



Journal of the Text Encoding Initiative

Issue 11 | July 2019 - June 2020
Selected Papers from the 2016 TEI Conference

Encoding Newton's Alchemical Library: Integrating Traditional Bibliographic and Modern Computational Methods

Meridith Beck Mink, Michelle Dalmau, Wallace Hooper, William R. Newman, James R. Voelkel and John A. Walsh



Electronic version

URL: <http://journals.openedition.org/jtei/2866>

DOI: 10.4000/jtei.2866

ISSN: 2162-5603

Publisher

TEI Consortium

Electronic reference

Meridith Beck Mink, Michelle Dalmau, Wallace Hooper, William R. Newman, James R. Voelkel and John A. Walsh, « Encoding Newton's Alchemical Library: Integrating Traditional Bibliographic and Modern Computational Methods », *Journal of the Text Encoding Initiative* [Online], Issue 11 | July 2019 - June 2020, Online since 18 February 2020, connection on 01 July 2020. URL : <http://journals.openedition.org/jtei/2866> ; DOI : <https://doi.org/10.4000/jtei.2866>

For this publication a Creative Commons Attribution 4.0 International license has been granted by the author(s) who retain full copyright.

Encoding Newton's Alchemical Library: Integrating Traditional Bibliographic and Modern Computational Methods

Meridith Beck Mink, Michelle Dalmau, Wallace Hooper, William R. Newman, James R. Voelkel, and John A. Walsh

ABSTRACT

The *Chymistry of Isaac Newton* (<http://chymistry.org>) project team has digitized and encoded, following the TEI Guidelines, the complete corpus of Newton's alchemical manuscripts, which total more than two thousand pages and over one million words. Newton cited more than five thousand published and unpublished works in these manuscripts; many of his annotations reference items in his own library, as he was an exceptionally dedicated reader of alchemical texts. Newton's extensive citations and annotations provide a window into his alchemical research and practices, and serve as the basis for our authoritative bibliography of his alchemical sources. The bibliography is being developed as both a stand-alone reference work and an integrated resource with the alchemical manuscripts, providing additional context for Newton's citations and

florilegia. Once finished, the bibliography will provide complete, structured citations—which often would appear very abbreviated or incomplete in the manuscripts—that can be formatted to comply with modern bibliographic conventions and bibliographic management systems. Our bibliography will also link to digitized online versions of the source texts available through Early English Books Online, HathiTrust Digital Library, and other digital repositories. The citations include quasi-facsimile title page transcription, a technique used for bibliographic description of rare books, to enable richer forms of citation analysis. By analyzing the citations, we will be able to date Newton's manuscripts, cluster manuscripts that cite the same or related sources, and, ultimately, generate network graphs that will reveal connections between the cited authors and texts and how they influence Newton's own ideas and work.

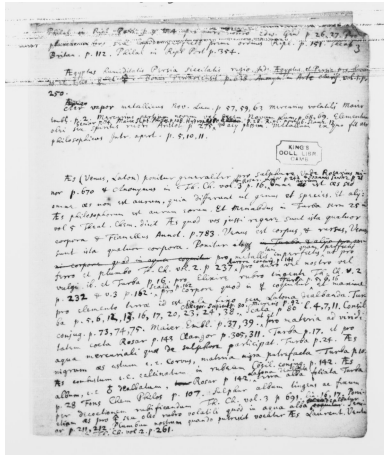
INDEX

Keywords: bibliography, alchemy, quasi-facsimile transcription, Zotero, latent semantic analysis

1. Introduction

- 1 Best known for his contributions to gravitational theory, calculus, and optics, Newton was also a serious student and practitioner of alchemy. His library was full of dog-eared alchemical books and manuscripts, and he wrote and transcribed close to a million words on the subject, although he never published any of them. His notes and unfinished manuscripts contained over five thousand references to alchemical texts and practices (figure 1). Newton even employed his own citation methods within his manuscripts and notes. It is unusual to see this level of specificity in citation practices in the seventeenth century.

Figure 1. Manuscript excerpt with citations from Isaac Newton, Keynes MS. 30/1 (King's College Library, Cambridge University), page image 3r in the *Chymistry of Isaac Newton*, edited by William R. Newman, 2005–, accessed December 13, 2019, <http://purl.dlib.indiana.edu/iudl/newton/ALCH00200>.



- 2 In an effort to better understand Newton's alchemical scholarship—as well as the study of alchemy in the seventeenth century more broadly—our team seeks to reconstruct, from the fragmentary citations in his personal papers, a comprehensive list, with more complete bibliographic information, of the hundreds of alchemical texts that Newton read and referenced. This work is meant as a complement to the larger *Chymistry of Isaac Newton* (<http://chymistry.org>) project, which began in 2003 with a focus on transcribing and TEI-encoding the complete corpus of Newton's alchemical manuscripts, which total more than two thousand pages and over one million words. Along with a scholarly edition of diplomatic and normalized transcriptions and facsimile page images of Newton's alchemy, the *Chymistry of Isaac Newton* project also includes pedagogical resources primarily focused on recreating experiments, including a lab unit that features video recordings of reenacted experiments, and online tools that include reference works and a Latent Semantic Analysis Tool to enable a deeper understanding of Newton's writings.
- 3 Our aim is to produce a comprehensive bibliography that accurately represents Newton's extensive alchemical reading and research, identifying his sources down to the specific edition of the printed texts he referenced. Once completed, Newton's alchemical bibliography will: (1) assist us in dating Newton's manuscripts; (2) allow for the clustering of manuscripts that cite the same or related sources; and, ultimately, (3) generate network graphs that will reveal connections between the cited authors and texts and how they influence Newton's own ideas and work.

2. The Bibliography

- 4 The methods for generating Newton's alchemical bibliography required traditional bibliographic research as well as compiling and encoding the bibliography following the TEI P5 Guidelines (TEI Consortium 2017).

2.1 Tracing the Bibliographic References

- 5 The team started the bibliography by simply identifying and tagging citations to printed works and manuscripts in Newton's alchemical papers. To date we have located over five thousand citations, which have been encoded with `<bibl>` elements that will all soon point to the full citation in the bibliography using the `@corresp` attribute (example 1). As part of the process of tracing the citations and adding the corresponding link to the main entry in the bibliography, the project team will also be checking to make sure all citations are tagged. We do have a few manuscripts that were published without the `<bibl>` tagging so we expect the total number of citations to grow to considerably over five thousand as we revisit the corpus.

Example 1. Example of encoded citations provided by Newton using the `<bibl>` tag.

```
<p><del rend="strike" hand="#in"><g ref="#UNx263f">ꝯ</g><hi
rend="super"><choice>
  <orig>ij</orig>
  <reg>ii</reg>
</choice></hi></del>
<add place="supralinear" rend="caret">lapidis</add> pro ejus solutione seu
liquefactione in decoctione<lb/> ab albedine ad rubedinem <bibl>Philal on Ripi.
```

G p.

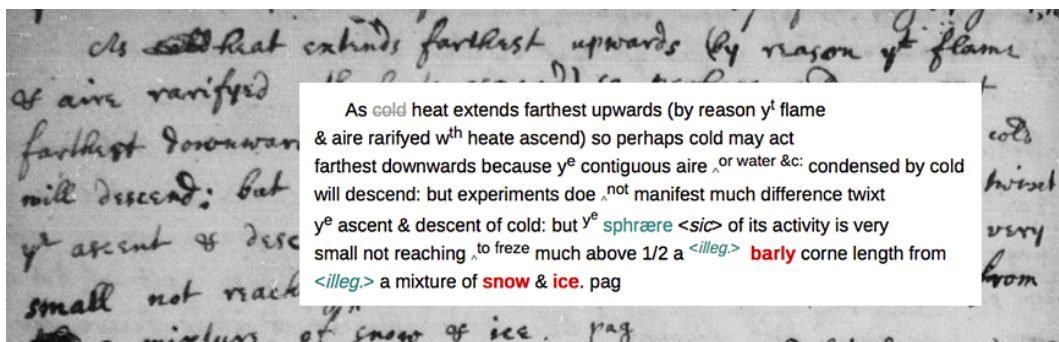
```
<add place="supralinear">61, 62,</add> 180, 365.</bibl>
<bibl>Artef p 5 lin 12</bibl><lb/></p>
```

```
<p><del rend="strike" hand="#in">lapidis</del> vel
Ceratio lapidis pro ejus liquefactione
<add place="supralinear" rend="caret">&amp; ablutione</add> post nigredinem
<bibl>Flammel annot<lb/> p 770.</bibl><lb/></p>
```

- 6 The tagging was the easy part. Next, identifying exactly what Newton was referring to in each of these citations was a meticulous process requiring detective work by several specialists—subject experts and rare books and special collections librarians. Newton's citations were often

fragmentary because he used abbreviated notes intended for himself. Considering that he was working before formal citation practices were developed, his references are remarkably consistent and clear to the modern reader. That said, in some cases we were able to see that Newton was referencing *something*—page numbers and abbreviations to titles—but exactly what he was citing, as in figure 2, is not immediately obvious.

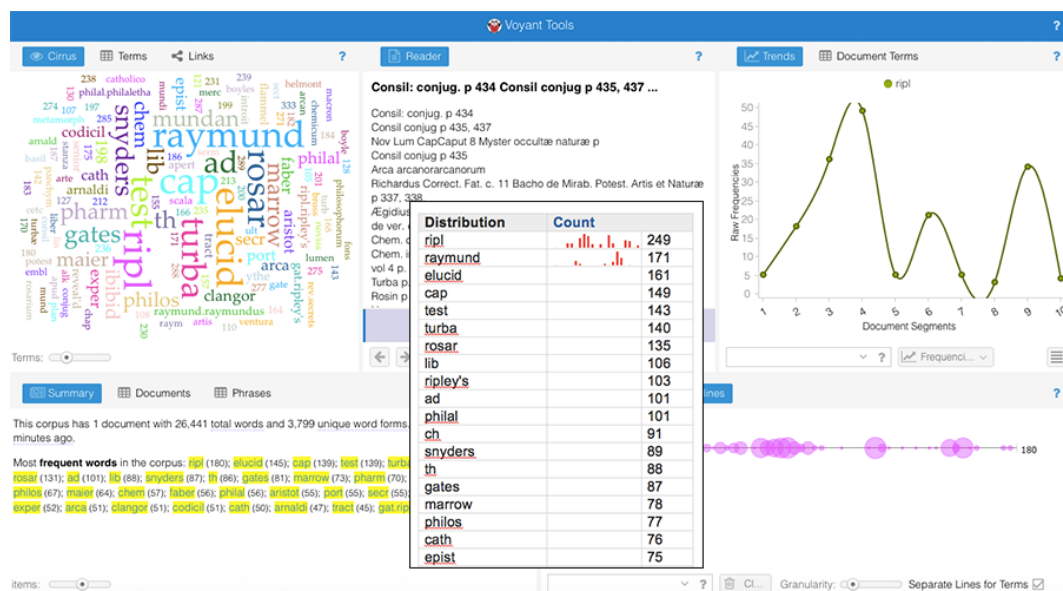
Figure 2. Manuscript with citation missing a page number, Isaac Newton, Portsmouth Add. MS. 3975, page image 16v in the *Chymistry of Isaac Newton*, edited by William R. Newman, 2005–, accessed December 16, 2019, <http://purl.dlib.indiana.edu/iudl/newton/ALCH00110>.



- 7 For example, Newton used the term “Th. Ch.” to refer to the *Theatrum Chemicum*, a multivolume compilation containing a multitude of alchemical tracts, which he cited numerous times throughout his manuscripts. Newton referenced a handful of other collections as well as the *Theatrum Chemicum*, such as the *Artis Auriferae*, published several times in ever-expanding form during the sixteenth and early seventeenth centuries, and the *Musaeum Hermeticum*, another work that grew over time as it was republished. We compiled the tables of contents for each of these collections to properly identify the individual tracts that Newton referenced. The project team agreed to enter referenced tracts as individual entries in the bibliography with a complete citation to the anthologized source. Newton occasionally cited “second hand” references in which he would attribute something to one author that was actually stated by another author. Clarifying this is critical for pointing to the correct reference from the alchemical manuscripts.
- 8 Bibliographic tagging of the manuscripts also allowed us to do a rudimentary text analysis to study the words that frequently occurred in the citations. After generating the output of the existing <bibl>s encoded in the manuscripts, we used the [TAPoRware Text Analysis Tool](#)¹ and the [Voyant Tools](#)² to check for frequency of terms and distribution of terms across the corpus. This allowed

us to determine that Newton's most frequently cited text was George Starkey's *Secrets Revealed*, published posthumously in 1669, a result which provided quantitative evidence that Newton had studied this work carefully. Starkey, writing under the pseudonym Eirenaeus Philalethes, was irregularly cited by Newton as philal.philaletha, philal, philos, and other variants. In addition, running the citations through the text analysis tools confirmed the degree to which name variants would benefit from normalization through the compilation of the bibliography. The text analysis also showed that Newton frequently cited George Ripley, a well-known fifteenth-century British alchemist, and Raymond Lull, a thirteenth-century philosopher, among others ([bad link to item:]).

Figure 3. Visualization generated with Voyant Tools and a concordance generated with the TAPoRware Text Analysis Tool of approximately 5,000 bibliographic citations encoded in the Newton alchemical corpus revealing most cited authors and issues with name variants.

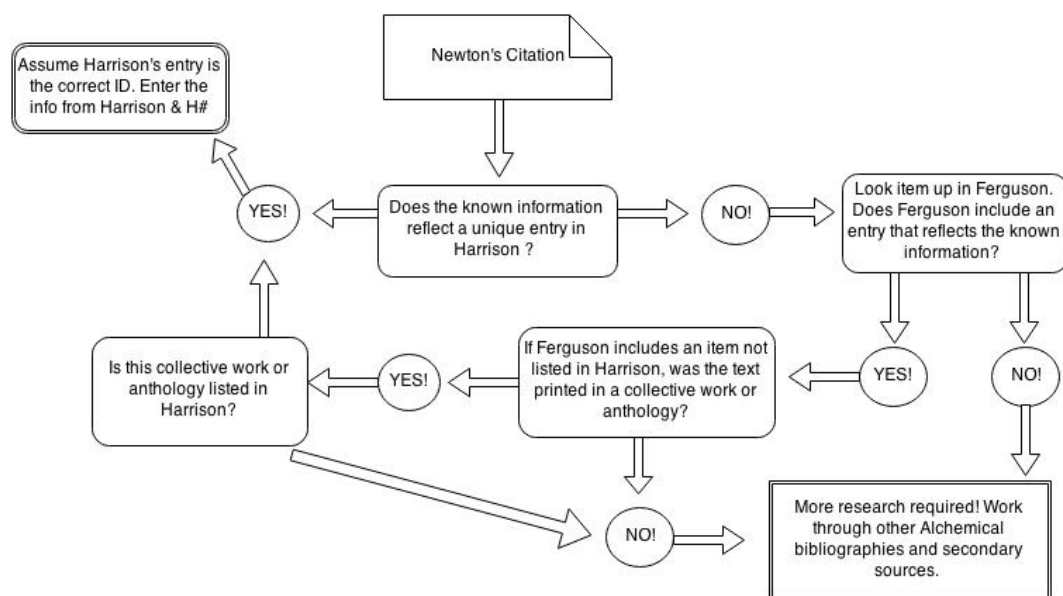


- 9 Newton's alchemical manuscripts reflect not only his own original work, but the work of other scholars, alchemists, and philosophers. By compiling an authoritative bibliography, we are able to correctly attribute the paraphrases, quotes, or transcriptions of long passages that appear in Newton's alchemical manuscripts, as well as the extent to which Newton drew from other authors.

2.2 Building the Bibliography

- 10 Owing to the iterative nature of the process of compiling the bibliography, which required extensive research, the project team decided to use [Zotero](#)³ because of the ease of data entry and availability of the Zotero-to-TEI XSLT stylesheet as an initial way to generate the bibliography.
- 11 A key resource for building our bibliography was John Harrison's *The Library of Isaac Newton* (1978), which is the most comprehensive catalog of Newton's working library. It lists the approximately 2,100 volumes of books and manuscripts in Newton's possession when he died in 1727. While the catalog captures the whole of Newton's library, Harrison did not necessarily record precise bibliographic information. Therefore, we also consulted John Ferguson's *Bibliotheca Chemica: A Catalogue of the Alchemical, Chemical and Pharmaceutical Books in the Collection of the Late James Young of Kelly and Durris* (1906) for clarification (figure 4). The citations compiled in Zotero retain a reference to Harrison by recording the identifier system Harrison himself devised (i.e., [H11]) and include supplemental information provided by Harrison when appropriate.
- 12 Once the correct editions were identified, metadata were often imported from cataloging systems, especially WorldCat records in addition to catalog records from the Chemical Heritage Foundation and the University of Wisconsin, both of which hold important early modern alchemical monograph collections, to ensure the most complete bibliographic metadata. The metadata were either corrected or enriched following the guidelines provided by *Descriptive Cataloging of Rare Materials* (2007) (known earlier as *Descriptive Cataloging of Rare Books*), or DCRM(B), Bowers's *Principles of Bibliographic Description* (2005, ch. 4), and Gaskell's *A New Introduction to Bibliography* (1972, 321–35).

Figure 4. Diagram representing the bibliographic research workflow for verifying Newton's citations.



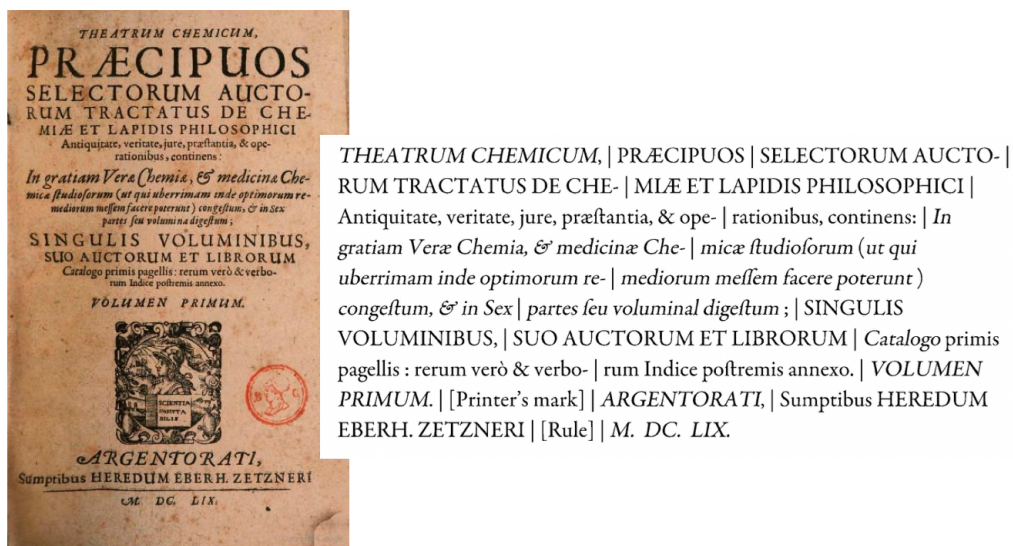
2.2.1 Use of Quasi Facsimile Transcription

- 13 Writing in the late seventeenth century, Newton typically referenced texts written and/or published during the fifteenth through seventeenth centuries. He also cited medieval sources, but these were usually reprinted in some of the contemporary printed editions and compilations in his library. According to print practice during the early modern period, all the bibliographic information about a work—such as author, date of publication, and place of publication—was contained on the title page. Title pages were critical to the Newton bibliography because we want to pinpoint as precisely as possible which edition or printing of a text Newton cited. This level of precision was important to the project team because the exact printing dates of the material Newton cited in his work allow us to better date when he was producing his alchemical manuscripts and to accurately identify his citations.
- 14 However, the fine detail of these title pages is frequently garbled by modern bibliographic protocols; it is not uncommon, for instance, for catalogers to replace the original punctuation with modern punctuation. Moreover, the titles commonly used to refer to books of this period may bear little resemblance to the title as printed on the title page. To give an obvious example, Newton's masterwork of gravitational theory is often referred to in brief as "the *Principia*," the

third word of its actual title, *Philosophiae naturalis principia mathematica*. Harrison's *The Library of Isaac Newton* frequently abbreviates long-winded seventeenth-century titles, undoubtedly in the interest of conserving space, but at the same time creating the potential for confusion.

- 15 In order to precisely record the fine nuances of an early-modern title page, bibliographers and catalogers have long used a method called *quasi-facsimile transcription* (QFT). The goal of QFT—as it was put by Fredson Bowers, who clarified and codified its rules in his magisterial *Principles of Bibliographical Description*—is “bringing an absent book before the eye of the reader” (2005). The method involves using a very specific set of rules to transcribe every letter, punctuation mark, rule, and page break on the title page, capturing as much detail as possible, down to the use of small caps and swash italics (figure 5).

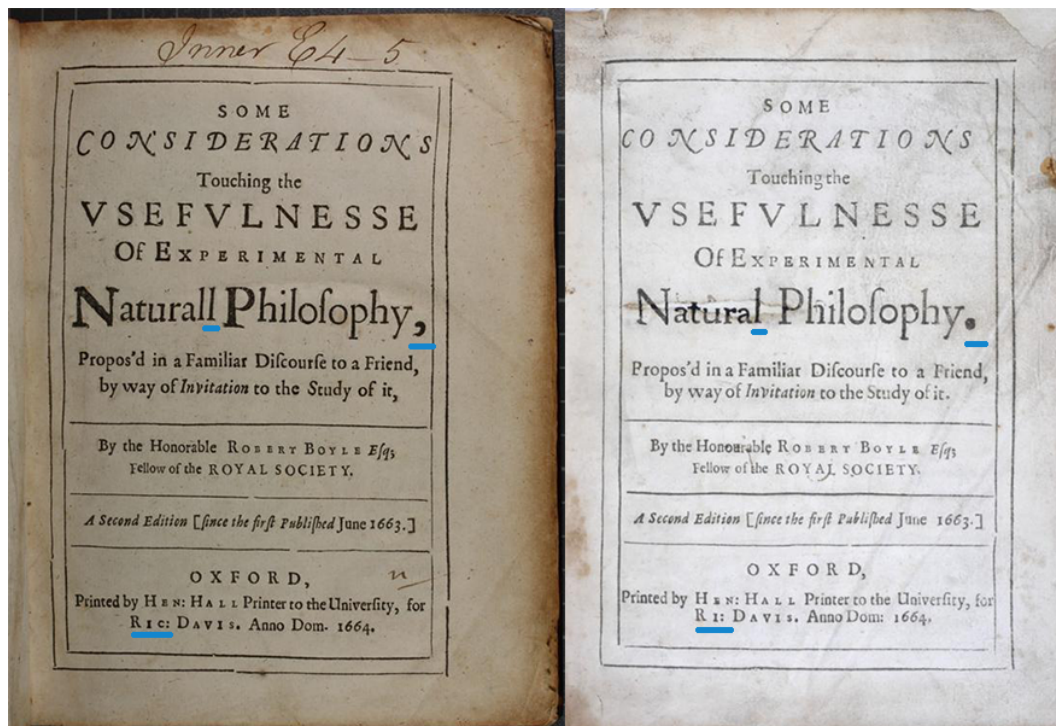
Figure 5. Example of a title from the *Theatrum Chemicum* and accompanying quasi-facsimile transcription.



- 16 Small variations in the title pages of books from this period are sometimes the only way to distinguish different editions or printings. Though there has long been debate among bibliographers over whether photographic title page facsimiles are superior to QFT, the method has undeniable advantages. For any encoding project, of course, QFT offers transparent searching. And there are in practice very few examples of title pages that were reset with such precision that QFT cannot distinguish one edition from another.

- 17 To give one example from the project, Newton frequently cited Robert Boyle's *Some Considerations Touching the Usefulness of Experimental Natural Philosophy*. The book was originally published in Oxford in 1663. It was reprinted in 1664, listing the same Oxford printer (Henry Hall) and publisher (Richard Davis), but there is a note in the book that the edition was committed to several presses—not an uncommon practice in seventeenth-century English publishing—and the details of the printing suggest that half if not all of the printing was done in London. In 1671, Boyle issued a second volume of the book, at which time a reprint of the second edition of the first volume was made but still with the original publication information: Oxford, Henry Hall for Richard Davis, 1634. So, there are two versions of the second edition (of the first volume) with identical publication information, one published largely in London (not Oxford) in 1664, and one published in Oxford in 1671 (not 1664). As Fulton (1961, 38–41) notes, to make the problem of identification acute, these two editions can only be distinguished by three inconsistencies in the spelling and punctuation as seen in figure 6, the one spelling “Naturall” with two l’s instead of one, with commas rather than periods after “philosophy” and “it,” and spelling “Ric: Davis” rather than “Ri: Davis.” Harrison’s citation for this book, as compiled in *The Library of Isaac Newton* (1978, 109), “Some considerations touching the usefulness of experimental naturall philosophy... 2 vols. 4°, Oxford, 1664–1671,” is utterly incapable of distinguishing which edition Newton might have owned.

Figure 6. Title pages from Robert Boyle's *Some Considerations Touching the Usefulness of Experimental Natural Philosophy* illustrating the nuances of different editions, and how the act of quasi-facsimile transcription assists in identifying the precise text that Newton referenced.



- 18 We used QFT in order to record the most accurate information possible about the texts Newton cited. We chose QFT over discrete TEI elements for representing bibliographic metadata found in the title pages mostly for practical reasons. It would have been too resource-intensive to reflect the typographic conventions of transcribing a title page from an early modern edition using TEI, and we did not want to break new ground given the well-established and widely accepted conventions of QFT. Using QFT consistently was essential to the bibliographic research process. Including the QFT in the TEI document, even if the title page elements were not granularly encoded, allows the team to maintain the TEI XML document as the authoritative source for the bibliography. The QFT transcription is encoded in the title element that is part of the <biblStruct> along with a supplied title to streamline metadata display for readability (example 2).

Example 2. Example of how QFT is captured in the TEI encoding.

```
<listBibl>
  <biblStruct type="book" xml:id="Boyle1672">
```

```

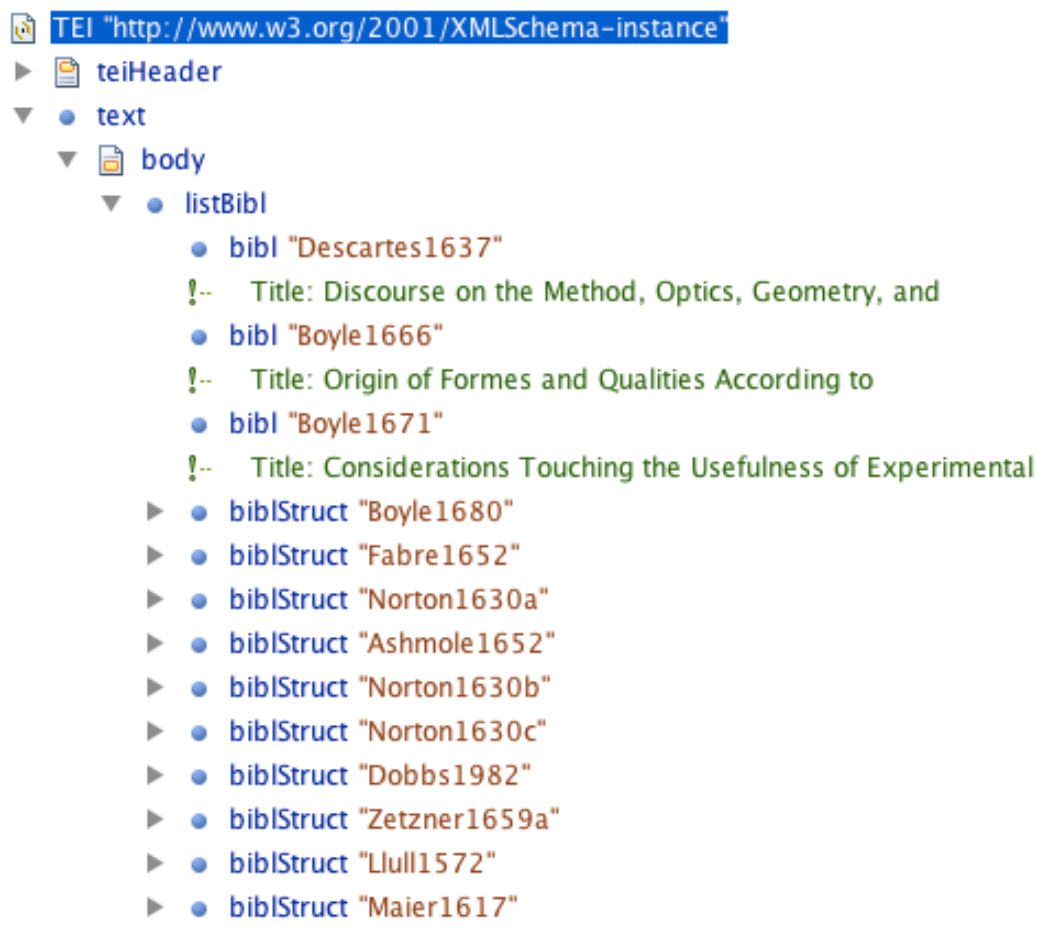
<monogr>
  <title type="short">New Experiments, touching the relation betwixt flame and
air</title>
  <title level="m" ref="https://quod.lib.umich.edu/e/eebo2/A29057.0001.001?
view=toc">TRACTS | Written | By the Honourable | Robert Boyle, | CONTAINING | New
EXPERIMENTS,
  touching the | Relation betwixt Flame and Air. And about | EXPLOSIONS. | An
HYDROSTATICAL Difcourſe oc- | caſion'd by ſome Objections of Dr. Henry More |
againſt
  ſome Explications of New Experiments | made by the Author of theſe Tracts: To
which |
  is annex't, An Hydroſtical Letter, dilucidating | an Experiment about a Way
of
  Weighing Water | in Water. | Of the Poſitive or Relative Levity of Bo- | dies
under
  Water. | Of the Air's Spring on Bodies under | Water. | About the Differing
Preſſure
  of Heavy So- | lids and Fluids. | [Double Rule] | LONDON, | Printed for
Richard Davis,
  Book-ſeller in Oxon | M DC LXXII.</title>
<author><forename>Robert,</forename><surname>Boyle</surname></author>
<imprint>
  <pubPlace>London</pubPlace>
  <publisher>Printed for R. Davis, book-seller in Oxon.</publisher>
  <date>1672</date>
</imprint>
</monogr>
<note>[H275] __ Tracts, containing new experiments, touching the relation
betwixt flame
  and air. And about explosions. An hydroſtical diſcourſe occaſion'd by ſome
objections
  of Dr. Henry More ... 8°, London, 1672. (A few ſigns of dog-earing.) Tr/
NQ.10.100
  CHF [Rare Book Storage QD27 .B695 1672]</note>
<note>not ſure if url is to correct edition, MA</note>
</biblStruct>
</listBibl>

```

2.2.2 Zotero to TEI

- 19 As mentioned earlier, we compiled the bulk of the bibliography using Zotero. Once the bibliography was close to completion, we exported the bibliography from Zotero to RDF, then used stylesheets provided by the TEI Community (available on GitHub⁴) to convert from RDF to P4. Finally, another stylesheet was used to conform to the most current version of the TEI Guidelines, P5.
- 20 The entries in the bibliography are grouped using a `<listBibl>` with individual citations in a `<biblStruct>` (figure 7). The bibliography is still a work in progress as new `<bibl>`s are encoded in the manuscripts that cite sources not yet compiled. Those newer citations are shorthand encoded with a `<bibl>` and identifier so that the linking mechanism from the manuscripts to the bibliography can continue smoothly. Entries tagged with `<bibl>`s in the bibliography will be collocated and individually traced following the methodology detailed earlier.

Figure 7. The entries in the bibliography are grouped using `<listBibl>` with individual citations in a `<biblStruct>`. As new citations are encoded in the manuscripts for which reference sources have not yet been compiled, they are encoded with an identifier attribute as part of the `<bibl>` tag.



3. Integrating the Bibliography with the Manuscripts

21 We envision Newton's bibliography as a standalone online reference and also as a resource tightly integrated with the alchemical manuscripts. At this point in the project, we have preliminary conceptual designs of how to display full citations in context in light of other critical apparatus conventions we are currently employing for the alchemical manuscripts. We have identified a couple of challenges regarding integration of the bibliography with the alchemical manuscripts that the project team needs to further consider: (1) contextualizing citations that reference longer quotes, and (2) properly attributing quotes that reference multiple authors. The standalone

version of the bibliography is still under development and is relying on [TEI Boilerplate](#)⁵ for online publication. Our goal is to include full text access via persistent URLs to the source materials hosted by HathiTrust, the Internet Archive, or EEBO, giving preference to the most optimal scans and open access resources.

- 22 To help us efficiently and accurately integrate the bibliography, the project team created a series of stylesheets to output the citation (contents within a <bibl>), the value of the @corresp attribute, and the manuscript source ([figure 8](#)). This serves two distinct purposes: (1) it provides the encoders with a quick way to reference whether an entry in the bibliography already exists, and (2) it facilitates review by the project editors to ensure that passages were properly cited.

Figure 8. XSLT output of citations encoded in the alchemical manuscripts that assists in the encoding and editorial review process.

| | |
|--|--|
| Eposit. on Pref: Ripl. p. 7. 58 | CHYM000001.xml#EIRENAEUS_PHILALETES1677 [[Portsmouth Add. MS. 3975, f.123v; ALCH00110]] |
| (Eposit. R. Ep. p 20) | CHYM000001.xml#EIRENAEUS_PHILALETES1677 [[Portsmouth Add. MS. 3975, f.123r; ALCH00110]] |
| Essays p 71, 72, 73. | no corresp [[Portsmouth Add. MS. 3975, f.47v; ALCH00110]] |
| Exper 24 | CHYM000001.xml#Boyle1660 [[Portsmouth Add. MS. 3975, f.24r; ALCH00110]] |
| Exper 40, 41 & Digress to Exper 41 | CHYM000001.xml#Boyle1660 [[Portsmouth Add. MS. 3975, f.26r; ALCH00110]] |
| Exper 6. pag 341 of fformes. | CHYM000001.xml#Boyle1666 [[Portsmouth Add. MS. 3975, f.40r; ALCH00110]] |
| Exper 7 of Forms pag 370. | CHYM000001.xml#Boyle1666 [[Portsmouth Add. MS. 3975, f.41r; ALCH00110]] |
| Exper 8 Boyle of Forms pag 380. | CHYM000001.xml#Boyle1666 [[Portsmouth Add. MS. 3975, f.41v; ALCH00110]] |
| (Exposit. Ripl. Epist. p 12, 17, 18, 20, 21, 26, 33. | CHYM000001.xml#EIRENAEUS_PHILALETES1677 [[Portsmouth Add. MS. 3975, f.123r; ALCH00110]] |
| ffrom pag 412 to pag 463 & from pag 741 to pag 803. | CHYM000001.xml#Boyle1665 [[Portsmouth Add. MS. 3975, f.18v; ALCH00110]] |

4. Next Steps

- 23 Once the bibliography is complete, the Newton project team, through careful analysis of the citations, will be better able to date Newton's manuscripts, to cluster manuscripts that cite the same or related sources, and, ultimately, to generate network graphs that will reveal connections

between the cited authors and texts and how they influenced Newton's ideas and work. The citation analysis will be combined and integrated with parallel work being done in other veins by this team to establish the order of composition of the alchemical manuscripts.

- 24 We have also been working on Newton's watermarks; on the evolution of his orthography; on the elemental composition of his inks by XRF spectrometry; and on mapping the overall semantic structure of the corpus through latent semantic analysis, with its observable patterns of reuse and reengagement.

4.1 The Newton Corpus and Latent Semantic Analysis

- 25 The team has had a conceptual map of the corpus in hand for several years, drawn from latent semantic analysis (LSA⁶), but the ideas themselves do not suggest an obvious order of progress. Newton's scholarly progression in topics like calculus, mechanics, and gravitation, for which we have well-founded intuitions, seems to unfold in his manuscripts in a discernible order. Yet, we still do not understand the directions Newton took in his alchemical studies because the ideas remain largely mysterious to us. As a result, we have a map of his alchemical ideas but we still need other clues to clarify their order of development, and the citations will constitute one of the foundations on which we can determine ordering and dating of manuscripts.
- 26 LSA is well established method in the field of information retrieval. It was originally designed to accomplish basic tasks in search (Berry, Dumais, and O'Brien 1995), and was subsequently used to try to model human cognition (Landauer and Dumais 1997). It starts with word counts from a set of documents, usually a large set, that are used to create a term-document matrix, which is a simple numerical representation of the corpus. Linear algebra and its vector-space methods give us a numerical model of the structure of Newton's alchemical manuscripts based ultimately on shared vocabulary and ideas. We have discovered in our work with Newton that the mathematical foundations of LSA make it particularly well suited to identifying the reuse of text passages and phrases in large corpora produced by one or more authors, and that makes LSA a valuable tool for structural text analysis of large corpora.
- 27 the *Chymistry of Isaac Newton* project has published the results of its LSA work in interactive, online component on its public website.⁷ The LSA component can produce a list of chunks or passages that are strongly linked by shared vocabulary and provide a measure of the strength of the relationship

using cosine similarities. More simply, LSA represents documents as “bags” or “buckets” of words with emphasis on how many times a word appears in a document. To identify concepts, since words have multiple meanings, LSA looks for patterns that group words together: for example, “sublimation,” “dissolve,” and “bodies” might appear in passages in which Newton is noting the transition of substances from solid to gas without passing the liquid phase (see figure 9).

Figure 9. Results from running the Latent Semantic Analysis Tool and two different manuscripts from Newton's alchemical corpus that reveal strongly correlated passages (denoted by the yellow highlighting).

Latent Semantic Analysis Tool

| https://webapp1.dlib.indiana.edu/newton/mss/dipl/ALCH00003/#f5r (new window) | https://webapp1.dlib.indiana.edu/newton/mss/dipl/ALCH00003/#f6r (new window) |
|---|---|
| ALCH00003 Keynes MS. 14, f.005r Artephius his secret Book | ALCH00003 Keynes MS. 14, f.006r Artephius his secret Book |
| <p>cold, the dry with the moist & the hard // with the soft; & thus is there a mixture made of contrary natures, that is of cold with hot & of moist with dry, an admirable connexion & conjunction of enemies. Then our dissolution of bodies which is made in this first water is no other thing then a killing of the moist with the dry because the moist is coagulated with the dry, for the moisture is contained terminated & coagulated into a body or into earth only by driness. Let therefore the hard & dry bodies be put in our first water in a vessel well shut, where they may abide until they be dissolved & ascend on high; & then they may be called a new body, the white gold of Alchimy, the white stone, the white sulphur not burning, & the stone of Paradise, that is the stone which converts imperfect metals into fine white silver. Having this we have also the body soul & spirit all together, of the which spirit & soule it is said that they cannot be drawn from the perfect bodies but by the conjunction of our dissolving water, because it is certain that the thing fixed cannot be lifted up but by the conjunction of the thing volatile. The spirit then by the mediation of the soul is drawn from the bodies, & the body is made no body, because at the same instant the spirit with the soul of the bodies mounteth on high into the upper part, which is the perfection of the stone & is called sublimation. This sublimation (saith Florentius Catalanus) is done by things sharp spiritual & volatile, which are of a Sulphureous & viscous nature, which dissolve the bodies & make them to be lifted up into the air in the spirit. And in this sublimation a certain part & portion of our first water ascendeth with the bodies; joining it self to them, ascending & subliming into a middle substance, which holdeth of the nature of the two, that is of the bodies & of the water; & therefore it is called the corporal & spiritual compound , Corsufle, Cambar, Ethelia, Zandarach, the good Duenech, but properly it is only called the water permanent because it flieth not in the fire, always adhering to the commixed bodies that is to the Sun & Moon.</p> <p>FOLIO[[5v]]FOLIO & communicating unto them a living tincture incombustible & most firm more noble & pretious then the former which these bodies had, because from hence forward this tincture can run as byle upon the bodies perforating & piercing with a wonderful fixion, because this tincture is the spirit & the spirit is the soule & the soule is the body, because in this operation the DEL[[spirit]]DEL body is made a spirit of a most subtile nature & likewise the spirit is incorporated & is made of the nature of a body with bodies, & so our stone contains a body a soul & a spirit. nature how thou changest the body into a spirit which thou couldst</p> | <p>surface of the dissolving water in the colour of whiteness & this whiteness is life: for in this whiteness the HI[[antimonial]]HI & // HI[[mercurial]]HI soule is by the appointment of nature infused with the spirits of the HI[[sun]]HI & HI[[moon]]HI which separateth the subtile from the thick & the pure from the impure, lifting up by little & little the subtile part of the body from the dreggs until all the pure be separated & lifted up. And in this is our philosophical & natural sublimation fulfilled. And in this whiteness is the soul infused into the body that is the mineral vertue which is more subtile then fire, being indeed the true quintessence & life which desireth to be born & to put off the gross earthly faeces which it hath taken from the menstruous & corrupt place of his original. And ADD[[in]]ADD this is our philosophical sublimation not in the naughty common ζ which hath no qualities like unto them wherewith our ζ drawn from his HI[[vitriolate]]HI caverns is adorned. But let us return to our sublimation.</p> <p>It is therefore most certain in this art that this soule drawn from the bodies cannot be lifted up but by putting to of a volatile thing which is of his own kind, by the which the bodies are made volatile & spiritual, lifting up subtiliating & subliming themselves against their own proper nature which is corporeal heavy & ponderous . And by this means they are made no bodies, but incorporeal & a quintessence DEL[[of the nature]]DEL of the nature of the spirit which is called Hermes his bird & mercury drawn from the red servant: & so the earthy parts remain below, or rather the grosser parts of the bodies; which cannot by any wit or device of man be perfectly dissolved. And this white fume this white gold that is this quintessence is also called the compound Magnesia, which as a man contains, or as a man is compounded of body soul & spirit. For the body is the fixed earth of the sun, which is more then</p> <p>FOLIO[[6v]]FOLIO most fine ponderously lifted up by the force of our divine water. The soul is the tincture of the Sun & of the Moon proceeding from the conjunction or from the communication of these two. But the spirit is the mineral vertue of the two bodies & of the water which carries the soule or the white tincture upon the bodies, & out of the bodies, as the tincture of Diers is carried by water upon the cloth. And that yal spirit is the bond or DEL[[tyall]]DEL tyall of the soul of the sun; & the body of the Sun is the body of fixion, containing with the Moon the spirit &</p> |

- 28 LSA also gives us numerical measures of the semantic similarity of any two passages in the whole corpus. Mathematically, that measure is a cosine calculated from vector representations of the two passages in an eigenvector space, and it has a value between zero and one. When two texts have a cosine nearly equal to one, it implies that the two are virtually identical, likely word-for-word from one end to the other. The cosines are a convenient measure of the degree of semantic entanglement of the two passages.

- 29 High cosine pairs, 0.8 and above (as seen in [figure 10](#)), point to promising locations where we are likely to find Newton reusing or rethinking text: working over the same ground, recalling or copying the same sentences or phrasing from one member of the pair to the other—and, always, one of the two must have been written before the other. In a mysterious corpus like these alchemical papers, large amounts of this kind of low-level information about otherwise hard-to-recognize shared structure can help us to see the shape of this work in much greater detail, and, perhaps, thereby make sense of larger trends in Newton's evolution as a practical chymist and a student of alchemy.⁸

Figure 10. Screen shot of the Latent Semantic Analysis Tool, available as part of the Chymistry of Isaac Newton project, revealing pairs of manuscript passages that highly overlap with cosine similarities of 0.9 and greater.

The Chymistry of
Isaac Newton

Text Symbols Search Manuscripts Advance Search

Home Browse Manuscripts Online Tools Educational Opportunities Academic Information Chymistry in the News Project Team

Site Index

Latent Semantic Analysis Tool

To begin, choose a search type and chunk size. After Step 3, hit the Continue button. [Instructions](#)

[Show Query Tool](#)

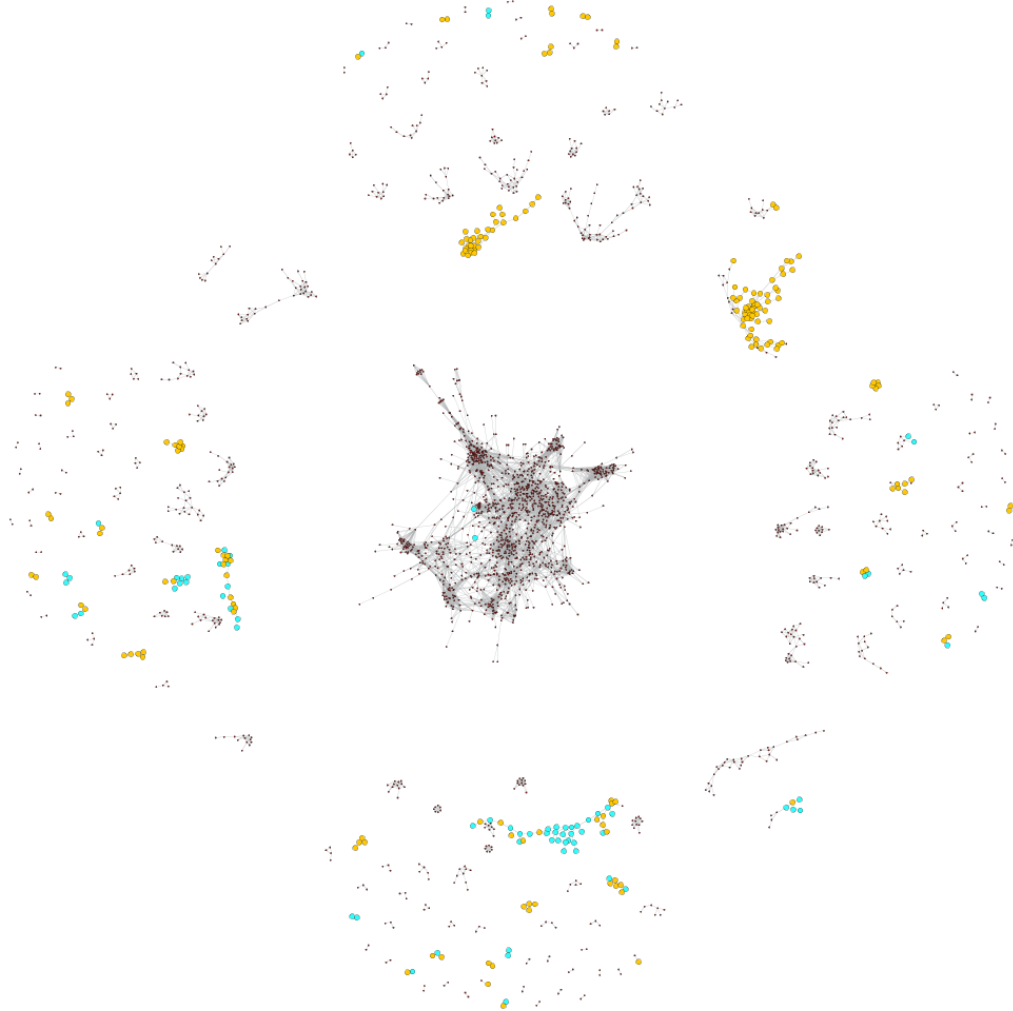
| | | |
|----------|----------------------------------|--------------------------------|
| 0.995290 | Babson MS. 417, f.014r_2 | Babson MS. 417, f.027r_3 |
| 0.994222 | Dibner MS. 1024 B SCDIRB, f.002r | Royal Society MM/6/5, f.003r_2 |
| 0.993009 | Babson MS. 417, f.017v_2 | Babson MS. 417, f.025v_2 |
| 0.992570 | Babson MS. 417, f.014r | Babson MS. 417, f.027r |
| 0.992047 | Keynes MS. 45, f.001r | Keynes MS. 45, f.001r_2 |
| 0.991551 | Babson MS. 417, f.012r_2 | Babson MS. 417, f.023r_2 |
| 0.989148 | Babson MS. 417, f.005v_2 | Royal Society MM/6/5, f.003r_2 |
| 0.988055 | Dibner MS. 1024 B SCDIRB, f.002r | Babson MS. 417, f.005v_2 |
| 0.986593 | Babson MS. 417, f.014v_2 | Babson MS. 417, f.024r |
| 0.985413 | Babson MS. 417, f.014r_2 | Babson MS. 417, f.024r_2 |

- 30 As the cosines decrease toward 0.7 and below, there can still be a fair amount of shared vocabulary in the two, but often less shared phrasing, if any at all. Inspection of these pairs can suggest that they belong to some subgenre because of the language, but Newton is clearly doing different work with the same language. In pairs much below 0.7, there may be apparent likenesses in the use of

one or two co-occurring terms that suggest a possible connection, but usually there is little else to support the idea. In LSA's spectrum-like vector representations of the text passages, even the co-occurrence of a few words in two passages must increase their cosine. It may be an indication of the general semantic similarity of these documents that the lowest observed cosine of any pair in the alchemical corpus was just above 0.4 and not lower.

- 31 LSA also gives us network graphs of all the passages as clouds of individual nodes, connected with other nodes only when their cosine exceeds a given threshold like 0.7, or 0.8, or 0.9, and these graphs help us to visualize the shape of the whole corpus, or pieces of it. The network graph (figure 11), for example, shows all the pairs of passages in Newton's alchemical manuscripts that have a cosine similarity of 0.7 or greater. It is a stable pattern because the underlying foundations—the collection of documents and the word counts in their tranches—do not change as a rule, but the graph shows that the whole collection does separate into many smaller semantic subnets. The graph can serve as a kind of map or atlas of locations where Newton worked with the same ideas across the entire corpus of 119 manuscripts.

Figure 11. Network graph produced by Chymistry of Isaac Newton's Latent Semantic Analysis Tool that shows pairs of passages in the corpus with a cosine similarity of 0.7 or greater.



- 32 The yellow nodes in [figure 11](#) come from Portsmouth Add. MS 3973 and the blue nodes come from Portsmouth Add. MS 3975, both of which record Newton's experiments in alchemy. They are central documents in the project's research program. The dense network of nodes at the center represents reading notes from traditional alchemical sources with their metaphorical language, while the outlying networks represent experimental notes and compositions written in the practical alchemical language emerging in laboratories in the late seventeenth century. We are interested in using the LSA tool to find recurring ideas, semantic structure, phrasing, and vocabulary. The list of results lets us work systematically through the pairs of documents, assessing their possible semantic relationships.

4.2 Using Latent Semantic Analysis to Track Citations

- 33 The passages shown side by side in figure 12 are Dibner 1024 B, f.2r, on the left, and Royal Society MM/6/5, f.3r2. The yellow highlighting identifies significant shared vocabulary and usually provides some sense of what might be shared. In this case, with a cosine of 0.989, there is a significant amount of overlapping text. The overlap is probably predictable from the titles alone, because both manuscripts address the work of French alchemist Pierre Jean Fabré, but here the LSA output shows us Newton referencing the same source materials in both documents, while only providing the citation in MM/6/5 (figure 12). The next question is whether there is obvious conceptual evidence to determine which of the two manuscripts is likelier to be the earlier composition based on the citations referenced.

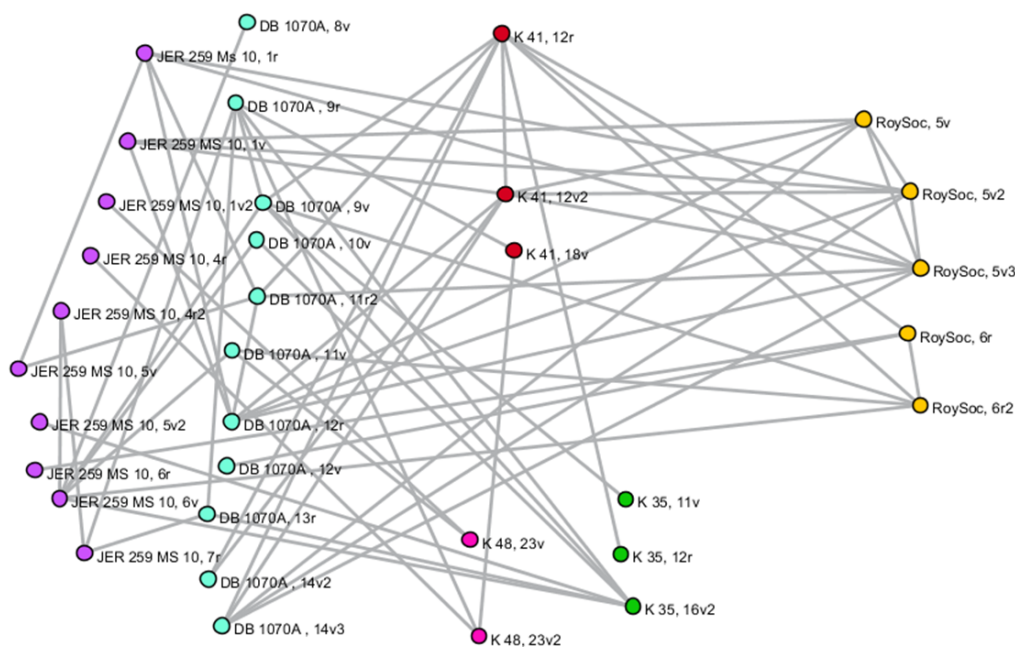
Figure 12. Results from the Latent Semantic Analysis Tool showing two passages from different manuscripts from Newton's alchemical corpus that have a similarity cosine of 0.989. The yellow highlighting indicates significant words that appear in both passages, which, in this case, is almost every word.

| | |
|---|--|
| <p>http://webapp-devel.dlib.indiana.edu/newton-dev/mss/dipl/ALCH00079/#f2r (new window)</p> <p>ALCH00079 Dibner MS. 1024 B SCDIRB, f.002r Ex Fabri Hydrographo Spagyrico.</p> <p>// Hæc ambo conjunguntur et ex conjunctione masculi huius & puellæ virginis nascitur semen seu fons noster. Ibi tunc temporis in matrimonio celebrando pondus est considerandum. Ex qualicumque enim pondere non oritur semen seu fons prolificus sed ex certo ac determinato pondere. Itaque par pondus utriusque materiae in nostro primo matrimonio celebrando est adhibendum. In Secundo autem matrimonio dum pars volatilis parti fixæ & permanenti conjungitur pondus etiam est perpendendum. Nam pars fixa necesse est ut dissolvatur ac diluatur et aqua fiat a parte volatili. Ideo major esse debet in quantitate pars volatilis quam ipsa pars fixa. Aliqui chemicorum ponunt decem partes aquæ super unam corporis fixi alii quatuor. Modò pars fixa dissolvatur a parte volatili hoc ego existimo vere sufficere. Nam deinde per coctionem perennem & continuam coagulaturs pars DEL[[fixa]]DEL ADD[[volatilis]]ADD a parte DEL[[volatili]]DEL fixa. Si sit maxima quantitas partis volatilis retardatur coagulatio unde his qui causam nesciunt contingit desperatio. Ego tamen non observavi pondus sed materias conjungi et maxima quantitate partis volatilis partem fixam dissolvi & deinde in balneo Mariæ aut levissimo cinerum calore superfluam volatilis materiae partem extraxi donec tenacem vidi materiam viscosam et valde nigram. Tunc vas repagulo satus firmo obsignavi et tempori ad coquendum tradidi, benedixitque Deus & incrementum dedit. Est etiam pondus observandum in multiplicatione. Nam opus alburni et rubrum perpotandum est et irrigandum suo spiritu seu fonte nostro crudo, sed puro et septies destillato. Ibi pondus est observandum ne elixir submergatur. Parcè admodum irrigandum ut materia cooperiatur tantum gladii spissitudine: quod multoties est</p> | <p>http://webapp-devel.dlib.indiana.edu/newton-dev/mss/dipl/ALCH00117/#f3r (new window)</p> <p>ALCH00117 Royal Society MM/6/5, f.003r_2 Extracts from "Fabre" (Pierre Jean Fabré) and other alchemical writers</p> <p>eorum substantijs. Poculum // ergo istud aqua est quæ in intimis eorum substantijs latitat. Hæc sola aqua fratrem nostrum et sororem nostram copulat et jungit ut ambo juncti progignant foetum nostrum Adonidem & Ganimedem qui enutriendus est ut ad perfectam ætatem deveniat. Faber Panctium p 674, 675</p> <p>Par pondus utri usque materiae in nostro primo matrimonio celebrando est adhibendum. In Secundo autem matrimonio dum pars volatilis parti fixæ & permanenti conjungitur pondus etiam est perpendendum. Nam pars fixa necesse est ut dissolvatur ac diluatur et aqua fiat a parte volatili. Ideo major esse debet in quantitate pars volatilis quam ipsa pars fixa. Aliqui chemicorum ponunt decem partes aquæ super unam corporis fixi alii septem alii quatuor. Modò pars fixa dissolvatur a parte volatili hoc ego existimo vere sufficere. Nam deinde per coctionem perennem & continuam coagulaturs pars volatilis a parte fixa. Si sit maxima quantitas partis volatilis retardatur coagulatio, unde his qui causam nesciunt contingit desperatio. Ego tamen non observavi pondus sed materias conjungi et maxima quantitate partis volatilis partem fixam dissolvi & deinde in balneo Mariæ aut levissimo cinerum calore superfluam volatilis materiae partem extraxi donec tenacem vidi materiam viscosam & valde nigram. Tunc vas repagulo satus firmo obsignavi & tempori ad coquendum tradidi, benedixitque Deus & incrementum dedit. Est etiam pondus observandum in multiplicatione. Nam opus alburni & rubrum perpotandum est et irrigandum suo spiritu seu fonte nostra crudo sed puro & septies destillato. Ibi pondus est observandum ne elixir submergatur. Parcè admodum irrigandum ut materia cooperiatur tantum gladii spissitudine quod multoties</p> |
|---|--|

- 34 The graph in figure 13 displays the location of textual similarities across six different manuscripts, including Royal Society MM/6/5 which we have just seen in part, where Newton worked over the same ideas at various times. All the connected pairs have cosine similarities of 0.9 or greater,

and share a considerable amount of text. Connected passages from the same document have the same color and are organized vertically in page order. The six different documents are arranged horizontally.

Figure 13. Network graph produced from the Latent Semantic Analysis Tool represents six documents that are found by LSA to share a large amount of text in certain sections of each of these documents. Each node represents a span of around 250 words of manuscript text, a lengthy passage with a quill and ink. In the passages shown in the graph, Newton rewrote the same material or revisited the same authors a number of times, and so this concatenation may represent a persistent locus of interest over a period of months or years.



- 35 Passages or nodes in [figure 13](#) that possess many connections will also likely contain direct quotations from the alchemical books that Newton was reading. The nodes or passages to which they are connected also often make the same citations, or paraphrase the quotations and contents found in the multiply connected passages. This graph therefore serves as a map of citation patterns across these six documents.

- 36 As it is everywhere else, the basic problem here is to discern the order of composition of these six documents. Sometimes Newton's editorial marks provide clues, but not as often as we would like. This is where we rely on the citations, bibliography, and the orthographic, watermark, and ink evidence to fill in the gaps in the analysis. The resulting clusters will not only have the benefit of showing the gradual increase of authoritative sources by Newton; they will also lay the groundwork for network analysis to reveal the connections that he saw among authors' works and ideas.
- 37 The citations constitute an independent order of evidence with its own rules that will have an impact on how to determine the order of composition of Newton's work in alchemy. When the improved and expanded citation analysis and the ink and paper evidence are all integrated with the semantically distinct clusters of passages and manuscripts that we have already discovered with our LSA tool, we should achieve a highly articulated view of how each cluster of related passages was constructed and gain a better sense of what Newton was doing in each.

BIBLIOGRAPHY

- Association of College and Research Libraries, and Library of Congress. *Descriptive Cataloging of Rare Materials*. 2007. Washington, D.C.: Library of Congress.
- Berry, Michael W., Susan T. Dumais, and Gavin W. O'Brien. 1995. "Using Linear Algebra for Intelligent Information Retrieval." *SIAM Review* 37 (4): 573–95. doi:10.1137/1037127.
- Bowers, Fredson. 2005. *Principles of Bibliographical Description*. New Castle, Delaware: Oak Knoll Press.
- Ferguson, John. 1906. *Bibliotheca Chemica: A Catalogue of the Alchemical, Chemical and Pharmaceutical Books in the Collection of the Late James Young of Kelly and Durris[...]* 2 vols. Glasgow: J. Maclehose and Sons.
- Fulton, John F. 1961. *A Bibliography of the Honourable Robert Boyle, Fellow of the Royal Society*. 2nd ed., Oxford: Clarendon Press.
- Gaskell, Phillip. 1972. *A New Introduction to Bibliography*, Oxford: Clarendon Press.
- Harrison, John. 1978. *The Library of Isaac Newton*. Cambridge: Cambridge University Press.

Landauer, Thomas K., and Susan T. Dumais. 1997. "A Solution to Plato's Problem: The Latent Semantic Analysis Theory of Acquisition, Induction, and Representation of Knowledge." *Psychological Review* 104 (2): 211–40. doi:10.1037/0033-295X.104.2.211.

TEI Consortium. 2017. *TEI P5: Guidelines for Electronic Text Encoding and Interchange*. Version 3.2.0. Last updated July 10. N.p.: TEI Consortium. <https://tei-c.org/Vault/P5/3.2.0/doc/tei-p5-doc/en/html/>.

NOTES

- 1 TAPoRware Text Analysis Tool: <http://tapor.ca/>.
- 2 Votantes Tools: <https://voyant-tools.org>.
- 3 Zotero: <https://www.zotero.org>.
- 4 Code for robustly converting RDF Zotero exports to TEI XML bibliographies, latest commit on January 25, 2012,
- 5 TEI Boilerplate: <http://teiboilerplate.org/>.
- 6 Latent semantic analysis (LSA) "is a mathematical method for computer modeling and simulation of the meaning of words and passages by analysis of representative corpora of natural text" (Thomas K. Landauer and Susan Dumais, "Latent Semantic Analysis," *Scholarpedia* 3, no. 11 (2008): 4356, revision 142371, http://www.scholarpedia.org/article/Latent_semantic_analysis.)
- 7 Chymistry of Isaac Newton LSA Tool: <http://webapp1.dlib.indiana.edu/newton/lisa/index.php>.
- 8 In the sixteenth and seventeenth centuries, the term "chymistry" was used interchangeably with "alchemy." Chymistry was a field that included not only the attempt to transmute base metals into gold and silver, but a host of other activities as well. Early modern chymists distilled alcoholic spirits from wine and beer, made mineral acids for use in metallurgy and mining, produced sophisticated pharmaceuticals, and fabricated pigments for artists, among other pursuits. One could almost say that chymistry combined pursuits linked nowadays to the disciplines of nuclear physics (at least in the case of transmutation), pharmacology, and industrial or technical chemistry.

AUTHORS

MERIDITH BECK MINK

Meridith Beck Mink is the Constellations program coordinator at the University of Wisconsin–Madison and a freelance consultant. She was a postdoctoral fellow specializing in data curation for early modern studies at Indiana University, where she contributed to the bibliography of the Chymistry of Isaac Newton project and consulted on digital scholarship in the Scholars' Commons. She has taught at Knox College and in Galesburg, Illinois, and was the lead researcher on the Council for Library and Information Resources' assessment of the National Digital Stewardship Residency (<https://www.clir.org/pubs/reports/pub173>).

MICHELLE DALMAU

Michelle Dalmau is an associate librarian and head of Digital Collections Services (DCS) at the Indiana University Libraries and co-director for the Institute for Digital Arts & Humanities (IDAH), a research center of the Office of the Vice Provost for Research, Indiana University Bloomington. As head of DCS, Michelle manages and coordinates digital library services for the Libraries and affiliated cultural heritage organizations across all IU campuses. As co-director for IDAH, Michelle fosters the development of digital arts and humanities infrastructure projects and initiatives through outreach, collaborative research and creative pursuits, consultation, professional development, and credit-bearing programs. Along with Beck Mink, Dalmau managed the bibliography component for the *Chymistry of Isaac Newton* project.

WALLACE HOOPER

Wallace Hooper is the project manager and programmer/analyst of the Chymistry of Isaac Newton Project, assistant scientist/scholar in the Department of History and Philosophy of Science and Medicine at Indiana University, and co-principal investigator with William R. Newman of NSF Project 1556846, *Multidimensional Chronological Analysis of Manuscript Corpora Using Isaac Newton's Chymical Papers as a Test Platform*. He holds a PhD in history and philosophy of science from Indiana University. He held two postdoctoral fellowships at the Museum of the History of Science in Florence, 1992–93 and 1995–96, where he attempted to establish the order of composition of the fragmentary notes on motion using analysis of concepts, orthographic changes, watermark distributions, and the elemental analysis of inks by proton-induced X-ray emissions (PIXE).

WILLIAM R. NEWMAN

William R. Newman teaches in the Department of History and Philosophy of Science at Indiana University. Most of his career has focused on the history of alchemy and chymistry (the expanded, early modern version of the field), premodern matter theory, and the long debate about the powers of art and nature that

alchemy helped to perpetuate. Newman is general editor of the Chymistry of Isaac Newton project (<http://www.chymistry.org>), which is editing Newton's chymical papers and replicating his processes where possible. His book *Newton the Alchemist* was recently published by Princeton University Press.

JAMES R. VOELKEL

James R. Voelkel is curator of rare books at the Other Library of Chemical History at the Chemical Heritage Foundation, and senior consulting editor at the Chymistry of Isaac Newton project. He holds a PhD in history of science from Indiana University, and a Certificate of Proficiency in Bibliography from the Rare Book School at the University of Virginia, Charlottesville. His research interests are in early modern science and the history of the book. He is the author of *The Composition of Kepler's Astronomia nova* (Princeton University Press, 2001) and *Johannes Kepler and the New Astronomy* (Oxford University Press, 1999).

JOHN A. WALSH

John A. Walsh is an associate professor in the School of Informatics and Computing at Indiana University. His research involves the application of digital and computational methods to the study of literary and historical documents. Walsh is an editor of digital scholarly editions (<http://petrarchiv.org>, <http://swinburneproject.org>, and <http://chymistry.org>). He developed the Comic Book Markup Language (CBML) (<http://www.cbml.org/>) for scholarly encoding of comics and graphic novels and TEI Boilerplate (<http://teiboilerplate.org/>) for publishing TEI documents on the web. Walsh's research interests include digital and computational literary studies; textual studies and bibliography; text technologies; book history; nineteenth-century British literature, poetry, and poetics; and comic books. <https://github.com/paregorios/Zotero-RDF-to-TEI-XML>.