Creating Time Capsules for Historical Research in the Early Modern Period: Reconstructing Trajectories of Plant Medicines

Wouter Klein
Freudenthal Institute,
History & Philosophy of Science,
Utrecht University
w.klein@uu.nl

Kalliopi Zervanou
Information & Computing Sciences
Utrecht University
k.a.zervanou@uu.nl

Marijn Koolen
Huygens ING
Amsterdam
marijn.koolen@huygens.knaw.nl

Peter van den Hooff
Freudenthal Institute,
History & Philosophy of Science,
Utrecht University
p.c.vandenhooff@uu.nl

Frans Wiering
Information & Computing Sciences
Utrecht University
f.wiering@uu.nl

Wouter Alink
Spinque B.V.
Utrecht
wouter@spinque.com

Toine Pieters
Freudenthal Institute, History &
Philosophy of Science, Utrecht
University
t.pieters@uu.nl

ABSTRACT

Historians argue that tracing early exchange, trade and uses of plant medicines (*materia medica*) can elucidate dynamics of drug trajectories in the early modern period. However, information on how these drug trajectories have evolved is hidden in large amounts of heterogeneous historical data. These data are in different formats, languages, genres and stages of digitization, thus creating huge challenges for i) accessing information, ii) presenting aggregations, trends and patterns, and iii) making the research process transparent and traceable for sharing, collaboration and validation. In this paper, we present the historical research platform of the Time Capsule system, from the user's perspective. We illustrate this with an extensive example, to show how the semantic integration of diverse historical data sources into a single linked data knowledge structure in our Time Capsule system has practical applicability for historical research.

KEYWORDS

linked data, historical research interfaces, historical information modeling, early modern history, drug trajectories

1 INTRODUCTION

Historical research requires analysis of disparate and incomplete data, where the building blocks for creating a narrative are scattered. The researcher aims beyond the most obviously relevant sources, towards discovering hints and unexpected connections in seemingly unrelated documents.

For this reason, we argue that for historians to profit from the full potential of information technology applications in their research, three general challenges should be addressed:

access relevant information is scattered across digital sources with limited metadata and content information. *The challenge is to enrich semantic metadata and integrate sources in a way that fits historical research methodology.*

presentation close reading of large and semantic interrelated digital data is not only constrained by the manual effort required, but may also obfuscate larger trends and patterns in such complex data. *The challenge is to establish the right form of distant reading that aggregates and summarizes large digital material in a meaningful and relevant way.*

validation digital information and digital research methods have accentuated the need for transparency. *The challenge is to establish methods for tracing and validating the research process and the arguments and claims based on it.*

Our work attempts to address these three challenges for digital historical research in a case study: the trajectories of botanical drug components in the Low Countries in the early modern period (ca. 16th-18th century). The reconstruction of these drug trajectories is vital for understanding the global exchange of knowledge and goods [20]. However, the respective historical evidence is hidden in large amounts of heterogeneous, dispersed and unrelated data.

In response to this need, Time Capsule[1] implements our solutions to information access, presentation and validation challenges. Time Capsule integrates a variety of botanical, linguistic, archaeological and historical data in a single historical knowledge graph. In this way, semantic links re-contextualize initially scattered data, while researchers are provided with a single access point for data

---

[1] http://www.timecapsule.nu

querying and analysis through multiple, interdisciplinary perspectives. Temporal and geographical visualizations present meaningful aggregates for researchers. Finally, system queries and results can be saved and shared in a *"time capsule"*, which allows users to retrace steps of the search process, to validate the usefulness of results and to share these with other researchers.

A general overview of the considerations, methodology and implementation of the semantic integration of our initially dispersed data sources, that form the knowledge backbone of the Time Capsule system, has been outlined in [40] and [41]. In this paper, we present the historical research platform of the Time Capsule system from the user's perspective. We provide an extensive example to illustrate how our semantic integration of diverse historical data sources, into a single linked data knowledge structure, has practical applicability for historical research.

This paper is structured as follows. In Section 2, we present current approaches to metadata and semantic integration of heterogeneous data. In Section 3, we provide an overview of our data sources and methodology for interrelating and enriching disparate data sources. Then we discuss Time Capsule's functionalities, from the user perspective in Section 4. The example, drawn from the history of pharmacy, concerns the reconstruction of a *drug trajectory* in the 17th and 18th centuries. Finally, in Section 5, we discuss conclusions and plans for future work.

2 BACKGROUND AND RELATED WORK

Data integration is a common issue in database research, because of the challenges associated with reconciling disparate database schemas and semantics, so as to support comprehensive data querying. An obvious solution is the integration of all data in a single repository, the so-called in-advance approach or data warehousing [37], but such static integration does not cater for updates in the original sources. More loose and dynamic approaches include database schemas mapping, [3, 29] and the use of semantic knowledge resources and ontologies to address potential semantic conflicts [4, 19, 25, 38].

For cultural heritage databases, different metadata standards are currently applied, both within and across different institutions. For example, bibliographic data is generally described using the MARC21 standard [28], while archival data is described by EAD [27]. There is an ongoing effort for standardization and integration of data formats and data sources through standard data models, such as CIDOC-CRM [9] and TEI [6].

Attempts for integration of diverse metadata schemas include mappings across schemas [5], mappings of metadata to ontologies, either ad-hoc manually developed ontologies [22], or cultural heritage standard ontologies, such as CIDOC-CRM [23], data conversion to a recommended metadata schema, such as SKOS [26], or conversion to a common data model, such as the Europeana Data Model [1]. However, metadata typically consists of bibliographic descriptions (e.g. author/creator, title), or physical descriptions (e.g. size, shape, material). Very often metadata related to the actual data semantics and content (e.g. persons or events mentioned), or metadata is missing, or inconsistent in the amount of detail provided. Thus, it is of limited use for non-curator experts to find the required information. Approaches that address such shortcomings

in metadata include substitution of the diverse metadata structures with automatic indexing [21] and enrichment of existing metadata with automatically acquired terms and relationships [42].

Another important challenge lies in data integration for access beyond warehouses. To address such a challenge requires an approach that allows integration and use on the Web of disparate data. Linked Data came as a response to this challenge [15], and it proposes the use of the RDF (Resource Description Framework), a general-purpose metadata language for representing information on the Web [36], often combined with a formal ontology specification in OWL (Web Ontology Language), a language built on top of RDF that is designed to represent complex knowledge about concepts, things and relations between things [35]. RDF and OWL provide a flexible way to describe abstract concepts and things in the form of RDF triples, which consist of a subject, an object and a predicate. The subject and object refer to things and the predicate specifies the relationship between them, thus making explicit the relationships among concepts and things, or things and their properties. For example, a botanical remedy is referred to in a medical book and is derived from a plant. In this way, the linked data approach manages to both integrate data sources and support data access beyond a single data repository, on the Web.

A detailed discussion on using linked data for cultural heritage is provided by Hyvönen [18]. An example of implementing content harmonization for a set of diverse resources has been investigated, for example, for a public online portal to cultural resources in Finland [24] and in the Europeana cultural multimedia platform [14]. Once data is structured according to linked data standards, it can be searched and queried semantically. Although this can be done with general-purpose tools (e.g. Hogan et al. [16]), domain-specific tools may be preferred in order to take full and direct advantage of linked data semantic annotations, as in the Time Capsule approach.

3 INTEGRATING DATA SOURCES

In Time Capsule we adopt a linked data approach for data source integration. According to Huggett [17], *"data are theory-laden and relationships are constantly changing depending on context [. . .] data are created by specific people, under specific conditions, for specific purposes, all of which inevitably leads to data diversity"*. For this reason, a particular challenge in this integration process lies in re-purposing and establishing explicit semantic relations among historical data in various digital formats, created for a variety of purposes and within the context of different research disciplines, such as linguistics, pharmacy, medicine, (ethno-)botany and history.

In our approach, as also illustrated in Figure 1, a domain ontology formally defining our set of concepts (e.g. drug components, plants, reference sources) and respective properties and relations is central. The individual data sources are then converted in RDF and key concepts are mapped to our domain ontology. In the conversion process, we semi-automatically identify ambiguities and spelling inconsistencies, filter out certain non-relevant information (e.g. curator information) and enrich the data with explicit relations among synonyms, term language, explicit temporal information and geographical coordinates information. The latter are semi-automatically acquired by geographical databases, such as
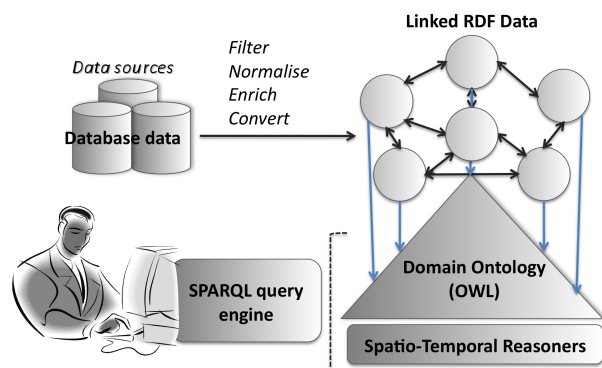
Figure 1: Time Capsule architecture overview.

Table 1: Time Capsule Data Sources

Dataset	Original format	Size (RDF triplets)
Pharmaceutical Historical Thesaurus	SQLite	72,339
Economic Botany Database	MS Excel	174,663
National Herbarium of the Netherlands	Darwin Core	44,348,662
Snippendaalcatalogus	proprietary DB	38,048
Boekhouder-Generaal Batavia	MySQL	9,468,155
RADAR	MS Access	2,494,117
Chronological Dictionary	CSV	19,296

GeoNames[2], or historical geographical resources, such as the list of the Dutch East India company trading posts[3].

In this section, we first give an overview of our data sources and their principal contribution to our knowledge structure and then, we discuss our domain ontology and the particular challenges related both to multi-disciplinarity and the historical nature of the domain that we had to address in designing it.

3.1 Data sources

In Time Capsule, we have semantically interconnected a variety of pharmaceutical, botanical and linguistic knowledge resources, primarily in the form of structured or semi-structured data. An overview of our current data sources is given in Table 1. In our system, these are also connected to relevant external linked data sources, such as DBPedia [2] and in principle these could be extended with any other available linked data source.

The *Pharmaceutical-Historical Thesaurus*, developed by the Foundation for Pharmaceutical Heritage, covers terminology related to botanical drug components, such as names and spelling variants of the various botanical remedies, plants of origin and parts of the plant used (e.g. seed, root) that were used in medicine in the past. The information is derived from pharmacopoeias, which are official pharmaceutical handbooks that apothecaries on a municipal level were obliged to use (issued between 1636 and 1795). These manuals dictated to apothecaries which drug components they were supposed to have in stock and the basic medicinal recipes they had to prepare with these. The thesaurus contains the lists of required botanical drug components from 10 pharmacopoeias. It also contains documentation information, such as the curator's name and various notes about the entries. It was developed using a proprietary thesaurus management tool[4] which internally stores the data as an SQLite database.

The *Economic Botany Database* of Naturalis Biodiversity Center encodes botanical information about approximately eight thousand plants in MS Excel format. The principal objective of the database lies in classifying information about uses (e.g. as dye, remedy, or food) of a given plant species. The data also includes botanical

---

[2]http://www.geonames.org/
[3]https://www.vocsite.nl/
[4]MultiTes Pro

taxonomic information, geographical origin and common names (English, Dutch and other), including descriptions of the physical specimen at Naturalis.

The data from *National Herbarium of the Netherlands* of Naturalis Biodiversity Center consists of three herbaria (the Wageningen, Leiden and Utrecht University collections), amounting to 2.8 million records of plant specimens. In addition to botanical plant name, family and genus, the National Herbarium provides information about a specimen's geographical origin, such as longitude, latitude and elevation. In the Time Capsule linked data structure, the National Herbarium provides the canonical, modern botanical name and taxonomic classification for a given plant species and the basis upon which other plant mentions in other data sources are resolved.

The *Snippendaalcatalogus* of from the Hortus Botanicus os Amsterdam encodes botanical information about Johannes Snippendaal's botanical garden in Amsterdam in 1646. This data source provides crucial historical botanical nomenclature and taxonomic information and mappings of Snippendaal's pre-Linnaean names, but also obsolete scientific names from the 19th century, and current scientific names.

The *Boekhouder-Generaal Batavia* is a mySQL database developed at Huygens ING, with information referring to the trade of commodities, as found in the accounting books (Boekhouder-Generaal) of the Dutch East India Company.

RADAR [34] is a relational archaeo-botanical database in MS Access format, developed at the Cultural Heritage Agency, with information about botanical macro-remains that were collected during archaeological excavations in The Netherlands.

Finally, the *Chronological dictionary* [33] contains information about a Dutch-language lemmas: their meaning, semantic class (based on a set of predefined, hierarchically structured concepts), etymology, chronology and references to their first written appearance in Dutch.

3.2 Time Capsule domain ontology

The Time Capsule domain ontology formally defines the concepts, their properties, and interrelations of the notions salient to our case study of drug trajectories. Nevertheless, it has been designed in such a way, that it should be extensible and applicable to other domains related to the so-called *Republic of Materials*, or historical study

Artemisia absinthium
Artemisia absynthium
Absinthium incanum, foliis compositis latiuscule multifidis
Gemeen alsem

Figure 2: Name variants for remedies, possibly derived from wormwood herb.

Figure 3: Historical concept mappings in the ontology.

Figure 4: Plant mention in references.

in the exchange and appropriation of materials and knowledge about them, in the early modern period. Currently, the principal concepts in the ontology revolve around three notions: *Naturalia*, or substances derived from nature, such as animals or plants and their parts; *Drug Components* and their properties; and *Reference Sources*, or sources of information for data, properties and relations stated in our instances, that originate from text documents or physical specimen references.

Apart from designing the ontology to be extensible and re-usable as much as possible based on input from experts on history, biology and botany, we had to address three particular challenges related to our domain: (i) variation, (ii) ambiguity and, (iii) scientific evolution of terms. *Variation* is a common phenomenon in scientific terminology (even more so in historical text sources), where spelling conventions are still uncommon and the scientific nomenclature is not yet standardised. Figure 2 illustrates an example of remedies found in our sources that may, or may not, refer to the same drug component. *Ambiguity* is related to variation and is exacerbated by underspecification, vagueness and uncertainty in historical sources. For example, a reference to *Radix Chelidonii* or *Chelidonii* does not clarify whether the remedy originates from the plant called Chelidonii majoris or Chelidonii minoris in the past. It might also be the case that a drug component is associated, under a single name, to different plants or plant parts (e.g. root or bark), in different reference sources. Finally, *scientific evolution* of concepts is a particular challenge, that is mostly related to the domain of the history of science. The evolution of human knowledge and science from e.g. the early modern period to modern times, inherently affects both the concepts and the terminology that is used, as well as the organisation and classification of knowledge itself. In designing our knowledge structure we therefore had to account for multiple "versions" of knowledge about the world at different moments in time.

Figure 3 illustrates our approach to the challenge of scientific evolution. The concepts associated with *Naturalia* (in dark red), either *Plantae, Animalia* or *Mineralia*, originate from the classification of knowledge in the early modern period, whereas the taxonomy *Kingdom > Order > ... Plant Species* is based on the modern botanical classification (in light green). In our ontology, we model both knowledge versions in the same knowledge structure and attempt to provide mappings/links, where possible, across time, so that a historian or other user can access all relevant data for longer time frames.

In order to address variation and ambiguity, we have adopted an ontology structure, as illustrated in Figure 4. This structure (i) separates statements about a given concept in a reference source
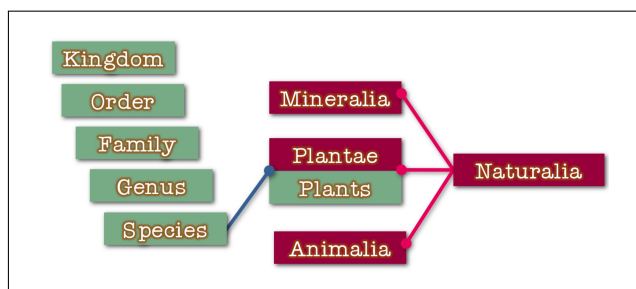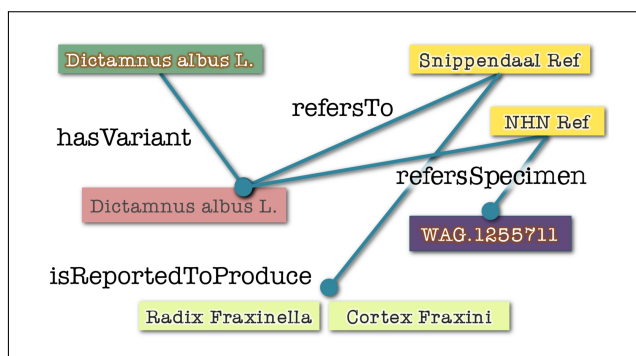
from the concept instance itself; (ii) does not resolve ambiguity, but rather follows a fuzzy classification, which allows an instance of e.g. a drug component to be associated to multiple plants of origin. In this way, we attempt to stay faithful to any information stated in a reference source, while reducing as much as possible the inherent bias in the knowledge modeling of our system and leaving ambiguity and variation open to a researcher's interpretation.

In the example in Figure 4, the instance of a plant concept (*Dictamnus albus L.*), relates to the mention Dictamnus albus L. in the Snippendaal Catalogue and the National Herbarium references, via the property hasVariant. According to the Snippendaal mention, this plant species, Dictamnus albus L. isReportedToProduce two remedies, one drug component originating from the root of the plant, *Radix Fraxinella*, and one originating from the bark, namely *Cortex Fraxini*. In this example, we can also see that there is a physical specimen of the plant mentioned in the National Herbarium, a specimen originating from the University of Wageningen collection, with ID WAG 1255711 (and respective geographical coordinates of the specimen origin, which are not included in this figure).

4 APPLICATION DOMAIN AND CASE STUDY

4.1 Drug trajectories

The domain of application for this work relates to research in the history of pharmacy. In particular, we focus on cultural heritage data sources that reveal (parts of) *trajectories* of botanical drug components in the Low Countries, from the sixteenth century, when natural drug components from the East and West Indies
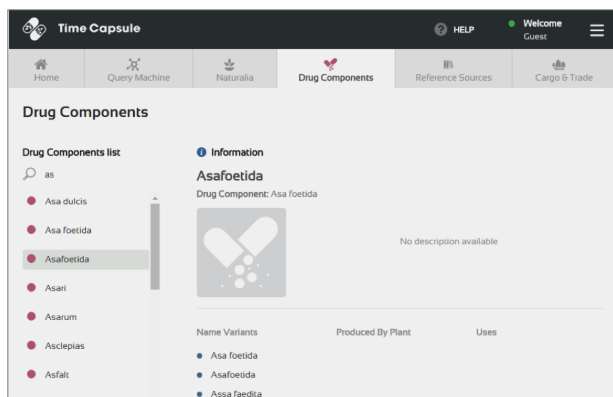
Figure 5: Results for Asafoetida in the `Drug components` **tab.**

Figure 6: Results for Asafoetida in the `Naturalia` **tab.**

started penetrating Europe, until the introduction of chemical and synthetic drugs, roughly in the mid-19th century.

Within historical research, *drug trajectories* denote developmental processes of remedies, such as growing economic importance of a drug because of a steady supply; scientific interest shifts because of ongoing research in therapeutic efficacy and side effects; drug acceptance or rejection by (segments of) society for financial, religious, ethical or other reasons. In other words, drug trajectories are specific aspects of the history of drugs, which can be studied diachronically, i.e. tracing one such aspect over a longer period of time) or synchronically, i.e. studying/comparing the interplay of different spatial contexts at a given point in time. Studying the history of pharmacy through the perspective of trajectories of individual drugs is a fairly recent approach that allows researchers to trace larger developments in the history of therapeutics. For example, the adoption of a given drug component may only partly result from its therapeutic qualities. Equally important in this process may be non-medical factors, such as public acceptance, marketing and availability, factors revealing the dynamics of the medical market during the historical time frame under consideration [11, 13, 20].

In our work, we are particularly interested in the trajectories of drug components from the East and West Indies and how these relate to networks of trade and knowledge circulation in the early modern Low Countries. In this period, natural history is considered the central focus of intellectual and commercial activity [7] and the circulation of knowledge related to medicinal components from across the world (e.g. South-East Asia, South Africa and the New World), to the Low Countries region, has drawn particular research attention in the history of pharmacy (see e.g. [7, 8, 10, 12, 31, 32, 39]).

Time Capsule is meant to provide the aggregated data required for new information to surface which is difficult for an individual researcher to produce by means of conventional historical research alone. Moreover, our system is aimed at assisting the researcher in analyzing these vast amounts of data, by providing spatio-temporal visualizations that give clues for a meaningful interpretation. In this way, existing historical narratives will be enriched.

In designing an interface and an online platform for information access, one should consider that the potential users may not be domain or system experts. It is often difficult for non-expert users to perform queries, either because they are unfamiliar with the
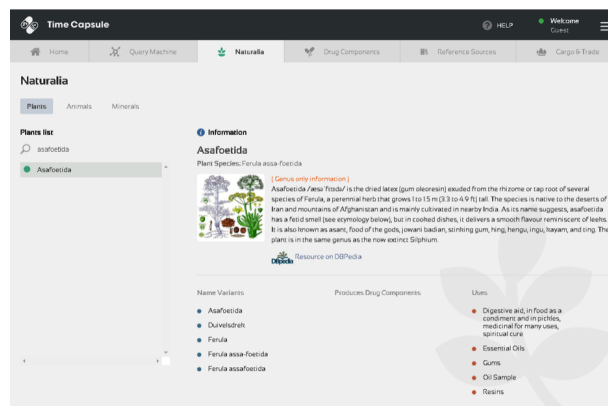
required terminology, or with the available information and the underlying data model. Our solution to this issue lies in providing two querying strategies in our Time Capsule research platform interface: one that supports *faceted exploratory search and browsing* of information by means of links, and keyword auto-completion suggestions; and one that supports the *creation of ad hoc queries*. The exploratory search mode is intended to engage a wider audience and reveal to both expert and non-expert users the content and structure of the underlying data. The user interface allows easy construction of ad hoc queries, by means of a user-friendly query wizard, which supports the creation of RDF SPARQL queries [30].

4.2 Researching Asafoetida

To illustrate the usefulness and user-friendliness of our system, we discuss an extensive example, where a researcher focuses on information for a drug component called *Asafoetida*, a resinous gum that was used in pharmacy from the 16th century onwards. Due to its strong smell, it was known as *Devil's dung* in many European languages. In this example, we want to know how its historical trajectories evolved over time.

The first and most obvious way of tracing information about Asafoetida in the Time Capsule system is by searching for it in the `Drug Components` tab (see Figure 5). The `Drug Components` tab is part of the faceted search interface which the Time Capsule platform provides for browsing, in this case, all system knowledge related to drug components. With the auto-complete function in the search text box (on the left hand side of the screen, indicated by the magnifying glass), several spelling variants for this drug component pop up (the list with red bullets below the search text box). Clicking on any of them will show the information found in the reference sources for this drug component: six different spelling variants; four *Reference sources* that mention this substance; a map that highlights the `Publication location` of the *References*; and a list with a total of 25 references to this drug component, which are all derived from one data set, the *Pharmaceutical-Historical Thesaurus*.

At this stage, no information is shown in the `Produced By Plant` header, which means that there is not any explicit information, i.e. relation, in our knowledge graph about the plant species
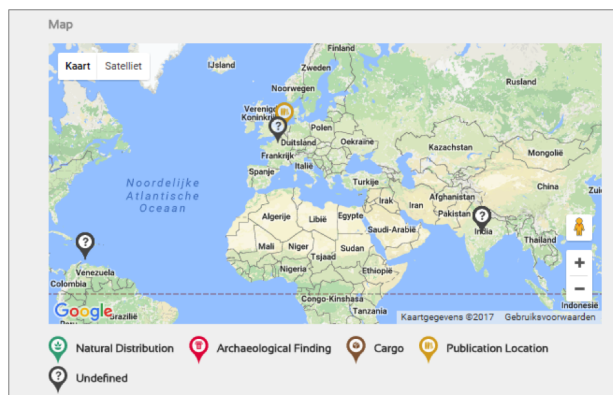
Figure 7: Map with geolocations for Asafoetida plant.

Figure 8: Results for Asafoetida in the `Cargo & Trade` tab.

from which the drug component Asafoetida is derived. However, the system provides other options to proceed. As illustrated in Figure 6, the query for *Asafoetida* also gives results in the `Naturalia` tab, namely the part of the faceted search in the Time Capsule platform that allows browsing and search of information related to all plant species information in our system. We discover in the `Naturalia` tab section that *Asafoetida* is also a spelling variant for the plant from which the drug component with the same name derives: *Ferula assa-foetida*. The result for this query shows several spelling variants for the plant, which are similar but not identical to the spelling variants for the drug component Asafoetida. Unsurprisingly, there are not any drug components for this plant mentioned under the `Produces Drug Components` header: we already saw that the explicit relation/link between this plant and its respective drug component is absent.

However, we discover additional information next to the `Name Variants`, namely a list of reported `Uses`. These data, originating from the Economic Botany Database, provide a number of possible contexts in which we might encounter Asafoetida. Apart from its use in medicine, Asafoetida is used in food, in spiritual cures and as an essential oil. Moreover, above the `Name Variants` we can find a text description about the plant genus and a photo, both resulting from linking our Time Capsule data to DBPedia linked open data. DBPedia informs us that the Asafoetida plant genus originates from Central Asia. This makes the historical use of Asafoetida in European pharmacy all the more interesting, because it would have required an intercontinental supply chain.

The information from DBPedia is corroborated by the integrated data in Time Capsule, as visualized in the map on the same page with results, illustrated in Figure 7. Apart from some expected locations in Europe, and one undefined pointer in the Dutch Caribbean, we also see one location pointer in India, where we would expect a `Natural Distribution` location, i.e. location of plant origin mentioned in our botanical specimen data.

At this stage, we need to explore in more detail the Asian connection that we came across in several of the data collections in the Time Capsule system, and investigate whether we can discover any additional information about the complex commercial trajectory of Asafoetida in the past. To do so, we can use once more the query *Asafo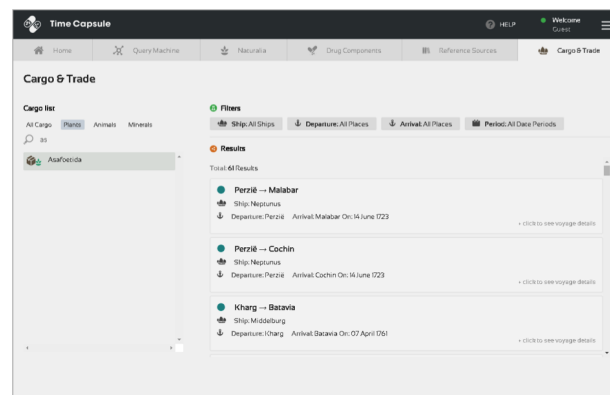etida* to browse in the `Cargo & Trade` tab. This is another part of the faceted search in the Time Capsule platform that allows browsing and search of information related to cargo carried by the ships of the Dutch East India Company in the 18th century. In this case, we want to investigate whether Asafoetida is found in any of these cargo records, as a trade item.

The results are illustrated in Figure 8. The green leaf next to the cargo icon in the search bar indicates that we are indeed dealing with a vegetable substance, namely a cargo item classified as a natural, plant substance. This leaf symbol is the result of the integrating and enrichment process that the respective *Boekhouder-Generaal Batavia* data set underwent in Time Capsule, that specifically marks cargo items such as *Naturalia* (among other classified materials recognised in this process). The query result shows 61 journeys where Asafoetida was part of the cargo. We know this result to be somewhat inaccurate, because journeys are sometimes listed more than once if more than one departure/arrival location is mentioned in the original data (e.g. a journey departing from Kharg in Persia is listed twice, once departing from Kharg and once from Persia).

These results show that most ships carrying Asafoetida follow one of two routes: either within the South-East Asian region, where only the place of destination is a Dutch settlement, or between South-East Asia and the Dutch Republic. In other words, the Trade & Cargo results provide convincing data for further analysis of Asafoetida's main supply route. The substance was apparently purchased in a region not controlled by the Dutch (India), then shipped to a Dutch emporium (usually Sri Lanka, or Batavia), and then shipped to the Dutch Republic. Details of one of these 61 journeys, as illustrated in Figure 9, departing from the Coromandel coast in India to Jaffna on Sri Lanka on 31 August 1767, reveals that this is an example of the South-East Asian trade route described above. The ship carried, among other things, a modest amount of 167 pounds of Asafoetida: modest, because other ships carried up to thousands of pounds of this substance.

We now know that Asafoetida was shipped regularly from South-East Asia to the Dutch Republic by the Dutch East India Company, in substantial quantities, over the course of the 18th century. We can now try to determine to what extent there was demand for this substance, not by browsing, but by using the `Query Machine`, (see second tab in Figure 5). The `Query Machine` implements, on the
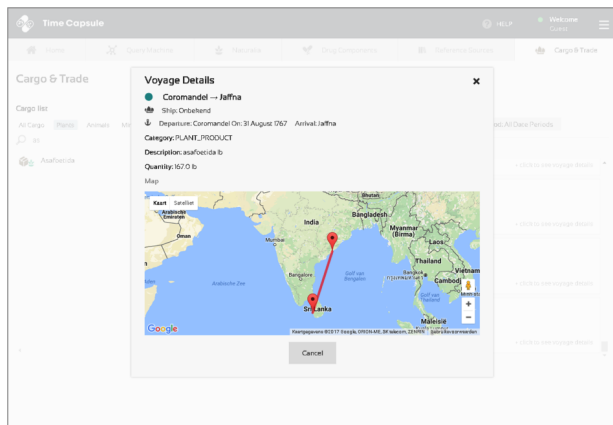
Figure 9: Single shipment detail for Asafoetida in the `Cargo & Trade` **tab.**

Time Capsule interface, the part where the user may form ad hoc RDF SPARQL queries, using a user-friendly query wizard.

In the `Query Machine` we formulate the custom query: *"List All Reference Source(s) that mention Asafoetida"*. The system gives six results, all of them from pharmacopoeia references, among which two editions of the Pharmacopoeia of Amsterdam (1636 and 1660). We know this result to be incomplete, because we know from our domain expertise that Asafoetida is mentioned in every early modern pharmacopoeia. For this reason, we would expect at least 10 pharmacopoeias to surface in our results, since the Pharmaceutical Historical Thesaurus data set contains data from these pharmacopoeias. However, although the results are not as precise as we thought, they still provide us with enough input to pursue our research further: the results indicate that apothecaries were supposed to use Asafoetida in Amsterdam in the mid-17th century. Since we know that the pharmacopoeia of Amsterdam was the official guide in many cities, our expectation is supported that both supply and demand were substantial.

In this example we jumped from one tab to another, using the results acquired at each step as input to extend our search strategy and knowledge about our main subject. In the process, we stumbled on deficiencies that inevitably show up in the data, but we never reached a dead end. Of course, the heterogeneous breadcrumbs we pieced together do not constitute a complete narrative about the history of Asafoetida. There are many more dimensions than can currently be traced in our system, such as the various Uses we came across, other than as a medicinal substance. But the fact that these bits and pieces could be found within one system has helped us. It allowed us to rapidly assemble very different kinds of information in a way that sustains the intuitive strategy of the researcher. It also showed connections that were not in the original data sources, and hence would not have surfaced if a researcher had sifted through the individual data sources manually. In that way, the Time Capsule system provides a substantial improvement of the researcher's digital toolkit, allowing him to save time while retaining an exploratory search strategy, and to stay in command of the research process, while outsourcing time-consuming parts to a digital search mechanism. The reduced effort in piecing this information together is made possible by the substantial effort of high-quality data integration, but this needs to happen only once to support many different explorations.

5   DISCUSSION AND FUTURE WORK

In this paper, we presented the Time Capsule system and research platform which contributes to digital history in several ways. It offers a new historical data model and knowledge resource and a novel approach for historical researchers. The system is built in such a way that it may not only be useful for researchers who are interested in early modern drug trajectories from different angles, but also for researchers interested in building, using and sharing similar research infrastructures.

Through semantic data integration, historians can query and analyze early modern historical data as a single knowledge graph and contextualize relevant information from various sources and aspects (botanical, linguistic, historical, ethnological, archaeological). Another important aspect in contextualisation lies in making explicit the relation of data to the dimensions of space and time, thus allowing the virtual spatio-temporal "reconstruction" of knowledge-transfer trajectories. Combining knowledge contained in different data resources requires a substantial research effort, due to the diversity and ambiguity of the original data. If different data sources are available to the researcher, but they are not interlinked, then the added value of having access to more data is rather limited, because the researcher has to continuously perform mapping and disambiguation when switching between resources. Domain knowledge modeling in the Time Capsule ontology and data source integration in an interlinked knowledge graph produces richer query results, that are less time-consuming to obtain; with transparent and retraceable data provenance; more trustworthy for a user than their original data constituents due to context; and thus better suited as building blocks for constructing a historical narrative.

Regarding the analysis of trends and patterns across data sets, distant reading methods stand or fall with the completeness and quality of the underlying data. Global patterns are hard to trace, given that the available data sets are limited to specific organizations that gathered information relevant and available to them. Obtaining and meaningfully adding information on the quality of digitization and data integration is far from trivial, as certain quality issues may remain hidden under the surface and there is currently little consensus on a proper methodology of digital sources and digital tool criticism. That is why domain expertise is essential in both digitization and integration stages. However, for data constituents that provide a fairly accurate and complete image (e.g. when an organization like the VOC applied a uniform, sustained method of describing the plants destined for shipment), the links among the data allow analysis of patterns beyond what would be possible for an individual researcher to query by himself. As more data sets are added, in combination with information about their completeness and their digitization and integration quality, the potential for pattern and trend analysis increases significantly.

In the current version of our system, structured database sources have been processed and integrated, and made accessible via a graphical interface for querying the aggregated data. We are finalizing a user study regarding the interface, and a detailed report on

the data modeling and integration process which will be published separately. Moreover, additional work is required in optimizing the various data visualizations, which are key to understanding such complex data structures.

Future work includes the extension of our system to additional collections. The Time Capsule knowledge base can be easily enriched with mineral and animal drug components and other substances beyond the domain of Naturalia, so as to cover extensive information about the *Republic of Materials*. The platform offers many possibilities to connect to other data sets in the future, such as the Sound Toll Registers, that could provide us information on how, for example, Asafoetida found its way across Europe after it came from Asia to the Netherlands. Finally, we plan to investigate ways to connect further with unstructured data sets. Historical documents provide a wealth of additional information, as well as valuable contexts for interpreting structured data. We plan to use our existing knowledge resource for processing and linking unstructured historical text to our data model.

ACKNOWLEDGMENTS

REFERENCES

[1] Europeana Network Association. 2016. *Europeana Data Model – Mapping Guidelines v2.3.* http://pro.europeana.eu/page/edm-documentation

[2] S. Auer, C. Bizer, G. Kobilarov, J. Lehmann, Ri. Cyganiak, and Z. Ives. 2007. DBpedia: A Nucleus for a Web of Open Data. In *Proceedings of the 6th International The Semantic Web and 2Nd Asian Conference on Asian Semantic Web Conference (ISWC'07/ASWC'07)*. Springer-Verlag, Berlin, Heidelberg, 722–735. http://dl.acm.org/citation.cfm?id=1785162.1785216

[3] C. Batini, M. Lenzerini, and S. B. Navathe. 1986. A comparative analysis of methodologies for database schema integration. *Comput. Surveys* 18, 4, 323–364. https://doi.org/10.1145/27633.27634

[4] J.A. Blake and C.J. Bult. 2006. Beyond the data deluge: data integration and bio-ontologies. *Journal of Biomedical Informatics* 39, 3 (June 2006), 314–320.

[5] L. Bountouri and M. Gergatsoulis. 2009. Inter-operability between archival and bibliographic meta-data: An EAD to MODS crosswalk. *Journal of Library Metadata* 9, 1-2 (2009), 98–133.

[6] TEI Consortium. 2014. TEI P5: Guidelines for Electronic Text Encoding and Interchange. Version 2.7.0.:16th September 2014. TEI Consortium. (2014). http://www.tei-c.org/Guidelines/P5/

[7] H.J. Cook. 2007. *Matters of exchange: commerce, medicine, and science in the Dutch Golden Age.* Yale University Press, New Haven.

[8] R. Dauser, S. Hächler, M. Kempe, F. Mauelshagen, and M. Stuber (Eds.). 2008. *Wissen im Netz: Botanik und Pflanzentransfer in europäischen Korrespondenznetzen des 18. Jahrhunderts.* Akademie Verlag, Berlin.

[9] M. Doerr. 2003. The CIDOC conceptual reference module: an ontological approach to semantic interoperability of metadata. *AI magazine* 24, 3 (2003), 75.

[10] S. Dupré and C. Lüthy (Eds.). 2011. *Silent messengers: the circulation of material objects of knowledge in the early modern Low Countries.* LIT Verlag, Berlin.

[11] C. Friedrich and W.-D. Müller-Jahncke (Eds.). 2009. *Arzneimittelkarrieren: zur wechselvollen Geschichte ausgewählter Medikamente: die Vorträge der Pharmaziehistorischen Biennale in Husum vom 25-28. April 2008.* Wissenschaftliche Verlagsgesellschaft, Stuttgart.

[12] E. van Gelder (Ed.). 2012. *Bloeiende kennis: groene ontdekkingen in de Gouden Eeuw.* Verloren, Hilversum.

[13] M. Gijswijt-Hofstra, G.M. van Heteren, and E.M. Tansey. 2002. *Biographies of remedies: drugs, medicines and contraceptives in Dutch and Anglo-American healing cultures.* Clio medica 66, Amsterdam: Rodopi.

[14] B. Haslhofer, E. Momeni, M. Gay, and R. Simon. 2010. Augmenting Europeana content with linked data resources. In *Proceedings of the 6th International Conference on Semantic Systems (I-SEMANTICS '10)*, A. Paschke, N. Henze, and T. Pellegrini (Eds.). ACM, New York. https://doi.org/10.1145/1839707.1839757

[15] T. Heath and C. Bizer. 2011. *Linked Data: Evolving the Web into a Global Data Space (1st edition). Synthesis Lectures on the Semantic Web: Theory and Technology.* Morgan & Claypool. 1–136 pages.

[16] A. Hogan, A. Harth, J. Umbrich, S. Kinsella, A. Polleres, and S. Decker. 2011. Searching and browsing Linked Data with SWSE: The Semantic Web Search Engine. *Web Semantics: Science, Services and Agents on the World Wide Web: JWS special issue on Semantic Search* 9, 4 (2011), 365–401. https://doi.org/10.1016/j.websem.2011.06.004

[17] J. Huggett. 2012. Promise and paradox: accessing open data in archaeology. In *Proceedings of the Digital Humanities Congress 2012 – Studies in the Digital Humanities*, Clare Mills, Michael Pidd, and Esther Ward (Eds.). HRI Online Publications, Sheffield, UK. http://www.hrionline.ac.uk/openbook/chapter/dhc2012-huggett

[18] E. Hyvönen. 2009. *Semantic Portals for Cultural Heritage.* Springer, Berlin Heidelberg, 757–778.

[19] V. Kashyap and A. Sheth. 1994. Semantics-based information brokering. In *Proceedings of the third international conference on Information and knowledge management (CIKM '94)*, N.R. Adam et al. (Ed.). ACM, New York, NY, USA, 363–370. https://doi.org/10.1145/191246.19130

[20] W. Klein and T. Pieters. 2016. The Hidden History of a Famous Drug: Tracing the Medical and Public Acculturation of Peruvian Bark in Early Modern Western Europe (c. 1650-1720). *Journal of the History of Medicine and Allied Sciences* 71, 4 (June 2016), 400–421. https://doi.org/10.1093/jhmas/jrw004

[21] M. Koolen, A. Arampatzis, J. Kamps, V. de Keijzer, and N. Nussbaum. 2007. Unified access to heterogeneous data in cultural heritage. In *Proceedings of RIAO '07*. Pittsburgh, PA, USA, 108–122.

[22] S.-H. Liao, H.-C. Huang, and Y.-N. Chen. 2010. A semantic web approach to heterogeneous metadata integration. In *Proceedings of ICCCI '10 (LNCS)*, Pan et al. (Ed.), Vol. 6421. Springer, Kaohsiung, Taiwan, 205–214.

[23] I. Lourdi, C. Papatheodorou, and M. Martin Doerr. 2009. Semantic integration of collection description: Combining CIDOC/CRM and Dublin Core collections application profile. *D-Lib Magazine* 15, 7-8 (2009).

[24] E. Makela, E. Hyvönen, and T. Ruotsalo. 2012. How to deal with massively heterogeneous cultural heritage data: lessons learned in CultureSampo. *Semantic Web* 3, 1 (January 2012), 85–109. https://doi.org/10.3233/SW-2012-0049

[25] E. Mena, A. Illarramendi, V. Kashyap, and A. Sheth. 2000. OBSERVER: An approach for query processing in global information systems based on interoperation across pre-existing ontologies. *Distributed and Parallel Databases* 8, 2 (2000), 223–271. https://doi.org/10.1023/A:1008741824956

[26] A. Miles and J.R. Pérez-Agüera. 2007. SKOS: Simple Knowledge Organisation for the Web. *Cataloging & Classification Quarterly* 43, 3-4 (2007), 69–83. https://doi.org/10.1300/J104v43n03_04

[27] Library of Congress. 2002. Encoded archival description (EAD), version 2002. (2002). http://www.loc.gov/ead/

[28] Library of Congress. 2010. MARC standards. (2010). http://www.loc.gov/marc/index.html

[29] C. Parent and S. Spaccapietra. 1998. Issues and approaches of database integration. *Communications of the ACM* 41 (May 1998), 166–178. https://doi.org/10.1145/276404.276408

[30] E. Prud'hommeaux and A. Seaborne. 2008. *SPARQL Query Language for RDF – W3C Recommendation.* Technical Report. W3C. https://www.w3.org/TR/rdf-sparql-query/

[31] L. Schiebinger and C. Swan (Eds.). 2005. *Colonial botany: science, commerce, and politics in the early modern world.* University of Pennsylvania Press, Philadelphia.

[32] S. Snelders. 2012. *Vrijbuiters van de heelkunde: op zoek naar medische kennis in de tropen 1600-1800.* Atlas, Amsterdam.

[33] N. van der Sijs. 2001. *Chronologisch Woordenboek.*

[34] H. van Haaster and O. Brinkkemper. 1995. RADAR, a Relational Archaeobotanical Database for Advanced Research. *Vegetation History and Archaeobotany* 4, 2 (1995), 117–125.

[35] W3C. 2012. OWL 2 Web Ontology Language Document Overview (Second Edition), W3C Recommendation 11 December 2012. (2012). http://www.w3.org/TR/2012/REC-owl2-overview-20121211/

[36] W3C. 2014. RDF Schema 1.1, W3C Recommendation 25 February 2014. (2014). http://www.w3.org/TR/2014/REC-rdf-schema-20140225/

[37] J. Widom. 1996. Integrating heterogeneous databases: Lazy or eager? *Computing Surveys* 28A, 4 (December 1996). https://doi.org/10.1145/242224.242344

[38] G. Wiederhold. 1992. Mediators in the Architecture of Future Information Systems. *Computer* 25, 3 (March 1992), 38–49. https://doi.org/10.1109/2.121508

[39] R. Wilson. 2000. *Pious traders in medicine: a German pharmaceutical network in eighteenth-century North America.* Pennsylvania State University Press, University Park.

[40] K. Zervanou, W. Klein, P.C. van den Hooff, M. Bron, F. Wiering, and T. Pieters. 2015. Creating Time Capsules for Colonial Botanical Drugs in the Early Modern Low Countries. In *Proceedings of the Annual Conference of the Alliance of Digital Humanities Organizations (DH2015)*. Sydney, Australia.

[41] K. Zervanou, W. Klein, P.C. van den Hooff, F. Wiering, and T. Pieters. 2017. Linking multi-disciplinary data sources for a historical research platform. In *Proceedings of the 4th DHBenelux Conference.* Utrecht, The Netherlands.

[42] K. Zervanou, I. Korkontzelos, A. van den Bosch, and Ananiadou S. 2011. Enrichment and structuring of archival description metadata. In *Proceedings of the Fifth ACL-HLT Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities (LaTeCH-2011)*, K. Zervanou and P. Lendvai (Eds.). Portland, OR, USA, 44–53.