



2019年新信息环境下 CASHL 资源与服务拓展设计研究

利用开源平台 ArchivesSpace 进行大型 特藏深度揭示与服务研究 ——以《希腊铭文集成》为例

课题单位：复旦大学图书馆

课题负责人：李彦霖

课题组成员：沈扬、任磊、宋志强、张春梅

结项时间：二零二零年六月

摘要：针对 CASHL 大型特藏在保持其系统性和完整性的前提下进行深度揭示和服务往往需要耗费大量的时间和经济成本搭建难以长期维护的专用平台的问题，为了更好地为全国读者提供这样相对唯一却权威的文献服务，缩短从采购成功到提供服务之间的时间、提高开放率和使用率，项目提出可以利用 ArchivesSpace 文献管理系统进行深度揭示和服务的解决方案。ArchivesSpace 系统为免费开源系统，用户可以免费部署使用，并根据自己的需要进行二次开发。ArchivesSpace 虽然主要为档案管理开发，但因为其起源地美国图书馆中档案和特藏常常属于同一个部门，馆藏指南的形式也就经常被应用在了美国图书馆的珍本特藏揭示中，有可以借鉴的描述规则，以补足 MARC 描述的细节和深度不足，并可输出世界通用格式 EAD 用于数据交流。在项目研究过程中，项目组总结出了将相关描述规则运用于 CASHL 大型特藏描述的建议。

关键词：CASHL ArchivesSpace EAD 特藏

目 录

1. 大型特藏的编目需求	4
2. 运用档案思维和档案届平台的设想	4
3. 应用过程	7
3.1 利用 ArchivesSpace 系统的深度标引	7
3.1.1 MARC 元数据映射	7
3.1.2 深度揭示	8
3.1.3 工作和使用流程	10
3.2 系统优化	10
3.2.1 汉化	10
3.2.2 防火墙	12
4. 结语	12
4.1 提高了特藏文献揭示的系统性和完整性	12
4.2 提高了特藏文献的利用率	13
参考文献	15

1. 大型特藏的编目需求

大型特藏，是文专图书的标志性收藏，具有 1.学科集中，有相对完整的专题；在国内（至少是高校范围内）具备相对唯一性；2.系统性和完整性，需要在一个地方收藏，无法拆分；3.平时经费很难采购；4.能够成为文专图书建设的标志性收藏；5.能够揭示、报道并为全国服务；6.由学者实名荐购的特点^[1]。

为了更好地为全国读者提供这样相对唯一却权威的文献服务，大型特藏文献对揭示的深度有很高的要求。

在复旦大学图书馆现行的流程中，大型特藏一旦采购成功，通常都会直接使用 MARC21 对大型特藏进行编目。但 MARC21 无法对特藏进行有深度有层次的描述^[2]，元数据无法深入到册以下的层次，即使对于著录了完整目录的书目信息，也无法提供章节甚至小节层面的检索点。

同时 MARC 著录缺乏层级继承。受限于该标准的局限性，使用 MARC 著录特藏往往存在过于关注单个馆藏的情况，而无法达到在特藏档案著录中形成由馆藏组合成系列，由系列组合成特藏，以凸显其系统性和完整性的结构。而目前的 MARC 著录数据无法突出特藏的层级结构不但导致在开放获取的便捷性受损，也无法直观揭示多个专题的大型特藏之间直接相互关系和所共同承载的内涵。

这导致目前的数据无论在对于大型特藏的描述性上还是实用性上都存在不足。数据条目的缺乏导致难以组织起有效的检索指南，降低了特藏的开放率和使用率，因此许多高校选择为每一套大型特藏定制专用平台。

但定制平台也造成很多问题：制作数据的工时以外，定制平台通常需要至少一个月以上的平台搭建时间，没有后台管理系统所有数据需要公司协助灌注，后期修改也需要公司配合造成数据维护困难，数据也和其他资源无法互通。更不用说因此带来的额外经济成本。

因此，亟需一套能够满足层级（套——单册——目次）著录需求的，且无需定制开发即能方便应用的进行大型特藏内容标引、管理和服务的系统，以便让大型特藏最大限度发挥应有作用。

2. 运用档案思维和档案届平台的设想

使用 ArchivesSpace 等 EAD 编辑系统制作馆藏指南（Finding Aid）作为国外高校档案和特藏机构常用的揭示体系，通过多年的使用和改善成为了比较成熟的图书馆档案特藏揭

示体系，海外诸多名校均已采用该体系多年^[3]。馆藏指南不单单描述馆藏题名、作者、日期、位置等常规编目数据，还描述馆藏档案的品相、流传历史和大型馆藏之间的层级关系等信息来帮助潜在的用户了解和分析特藏档案的内容，解决国内档案中缺乏层级的著录导致的“只见树木，不见森林”问题^[4]。

常规出版物所具有的元数据（MARC21）往往是非常格式化的，而档案馆、特藏馆及博物馆中藏品通常无法总结出一个通用的编目格式，且此类机构馆藏往往具备组合（Collection）比单件（Item）价值高数倍的情况。馆藏指南最早是针对这种特殊性为揭示档案类资料所设计的方案，如相对于 MARC21，馆藏指南更注重多件馆藏之间所形成的补充关系、层级关系，逐渐被档案以外的领域所利用。

馆藏指南的制作一般根据美国档案馆员协会 SAA（Society of American Archivists）出版的 DACS 手册（Describing Archives: A Content Standard）；在美国图书馆中，档案和特藏常常属于同一个部门，这些部门中又以档案为主导的占大多数，很自然的，馆藏指南的形式也就经常被应用在了美国图书馆的珍本特藏揭示中（多见于成规模的档案资料中含有少量书籍或其他本册如日记的），因而已有可以借鉴的描述规则^[5]。馆藏指南的电子化同样由 SAA 制定了基于 XML 语言的 EAD（Encoded Archival Description）格式^[2]。XML 是可扩展的结构化标记语言，现在图书馆广泛认可的数据连接技术 RDF 资源描述框架也同样是基于 XML 语言。

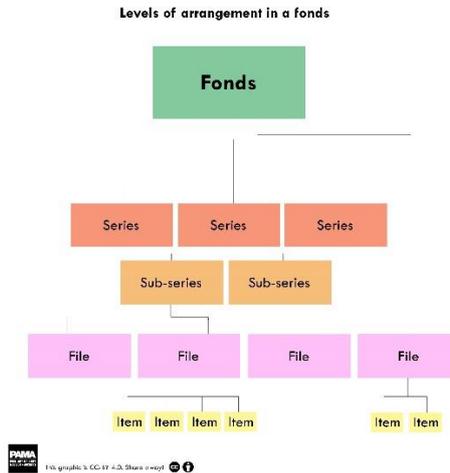
XML 作为 EAD 使用的语言基础，允许用户对标记语言进行自定义，以此制定相关标准以实现在网络不同平台或应用间的结构化数据的描述和交换。EAD 在 XML 的基础上定义档案特藏著录所需的关键标签，形成统一的格式与标准以此保证各馆藏机构能够对特藏信息有效的进行跨平台查询、检索、显示和交换，是符合当前网络环境的元数据标准^[6]。同时，XML 也赋予 EAD 一定的灵活性，可以不同国家不同机构根据实际情况增加著录标签，以最大化特藏档案发现的可能性。这也正是该标准目前不但在美国、英国、澳大利亚、德国、葡萄牙、荷兰等国家能够推行的基础，也是 EAD 能够在中国顺利进行本土化的保证^[6]。目前，虽然 EAD 在我国研究较少，但吉林大学教授王萍^[7]的相关研究《档案著录国际标准(EAD)的推广应用》已经证明了 EAD 在中国推行和应用的可行性。

同时，相较于传统针对内容和形式进行著录的标准，EAD 在保持对于内容关注的同时，添加了藏品背景信息标签的著录，这对于档案特藏研究者是十分有用的信息。多级著录是 EAD 的另一个特点，通过标签之间的继承关系，有效利用 EAD 标准，为特藏进行分级以及对各级进行著录，避免重复著录，增加检索点提高检索效率^[4]。因此为所藏档案构

建一套具有等级结构的检索指南框架，能够更好地支持目前的网络检索要求。

EAD 也具有在未来形成关联数据进行数字人文研究的能力。在 2012 年由英国 JISC 的档案中心发起的 LOCAH 项目就将 JISC 档案中心拥有的数据同 DBPedia、BBC 和 OCLC 相联系，制作出了基于数据可视化的关联数据地图^[8]。

档案指南的最重要的部分是一组档案的树形结构：



档案结构示例 [9]

而在档案届也有讲顶层元数据输出成 MARC 格式以便读者在 OPAC 中检索的传统。因此如果将一组档案看做一本书，而树形结构是这本书的目录的话，可以很容易得出，档案的揭示需求和大型特藏的揭示需求高度相似。

2019 年初，复旦大学图书馆正式引进开源平台 ArchivesSpace 用于制作馆藏指南：ArchivesSpace 是整合了 Archivist' Toolkit(AT)和 Archon 两种用于制作机读馆藏指南的工具所开发的开源档案管理平台。平台依据 DACS，并有所扩展。其他同类平台通常需要上传制作完成的 EAD 代码文件，而 ArchivesSpace 提供表单式填写界面，操作更便捷。平台本身提供读者界面，有简易的请求功能，自动生成引用功能等常规功能，也提供 EAD 文件导出功能可用于在其他自有网站发布，提供导出 RDA 格式的 MARCXML 供导入本馆自有系统或跨馆联合发现平台，功能相对其他同类平台完善^[3]。

研究者设想可以利用开源工具 ArchivesSpace，套用档案编目规则，对大型特藏《希腊铭文集成》、《希腊铭文补编》进行以制作馆藏指南的方式进行深度揭示：

1. 根据特藏资源的特点进行更科学的深度揭示。
2. 提高特藏文献的利用率。
3. 缩短从采购到服务中的数据和平台制作经济、时间成本。
4. 探索档案指南在中国高校图书馆的应用以及 EAD 的本土化

3. 应用过程

3.1 利用 ArchivesSpace 系统的深度标引

3.1.1 MARC 元数据映射

首先研究者手动导出用于实验的已有《希腊铭文集成》相关书目 marcxml 文件，手动导入 ArchivesSpace，制作成简单顶层数据，并根据书籍的丛编构成初步整理出层级-每个单册视为一个 Item，每个丛编视为一个 Series，有更细化的丛编结构的也可以利用 Sub-Series 等层级来体现，而整套大型特藏视为一个 Collection。在单册层面及以上大部分编目规则都可以直接套用自 DACS，以揭示档案的思维和工作方式为大型特藏书目编录丰富而有深度的元数据。

ArchivesSpace 中字段/EAD 和 CNMARC 及 MARC21 中常用字段的映射如下：

CNMARC	子字段/ 指示符 号	MARC21	子字段 /指示 符号	ArchivesSpace 字段	EAD
010 国际 标准书号	\$a ISBN/\$d 定价	020 国际 标准书 号	\$a ISBN/\$c 定价	N/A	N/A
200 题名 与责任	\$f 第一 责任人	100 人 名 110 机 构名 111 会 议名		Agent (Role: Creator)	<origination><persname> <origination><famname> <origination><corpname>
200 题名 与责任	\$a 正题 名	245 题 名	\$a 正题 名/\$c 责任者 说明	Title	<unittitle>
210 出 版发行项	\$c 出版 发行者 名称/\$d	260 出版 发行项	\$b 出版 发行者 名称/\$c	Dates (Label: Publication)	<unitdate>

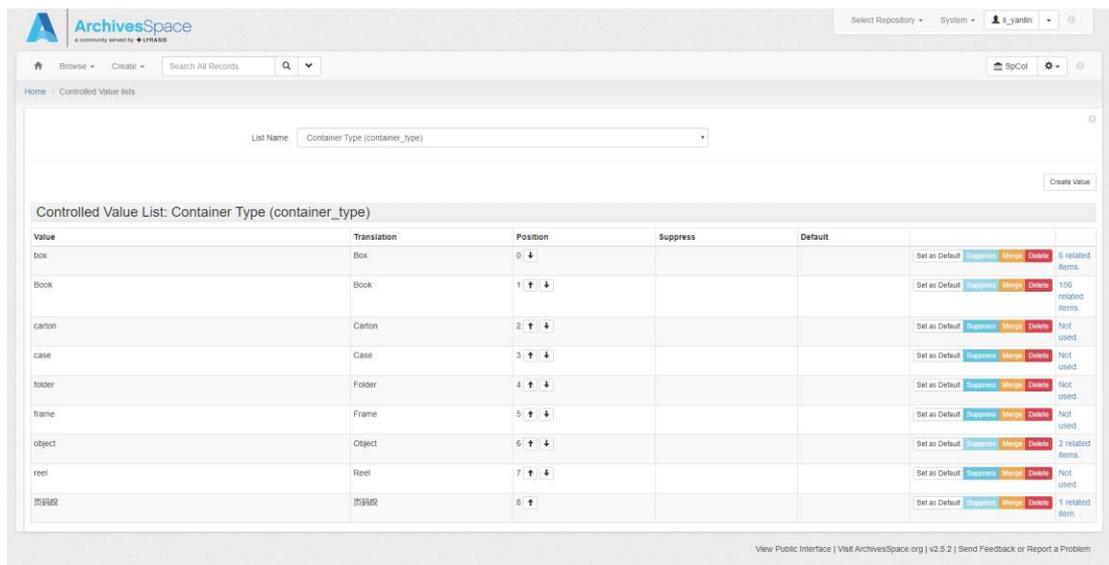
	出版、 发行日 期		出版、 发行日 期		
215 载 体形态项	\$d 尺寸 及开本	300 载体 形态	\$a 内 容/ \$c 尺 寸	Extents (Physical Details/ Dimensions)	<physdesc> <dimensions>
905 馆 藏信息	\$a 收藏 馆代码 /\$d 分 类号/\$e 书次号	852 馆 藏位置	\$a 收藏 机构或 个人 /\$h 分 类号/\$i 书次号	Instances (Top Container)	<repository> <physloc>

书籍元数据	Collection
Inscriptiones Delii / consilio et auctoritate Academiae inscriptionum et humaniorum litterarum francogallicae editae. 1961	Series
fasc. 2. Inscriptiones Delii liberae. Tabulae archontum, Tabulae hieropoconum annorum 314-250 / editi Felix Dürnbach	Item Books Book: K885.457.42 :f59:10 /fasc.2 :f0 [B0905091701303]
fasc. 3. Fasciculi tertii tabulae / adnotavit F. Hiller de Gaertringen (1927)	Item Books Book: K885.457.42 :f59:10 /fasc.3 :f0 [B090509170139C]
fasc. 4. Inscriptiones Delii liberae, Decreta foederis, Catalogi dedicationes varia / editi Petrus Rousset (1914)	Item Books Book: K885.457.42 :f59:10 /fasc.3 :f0 [B090509170139C]

3.1.2 深度揭示

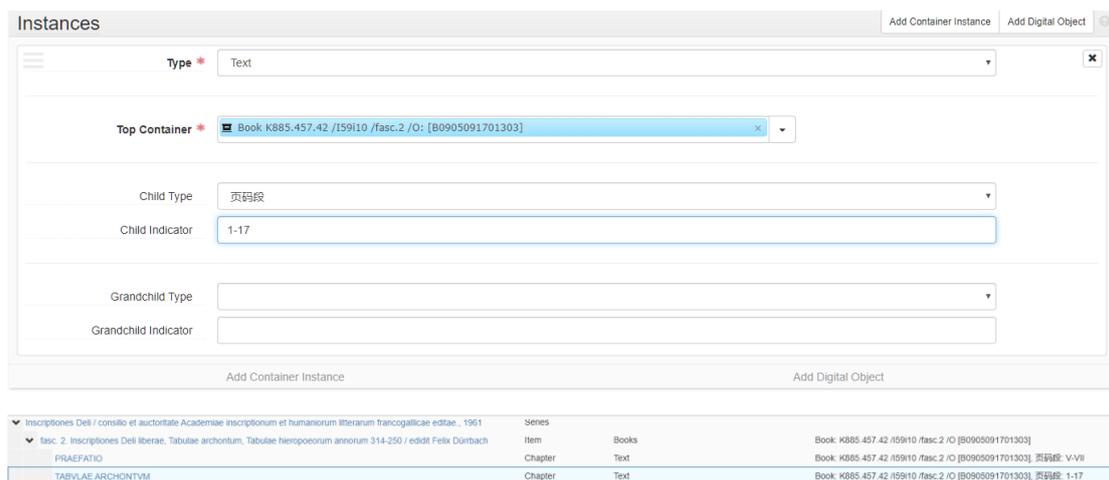
为了深度揭示章节及以下层级元数据，在书籍中建立子目录，使用 Other Level，描述为 Chapter，录入章节进行深度揭示，

因为 ArchivesSpace 系统为揭示未装订成册的，为了体现章节的页码信息，需要在 Container Type 中增加一个容器选项：页码段。



下拉列表字段选项编辑界面

在容器著录时，选择该单册的条码，并将页码段作为子容器录入。



ArchivesSpace 系统中，collection 以下层级均提供相同字段，可以根据需要进行深度标引，而得益于树形的结构，所有条目也都会继承上级条目的描述从而减小深度标引的工作量。

每层 ArchivesSpace 都提供丰富的不限数量的检索点和备注字段，从而实现深层次的描述：



检索点衍生到章节后可以实现跨系列的章节层面检索/对比等高效查询和利用。

复旦大学图书馆特藏部所收藏的部分手稿应保护隐私等原因不方便提供全文查看，为方便读者检索将原文作为备注录入在了条目对应的 Notes 字段中，并勾选不发布选项，以达到在不发布全文的前提下提供全文检索。研究者认为此方案同样适用于为出于版权原因不方便提供全文查看的大型特藏提供全文检索并可以进一步为文本分析服务（如词频分析、情感分析等）提供素材。Note 字段同时提供丰富的属性标签用于细化备注，字段内部对 mark down language 标签全面支持且提供更多用于系统内部关联的标签，深度使用用户可以应用于各种描述、备注场景。

应用 MPLP 的最小化发布前工作量最大化发布馆藏量的工作流程规划理念^[10]，在本章节元数据制作完毕后既可以发布上线服务。其后的数字化和数据深化（内容标引、小节目录等）则可以在发布后再逐步进行，甚至在更多考虑性价比或人手不够等情况下可以视每个部分受读者需求的程度进行优先加工。

3.1.3 工作和使用流程

馆员工作流程：

采购→（MARC 编目）→EAD 元数据制作→（数字化）→上线服务→按既定优先级细化编目和数字化/按读者需求细化编目和数字化

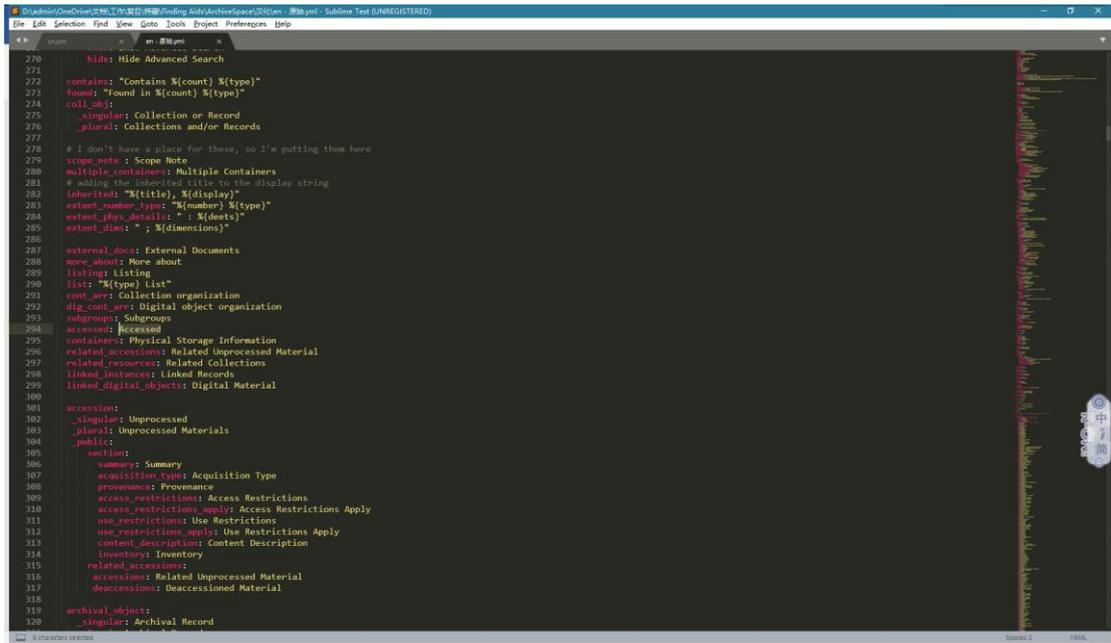
CASHL 读者使用流程：

OPAC（MARC）记录→借阅/ArchivesSpace 记录→使用电子版/提交文献传递申请→挂接文献传递产生的电子版（如无版权限制）供后续读者使用

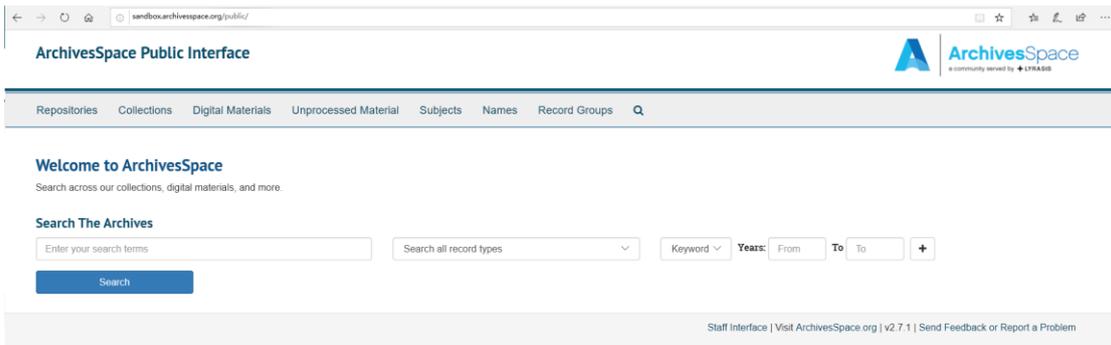
3.2 系统优化

3.2.1 汉化

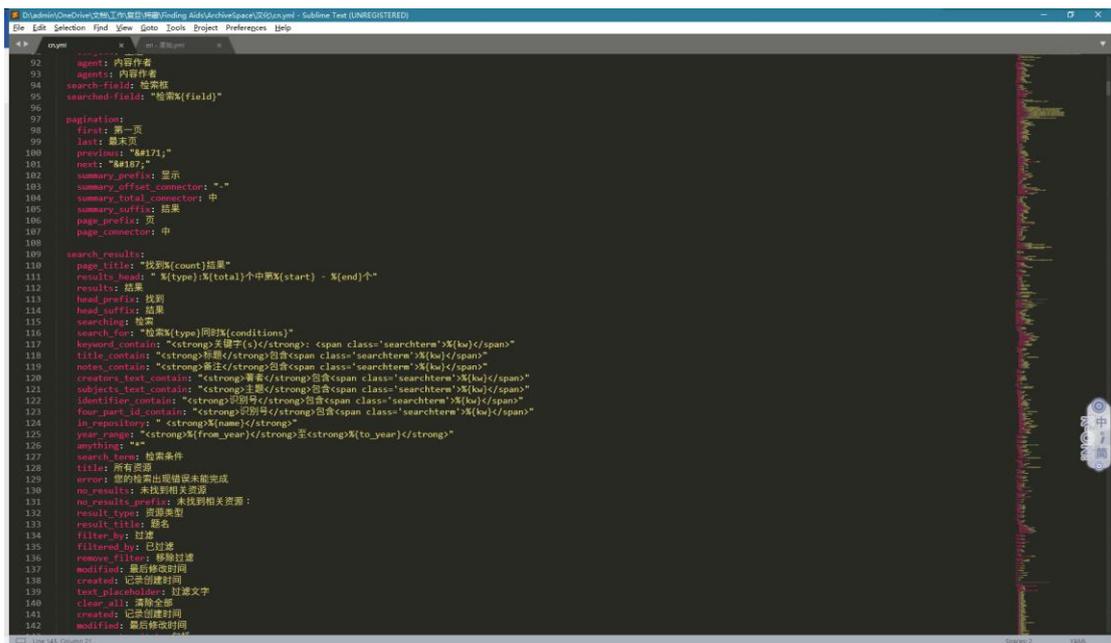
为方便用户使用，研究者对用户端进行了汉化，直接修改 yml 文件并替换即可，无技术难度：



原版英文 yml 文件



原版英文读者端界面



复旦大学图书馆汉化后的 yml 文件

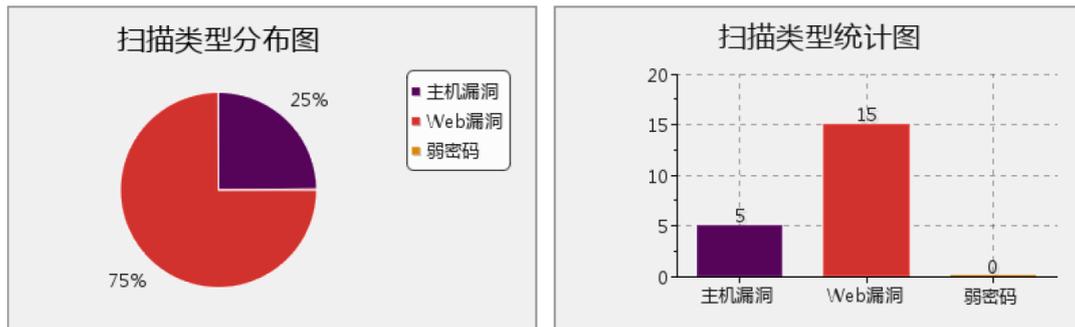


复旦大学图书馆汉化后的读者端界面

ArchivesSpace 是一个开源软件，用户也可以对前端自行定制。

3.2.2 防火墙

申请读者端界面的域名时，根据复旦大学的有关规定扫描检测未优化的平台时发现了 20 个风险，简单优化后现已上线。



这部分问题在前端用 WAF (web application firewall) 进行了保护后均已解决，复旦大学图书馆所使用的 ArchivesSpace 系统前台 (tecang.fudan.edu.cn) 现已上线。ArchivesSpace 的前台 (读者端) 和后台 (馆员端) 用了一个域名下的两个端口，出于数据安全考虑，复旦大学图书馆只上线了前台，保持了后台仅可校内访问的状态。这样的状态导致防火墙在设置时需要开放相应端口，否则后台将无法通过服务器以外的机器访问。

复旦大学图书馆现使用 Aleph 系统管理书目数据，ArchivesSpace 系统则处于补充信息的定位。两个系统中都有通过记录 id 获得数据的接口，只要录入了对方的对应 ID，或则是能做到 ID 映射，基本上可以做到数据互访，从而解决需要手动导出导入 MARC 记录，以及需要手动修改 MARC 挂接相关 ArchivesSpace 链接的问题。

4. 结语

4.1 提高了特藏文献揭示的系统性和完整性

根据特藏资源的特点进行了更科学的深度揭示。实现了章节甚至更深层次的元数据标引、展示和检索。更好地对大型特藏进行系统和完整的展示。

复旦大学图书馆 OPAC 中元数据展示界面

复旦大学图书馆 ArchivesSpace 中元数据展示界面

4.2 提高了特藏文献的利用率

利用该平台可以略过漫长的平台搭建过程的同时时间深层次揭示和服务, 只需下载部署立刻可用; 标引和扫描等深度揭示工作可以拆解, 按需求调整优先级等, 大大缓解人手和工期问题。如果可以得到推广, 图书馆之间的深度标引规则将统一为国际通行的 EAD 格式, 而不是每个平台的独立规则, 有利于资源的交流和共享。

本方法致力于为大型特藏的揭示提供一个快速而经济又略优于普通 MARC 揭示的方案, 但由于 EAD 格式对数据属性支持不够丰富, 还停留在检索点的层面, 在需要进一步利用数据(如数据挖掘等)时可能还需迁移到更专业的平台; 对于对数据的丰富性和深度有更高要求的项目, 研究者推荐可以先用本文的方法进行快速的揭示并上线, 以实现在采购

成功后第一时间开始提供服务，之后再对资料进行更深度的挖掘、分析，这样无论是在制作高质量数据还是搭建更合适特定大型特藏的专业平台的漫长的过程中，都不影响读者对文献的利用，也能缓解数据制作者、平台建设者的时间焦虑。

这个模式同样可以延展到普通馆藏的文献传递，既利用 ArchivesSpace 发布文献传递中产生的没有版权限制的文献资料的不完全电子版供读者自行查看，节省人力的同时也方便读者。

参考文献

- [1] cashl, “CASHL_开世览文,” [联机]. Available: www.cashl.edu.cn. [访问日期: 15 2020].
- [2] P. Carini and K. Shepherd, "The MARC standard and encoded archival description," *Library Hi Tech*, vol. 22(1), pp. 18-27.
- [3] R. Toov and A. Wick, "Making it Work - Understanding and Expanding the Utility of ArchivesSpace," *Journal Archival Organization*, Vols. 14:1-2, pp. 35-54.
- [4] 马寅源, “国内外档案多级著录的比较研究,” *档案学研究*, 卷 02, pp. 51-56, 2017.
- [5] S. E. Berger, *Rare books and special collections*, Chicaco: Neal-Schuman, 2014.
- [6] 宋凯, “EAD(编码档案著录标准)的应用调查与分析,” *兰台世界*, 卷 14, pp. 15-16., 2013.
- [7] 王萍, “档案著录国际标准(EAD)的推广应用,” *档案学通讯*, 卷 02, pp. 89-93, 2010.
- [8] “Use Case LOCAH (2010) Retrieved from,” [联机]. Available: https://www.w3.org/2005/Incubator/1ld/wiki/Use_Case_LOCAH#Use_Case_Scenario.
- [9] D. Meissner, "Implementing More Product, Less Process," Fayetteville, AR, 2016.
- [10] M. A. Greene and D. Meissner, "More Product, Less Process: Revamping Traditional Archival Processing," *The American Archivist*, vol. 68, pp. 208-263, 2005.