

## 论郭店楚简数字化工作之得失：兼及出土文献数据库的开发标准问题<sup>①</sup>

张昊苏 / 南开大学文学院

**摘 要：**以郭店楚简的数字化工作为研究个案，可以透视出土文献数据库建设的得失。今所知与郭店楚简有关的代表性数据库约十余种，已完成了郭店楚简基本数据的建设。但各库的数据来源、功能开发及用户友好性等方面都存在一定缺陷。其核心问题是各数据库均未使用最新的图版、释文作为来源数据，导致数据存在遗漏、误编、误缀、误释等问题。理想的郭店楚简数据库，应该尽可能设定著录体例，通过数字技术保存全部文物信息、引入研究论著资料，并结合研究需求开发更多功能。约言之，郭店楚简数据库有必要由数字化向数据化进一步发展，即在基础性的全文阅读和检索上，探讨提供更多服务的方式。这一经验适用于各类出土文献数据库建设。

**关键词：**郭店楚简 出土文献 数字人文 数据库 低代码平台

1993年10月，湖北荆门郭店一号楚墓出土了大量竹简，其中有字简共730枚，字数达13,000余字，此即著名的郭店楚简。郭店楚简既是考古挖掘出的珍贵历史文物，又为古典文献、思想史、简帛学、古文字学、书法史等学术领域提供了非常重要的学术资料。据不完全统计，三十年来的有关学术研究著作超过数

<sup>①</sup> 本文为教育部人文社科规划青年基金项目“汉代礼乐文献辑校与研究”(23XJCZH002)、2022年度南开大学文科发展基金科学研究类项目“乾嘉时期小说文献与文学思想研究”(ZB22BZ0301)、南开大学2023年本科教育教学改革项目“数字人文背景下古代小说文献学课程思政与一流课程建设”(NKJG2023089)、广西师范大学教育教学改革A类重点项目“新文科建设为导向的古代文学课程改革”(2021JGA04)的阶段性成果。文中引用与郭店楚简释读有关之例证，均由荆门市博物馆范晓佩提供信息或审订观点，谨致谢忱。

百种，论文以数千篇计，实为近年来之显学。<sup>①</sup>

允称“显学”的另外一端，是其研究过程与古代文献，尤其是出土文献的数字化进程密切相关。从考古全过程来看，发掘、修复、记录、展陈等工作均需用到数字化手段，近些年科技考古的兴起，更代表着新科技对考古工作的有效助力。从研究层面看，在新的数字环境下，为了更好地对郭店楚简原始文物、文献信息及相关学术资料加以充分利用，运用数字化、数据化手段已成必备趋势。以《郭店楚墓竹简》<sup>②</sup>和《楚地出土战国简册合集（一）：郭店楚墓竹书》（以下简称《郭店楚墓竹书》）<sup>③</sup>等为代表的文献整理成果，代表了对郭店楚简保护、摄制、释读、注解的杰出成就，也成为相关领域学者的必读专业书籍。迄今为止，已有诸多数据库能够在这些研究成果的基础上，提供郭店楚简的检索、阅览及图文对照等功能，亦有多位学者就相关问题展开相应研究，其梳理、评鹭均翔实而见功力。<sup>④</sup>

然而类似研究既然多以描述为主，就往往归于一种“宿命”，即其公开发表、出版之时，已很可能不及更正最新信息，致有“过时”之嫌。因此，每隔一段时间，都对已见于著录的数据库加以重新观察，关注其能否正常使用、使用权限是否产生变化、有无数据及功能更新；并随时增补新开发问世的数据库，应有一定价值。同时，由于相关数据库所著录之内容相当丰富，故从细密分析郭店楚简这一单项，尤其是从数据库建设与学术研究结合的角度，也还有发覆空间。此外，既往研究对出土文献数据库的讨论，较为注意字符集规范标准与数据库建设、使用的相关问题，但对于其他领域讨论较少。本文希望在此基础上讨论的是：有相应规范标准，但数据库建设中不够缜密的操作性问题；以及相关数据库建设者可

①以“郭店”为主题词在“中国知网”检索，可得学术期刊1,889篇，学位论文790篇。在“读秀”中可检得中文图书267种。此外还有大量以楚简、出土文献为研究对象，但以郭店楚简为核心材料的研究，其具体数量难以精确统计。

②荆门市博物馆：《郭店楚墓竹简》，北京：文物出版社，1998年。下文出现的《郭店楚墓竹简》，无特殊标注者，均为此版本。

③陈伟、彭浩主编：《楚地出土战国简册合集（一）：郭店楚墓竹书》，北京：文物出版社，2011年。

④早期的文献检索教程很少涉及出土文献的检索问题。系统介绍中国古典文献数据库，并有相当篇幅论及郭店楚简的研究，参见张三夕、毛建军《汉语古籍电子文献知见录》（北京：世界图书出版公司，2015年），王国强《网洋撷英：数字资源与汉学研究》（南昌：江西高校出版社，2020年）等。专门评论出土文献数据库建设及郭店楚简文献数字化问题的研究，参见梁松涛、赵艳平《浅析我国出土文献数据库建设的制约因素及对策》（《图书馆工作与研究》2008年第4期），毛建军《简帛文献数字化述论》（《兰台世界》2011年第8期），尤建辉《出土文献数字化整理简述》（《黑龙江史志》2014年第21期），刘春雪《我国出土文献数据库出版述论》（《出版与印刷》2019年第1期），王建欣《出土文献数字化建设现状研究》（《古籍整理研究学刊》2023年第1期），匡陕勇、苏展《郭店楚简的数字展示技术开发与应用》（《江汉考古》，2019年S2期）等。网络资源还可参考“奎章阁”网站（<https://www.wenxianxue.cn/>）之“出土”类目及部分介绍性文章。

能在功能设计时并没有特别在意、但可能很重要的数据问题。不论内容储存标准也好，数据处理方式也好，都有必要考虑服务于此——即研究者，尤其是非古文字专业研究者的多种实际需要。

本文试就个人所见（2023年7至9月检索及亲身体验），对既有之代表性郭店楚简数据库工作略加分类、介绍、评鹭，并尝试结合现有数据库开发方向及实际研究需求，为郭店楚简数据库建设工作提出部分建议。在梳理郭店楚简这一数据个案的基础上，我们可以更清晰地看出当下出土文献数据库建设中可能存在的一些问题及发展方向，即同一文本可以服务于研究者的不同需要，这一思路的重要性、可行性还没有得到充分关注和讨论。

## 一、述略与补遗：代表性的郭店楚简数据库

区别于近世文献动辄以亿字计数之“大数据”，郭店楚简实属吉光片羽，所容纳的文字数量有限，数字化本身工作量不算特别大，而且图版、释文均已公开发表，故已有若干机构及个人均在建设与郭店楚简有关之数据库。其好处当然是可以最大限度地降低资料获取成本，研究者可从不同渠道获取电子资源。但坏处及缺陷则有以下两点：第一，其数据来源相对单一，且通过翻拍图版建立的数据库，在资源清晰度及准确性上都有若干瑕疵；第二，类似数据库数量较多，但区分度并不特别明确，有“重复发明轮子”之嫌疑，尤其是对于普通使用者来说不易上手。

学者使用不同数据库检索同一类型文献，是否能够得出同样的结果？这样的工作流程是否可能造成工作无法被标准化，以及时间上的浪费？当获取原始图版及全文本身并不成为特殊难题的时候，文献数据化工作有必要基于现有数据库建设的情况，根据研究者的实际痛点，满足实际使用需求，从数字化水准上进一步提升，并挖掘开发新的数据处理功能。

就笔者知见，与郭店楚简有关，较有代表性的公开数据库有以下几种：

1. 华东师范大学“战国楚文字检索系统”。<sup>①</sup>该系统需提交申请书后可免费开通使用权限。其早期数字化成果名为《战国楚文字数字化处理系统》（光盘），包括“战国楚文献检索系统”“战国楚文字字库”“楚文字输入法”“楚楷对应转换程序”四大部分。据张三夕、毛建军《汉语古籍电子文献知见录》著录：“‘战国楚文字检索系统’由华东师范大学中国文字研究与应用中心研发。该检索系统是

<sup>①</sup><https://wjwx.ecnu.edu.cn/wenzidb/Home/Login.aspx>（以下所引述的网络数据，部分信息由东洋文库段宇提供，谨此致谢。）

在‘战国楚文献检索系统’的基础上审核、增补完成的。‘战国楚文字数字化处理系统’于2003年由华东师范大学中国文字研究与应用中心开发，上海教育出版社出版。”<sup>①</sup>该库标明所据参考资料为《郭店楚墓竹简》。现阶段则有较明显的数据更新，为“中国文字智能检索网络数据库”之一，名为“战国楚简数据库”，安装字体文件后可在线访问，包含“全文检索”（按文献出处检索、字词检索、字形检索、字际关系检索）、“考释检索”（被释字检索、考释论著检索）、“偏旁检索”“通假检索”“字体分类检索”“逐字义类检索”等检索路径，并关联原简、释文及相关考释信息。<sup>②</sup>

2. 武汉大学简帛研究中心“简帛网”。<sup>③</sup>该网站提供简帛字形辞例检索，可通过检索之单字、偏旁，对应楚简原文及图版。但游客至多只能浏览10条记录。此外，该网站的“简帛文库”“简帛论坛”等板块亦收录大量有关研究，可供阅读及检索。该库未标明参考资料出处，根据图版内容判断应出自《郭店楚墓竹简》。此外还设计有“简帛文献目录检索”，<sup>④</sup>但目前并无数据，使用者只能通过“简帛文库”的“楚简”分类阅览其中选录的部分文章。

3. 香港中文大学“郭店楚简资料库（2001）”。<sup>⑤</sup>该数据库由香港中文大学历史系张光裕教授开发，内容基于氏著《郭店楚简研究·第一卷·文字编》<sup>⑥</sup>释文部分，先将文字资料制作成在线全文检索数据库，后又增入“郭店楚简主要参考论著”检索系统。近年，“由于图书馆网站技术原因，光裕乃将该资料库转换为word档，转存于历史系网站”，于2022年11月1日起，免费向读者提供下载。“将730枚郭店楚简按竹简编号、篇目名称排列，读者可用竹简编号、篇目名称或个别字词，作全文检索。”<sup>⑦</sup>文件题名为“郭店楚简十六篇释文修订稿全文检索”。从现有状态看，该数据库已实际上归于寂灭。

4. 台北“中研院”历史语言研究所“先秦甲骨金文简牍词汇资料库”。<sup>⑧</sup>该数据库“收录甲骨文、金文、简牍三种文字材料，依词类、材质作为搜寻条件，展示词汇的词性、时代、书目来源、释文段落。特色在于打破材质作为前提的预设，以词汇本身为焦点，横跨不同时代观察词汇的流变”，“提供词汇检索及全文检索……每个词汇包括出处、编号、上下文，便于使用者回查与检阅，帮助理解

① 张三夕、毛建军：《汉语古籍电子文献知见录》，第75页。

② <https://wjwx.ecnu.edu.cn/guwenzi/ChuJiandb/Home>

③ <http://www.bsm.org.cn/>

④ <http://www.bsm.org.cn/?wenxian/>

⑤ <https://www.history.cuhk.edu.hk/tc/elkt/guodian/>

⑥ 张光裕：《郭店楚简研究·第一卷·文字编》，台北：艺文印书馆，1999年。

⑦ <https://www.history.cuhk.edu.hk/tc/elkt/guodian/>

⑧ [https://inscription.asdc.sinica.edu.tw/c\\_index.php](https://inscription.asdc.sinica.edu.tw/c_index.php)

词汇意义”。该库的优长是对于词类的分类尤其详尽，并且能提供严式、宽式释文，并提供编号以便使用者回查原简。该库声明所据参考资料为《郭店楚墓竹简》。“中研院”的类似数据库还有“文物图像研究室资料库”，<sup>①</sup>其中“简帛金石资料库”收录了可全文检索的《郭店楚墓竹简》。此外铭传大学资讯工程学系亦声称建立了郭店楚简数据库，但网络上未能检索到相关信息。<sup>②</sup>

5. 日本早稻田大学“楚简数据库”。该数据库由早稻田大学文学部工藤研究室研制。刘春雪介绍“该数据库由楚系文字资料库和简牍资料库两部分组成，设计了检索功能，使用者可选择字词、篇章等进行检索”。<sup>③</sup>经笔者检索、访问，暂未查询到该数据库的公开上线信息，仅知其“读解”曾在系办杂志《史滴》中逐期刊行。鉴于工藤元男先生主持的长江流域文化研究所主页同样下线，似证明该数据库早已无法使用，令人深感遗憾。此外，池田知久主导郭店楚简研究会时据说曾建设过相关数据库，挂在大东文化大学数码藏品网站下，但网络上未能检索到相关信息。

6. 美国科罗拉多大学“郭店与马王堆原件选集数据库”(Database of Selected Characters from Guodian and Mawangdui Manuscripts, 2006)。<sup>④</sup>该库由李梦涛(Matthias L. Richter)教授开发，内容包括郭店楚简及马王堆帛书，但因技术问题，该库目前仅能提供Excel文件。除提供文字释文、释音、图版及各家解释外，李梦涛教授还尝试将郭店楚简变成结构化数据，如统计了文字出现频率及组成构件等。其所据参考资料为《郭店楚墓竹简》。

7. “古音小镜”网站。<sup>⑤</sup>该网站由计算机行业的从业者顾国林先生基于个人兴趣建设，该数据库的前端呈现与后端架构都与其他数据平台有明显的差别，在资料共享、检索使用与可视化呈现等方面都有独到的探索。<sup>⑥</sup>该网站的“古文字”栏目旨在解决普通学生及公众查询困难的问题，其中设有“楚简字形库(工程版)”，但目前部分功能尚未完工，其中郭店简为“初步完成”状态，主要参考资料为荆门市博物馆编《郭店楚墓竹简》(书法大图版)和《楚地出土战国简册合集》两书。该字形库可同时查询简繁、变体，点击单字可直接跳转原简大图。

上述介绍仅属挂一漏万。但足以看出，除挖掘、收藏郭店楚简的荆门市博

① <http://saturn.ihp.sinica.edu.tw/~wenwu/www.htm>

② 项目名为 Guodian Bamboo Slips Database Building and Research Tool Development，主持人为 Shieh, Yuh-Pyng。

③ 刘春雪：《我国出土文献数据库出版述论》，《出版与印刷》2019年第1期。

④ <https://www.colorado.edu/faculty/richter-matthias/database>

⑤ <http://www.kaom.net/index.php>

⑥ 盛一涵、田雨娇：《历史语言学的数字成像：“古音小镜”网站——兼议数字人文项目生长路径》，《数字人文研究》2022年第1期。



物馆外，海内外均对郭店楚简的数字化进程有颇多贡献。此外还应注意的，从更多古文字数据库及相关网站中亦能获取与郭店楚简有关的信息，如陈信良于2012年创建的古文字索引资料库“引得市”，其资料来源除包括《郭店楚墓竹简》《郭店楚墓竹书》外，还包括《郭店楚简文字编》<sup>①</sup>《楚文字编》<sup>②</sup>《楚系简帛文字编》<sup>③</sup>等书，及刘钊《郭店楚简校释》<sup>④</sup>、李零《郭店楚简校读记》<sup>⑤</sup>等十余部学术成果，研究者可通过该数据库之免费及付费功能，检索出土文献与古文字相关的各类资料。以郭店楚简为例，即可搜索简号、多种图版及释文的对应页码，形成电子通检。但因版权问题，该数据库仅有“引得”(index)，不提供任何形式之电子文档，使用者必须根据个人藏书回查全文。<sup>⑥</sup>著名简帛研究重镇如清华大学出土文献研究与保护中心、复旦大学出土文献与古文字研究中心等亦设计有相关数据库及提供文献资料分享。类似资料库尚有多种，限于篇幅不遑逐一列举。

## 二、数据来源与信息更新

数据库资源建设对于郭店楚简的具体应用提供了诸多便利，且不同机构及个人的学术旨趣和数字资源可以形成互补，为郭店楚简的系统研究提供了更多的可能性。但大量类似数据库的开发同时也会给研究者带来困扰，比如具有同质化现象、功能相对单一等问题。在今天的数字环境下，用户友好型的数据库应该重点考虑数据可靠性、功能多样性及内容更新等问题，方能更好地服务于使用者，为学术研究效率提升作出更大贡献。

如前文所述，公开出版的郭店楚简整理成果已有多种。目前各数据库主要采取1998年出版的《郭店楚墓竹简》为基础来源数据，将其图版、释文信息加以数字化，或在此基础上进一步增补学界有关郭店楚简的代表性研究。然而，似乎各数据库开发者都忽略了以《郭店楚墓竹书》订正《郭店楚墓竹简》数据的重要性。作为首次发表郭店楚简全部内容的代表性研究，《郭店楚墓竹简》内容包括郭店楚简的全部图版、释文及注释。图版以原大形式黑白影印，释文摹写原型，但删去原简符号，篇末附有竹简整理号与出土号对照表，非常方便学者使用。但需要说明的是，与学术研究后出转精的一般规律相符，该书并非今天意义上的

① 张守中、张小沧、郝建文：《郭店楚简文字编》，北京：文物出版社，2000年。

② 李守奎：《楚文字编》，上海：华东师范大学出版社，2003年。

③ 滕王生：《楚系简帛文字编》，武汉：湖北教育出版社，2008年。

④ 刘钊：《郭店楚简校释》，福州：福建人民出版社，2005年。

⑤ 李零：《郭店楚简校读记》，北京：北京大学出版社，2002年。

⑥ <https://www.mebag.com/index/>

完美数据，有少量细节需要根据最新研究加以补充修正。除1998年版的图版清晰度稍微逊色于新版本外，《郭店楚墓竹书》的整理者亦已经提出此前有部分不同照片、遗漏简及未公布的简背文字，并有部分释文、编联应加以修正。<sup>①</sup>作为集成性的新成果，《郭店楚墓竹书》在序言、凡例及篇前整理说明中提示的信息，理应成为数据库建设者关注的重要问题。

将这些用例放在上述各数据库中加以检验，我们可以印证现有数据库的基本情况。总的来说，现有数据至少存在两个比较重要的问题：一是所使用的数据本身存在一定瑕疵，未能采用学界更新的研究成果和较优数据；二是部分数据库虽意识到此问题而有所修正，但却导致了数据本身与数据来源未能完全统一，且普通使用者无法就此产生查验意识。<sup>②</sup>

**分类例。**学界熟悉的郭店楚简，主体部分均为整篇文字。但除此之外还别有残简27片。这些残简当然原本归属于某一文章，但因各种原因还有待缀合。《郭店楚墓竹简》在图版之末附录了残片，但未另外做出释文，亦未在目录中加以标记。<sup>③</sup>《郭店楚墓竹书》则配以释文，并将其中三个残片与对应竹简缀合。以《郭店楚墓竹简》为底本，带来的问题是，数据库制作者极易忽略“残片”的存在，如“简帛网”就遗漏了这部分内容。

**补漏例。**《郭店楚墓竹书》指出：“龙永芳披露了在竹简养护过程中发现的一枚先前未曾刊布的新简。”该简属于《语丛·三》，释文为“从所少好，与所少乐，员（损）”。<sup>④</sup>经检验其他数据库，“郭店楚简资料库（2001）”“简帛网”等数据库并未著录。仅“先秦甲骨金文简牍词汇资料库”著录了本简信息。此遗漏可能与该简编号为“遗简”，故在数据统计中易被忽视有关。

**编联例。**《郭店楚墓竹书》有多篇文章在编联时，根据新的研究成果，更改了《郭店楚墓竹简》的编联顺序。至少《太一生水》《穷达以时》《唐虞之道》《成之闻之》《尊德义》《性自命出》《六德》《语丛·一》均出现了“学者间分歧较大”的情况，其他各篇亦或多或少存有部分的简序调整。以《太一生水》为例，学界对编联顺序历来有甚多争鸣，《郭店楚墓竹书》采取了裘锡圭的意见，将该篇分

① 陈伟、彭浩主编：《楚地出土战国简册合集（一）：郭店楚墓竹书》，《序言》第iv—v页、《前言》第1—2页。

② 类似情况在古籍数据库中大量存在。带来的潜在危机是，学者研究视野受到数据库基础设施的限制，且不易意识到并摆脱之。

③ 荆门市博物馆：《郭店楚墓竹简》，第108页。

④ 龙永芳：《湖北荆门发现一枚遗漏的“郭店楚简”》，《中国文物报》2002年5月30日，第二版。释文见陈伟、彭浩主编：《楚地出土战国简册合集（一）：郭店楚墓竹书》，第159页。

为三组，即简1—8、简10—13与简9与14。<sup>①</sup>即，作为文物著录标准的编号顺序，与需要阅读的文字顺序是不同的，其背后涉及编联方式与上下文脉络等问题。既有数据库或因其重点立足于文字检索，多不能很好地呈现这一问题。

**缀合例。**《五行》18号简下段至19号简上段残缺，<sup>②</sup>“终也”上脱去五字。因《五行》文本别见于长沙马王堆汉墓，据帛书本可推知原文为“君子之为德也，有与始，亡与终也”。通过比对，《郭店楚墓竹书》整理者判定21号残片“始亡”二字恰可与之拼合。<sup>③</sup>但经检验其他数据库，不仅并未采纳，且因“始亡”为残片，故在以篇目为检索前提的方式中，常常无法被筛选出来。

**简背文字例。**如《成之闻之》简13背面有数字“七十二”，可能为对简的清点计数，此在《郭店楚墓竹简》中并未提及，《郭店楚墓竹书》补充了图版及释文，而且标明了其对应简正面的文字。此亦往往为诸数据库之所无的信息。

**订误例。**《郭店楚墓竹书》在释文方面即有不少后出转精之处，对读者披阅自见，无需逐条讨论。此处仅举学界人所共知的一例。《老子甲》1号简“绝伪弃”下之字，《郭店楚墓竹简》释读为“诈”，《郭店楚墓竹书》订正为“虑”，目前已成为学界共识。这里涉及两个问题：一是在文字释读存在争议时，理想的数据库应该不断结合学界研究成果加以更新；二是在实际工作中，应用更好的设计方式以方便读者检索。经检验其他数据库，“简帛网”“先秦甲骨金文简牍词汇资料库”，以“虑”字可以搜到对应图版，以“诈”字则无匹配检索，证明数据库对《郭店楚墓竹简》有所修正。但从使用者的角度看却有诸多不便——如使用者通过释读为“诈”的研究成果搜索，很可能在多种数据库中无法找到对应图版，研究效率受到限制。此外，随着学术研究的日趋细密，与更多出土文献资料的发现，《郭店楚墓竹书》的部分释文、注释同样有必要随时加以修订、补充。比如，《老子》甲简6“化（祸）莫大（乎）不足”，释文“乎”字前脱隶定字“虐”。《老子》乙简8“尾（托）天下矣”，“尾（托）”字前脱“若可以”三字；《穷达以时》简15“壘（幽）名不再”，“名”系“明”字之讹。上述盖均系手民之误。此外，部分新的释文、注解内容或已成为学界共识，则对应数据亦应随之更新。从现有数据库多依赖单一数据来源这一点看，其本身边界清晰，并非特别问题；但如果从研究者的角度考虑，数据库应该为研究者检索、跟踪最新成果提供方便，与之相关的数据收集与功能设计亦应纳入思考视野。

① 裘锡圭：《〈太一生水〉“名字”章解释——兼论〈太一生水〉的分章问题》，载安徽大学古文字研究室编：《古文字研究（第二十二辑）》，北京：中华书局，2000年，第219—226页。

② 荆门市博物馆：《郭店楚墓竹简》，第32页。

③ 陈伟、彭浩主编：《楚地出土战国简册合集（一）：郭店楚墓竹书》，图版第41页。



郭店楚简作为一个代表性的数据个案，对类似出土文献数字化工作同样具备启示意义。冯大建指出，好的文献数据库在底层数据方面，至少需要包括三个层次，即原始图版层、原始文本层与通行文本层。<sup>①</sup>出土文献面临的数据标准可能更为复杂。首先，出土文献的“原始图版层”，其要求与通常的古籍书影存在差异。出土简牍由于其文献特殊性，出土时多为散乱断简，经研究者编联、修复后方成为可阅读之文本，校验时需要获取文物高清细节才能细致分析编联、缀合；又由于其文物特殊性，出土后存在变色、饱水等现象，需要及时著录方能保全文物原始信息。其次，严格按照原始文献所生成的摹写、隶定字形往往并无规范繁体汉字与国标字符集对应，有些字究竟应如何释读也有争议，导致在显示、检索等方面多有不便。再次，对同一文字的释读往往存在争议，故研究者可能既需要在学术写作中方便引用的通用文本，也需要同时检索多种版本的释读注解，并随时关照学术论著中的最新动态。这就需要学术研究者与数据开发商在底层设计层面建立良好数据关联并设计阅读界面，沟通研究需求、检索方式、阅读模式、引用规范等具体使用场景，在建立技术规范与著录标准的前提下，不断更新数据，庶能方便学者之阅读、引用。这也是数字化文献相较于传统纸质文献的优越之处。

### 三、功能设计与反思

现有的郭店楚简数据库，主要由古文字研究机构或研究者开发，以因应其研究需求。故多数数据库均可以提供全文检索及图文对照功能，有的还可以采用高级检索方式，以便筛选特定辞义、形制、字体及展开对比。但郭店楚简的价值绝不仅限于古文字学、简帛学的研究，相信考古学、历史学、思想史等专业的研究者，亦对此有不同的需求，服务于这些学者的信息保存和文本挖掘功能也有极高的开发价值。

**数据库类型开发的选择。**从数据类型来说，常见的文史类数据库包括参考数据库及源数据库。前者指包含各种信息、数据或知识原始来源和属性的数据库，数据库中的记录是通过对信息、数据或知识的再加工和过滤（如编目、索引、提要、分类等）所形成的二次文献。即，参考数据库的信息是不自足的：用户从中获取信息源信息后还需进一步去查找原文，典型的为索引与目录数据库。源数据库则指能直接提供原始资料或具体数据的自足性数据库，用户不必再查其他信息

<sup>①</sup> 冯大建：《古代小说文献研究与数字化之思考》，《文学与文化》2023年第4期。

源即可满足需要，典型的为全文数据库。针对同一研究对象，数据著录、加工标准不一，数据库也会有极大差别。理想的数据库应具有集成性、兼容性的特点，可满足使用者的各类需求。就现有情况看，数据库之间的兼容性似存问题，且颇多数据库无法满足全文阅读的使用需求。

**原简形制等文物信息的重要性。**经由数字技术赋能，采用高清方式全面记录文物信息，并采取3D、AR、VR等方式加以呈现，仅在国内博物馆系统中就已有相当多的成功实践。匡陝勇、苏展已经论述了郭店楚简的数字展示技术开发与实践工作。<sup>①</sup>从博物馆展陈的角度，采取数字展示技术的主要优势是提供良好交互体验，并降低楚简原件实物展出的特殊困难；从学术研究的角度，研究者同样有类似的需求：获取现存文物的细节信息有助于深入研究。最典型的议题当然是通过出土简的形制来判定书籍性质，并与同时期的其他简做对比研究。类似信息还包括残断情况、编联痕迹、简背信息、墨块墨线标记符号等，可判断编联的合理性、原文的分章方式等问题。学者已经发现，《郭店楚墓竹简》整理时多处脱去了重文符号。在以往研究中，这些问题主要由考古专业，尤其是具备观察原简条件的学者加以描述。但数字化手段可以全面呈现这些文物信息，即可以进一步推动相关研究，降低文献门槛。

**原始文献与研究论著之关联。**数据库的最大优势是可以处理人力难以解决的问题。具体到郭店楚简这一方向，竹简本身所有的文字是相对有限、易于被人工细读掌握的，但围绕竹简的相关研究，则相当丰富，不易穷尽。尽管学界普遍公认《郭店楚墓竹书》能够汇集众长，研究水准较高，但毕竟仍是部分学者合作的学术产物，容偶有白璧微瑕，或有需要后出文献例证转加补正之处。由此则需要充分把握相关领域的研究动态。此外，学界不少研究观点、思路，与《郭店楚墓竹书》之释读不相同，这些争议性歧说同样是学术史的重要组成部分，对于后来学者开展研究颇有裨益。<sup>②</sup>以郭店《老子》分组为例，李零对《老子》甲的分篇与通行整理观点不同，将简21—23、24、33—39为上篇，1—20、25—32为下篇。但现有数据库建设中，或因各种受限条件，常常将全文数据库与论著数据库分别建设。新的数字环境呼唤更为全面完善的数据建设方案，使学者可以多角度理解、分析现有资料。

**多版本的文献比对与挖掘。**如果粗略分类，郭店楚简所涉篇目、文字有两类。一类是存在其他传世本或出土本的经典文本，如《老子》《缁衣》《五行》等。

① 匡陝勇、苏展：《郭店楚简的数字展示技术开发与应用》，《江汉考古》2019年第S2期。

② 哪怕是部分“误说”，其立论依据和学术影响也都值得重视，学者需要在全面掌握这些信息的前提下才能得到更深入的知识。

这类文本由于可与传世文献参照对读，并由此探索先秦思想史、文献发展史等议题，故相关研究者对这类出土文献亦有相当高的数字化需求。最典型的显学当然首推《老子》。郭店楚简《老子》甲、乙、丙出土后，就提示研究者：《老子》一书的争议性并不仅仅在于上下经的顺序、具体分章与异文的差别，而在于其本身有可能是多种道家文本重新组合成的思想论著。<sup>①</sup>传统研究中，刘笑敢《老子古今：五种对勘与析评引论》<sup>②</sup>将《老子》竹简本、帛书本、傅奕本、河上本、王弼本按八十一章顺序逐句对照排列，展示了多版本比对的重要性，类似的汇校研究也还有多种。然这一分章、排序，本质上即秦汉以来研治通行本之思路，其研究对象往往也是“流”而非“源”。郭店楚简《老子》甲、乙、丙三组简文，其章数、章序与八十一章本差异极大，甚至有可能与八十一章以外之其他文字组成新的文献。如《老子》丙与《太一生水》的抄写形制、字体一致，故有不少学者都怀疑其可能是同一篇或关系密切的文字，惟因受制于对通行本《老子》的认知，不易以此思路展开详细论证。又如今本《老子》第六十四章，全文为：“其安易持，其未兆易谋。其脆易泮，其微易散。为之于未有，治之于未乱。合抱之木，生于毫末；九层之台，起于累土；千里之行，始于足下。为者败之，执者失之。是以圣人无为，故无败；无执，故无失。民之从事，常于几成而败之。慎终如始，则无败事。是以圣人欲不欲，不贵难得之货。学不学，复众人之所过。以辅万物之自然，而不敢为。”<sup>③</sup>但在郭店楚简中，“其安……足下”一段对应《老子》甲简25—27，并有句读符号，可看作分章标志。“为者败之……以辅万物之自然而不敢为”对应《老子》甲简10—13，且与“道恒无为也”连写（对应今本《老子》第37章）。同时这段内容又单独出现在《老子》丙中。可见今本的第64章在郭店楚简本中是分在两处的不同章，而且有不同的组合方式。这就需要郭店楚简的研究者扬弃秦汉以来通行本的成见，对出土文献原貌加以重新审查。在数字化的背景下，可以通过“切换底本”的方式，以郭店楚简本为核心底本（鉴于《老子》甲、乙、丙本既有交叉，又有区别，实际上是三种不同的“底本”），其他版本参校，从而呈现出不同的文献对读形态。这既能使研究者顺畅地使用对勘方式研读出土文献，又可以避免研究思路在不经意间受到传世文本之固化。

① 如部分学者认为郭店本《老子》只是与今本《老子》有着相同内容的不同文本，亦有学者指出《太一生水》很可能即郭店《老子》之一部分。参见谷中信一《郭店楚简〈老子〉及び“太一生水”から見た今本〈老子〉の成立》（《楚地出土资料と中国古代文化》，东京：汲古书院，2002年）、池田知久《郭店楚简〈老子〉新研究》（南京：江苏人民出版社，2022年）等研究。

② 刘笑敢：《老子古今：五种对勘与析评引论》，北京：中国社会科学出版社，2006年。

③ 《老子》版本异文甚多，这里仅区别大致起止，不另外校勘标明细节。王弼注、楼宇烈校释：《老子道德经注校释》，北京：中华书局，2008年，第165—166页。

思想史互文视野下的郭店楚简文本。“互文”为文学批评领域近些年常用词语，素有广义、狭义之别。简单来说，即任何文本都不可能是凭空虚造的，其表达、观念、思路都与其他文本存在各种各样的网络联系。<sup>①</sup>如前面所探讨的《老子》第六十四章之问题，假如仅就《老子》诸版本比勘的话，那么有些问题无法得到判断。但该章恰好在先秦两汉的若干文献中得到引用，则可以作为旁证判断先秦思想家能够见到的文本原貌可能如何。可见研究这一问题还需要以其他传世文献作为参照。此外还有一些是并无其他文献可供参考的独特文本。但为这类文本锚定思想史地位，同样需要以传世文献作为参照。比如，日本学者池田知久指出郭店楚简《穷达以时》是和《吕氏春秋·慎人》篇、《荀子·宥坐》篇等大致同时代成书的文献，并由思想史人物、文本之对应时间，以判断郭店楚墓的下葬年代可能在战国晚期。<sup>②</sup>即使我们不认可池田知久的观点，也存在另一种值得思考的可能性：部分传世文献的思想甚至表述有可能在更早之前的文本中就已出现，在此基础上可以对思想史关键概念之演进、篇章之生成方式，建立更深入的认知。对此，既需要学者高度关注传世文献与出土文献的对读，也要求将学者需求与现有技术进行良好对接，进一步开发数据库的检索、匹配与计算功能。<sup>③</sup>在以往的文献学工作中，出土文献由于性质相对特殊，故容易在文献学理论探析和具体整理工作中被学者忽视。有必要通过数据库的建设来抹平知识鸿沟，加强相互交流。

#### 四、展望：数据库的深度加工

荆门市博物馆未公开的“郭店楚简信息资源库”，因其独具地利人和，有可能在数据库质量层面超越既有之数据库。据介绍，该数据库主要内容包括数字资源、文字信息、学术资源三部分。其中数字资源包括高精度三维点云、高清纹理影像、高精度三维模型、三维交互程序、高清红外影像、数字拓片影像。文字信息包括原文、释文、注释、出处；学术资源包括研究论文、著录文献。但目前遗憾的是相关资源并未公开发布，故学者对其情况还不能建立更全面的认知。但仅就能了解到的信息来看，该数据库旨在收录全部与郭店楚简有关的信息，且在数据的精度和呈现类型方面都有明显进境，颇值期待。

① 李玉平：《互文性：文学理论研究的新视野》，北京：商务印书馆，2014年。陈洪、陈千里：《寻绎文化的血脉：互文视角下的中国文学经典》，北京：高等教育出版社，2019年。

② 说详池田知久：《郭店楚简〈老子〉新研究》，南京：江苏人民出版社，2022年。

③ 笔者曾经想象过能够更深度利用数据库资料和文本计算的“e互文”，近年来的一些数据平台已经具备相应功能。

资源公开与共享，是建设学术数据库的重要意义，也是研究者的首要需求。以往所受到的文献环境制约，在相当程度上已可通过研究者良好的信息素质加以弥补。但弥补的前提是具备可供学者持续使用的优秀数据库，而这一问题在类似领域也屡有出现——前文已述及部分重要数据库的湮灭。数据库需要后续维护以确保正常使用，亦需要不断更新数据以配合学术最新进展，而其背后都应有相关建制、人员持续付出努力。如果仅就“郭店楚简”本身而言，其信息是相对有限、且已被全面著录的，但其“衍生文献”，如新的释文、注解与研究，则必然不断增加。相关文献的类型至少包括著作、学术论文及网络论文，语言亦不仅限于中文，这就需要研究者持续收集资料，并根据研究成果调节数据库功能与数据信息。尤其是通过网络论坛发表最新研究观点，在出土文献研究领域非常流行。前文所述及的几个代表性的古文字、简帛研究网站，均发表或收录了大量的郭店楚简研究成果，其中不少属网络首发，并为学术共同体所认可。

**以数据结构化及深度加工提升数据库水准。**“数字化”(Digitalization)是数据库设计的基础阶段，它主要强调的是将原始信息转化为电子形式，以方便存储、处理和检索。在这个阶段，数据库建设者会将纸质资料、图像、音频、视频等各类信息转化为电子文件，并采用各种数字化技术进行存储和管理。然而，虽然“数字化”提供了极大的便利，但它并不足以满足所有需求。部分出土文献数据库，尽管整理了大量原始资料，能够提供的功能却相对有限，有的甚至仅仅提供相关数据的阅读与检索功能。在此基础上将数据结构化并建设多模态数据集(Multimodal dataset)，则是一个更高层次的要求和过程。它不仅要求将信息数字化，更要求将这些信息以特定的方式进行结构化标记，以便于更深入的分析和研究，为量化研究、人文计算、可视化等工作提供基础数据，研究者、开发者可以在此基础上不断开发新的功能。近年来，中文古籍研究界已经开发了多种具有丰富文本挖掘功能的数字人文平台，其思路可资借鉴。<sup>①</sup>

作为数字人文领域和出土文献的关注者，笔者建议，理想的郭店楚简数据库建设，还应该结合学界（尤其是不同领域研究者）与业界的需求，尝试拟定某些核心标准问题，以提升数据库建设的质量，并为学界提供理想通用数据平台的参照。本文所提的部分建议与思路，有的在相近领域中已有成功先例，有的在技术上已不存太大瓶颈，有的则属于未来数据建设的可能性之一。有关共同标准的具体建设工作，期待各学科从业者共同探讨，有所创造。不论是传世文献还是出土文献，都是研究者、爱好者认知中华优秀传统文化的重要文化资源，我们期待

<sup>①</sup> 刘石：《文献学的数字化转向》，《文学遗产》2022年第6期。



已进入公共版权领域的文献资源可以得到充分的整理、共享与利用。比如未来可考虑建设低代码平台，集成文献数据与基本功能，以满足不同使用者的多元研究需求。

## On the Gains and Losses of the Digitalization of Guodianchujian and the Development Standard of the Database of Unearthed Documents

Zhang Haosu

**Abstract:** Taking the digitization work of Guo Dian Chu Jian as a case study, we can see the gains and losses of the construction of unearthed literature database. At present, there are about ten kinds of representative data databases related to Guodianchujian, and the construction of basic data of Guodianchujian has been completed. However, there are some defects in the data source, function development and user friendliness of each library. The core problem is that each database does not use the latest chart and text as the source data, resulting in data omission, miscompilation, misfix, misinterpretation and other problems. The ideal database of Guodianchujian should set the description style as much as possible, save all the cultural relics information through digital technology, introduce the research and thesis data, and develop more functions according to the research needs. In short, it is necessary for Guodianchujian database to further develop from digitalization to data, that is, to explore ways to provide more services on the basis of basic full-text reading and retrieval. This experience is applicable to the construction of various unearthed document databases.

**Keywords:** Guodianchujian; Unearthed Literature; Digital Humanities; Database; Low-code platform

(编辑: 许可)