all form an ecosystem of discovery and access to supplement PAL.

## **FUTURE STEPS**

### ***Using Data Outside of Aeon***

We have begun exploring options for using the recommendation data outside of Aeon. One early prototype surfaces a link in our primary search interface. For example, searching for the William

Parker Foulke Papers shows a link of what people who requested from this collection also looked at. See figures 8 and 9.

### Generalizing for Other Repositories

There are ways to integrate the use of Aeon with EAD finding aids. The systems that the APS has developed to collect data for automated recommendations takes advantage of our infrastructure. We'd like for other repositories to be able to use PAL. It is our hope that an institution using Aeon in a different way will help us generalize this system.

### Generalizing beyond Aeon

PAL is currently configured to pull data out of the Microsoft SQL database used by Aeon. However, all the manipulation is done outside of Aeon and is therefore generalizable to data collected in other ways. Because archives and special collections have long-held statistics in different types of systems, we hope to be able to generalize beyond the Aeon use case if there is any interest in this from other repositories.

### Integrating PAL into Aeon

Conversations with Atlas staff about PAL have been positive, and there is interest in building many of the features into future releases of Aeon. As of this writing, an open Uservice forum topic is taking votes and comments about this integration.<sup>7</sup>



**Format**

- Book (346)
- Manuscript Collection (58)
- Analytic (10)
- Pamphlet (4)
- Periodical (1)

**Subject**

- Baccalaureate addresses (18)
- Photographs (11)
- Early National Politics (8)
- Functions (8)
- Miscellaneous (8)
- National Academy of Sciences. (U.S.) (8)
- Philadelphia History (8)
- Physics (8)
- Astronomers (7)
- Business Records and Accounts (7)
- Diaries. (7)
- Electricity (7)
- Gelatin silver prints (7)
- Anishinaabe (6)
- Biology (6)
- Business and Skilled Trades (6)
- Dobzhansky, Theodosius Grigorievich, 1900-1975 (6)
- Education (6)
- Einstein, Albert, 1879-1955 (6)
- Evolution (6)
- Family Correspondence (6)
- General Correspondence (6)
- Houston family (6)
- Paulina. I inus. 1901-1994 (6)

**MANUSCRIPT COLLECTION**  
**1** Title: **Henry Howard Houston, II Papers**

**Creator:** Houston, Henry Howard, II, 1895-1918  
**Dates:** 1903-1918  
**Abstract:** Henry Howard Houston II (1895-1918) was the son of a prominent Philadelphia family, the Houstons. He enjoyed a privileged upbringing, graduating from Chestnut Hill Academy in 1912 and from the University of Pennsylvania in 1916. Upon his graduation date, he enlisted in Battery C of the First Field Artillery, Pennsylvania National Guard, and was sent to the Texas and New Mexico borders in response to Pancho Villa's raids from Mexico. In January 1917, he volunteered with the American Field Ambulance Service in France during World War I, stationed near Verdun, where he was constantly under fire when performing his duty to pick up wounded soldiers from the trenches. Henry took a commission in the French Army and enrolled in its transportation school. The French government awarded him the Croix de Guerre for gallantry under fire. When the U.S. entered the war, Henry resigned his commission in the French army and rejoined the Pennsylvania National Guard, his unit being the 53rd Field Artillery Brigade, 28th Division. He returned to the U.S. and received training as an aerial observer at Fort Sill, Oklahoma. Early in 1918 he returned to France as a staff member of the 53rd Field Artillery Brigade, commanded by General William Price. In August 1918, Lt. Houston was killed south of Arcis-le-Ponsart while returning from an air field. He was 23. The bulk of this eight linear feet collection reflects Henry's personal life as a son, soldier, friend and lover, during the last three years of his life 1916 to 1918.

**Call #:** Mss.Ms.Coll.76.17  
**Extent:** 8.0 Linear feet  
**Subjects:** Battlefields--France--Rheims | Battlefields--France--Verdun | Belgium | Houston, Henry Howard, II, 1895-1918 | Maps | Pennsylvania. National Guard. Infantry Division, 28th... -- Correspondence | Philadelphia. Citizens | United States. Army--History--Punitive Expedition into Mexico, 1916 | United States. Army--History--World War, 1914-1918 | World War I -- France | World War I -- Germany | World War, 1914-1918 | World War, 1914-1918 --Regimental histories --United States --28th (Iron) division | World War, 1914-1918--France | World War, 1914-1918--Regimental histories--United States

[View Finding Aid and Request Materials](#) | [View PDF](#) | [View XML](#)  
 PAL Recommendations: Patrons who used Henry Howard Houston, II Papers also used these collections.

**Figure 8.** A link in the search returns that leads to recommendations based on finding aid search. Clicking on the link “PAL Recommendations: Patrons who used Henry Howard Houston, II Papers also used these collections” will open an HTML page with a list of links to finding aids.





Recommendations Based on *Henry Howard Houston, II Papers*

This is a list of collections that researchers who looked **Henry Howard Houston, II Papers** also looked at

- [Mss.Ms.Coll.200 - Miscellaneous Manuscripts Collection, 1668-1983](#) (98 researchers)
- [Mss.B.P31 - Peale-Sellers Family Collection, 1686-1963](#) (87 researchers)
- [Mss.B.B284d - Violetta Delafield-Benjamin Smith Barton Collection](#) (69 researchers)
- [Mss.Ms.Coll.52 - Wyck Association Collection](#) (55 researchers)
- [Mss.B.F826 - William Parker Foulke Papers, 1840-1865](#) (34 researchers)
- [Mss.Ms.Coll.3 - Charles Coleman Sellers Collection](#) (32 researchers)
- [APS.Archives.IId - American Philosophical Society Archives, Record Group IId](#) (26 researchers)
- [Mss.B.H142.1-3 - David Hall papers, 1745-1822](#) (21 researchers)
- [Mss.974.811.Ea7 - Eastwick Collection, 1746-1929](#) (20 researchers)
- [Mss.B.Sh61f - Shippen family papers, 1733-1878](#) (20 researchers)
- [Mss.Ms.Coll.72 - Gratz Family Papers](#) (17 researchers)
- [Mss.B.R124 - C. S. \(Constantine Samuel\) Rafinesque correspondence and writings, 1808-1840](#) (15 researchers)
- [Mss.Ms.Coll.115 - Kane Family Papers](#) (13 researchers)
- [APS.Archives.IIf - American Philosophical Society Archives, Record Group IIf](#) (11 researchers)
- [Mss.B.Or2 - George Ord Collection](#) (10 researchers)
- [Mss.B.L493 - John L. \(John Lawrence\) LeConte papers, 1812-1897](#) (9 researchers)
- [Mss.B.Sa95.g - Thomas Say papers, 1819-1883, 1955](#) (8 researchers)
- [Mss.B.H39p - Joseph Henry letters, 1836-1878](#) (7 researchers)
- [Mss.B.L437 - Lawrence, George N. \(George Newbold\) Papers, ca. 1860s-1870s](#) (7 researchers)
- [Mss.B.P3825 - C. W. \(Caspar Wistar\) Pennock papers, 1829-1891](#) (7 researchers)
- [Mss.B.W692 - Alexander Wilson papers, 1806-1813](#) (4 researchers)
- [Mss.580.P97 - Journal of a botanical excursion in the Northeastern parts of Pennsylvania and in the state of New York, 1807](#) (3 researchers)
- [Mss.B.Ag12.ms - Louis Jean Rodolph Agassiz papers, 1833-1873](#) (3 researchers)
- [Mss.B.Au25 - John James Audubon Papers](#) (3 researchers)
- [Mss.B.Sh81 - Charles Wilkins Short correspondence, 1813-1867](#) (3 researchers)

Figure 9. HTML link of recommended finding aids based on search.

CONCLUSION

The APS is trying to add to the already robust options for users to find relevant manuscript collections. In addition to traditional reference, web exhibitions, and online search and browse tools, we have started reusing circulation data and self-identified user interests to automate recommendations. This new system fits within the ecosystem of tools we already supply. This is a snapshot of where the PAL recommendation project is as of this writing, and we hope to work with other special collections libraries and archives to continue to grow the tool. If you are interested, we hope you reach out.

ENDNOTES

<sup>1</sup> "Subject Guides and Bibliographies," American Philosophical Society, accessed February 27, 2018, <https://amphilsoc.org/library/guides>; "Exhibitions," American Philosophical Society, accessed February 27, 2018, <https://amphilsoc.org/library/exhibit>; "Galleries," American Philosophical Society, accessed February 27, 2018, <https://diglib.amphilsoc.org/galleries>.

<sup>2</sup> "Aeon," Atlas Systems, accessed February 27, 2018, <https://www.atlas-sys.com/aeon/>.



- 
- <sup>3</sup> Michael Mönnich and Marcus Spiering, “Adding Value to the Library Catalog by Implementing a Recommendation System,” *D-Lib Magazine* 14, no. 5/6 (2008), <https://doi.org/10.1045/may2008-monnich>.
- <sup>4</sup> Matthew Reidsma, “Algorithmic Bias in Library Discovery Systems,” *Matthew Reidsma* (blog), March 11, 2016, <https://matthew.reidsrow.com/articles/173>.
- <sup>5</sup> “AmericanPhilosophicalSociety/PAL,” American Philosophical Society, last modified September 11, 2017, <https://github.com/AmericanPhilosophicalSociety/PAL>.
- <sup>6</sup> “William Parker Foulke Papers, 1840–1865,” American Philosophical Society, accessed February 27, 2018, <https://search.amphilsoc.org/collections/view?docId=ead/Mss.B.F826-ead.xml>.
- <sup>7</sup> “Recommendation System to Suggest Items to Researchers Based on Users with the Same Research Topic,” Atlas Systems, accessed February 27, 2018, <https://uservice.atlas-sys.com/forums/568075-aeon-ideas/suggestions/18893335-recommendation-system-to-suggest-items-to-research>.

