<table>
<thead>
<tr>
<th><strong>Title</strong></th>
<th>Mapping Drugs across Epistemic and Geographic Domains: A case study for Early Medieval China</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Author(s)</strong></td>
<td>Stanley-Baker, Michael; Ho, Brent Ho-leung</td>
</tr>
<tr>
<td><strong>Date</strong></td>
<td>2015</td>
</tr>
<tr>
<td><strong>URL</strong></td>
<td><a href="http://hdl.handle.net/10220/44783">http://hdl.handle.net/10220/44783</a></td>
</tr>
<tr>
<td><strong>Rights</strong></td>
<td>© 2015 The author(s). This is the author created version of a work that has been peer reviewed and accepted for presentation in the 6th International Conference of Digital Archives and Digital Humanities. It incorporates referee’s comments but changes resulting from the publishing process, such as copyediting, structural formatting, may not be reflected in this document.</td>
</tr>
</tbody>
</table>
Mapping Drugs across Epistemic and Geographic Domains in Early Medieval China

以電腦地圖科技跨越六朝草藥於地理及知識領域的分界

Paper for presentation at 6th International of Digital Archives and Digital Humanities, December 2nd, 2015.

By: Michael Stanley-Baker & Brent Ho-leung Ho

Chinese Abstract:

我們如何更好的理解傳統中國醫學市場的複雜性和多樣性，以及醫學知識的各種不同的轉化和轉交方式？如何比較道家與佛家等不同團體在中國史醫學中扮演的角色及它們彼此競爭的方式？大多數過去的研究是基於軟事性的例子來探究，因為總體性地分析道藏與大藏經等大量文獻需要很多專業知識，很少醫學史家能憑一己之力嘗試。

馬克斯普郎克科學史研究所正在建設的平台將運用現有的數位方法來總體性地分析道藏和大藏經文獻。它將能在此數據文庫檢索多詞詞集，產出時間、文類、和地區上的統計分析。此外，該平台將能在時空的維度上視覺化呈現數據，從而使學者能夠識別出藥物知識被創造的區域和時間。為了準備文獻供此平台分析，我們也將運用一個新型半自動文本標記系統的MARKUS來減少標記的時間。

Project Questions and Rationale

It’s a long-held cultural belief that Daoists, or more specifically, transcendents were among the primary stakeholders in the early Chinese drug market. They held secret drug recipes, they made money by picking plants in the mountains and selling them in markets, and they used this knowledge not only to heal individuals, but to achieve miraculous longevity. How can we assess this claim, and come to a better understanding of the roles Daoists played in the medical marketplace, and the medical practice of religious figures generally?

So far, evidence used in these arguments has been anecdotal. When we read histories of pharmacology, or drug knowledge, they rarely include materials from the Daoist canon. Pharmacological history begins with the hazy formation of the Shennong Bencao jing 神農本草經 (SNBCJ) in roughly the second century CE, and moves on up through history through a series of centralized, standard texts. Arguments about Daoism are usually based on the fact that the earliest layer, the SNBCJ, is organized into Upper, Middle and Lower quality drugs 上中下, for nourishing life 養命, nourishing inner nature 養性, and for curing disease 治病. Many scholars contend that this reflected a “Daoist” goal of searching for longevity. However, close studies of the Daoist canon show that Daoists and transcendents used such drugs in ways that did not correspond to this hierarchy.
Furthermore, this way of approaching drug culture constructs its development through only a very narrow lineage of texts, leading us to imagine that drug knowledge circulated in China within a systematic, serially produced set of texts, and that the development of this knowledge took place over time through editorial processes, the revising and adding to the Bencao. Yet this represents just a fraction of the drug culture of the time, accessible to only a very limited number of people. Such an analysis shows us nothing of the complex interactions between sellers, doctors, Daoists, Buddhists, homemakers, and many other characters involved in the drug trade.

Other studies of the wider circulation of drug knowledge are largely anecdotal, focusing on recipe texts which have been recovered from excavated sites like Dunhuang, Mawangdui, Wuwei and other places. These have received a lot of attention in recent years. However, these studies do not make systematic arguments about the entirety of the drug record, and the Daoist and Buddhist canons remain largely overlooked. The problem here is not a matter of smallness of resources, but of largess. References to drugs appear in many different shapes and forms in the Canon, and we cannot treat them all with the same analytical lens, or trust them equally as reliable sources of historical evidence. There are drug recipes for fasting, for yangsheng, for healing. There are hagiographies describing Transcendents using plants, animals and minerals for healing and to attain miraculous powers. Different texts each use and construct drugs in different ways. Who would ever think of a fabled story about someone rising to heaven in broad daylight as a reliable source for how to use a plant or mineral? Yet, these are nevertheless a cultural record of some kind, about drugs and popular cultural imaginings of them, and they testify to the circulation of these drugs through broad cultural domains. Such a reckoning of the diversity of drug lores in the canons, and a theoretical framework for how to study them has never been made.

There remains to be produced a systematic overview of the drug contents in the Daoist and Buddhist canons against which we can compare the genealogical development of the Bencao literature. This is the goal of the project at the Max Planck Institute for the History of Science, run by Chen Shih-pei and myself, with much input from Brent Ho-Leung Ho of Leiden University.

Project Structure

We will proceed in three basic processes. First, we will develop digital tools to analyse the Buddhist and Daoist canons and identify all occurrences of drug names. We will organize the results according to textual categories, and produce statistical analyses about the types of drugs, and the time periods and locations in which these texts were produced (so far as can be done with existing catalogues). Secondly, on the basis of these results, we will then do detailed analysis of significant drug texts, using a semi-automated text marking software, MARKUS, developed by Brent Ho-leung Ho as part of the Communication and Empire project at Leiden. With the help of a text-marking team, we will digitally mark up the most significant texts for all of the drug terms, as well as the people associated with them, the places they were grown or sold, and associated uses and disease terms. Third, we will insert GIS tags for the associated place names, thus enabling us to map these references across time and across space.

This approach has the potential to produce many answers. First, we may find a very different chronology of when certain drugs entered into China than we get from the pharmacopoeic record. We can learn which sects possessed knowledge of which drugs, how they used them, and when, how and where
was such knowledge transmitted. **Second,** we will be able to come to conclusions about how drugs were described in different types of literature within the canons, and develop a theoretical framework for understanding the multiple ways in which drugs were conceived of and used at a given time. **Third,** drugs as material objects have a distinctive property, which is that many of them grow only in specific regions, and circulated through specific channels in the marketplace. To what degree was the early understanding of drug function conditioned by the kinds of diseases prevalent in areas where they grew? Daoist hagiographies are quite specific about where specific drugs were eaten, but almost no comparisons have been done with the *Bencao* literature concerning their sites of production, that is, the different *regional imaginations* of drug lores. **Fourth,** and hypothetically, space-time distributions of drug-texts may reveal texts produced in the same time and place that have not yet been closely compared, because they were produced within different genres of literature that are not normally compared. We may be able to identify where or when specific drugs became “hot topics” of conversation, generating the production of new drug literature knowledge.

Perhaps most long-lastingly, this tool for analyzing the Buddhist and Daoist canons could then be used in the future for many other kinds of research in the future – for many other kinds of material culture, and intellectual history – enabling researchers to analyse and compare the canons in new ways.

**Stage 1: Text Acquisition, Identification and Analysis**

The first stage begins with big corpus analysis. Not only do we need accurately typed digital texts, and ones based on best published editions, but they must be free of copyright restrictions. Our research results must be open access, as stipulated in the Berlin Declaration on Open Access and as required by the Max Planck Society. Digital Humanities thus brings in very different considerations to the question of textual filiation, or 版本學, than textual scholars are used to. Further features which mark a text as good for digital analysis include how regularly the text is structured, how well small units of text are marked off, and others. Thankfully we have the excellent edition of the Buddhist canons in Dharma Drum’s CBETA, and recently, good copies of the Daoist Canon came to light both on Ctext.org and Kanripo.

Our primary medical text for comparison will be the late 5th century BCJJZ. We have had a copy typed up of the best current edition by Shang Zhijun 尚志君. This work has the advantage that layers in the text are also the primary source for earlier editions in the *Bencao* genealogy, such as the SNBCJ, the *Mingyi bielu* 名醫別錄, the *Cai Yong bencao* 蔡邕本草, the *Wupu bencao* 吳普本草 and the *Li Danzhi yaolu* 李當之藥錄. So when we mark up the BCJJZ, we will also be marking up all of those textual layers as well. This text functions as a basic source of comparison for the entire project, because it has the most complete and systematic geographical data on drugs, but also because this text really had one of the greatest impacts on pharmacology in the Six Dynasties. In addition we will include privately transcribed versions of the excavated literature from Mawangdui 馬王堆, Zhangjiashan 張家山, Wuwei 武威 and other locations, as well as the major classics from the Han dynasty.
Having gathered these sources, we will then do statistical analysis of the appearance of drugs terms in these corpuses. We have a series of lists of drug names – from the *Bencao gangmu*, from medical dictionaries, and from scholars like Chen Ming 陳明 and Catherine Despeux. With these lists of drug names, we will do corpus-wide searches to identify the texts in which they appear, in what quantity, and how often.

As many of you know, the best Sinological database currently available, Academia Sinica’s Scripta Sinica 漢籍電子文獻, and CBETA, only allow very limited single or two-term searches. It is not possible to search for occurrences of groups of 10 or 20 words in a text, much less 1,500 or 9,000 terms. So we will have to develop means to process the Daoist and medical canons to find out how many terms appear in which texts at what frequency. We are talking closely with Professor Hisang Jieh 项潔 and Du Xiechang 杜協昌 about using their database development platform, Docusky (used as the foundation for the Taiwan Digital History Library) for doing this analysis.

As for the Buddhist canon, we have already gotten very useful results with the help of Michael Radich, and his TACL platform, which uses an n-gram system to search the *Taishō*. With this, he’s been able to produce reports of how many different drug terms appear in which texts, and we have preliminary results of searches for collective drug terms. These have successfully identified the medical texts known to historians, but also a few other surprises, which we will look into.

Nevertheless, with such large-scale automated searches, a number of errors of course creep into the results, so these further take time by highly skilled text-readers to clean them up. This takes time and man-power.

In order to limit our texts to the Six Dynasty period and earlier, we further need to produce a good bibliography, and meta-data for these texts. There are quite good catalogues out there, which are “good enough” for the purposes of this kind of general survey, including Kristofer Schipper’s *Taoist Canon*, and the WWW Database of Chinese Buddhist texts (DCB). The *Taoist Canon* uses scholarship that was up-to-date in the 2000’s, and the DCB is based on a number of scholarly catalogues which in the main rely on traditional ascriptions. However, as every Buddhist and Daoist scholar knows, the dates and authorship of many of these ascriptions are hotly contested through extensive scholarship, which is beyond the scope of this project to come to a thorough accounting of these dates. So we will coordinate with other projects, such as the Chinese Buddhist Canonical Attributions Database, and where available, draw on their data.

Ultimately, it will be up to the textual scholar using the data we collect to confirm the provenance of any important texts during the research and writing phase, and perhaps at that time go back and refine the data in the database, perhaps through academic crowd-sourcing. Nevertheless, a large number of texts in both canons are can be considered to have fairly reliable dates, authorship or translators, and even locations of their production.

With this textual meta-data, we will already be able to say a significant amount about the flow of drug knowledge through different textual genres, periods and geographical regions, which has never been said before. As we get these results, we will identify core texts that we want to examine more closely. We will identify them based on different criteria. Some may be because of the sheer volume of drugs described in them: the major drug and recipe texts that have been studied in the few existing studies on religious medicine. Some may represent distinctive genres of drug use, such as the alchemical literature,
hagiographies, healing narratives & so on. And others may prove interesting because they were produced in the same time and place.

**Stage 2: Text Marking**

Once we have identified these texts, we will then send them out for text-marking and review. We will use MARKUS, a semi-automated text marking platform that has been developed as part of the Communication and Empire project by Brent Ho.

Here is an example for how to extract drug properties and place names from *Annotated and Collated Materia Medica (Bencaojing jizhu) by Tao Hongjing* in MARKUS. *Bencao jing jizhu* contains comprehensive information of drugs and it recorded drugs information systematically. Six of seven chapters record drug information in a regular written format: Drug names, flavor, temperature, location grown, harvesting times, alternative names and other properties are described in a regular order which can be recognized by computer analysis.

Before the tagging process, the BCJJZ (chapter 2 to chapter 7) must be prepared as a single TXT file in UTF-8 format. Each drug record must be separated by an empty line because MARKUS uses the empty line to distinguish between paragraphs. After the pre-processed *Bencaojing jizhu* is loaded into MARKUS through the front page of the website, you can choose from three markup interfaces. The **automated markup** interface shows a list of built-in historical entities. Select “paragraph division” and “place name” and then MARKUS will starts scanning through the text once the “start markup” button is clicked.

After a while, paragraph divisions are detected and place names are tagged automatically in blue color (See fig.). However, the result is not perfect. For example, “政和” has been tagged as place name. The result needs to be corrected manually in the **manual interface**.
The **automated** and **keyword interfaces** provide a “web reference” function when a tag is clicked. In the **manual markup** interface, there are editing functions such as “save” (disk icon), “scan” (lens icon) and “delete” (Trash bin icon) appear in the popup menu. (see fig).

For example, to remove the place name tags “政和” from the text, we can either click the “trash” icon to remove an individual item, or click the “lens” icon to scan all the same tags in the text. After all the “政和” instances are scanned, we can go through the tag list to make individual decisions (by pressing lock or delete). If all the tags are errors, they can be removed as a batch by clicking the “remove all except locked” button. It is much more efficient than removing them one by one.

On the other hand, not all place names are discovered automatically. Place names not included in the built-in list need to be tagged manually by selecting the text and clicking the “disk” button to save it. After a new tag is added manually, a “lens” icon will appear, which can list all the possible tag instances in the text. With the similar batch-editing interface, possible instances can be checked one by one or be tagged in batch by clicking the “Apply to all” button.
Unfortunately, the built-in entities lists only cover place names drawn from the Harvard Historical GIS Project. For other terms, like drug names, taste, property and temperature, we tag them semi-automatically by using the regular structures inherent to the text. We can describe these regularities to the software using “Keywords helper” in the keyword markup interface. For example, to identify the pattern by which the drug name regularly appears in the text, we can leave the left keyword box “empty”, and input a “space” in the right keyword and the number of characters in between are from 1 to 5 (”,1,5,”).(See fig.). This will identify all instances where a line begins, and is followed by one to five characters, and then a space.

Based on the written description, MARKUS will search through the text and list all the keywords that fit the described regular pattern. Obvious errors can be removed from the list, like “毒・三蟲・”, as this is a drug properties and disease name, which the computer has mistakenly identified as a drug name, based on the programmed analysis. When the button “ready to markup” is clicked, a pop up window will appear and we can assign a tag name and a color for the text and the tag. MARKUS will then start tagging discovered keywords with assigned tag name with selected color.
We can apply the same procedure to tag other drug properties, for examples, taste with temperature (“味”,1,20,”’”) is tagged in green, alternative name (“一名”,1,5,”’”) is tagged in purple and action (“主”,1,20,”’”) is tagged in brown.

These results can be exported into an excel format (see fig.), once all tags have been checked in the manual interface.
Stage 3: Database Production, Time-Space Visualisation

The third stage of analysis is time-space visualization. To help us visualize the distribution of drug lore, we will use PLATIN, a time-space display software developed at the Max Planck Institute. This software can easily display datasets using different dot sizes to represent populations at different scales, while also letting you manipulate the datasets at the bottom of the screen. As we accrue marked texts that describe different features of drugs, such as their names, properties, geographic location, associated people, and their curative or cultivation functions, we will export these into Docusky. Docusky will produce miniature databases structured around inherent features of each individual text. Each text thus informs the structure of its own mini-database. These results can then be exported as excel files into PLATIN, which can display different results across the map, and also within different time scales. This is extremely powerful, because through this process, we can search not only the Buddhist and Daoist canons, but also search the time and places where those words were written, and visualize those results. From this we can begin to consider much closer histories of religious and medical interaction.

Conclusion

This project is in its nascency, but we hope that scholars in other fields will find this potentially useful for their own work. Right now we are developing the processes to do this work, and learning how long it takes to process these texts. Next year, after selection of texts, we will send off a batch to be marked by a separate company. This process of analysis promises to be much faster than earlier methods of solely human-scanned texts, but we don’t know how long it will take. As we find out with time, we can then share this knowledge and our workflow with other interested scholars, enabling them to better calculate how much work can reasonably done in what kind of time and with what kind of budgets. At the beginning of year 2017, we will host a pilot training workshop, sharing the technology with a few interested scholars who will be invited to use the tool to do their own research, and then share the results and the process at a conference in the summer of 2017. At the conference we will also host a larger training workshop for anyone who would like to attend. From this, we hope that collaborative projects can be built, so that teams of scholars can write grants together to do future research, using the technology, infrastructure and budgetary knowledge we have gained through our pilot project.