

名人手稿馆元数据方案的设计和实现

上海图书馆名人手稿数字图书馆课题组¹
(上海图书馆 200031)

名人手稿是上海图书馆的特藏之一，上海图书馆自 1992 年开始恢复向社会各界征集名人文献的工作，1996 年新馆开放后，更是加快步伐、增强力度进行收集整理，目前藏品已逾 5 万 2 千余件。建设一个名人手稿数字图书馆是上海图书馆数字图书馆建设的一项重要工程。

根据实际情况，名人手稿数字图书馆的建设将分期进行。在大规模地数字化之前，名人手稿数据库的建设将先期完成，同时进行数字图书馆系统的设计和试验性数字化，形成名人手稿数字图书馆的一个原型系统，供研究、演示和进一步修改完善后投入使用。

项目开始之初，课题组从元数据标准规范和数字化项目两个角度进行了全面考察。国际上类似的项目只有近期完成的“欧洲手稿与书信网络集成(MALVINE)”²项目在需求、复杂程度和技术环境方面最为接近，于是我们与该项目的人员取得了联系，获得了其完整的元数据元素列表。但由于该项目是一个手稿与书信的联合目录项目，采用的属性元素集合没有考虑直接执行有关的元数据标准，因此我们仅仅从属性元素的选取方面参考了该项目，在体系结构和方案设计上没有太多的借鉴。

名人手稿数字图书馆的开发目的是对名人手稿馆所涉及的所有资源进行有效的管理和利用。元数据方案是数字图书馆需求分析和系统设计时需要首先考虑的因素，是数据加工制作、藏品数字化和系统设计与功能实现的基础。制定名人手稿馆元数据方案的目的是提供其所涉及的所有资源类型的属性定义和置标方案(也就是相当于传统上的制定编目规则和目录格式)，为名人手稿数字图书馆提供资源管理(又称为内容管理)方案。元数据方案基本上决定了系统的整个架构，以及系统设计(包括各类著录系统、检索系统、管理系统等)的需求。

名人手稿馆馆藏在许多方面不同于图书馆传统的资源，它是文献和文物的结合。其元数据方案也不同于一般博物馆的元数据方案。与通用元数据标准的制定所不同的是，后者考虑更多的是互操作性和通用性，需要进行许多妥协和折衷，会牺牲许多个性和细节，而“名人手稿馆元数据方案”必须详尽地考虑和满足资源管理、保存、揭示、检索、利用等各方面需求，同时也要在实现众多个性化的需求之外，保证与上海图书馆的整个数字图书馆架构兼容，甚至与正在制定中的国家标准兼容，并符合国际上通行的标准和做法。只有这样，才能实现较高的互操作性、灵活性和可扩展性。

总体考虑

由于名人手稿馆的馆藏资源丰富，种类繁多，本方案所针对的对象不仅仅是名人手稿，确切地说应该是“名人手稿馆信息系统元数据方案”，包括创作手稿、信函、书画篆刻作品、照片、签名本、日记、录音/录像资料、奖状奖章等各方面的资料，并且有可能还会增加。对于这样一个复杂的系统，系统设计的开放性、灵活性、可互操作性和可扩展性就显得非常

¹ 本课题成员有：祝均宙，刘炜，赵亮，浦纯，楼向英，张春景，夏翠娟，孙秀娣，王洪治，徐频，朱小灵，王恺顺。本文执笔：刘炜，楼向英，张春景

² 参见：<http://www.malvine.org/>

重要，而且还要兼顾其永久保存的特性和整个信息处理流程易管理性的考虑。

按照传统的方法，可以请负责系统设计的计算机专家、资源管理部门的文献专家和系统将来的用户一起，对每一种资源类型乃至整个系统提出详尽的著录、检索和其他功能需求，然后通过开发专用的数据库系统来实现。这样做的问题是，系统一俟开发出来，就可能是一个“遗留系统”，虽然能够实现名人手稿馆对于信息管理的需求，但是不具有数字图书馆所要求的开放性、互操作性和可扩展性，将对上海图书馆现有的计算机信息管理系统带来新的互操作问题，也很难在资源内容上方便地与其他“数字图书馆”系统互通互联。整个系统将成为一个典型的“封闭系统”而不是“数字图书馆”。

或者可以按照目前“数字图书馆”的一般设计方法，制定一个包含“核心”元素和所有扩展元素的并集，作为元数据方案。但这样一个方案有一个危险，它无法保证与将来的国家标准，甚至上海图书馆自身的元数据标准相统一，因为这些标准尚在讨论之中，还未定稿，但是由于项目实施尚有两年甚至更长的过程，这些标准有可能在项目实施过程中就会逐步制定出来，这个过程中碰到的兼容性和互操作性问题很难处理。

为了尽可能避免上述问题，我们考虑制定一些基本原则，依据这些原则制定具体的工作手册，以求最大限度地获得灵活性和可扩展性。首先在体系结构方面尽可能参照一些成熟的参考模型、分析模型来做，例如 OAIS³所提供的信息系统参考模型在数字资源的永久保存方面提供了一个思考框架，国家图书馆已经在它的基础上有一些元数据方案的探索⁴；FRBR⁵对数字对象整个生命周期不同过程和形态的关系建立了一个思考框架，对于建立复杂的数字对象之间的 ER 模型以及不同阶段知识产权属性的管理非常有用，这个模型还可以看成是一个初步的名人手稿资源的本体模型。在元数据描述语义方面我们考虑尽可能“复用”现有的方案和标准，而尽可能少地“创造”新的元素；整个元数据方案按照 DCMI 对于元数据应用概要（Application Profile）的抽象模型⁶来建立。在置标方面尽可能采用标准的方案或者灵活的 XML/RDF⁷模式。

因此名人手稿数字图书馆的元数据方案是一种“混合”型元数据应用概要的形式，即借鉴 OAIS、FRBR 以及 DCMI 目前正在形成的 Abstract Model 作为方法论，采用以 DC-Lib⁸为基础的“上海图书馆元数据方案”作为核心元数据⁹，并从多种元数据标准、方案中“复用”元素，对所有元素的语义强调严格遵从，但在著录规范中对在每种特定资源类型中的具体含义进行补充说明，限定或扩展方式也强调尽可能采用现有的框架、体系和规范，并充分采用 XML Schema（METS¹⁰和 MODS¹¹等）和 RDFs（WSDL¹²）提供的结构限定方式，最后再考虑增加子元素或元素。

本文是对名人手稿馆元数据方案的总体介绍，包括原则、流程、框架、模型、元素集（包括核心集和扩展集）及其置标的考虑等等，限于篇幅，不可能介绍得非常详细，规范控制和系统的需求与设计将在以后另文阐述。当然本项目是一个具体实践，限于条件和水平，在实

³ Reference Model for an Open Archival Information System (OAIS). Consultative Committee for Space Data Systems (CCSDS). URL: <http://www.classic.ccsds.org/documents/pdf/CCSDS-650.0-B-1.pdf>(检索日期: 2004-2-1)

⁴ 中文元数据方案(征求意见稿). 国家图书馆. 内部资料, 2001年6月

⁵ Functional Requirements for Bibliographic Records (FRBR). IFLA Study Group on the Functional Requirements for Bibliographic Records. URL: <http://www.ifla.org/VII/s13/frbr/frbr.pdf>(检索日期: 2004-2-1)

⁶ Dublin Core Abstract Model. Andy Powell. URL: <http://dublincore.org/documents/abstract-model/>(检索日期: 2004-1-14)

⁷ RDF Vocabulary Description Language 1.0: RDF Schema. W3C Proposed Recommendation. URL: <http://www.w3.org/TR/rdf-schema/>(检索日期: 2004-2-1)

⁸ Library Application Profile. Rebecca Guenther. URL: <http://dublincore.org/documents/library-application-profile/>(检索日期: 2004-1-14)

⁹ 这一核心集也包含了 IFLA 提出的 10 项“核心记录元素 (Core Metadata Elements)”集合(草案)。参见: <http://www.ifla.org/VII/s13/guide/metaguide03.pdf>

¹⁰ Metadata Encoding and Transmission Standard (METS). URL: <http://www.loc.gov/standards/mets/>(检索日期: 2004-1-14)

¹¹ Metadata Object Description Schema (MODS). URL: <http://www.loc.gov/standards/mods/>(检索日期: 2004-2-1)

¹² Web Services Description Language (WSDL) 1.1. W3C Note. URL: <http://www.w3.org/TR/wsdl>(检索日期: 2004-2-1)

现时必然有许多妥协和折衷，缺憾之处在所难免，敬请批评指正。

设计原则

设计原则是设计思想的具体体现，贯穿整个设计过程，对项目的后期实施也会产生巨大影响。名人手稿馆元数据方案的设计原则一部分可以是元数据方案设计的通用原则，但在具体尺度的把握上有自己的特色，另一部分属于具体原则，专门针对本项目而制定。

通用原则包括如下六项¹³：

1. 简单性与适用性原则

简单性要求元数据方案尽可能采用精简的“核心”元素集，以便于实现，降低成本，加快实现进度，以及有利于互操作的实现；适用性要求数据元素必须“够用”，必须能够完全实现系统需求。简单性和适用性是一对矛盾，参与方案设计的各方人员往往会有不同意见，需要仔细斟酌。对于名人手稿馆元数据方案，我们以适用性原则为重点，同时从技术实现的角度删繁就简，满足最低“核心”元素集的要求。

2. 专指度与通用性原则

专指度指对于特殊领域资源描述所提出的特殊要求的满足，例如名人手稿馆元数据方案中的“捐赠人”、“捐赠日期”、“书写工具”、“誊写人”等描述要求；通用性原则要求考察是否有更一般的或这些“专指概念”的上位概念能够满足描述要求，例如考虑某一“专指”元素到一个一般“核心”元素的映射或考虑如何进行“dump down”的方案；决定是用“专指”元素还是“通用”元素的过程，就是权衡专指度与通用性的过程。这两个原则也是一对矛盾，其实这也是考虑互操作问题。在名人手稿馆元数据方案中典型的例子是：如果考虑通用性，对于不同资源类型的相应元素定义统一的元素名称（如“责任者”），这样对于某些类型的资源会显得非常别扭（例如对应“书信”中的“下款名”），这就需要权衡，是增加元素，还是增加修饰，还是在系统实现时进行处理，等等。

3. 互操作性与易转换性原则

因为元数据方案的立足点就是解决互操作问题，这里的许多原则实际上都是在一个侧面或从一定程度上解决互操作问题。可以看出“互操作性”原则是元数据方案设计和实现中需要遵循的最重要的原则，通过尽可能地复用标准方案、复用元素、或复用修饰词及扩展方式，以及建立映射、转换机制等方式来达成互操作性。易转换性原则指元素的含义应该尽可能符合“原子性”要求，以便于向其它元数据方案（一般是标准的或“核心”的方案）映射或转换，尽可能保证在映射和转换过程中语义不损失。

4. 灵活性与可扩展性原则

强调标准性和专指性常常都意味着灵活性和可扩展性的损失。灵活性和可扩展性都是指元数据方案对于未来的适应性，常常要求总体的平衡，不能在某一方面强调过度。例如对于限定，应该支持多种限定方式，同时个别元素的限定级别不宜过深；对于现有标准的遵循，不宜过于严格以至于标准的未来版本扩展了元素而自己制定的方案却扭曲地局限在“核心元素”的限定上（有些方案将许多扩展都置于 DCMES 的“Description”下，以至于这个元素过于臃肿，同时增加了限定级别，方案显得非常不灵活）。

5. 用户需求原则

这一原则是不言而喻的。但是其前提是必须分清谁是用户。名人手稿馆元数据方案的用户首先是其工作人员，因为整个系统首先作为馆藏管理系统，然后是专业用户。普通读者的需求可以包含在专业用户中。

¹³ 六项通用原则的前五项参考 2003 年 7 月《专门数字对象描述元数据规范研制工作手册（修订稿）》并根据名人手稿馆的应用进行了扩展与细化后拟定。

6. 遵循现有标准原则

通过符合元数据标准或协议而达到“互操作”是根本意义上的互操作，因此遵循现有标准对于实现互操作是至关重要的。但是对于标准也要有所选择，元数据领域有很多“标准”或事实标准，目前大多数标准对于具体应用来说没有能立即照搬使用的，必须遵循一定的方法论和应用流程进行选择取舍，这是元数据应用的难点，也是目前研究多应用少的原因之一。

除上述通用原则之外，我们考虑的具体原则是指在元数据方案的设计和执行的流程中需要掌握的原则，包括资源分析原则、扩展原则（元素扩展原则和修饰限定原则）、元素定义原则、置标原则和系统实现和应用原则等，在此不作详述。

设计流程

整个设计流程以下图简略表示：

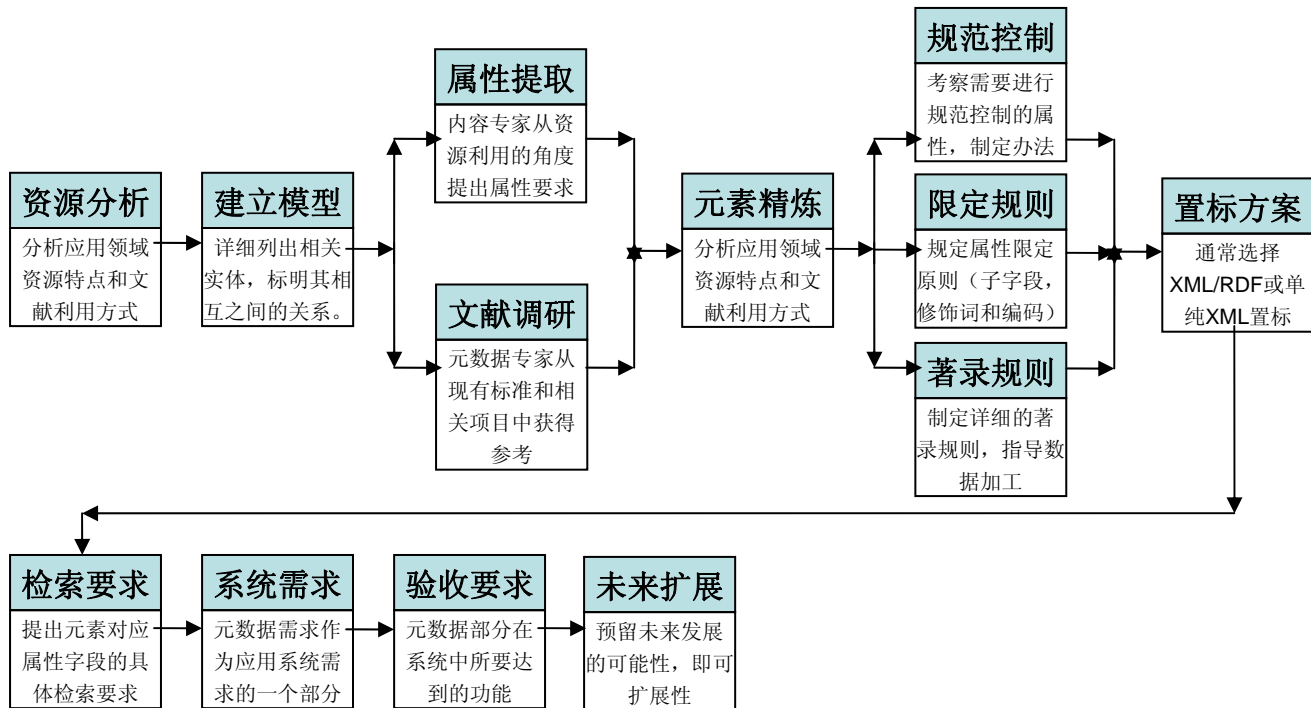


图 1：名人手稿馆元数据方案设计流程图示

资源特点

名人手稿馆的收藏原则与一般图书馆和博物馆都有所不同，立足于文献，但却是“大文献”概念，具有一定历史人文价值的任何形式的载体都可能是收藏对象。名人手稿馆的馆藏资源类型可以看成是一个其特有的内部本体，包括创作手稿、书信、日记、照片、书画篆刻作品、签名本、笔记、账簿、录音录像资料、证章奖状以及各类小件实物等，将来数字化对象也会直接成为名人手稿馆的馆藏，而不需要有物理馆藏与之对应。名人手稿馆的工作人员作为领域专家，倾向于上述直观实用的分类，这种分类实际上是内容和形式两方面的属性混合起来加以区分的。系统分析人员希望要么按照内容属性区分，要么从形式上区分，当然系统可以支持多种分类方法同时存在，并提供满足上述“混合”分类的用户视图，但前提是必须彻底弄清每种资源的属性和资源之间的相互关系。由于名人手稿馆资源类型复杂的实际情况，系统必须支持分类的灵活性和可扩展性，因为谁也无法预料将来会入藏哪些藏品，馆藏极有可能超出原有的“本体”范畴。

“关于名人”的资源是名人手稿馆不同于图书馆或者博物馆馆藏的的另一个重要特点，除按照所需揭示的属性客观著录外，与名人的相关属性也应在资源描述方案中予以充分揭示，否则就失去了其作为名人手稿馆馆藏的意义。

明确定义资源类型是制定具体的元数据方案中非常重要的第一步，对于资源类型的界定和说明以资源分析报告的形式固定下来，有助于明确项目的边界，更好地理解馆藏，在提取“核心”属性方面达成一致，确定描述资源对象及其属性之间的关系，并初步形成对于著录的许多规定。资源分析通常需要确定的内容是：资源类型的定义和范围、资源对象之间的关系、对象粒度（著录级别和著录单元）、属性语义（具体内容）以及对于具体属性的检索需

求。

资源类型

通过资源分析报告可以归纳出每种资源类型的特点，从而确定满足其需求的属性元素，元素之间的关联以及结构，从而确定每种资源类型的元数据方案，达到对名人手稿馆资源细致揭示的目的。下面就以上各方面对每种资源类型进行说明。

创作手稿

创作手稿指的是名人在创作或研究过程中所产生的手书文字记录，可以是草稿、未定稿、校改稿（包括打印稿的手书校改稿）或最终的完成稿。创作手稿的体裁不限，可以是任何一种文学作品，也可以是论文，但不包括书画作品和其他艺术作品。鉴于创作手稿是未出版的公开发表物，具有不确定性，所以创作手稿的每一次修订都作为一个版本，因此名人手稿馆对其著录到每一种创作手稿的修订稿。创作手稿的价值就在于名人的笔迹和手稿所作的一系列修改，所以我们将打印稿不算作创作手稿的一部分，在重点揭示创作手稿的作者的同时，并对在创作手稿的过程中所使用的书写工具，文字色泽和稿件的修订作为描述的重点。手稿的抄写人虽然对手稿的“内容”没有贡献，但鉴于手稿的抄写人与手稿作者之间的关系一般会比较密切，因此我们将手稿的抄写人作为“誊抄者”元素放入元数据方案。创作手稿中有一部分是翻译稿，即翻译国外作品的稿件，针对此类手稿，我们还特别描述了稿件的原作题名，原作的作者，原作的语种等内容。

信函

信函指的是指源自名人的各类通信，包括信件、明信片、贺卡、请帖等。我们将每一份信函作为一个著录单位，普通信件可以包括信封、信纸或附件。信件同时包含信封与信纸者，作为一个完整的著录单位著录。仅有信封或仅有信纸者，信封或信纸作为一个单位著录。对于一个信封包含多封信件，著录时要在“关联”字段说明关系。并把信封和每封信件都作为单独的著录单位。收信人对写信人的回信，这个过程中两封信件之间存在一定的继承关系，两者的关系也需在“关联”字段中说明。信函也是名人的手迹，因此信函的描述中也包括了书写工具和文字色泽元素，由于信封和信纸会出现使用不同的书写工具和不同的文字色泽，因此我们将信纸和信封都分别用书写工具和文字色泽进行了描述，并对信纸的风格特征也进行了揭示。由于信封是作为信函整体整体的一个部分，因此除了对信纸内容进行充分揭示以外，还增加了收信人地址栏，寄信人地址栏，收信人邮编，寄信人邮编，寄信邮戳日期，收信邮戳日期，收信人，寄信人这些元素。

日记

日记指的是名人按照日期记载事物或个人感受的文字记录，包括日记本、记事本、备忘录、日历本、帐册等。每册日记，每本记事本和记帐本都单独作为一个著录单位。日记总册中包含若干本单册日记，所以在著录时，需要对单册日记著录的同时也需对日记总册进行著录，并说明两者之间的关系。

照片

照片是指以摄影形式记录名人及其家属，友朋生活、工作的文献载体。每张照片为一个著录单位。对于照片总集中包含若干照片的情况，原则是对每张照片著录的同时也对照片总集进行著录，并在著录时，在“关联”字段说明单张照片出自哪一本照片总集。而对于来自原照或者报刊的模拟照和翻拍照，杂志，著录时原照和模拟照，翻拍照分别著录，并说明模拟照和翻拍照的出处；还存在照片与底片的衍生关系，著录时对底片不予著录，只说明底片的类别和规格。由于照片的著录不是针对照片本身，而是针对照片中的人物，因此我们重点揭示的是照片中的人物，从而设置了人物和位置元素，由于人物和位置是一一对应的关系，

所以在著录时，人物和位置需要成组出现。

书画篆刻作品

书画篆刻作品是指以刀、笔等作为工具，利用墨或各类颜料，在纸、石或其它载体上创作的平面视觉艺术作品。每幅书画、每件作品都单独作为一个著录单位。书画篆刻作品是名人手稿馆资源中描述和著录难度最大的一类资源，由于书画篆刻的组成部分比较多，如一幅书画篆刻作品主要由书，画，题记和印章四部分组成，所以描述的元素也很多，不仅要描述书法的行款，书体，内容，书写工具和文字色泽，还要描述画的技法，画的材质，题记的作者，题记内容，题记的类别，印章所有者，印章的位置，印章的类型，印章的印文，印章的印文类型等。并且一副书画作品中可能有多个题记者，多个题记，也可能有多个印章，所以题记的所有内容都需要成组出现，印章所有者，印章类型，印章位置，印章印文，印章印文类型也需要成组出现。

签名本

签名本指的是带有名人签名、题词、印章或其它印记的纸质印刷本或出版物。对于正式的签名本，著录时以“种”为单位，即一位名人捐赠的一套书为一种，捐赠的单本书也可以为一种，对于一套书可以做一条单独的记录，但包括的书籍需做分析著录。由于签名本在名人手稿馆的价值主要在于其签名部分，并且上图馆藏中已经有其相应的书目记录，所以在描述时，我们主要描述签名部分，此处跟书画篆刻的题记部分比较类似，主要有题记，藏书票和印章三部分，并且由于存在一一对应关系，每个部分都各自成为一组。题记主要有题记作者，题记全文，题记位置，题记类别等元素，藏书票主要有藏书票尺寸，藏书票文字，藏书票图案描述三个元素，印章有印章所有者，印章类型，印章位置，印章印文，印章印文类型五个元素。

纸质资料

纸质资料指由名人捐赠的、不带有名人签名或印章的打印或复印文献，主要是指与名人相关的，或有关名人参与的重大历史事件的打印资料或剪报、期刊资料等。每一种纸质资料为一个著录单位。这些与名人有关的打印资料，剪报或者期刊资料一般都是由与名人关系比较密切的朋友或者名人的晚辈收集、捐赠，因此在元数据方案中我们增加了“收集者”这一元素。

音像资料

音像资料指以磁光等介质记录的有关名人或名人捐赠的视频、音频资料。鉴于音像资料比较特殊，比如一个采访内容可以分几个光盘，一个采访内容可以有光盘，磁带，录像等多种载体形式，所以音像资料的著录级别不以音像资料的存储介质来划分，而是以一个完整的内容为一个著录级别，不同的存储介质可在复本管理中进行说明。一个长达半小时的完整的音像资料有可能只有当中的几分钟是最关键的，因此我们在描述时，设立了关键片断开始码，关键片断结束码，关键片断描述等元素对关键的内容进行描述。一部音像资料的完成所涉及的人物也是较多的，例如有采访者，口述者，解说者，制作者等，因此对这些人物的说明非常重要，为此我们设立了“人物”和“出现时码”元素，并且这两个元素是一一对应的关系，需要成组出现。

笔记本

笔记指的是名人在听课、听报告、读书等情况下所做的摘抄、随想等文字记录，没有确切时间或者不是以日期作为记录顺序的。每册笔记都单独作为一个著录单位。笔记资源类型中也存在与日记类似的总册的情况，所以有包含关系，可以在关联字段中予以说明。笔记也是名人手迹的一种体现，因此在描述笔记时，我们也使用了“书写工具”和“文字色泽”元素。

实物

实物指由名人创建或使用过的，或其它与名人有关的物品，除已归入上述各类的之外，比如印章、证章等。每件实物单独作为一个著录单位。实物是比较特殊的一类资源，对它的描述就有点类似博物馆对其馆藏的描述，由于我们不是专业的博物馆，所以只选择了一些与名人很相关的元素来对其描述。由于实物可能是名人自己制作，也可能是名人出于爱好收集，也可能是一个名人（或团体）赠与另外一个名人，因此我们除了设立“创建者”元素外，还设立了“收集者”、“颁发者”、“颁发对象”等元素，并同时实物本身的“材质”和“色泽”进行说明。

证件指名人用于各种场合、代表其身份的文件，在范围上包括了出入证、医疗证、工作证、代表证、选民证、挂号卡、小名片等。每个证件都单独作为一个著录单位。

证书指由机关、学校、社团等颁发给名人以表彰其成就或授予其荣誉的文件。每张证书单独作为一个著录单位。

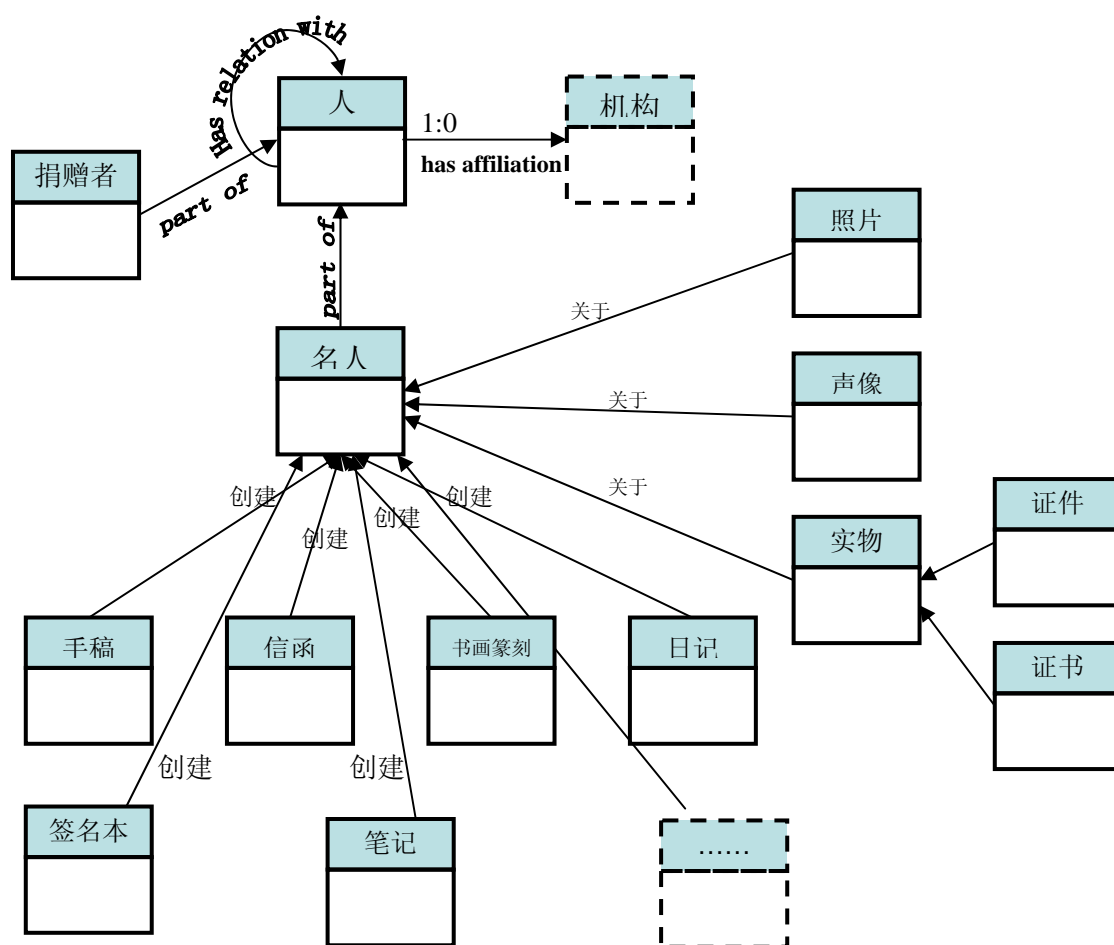


图 2：名人手稿馆资源实体关系图

元数据方案

方案的性质

从整体上看，元数据对于信息系统来说，具有描述、定位、发现、证明、评估，检索等作用（FRBR¹⁴定义了元数据最主要的四个功能：查找 Find、标识 Identify、选择 Select 和获取 Obtain）。元数据方案应该能够全面描述信息系统中与信息资源相关的各类特征和属性，

¹⁴ 参见：<http://www.ifla.org/VII/s13/frbr/frbr.pdf>

包括静态特征、状态特征、生命周期特征、管理特征以及系统在实现功能过程中所要求的其他语义信息等等，而不仅仅是描述性信息，而实际上人们研究和使用的较多的大都是描述性元数据，对于其它类型的元数据¹⁵，例如保存型元数据可以参考一些现成的参考模型或国际标准（例如 OAIS），或者一些大馆的元数据方案（例如美国国会图书馆的“美国记忆”项目的方案）来制定；管理型元数据和技术性元数据，大多交给系统分析和设计人员提供解决方案，这些数据元素大都可以在系统运行过程中自动生成，或在进行数字化、数据转换、或数据装载的过程中进行机器辅助人工添加。因此，名人手稿馆的元数据方案主要是一个描述性元数据的方案。

方案的组成

制定具体的元数据方案不同于元数据标准规范的研究，后者只需要少数一些推荐文件，甚至只有一组元数据元素集就可以了，而元数据方案需要一整套指导实施的文档，可以按照前文的流程确立文档，也可以根据实际需要制定。本方案主要包括以下 6 个文档，然而本节主要介绍方案中最重要的元素集方案、置标方案和著录规则，其中元素集方案参考 ISO11179 的要求名称，选取了 14 个属性进行定义。这 14 个属性是：中文标识、英文标识、命名域、版本、注册机构、语言、应用规则、定义、注释、数据类型、最大出现次数、元素修饰词、编码体系修饰词和最佳操作。

表 1：名人手稿馆元数据方案包括的文档列表

| | |
|------------|--|
| 资源分析文档： | 定义资源类型，确定资源类型的内涵和外延，确定著录级别和著录单位，属性提取，确定著录对象之间以及属性之间的关系，属性的检索要求，核心属性的支持 |
| 元数据元素集及定义： | 按照规范格式定义属性元素，定义限定和扩展规则，定义子元素和编码体系，定义元素之间的关系 |
| 著录规则： | 元数据方案应用于具体资源类型著录时的细节描述 |
| 置标方案： | 采用 RDF/XML 置标的规则，命名域规定，标签定义，编码规则 |
| 规范控制流程： | 人名规范档的建立、维护、更新、应用流程和规则 |
| 元数据著录系统需求： | 开发具体的著录系统所要求的通用和专用需求规范 |

属性元素集——核心集及基本扩展

如前文所述，本方案的核心集为以 DC-Lib 为基础的《上海图书馆元数据方案》，选取其中 18 个元素中的 17 个（不包括其中的“用户对象(Audience)”元素）以及多种资源类型共有的元素作为核心集，但核心集并不是必备集。整个方案采用混合型的元数据概要（Metadata Profile）形式，包括核心集和扩展集两个部分，复用了 CDWA¹⁶的 4 个元素（材料与技术、制法、相关文本信息、题记或标记）和 REACH¹⁷的 1 个元素（出处），并且针对名人手稿馆资源的特殊性，扩展了 5 个个别元素（获得方式，捐赠日期，捐赠者，书写人，载体形态附注）。方案允许每个元素对于不同的资源类型有不同的“显示名”和不同的元素修饰词与编码体系修饰词，限定方式原则上限制在两层。由于名人手稿馆所要求的资源类型的开放性，我们考虑其属性元素集为各已知资源类型所需元素的并集（即目前的核心集和扩

¹⁵ 对于元数据的类型并无统一公认的划分，IFLA 在一份元数据应用指南文件（*Guidance on the Structure, Content, and Application of Metadata Records for Digital Resources and Collections* a draft for review, 参见 <http://www.ifla.org/VII/s13/guide/metaguide03.htm>）中将元数据分为管理型、描述性、分析型（Analytical）、版权管理和技术性五种，传统上可以分为管理型、描述性、保存性、技术型、使用型等五类。

¹⁶ Categories for the Description of Works of Art(CDWA). URL: http://www.getty.edu/research/conducting_research/standards/cdwa/(检索日期：2004-2-1)

¹⁷ RLG REACH Element Set for Shared Description of Museum Objects . URL: <http://www.rlg.org/reach.elements.html>(检索日期：2004-2-1)

展集), 新增的资源类型原则上只能从已有的属性元素集合中选取元素, 否则增加或修改属性元素都会给已经实现的应用系统构成挑战。这样元数据方案才能保持必要的稳定性。

1. 核心集

| 元素名称 | 修饰词名称 | XML/RDF 标签 |
|--------|-------|---|
| 标识符 | | dc: identifier |
| | 索取号 | callNo |
| | 排架号 | shelvingControlNumber |
| 题名 | | dc:title |
| | 交替题名 | dcterms:alternative |
| 创建者 | | dc:creator |
| | 责任方式 | creator.role |
| 其他责任者 | | dc:contributor |
| | 责任方式 | contributor.role |
| 出版者 | | dc:publisher |
| 版本 | | dc:edition |
| 日期 | | dc:date |
| | 创建日期 | dcterms:created |
| | 修改日期 | dcterms:modified |
| 格式 | | dc:format |
| | 大小 | dcterms:extent |
| | 媒体 | dcterms:medium |
| 主题 | | dc:subject |
| 说明 | 内容提要 | dc:description |
| 来源 | | dc:source |
| 覆盖范围 | | dc:coverage |
| 语种 | | dc:language |
| 类型 | | dc:type |
| 权限 | | dc:rights |
| 关联 | | dc:relation |
| | 参照 | dcterms:reference |
| 馆藏位置 | | mods:location |
| 获取方式 | | marc:availability |
| 载体形态附注 | | marc:notesPertainingToPhysicalDescription |
| 工具 | | cdwa:implement |
| 颜色 | | cdwa:color |
| 创作地 | | reach:placeoforigin/discovery |
| 个人名称 | | person |
| 团体名称 | | corporate |
| 重要事件主题 | | events |
| 地名主题 | | place |
| 个人捐赠者 | | personaldonator |
| 团体捐赠者 | | corporatedonator |

| | | |
|------|--|----------------------|
| 捐赠日期 | | donated |
| 装订形式 | | descriptionOfBinding |

2. 扩展集

表 2:

| 资源类型 | 著录单位 | 扩展元素 | XML/RDF 标签 | 是否必备 | 是否可重复 |
|--------|-----------------|---------|----------------------------------|------|-------|
| 创作手稿 | 手稿的每一个版次。 | 誊抄者 | Scribe | 供选择 | 可重复 |
| | | 学科名称主题 | topicOfSubject | 供选择 | 可重复 |
| 信函 | 单一信函 | 发送邮戳日期 | postmarkFrom | 供选择 | 可重复 |
| | | 接收邮戳日期 | postmarkTo | 供选择 | 可重复 |
| | | 信函完整性附注 | integralityAnnotation | 供选择 | 可重复 |
| | | 信纸风格特征 | styleOfLetter | 供选择 | 可重复 |
| | | 信封收信者 | acceptor | 供选择 | 可重复 |
| | | 信封寄信者 | sender | 供选择 | 可重复 |
| | | 寄信人地址栏 | addressOfSender | 供选择 | 可重复 |
| | | 寄信地邮编 | postalcodeOfSending | 供选择 | 可重复 |
| | | 收信人地址栏 | addressOfAcceptor | 供选择 | 可重复 |
| | | 收信地邮编 | postalcodeOfAccepting | 供选择 | 可重复 |
| 日记 | 每册日记, 每本记事本和记帐本 | 丛编题名 | dc:isPartOf | 有则必备 | 不可重复 |
| 照片 | 每张照片 | 拍摄质量 | quality | 有则必备 | 不可重复 |
| | | 底片类别 | typeOfFilm | 有则必备 | 可重复 |
| | | 底片规格 | sizeOfFilm | 有则必备 | 可重复 |
| | | 人物 | person | 必备 | 可重复 |
| | | 位置 | location | 必备 | 可重复 |
| | | 照片总集名 | dc:isPartOf | 有则必备 | 不可重复 |
| 书画篆刻作品 | 每幅书画、每件作品 | 签名者 | signatory | 有则必备 | 可重复 |
| | | 印章所有者 | ownerOfSeal | 有则必备 | 可重复 |
| | | 行款 | NumberOfRows | 有则必备 | 不可重复 |
| | | 技法 | facture | 有则必备 | 可重复 |
| | | 题记位置 | cdwa:inscriptions/marks.location | 有则必备 | 可重复 |
| | | 印章位置 | locationOfSeal | 供选择 | 可重复 |
| | | 题记类别 | cdwa:inscriptions/marks.type | 供选择 | 可重复 |

| | | | | | |
|------|------------|---------|--|------|-------------|
| | | 印文类型 | InscriptionsType OfSeal | 供选择 | 可重复 |
| | | 印章类型 | | 供选择 | 可重复 |
| | | 书体 | cdwa:inscriptions/marks.typeface | 供选择 | 可重复 |
| | | 下款内容 | subscript | 供选择 | 可重复 |
| | | 印章印文 | inscriptionsOfSeal | 供选择 | 可重复 |
| | | 上款内容 | superscript | 供选择 | 可重复 |
| | | 题记全文 | cdwa:inscriptions/marks.transcriptionOrDescription | 供选择 | 可重复 |
| | | 材质 | cdwa:materialsAndTechniques.materials | 供选择 | 可重复 |
| 签名本 | 每一种签名本和捐赠本 | 签名者 | signatory | 有则必备 | 可重复 |
| | | 印章所有者 | ownerOfSeal | 有则必备 | 可重复 |
| | | 藏书票文字 | inscriptionsOfStamp | 有则必备 | 不可重复 |
| | | 藏书票图案描述 | descriptionOfStamp | 有则必备 | 可重复 |
| | | 藏书票尺寸 | sizeOfStamp | | |
| | | 题记位置 | cdwa:inscriptions/marks.location | 有则必备 | 可重复 |
| | | 印章位置 | locationOfSeal | 供选择 | 可重复 |
| | | 印文类型 | InscriptionsType OfSeal | 供选择 | 可重复 |
| | | 印章类型 | | 供选择 | 可重复 |
| | | 印章印文 | inscriptionsOfSeal | 供选择 | 可重复 |
| | | 题记全文 | cdwa:inscriptions/marks.transcriptionOrDescription | 供选择 | 可重复 |
| | | | | 丛编题名 | dc:isPartOf |
| 纸质资料 | 每种纸质资料 | 收集者 | collector | 有则必备 | 可重复 |
| 音像资料 | 一个完整的内容 | 开始时码 | timeOfStart | 有则必备 | 可重复 |
| | | 结束时码 | timeOfEnd | 有则必备 | 可重复 |
| | | 出现时码 | timeOfApperance | 有则必备 | 可重复 |
| | | 关键片断描述 | descriptionOfKe | 供选择 | 可重复 |

| | | | | | |
|----|------|---------|---------------------------------------|------|-----|
| | | | ysegment | | |
| | | 关键片断开始码 | startOfKeysegment | 供选择 | 可重复 |
| | | 关键片断结束码 | endOfKeysegment | 供选择 | 可重复 |
| | | 声音描述 | descriptionOfVoice | 供选择 | 可重复 |
| 笔记 | 每册笔记 | | | | |
| 证件 | 每本证件 | | | | |
| 证书 | 每张证书 | | | | |
| 实物 | 每件实物 | 收集者 | collector | 有则必备 | 可重复 |
| | | 材质 | cdwa:materialsAndTechniques.materials | 有则必备 | 可重复 |

置标方案

元数据的置标本质上是元数据方案的形式化,主要是为了计算机对元数据信息进行存储和处理而用的,比如 2709 格式就是 MARC 属性字段的一种表达(即形式化)。与传统的机读目录应用系统所不同的是,目前以互联网技术为代表的信息处理技术非常强调开放性、可扩展性和互操作性,不像传统的系统只需考虑系统内部的问题,实现独立的功能即可。从长远看这将带来很大的好处,然而目前由于技术尚未成熟,开放环境下的一整套技术架构尚未完全建立起来,应用所需的许多标准规范也正在研究探索之中,因此实现目前的应用在采用的技术方面显得非常复杂,技术实现的每一步都是独立的,都有许多方案/标准/规范可供选择,很多时候鱼和熊掌无法兼得。

属性元素集定下来之后也可以直接采用关系数据库系统来实现具体的查询、管理功能,当然作为数字图书馆应用来说最好中间加一层基于 XML 的内容管理层,以便于将来独立于系统的长期保存、与其它系统的互操作等,这就有必要对元数据进行基于 XML 的置标。元数据方案定义的许多属性,并不是简单的采用某种置标语言就能被机器理解,不同的置标方案的能力是不同的,例如单纯的 XML 不支持数据类型的定义,而必须用到 XML Schema,如果将来需要进行语义互操作,最好采用 RDF 置标等等。就目前而言,还没有哪种置标方案能够将元数据方案中所定义的所有约束表达出来,特别是语义方面的约束。这将有赖于将来基于本体的置标语言的标准化,结合多种置标标准共同实现丰富而灵活的功能。

名人手稿馆元数据的置标采用 XML/RDF 形式,然而由于元数据著录系统支持任何以 DTD 或 XML Schema 方式载入元数据方案及相应的约束¹⁸,不同元数据格式之间的转换(不论是置标方式的转换还是元素之间的映射)变得相对容易,系统还支持 XML/RDF 格式与 CNMARC (ISO2709) 格式之间的转换,因此具体采用何种置标格式甚至是无所谓了。当然本系统物理上存有一套基于本元数据方案的 XML/RDF 置标元数据,同时研制了大量基于标准的置标方案以及其相互转换的映射规范,这些“元元数据”大都以 XML Schema 或 DTD 文件的形式存放,必要时调入著录系统。当然元数据库的存储和管理还是采用商用的关系数据库系统。具体的元数据著录与管理系统的功能详见下文具体介绍。

¹⁸ 这种技术方案借鉴了台湾的 Metalogy 系统的设计思想。

著录规则

由于名人手稿馆的资源种类繁多，相互差异性较大，具体的标引人员对不同资源类型的熟悉程度一般都是不同的，因此我们在名人手稿馆元数据方案的指导下，对每种资源都编制了详细的著录规则，对每种资源的个性之处和揭示要求进行了详细充分的说明。

规范档

利用规范档进行规范控制在情报检索中属于事先控制行为，如果进行得彻底，能够按照概念语义组织信息资源，通过概念的联系建立资源之间的关系，一方面大幅度提高信息系统的查准率和查全率，另一方面启发和引导用户的查询行为。然而“先组式”情报检索系统所花费的代价也是非常大的。首先建立一个准确的规范档需要领域专家们呕心沥血的工作，其次利用规范档进行标引工作也是非常费时耗力的工作，即使经过大量的培训，其一致性也难以把握。在名人手稿馆著录系统中，规范档的主要作用在于对元数据库进行规范控制，以保证元数据库的一致性，同时在未来的系统中起到聚集、规范、扩展、关联的作用。由于名人手稿馆的目标就是收集关于“名人”的各种资源，因此名人手稿馆著录系统目前只实现了关于人名的规范档控制，包括规范名，出处，国籍，民族，生年，卒年，籍贯，出生地，性别，小传，专长，著作，职衔，关系人等字段。在以后的工作中，根据实际情况将逐步实现对团体名称，地理名称，主题，作品，事件等的规范控制。这将是一个漫长而细致的过程。

编码体系

编码体系是对元数据取值的约束和规范，是修饰词的一种，用来说明元素所表达属性的值，包括控制词表及正式的符号体系或解析规则两类。用某一编码体系表示值意味着这个值来自于某一控制词表（包括控制词表的语义符号，如分类体系或主题词表中的术语），或者其格式符合某种正式编码标准的字符串（如“2000-01-01”作为一个日期的标准表达）。

编码体系的采用有助于提高数据的质量、通用性和互操作性。从一个更广泛的意义上看前文所述的规范档也是一种编码体系，例如人名规范档就在人名取值时作为规范名使用。在名人手稿馆的各类资源中我们经过研究和讨论，自定义了大量的编码体系，如下表所示。其中有一些复用了相应的标准，但更多的则因为没有现成的标准，我们不得不从实用的角度出发进行定义，没有作严格的规范，可能在今后的应用中还需要修改。

| Scheme 中文名称 | Scheme 英文名称 | Scheme |
|-------------|----------------------|----------------------------------|
| 获取方式 | Availability | 捐赠/翻制/自拍/移交/购买/竞拍/原存/征集奖励/不详 |
| 权限 | Rights | 公开/绝密/机密/秘密/内部 |
| 装订方式 | descriptionOfBinding | 精装/平装/线装/未装订 |
| 创作手稿文献类型 | typeOfScript | 原稿/校改稿/复写稿/清样校勘/手写复印件/打印稿/复制件/抄稿 |
| 书写工具 | Implement | 毛笔/钢笔/圆珠笔/铅笔/水笔 /油画笔/硬笔 |
| 文字色泽 | color | 黑色/蓝色/红色/蓝黑色/绿色/紫色 |
| 书写工具 | Implement | 毛笔/钢笔/圆珠笔/铅笔/水笔 /油画笔/硬笔 |
| 日记文献类型 | typeOfDiary | 日记/记事本/帐册/备忘录/日历 |

| | | |
|--------|--------------------------------------|---|
| 书体 | styleOfCalligraphy | 隶/篆/楷/行/草/鸟虫 |
| 材质 | materials | 绢/宣纸/绵纸/卡纸/油光纸/彩画纸/书写纸/其他纸/竹/布 |
| 装裱形式 | Mount | 立轴/手卷/屏/横批/扇面/册页/对联/镜片/斗方/无装裱 |
| 载体形态附注 | notesPertainingToPhysicalDescription | 册面/函套/木盒/纸盒/布套/夹板 |
| 题记类别 | typeOfInscriptions | 一般题记/题识/题端/题跋/外题签/内题签 |
| 印章类型 | typeOfSeal | 名章/斋馆印/别号印 |
| 印章印文类型 | typeOfInscriptions | 朱文/白文/朱白文 |
| 书画文献类型 | typeOfDrawing | 按是否原件分:原件/复制件 |
| | | 按体裁分:书法/绘画/设计稿/篆刻作品/书画综合作品 书法 按书体分为:隶/篆/楷/行/草/鸟虫 |
| | | 绘画按技法分为:国画/版画/油画/素描速写/水粉水彩画/其他 绘画按用途分:宣传画/漫画/年画/连环画/其他 |
| 信函文献类型 | typeOfLetter | 按内容分:友朋/家书/公事/公函/情书 |
| | | 按载体分:一般信函/贺卡/明信片/请柬 |
| | | 按形式分:手写/复写/打印/复印 |
| 音像文献类型 | typeOfAV | 按内容分:口述历史/名人传记/文献展览/名人活动 |
| | | 按形式分:视频/音频 |
| | | 按是否原件分:原件/复制件 |
| 拍摄质量 | quality | 优/一般/差 |
| 照片颜色 | PictureColor | 黑白/彩色 |
| 底片类别 | typeOfFilm | 正片/负片/反转片/数码 |
| 底片规格 | sizeOfFilm | 120/135/127/APS |
| 照片文献类型 | typeOfPicture | 按人物数量分:单人照/双人照/多人照 按人物关系分:友朋照/夫妻照/家庭成员照/师生照/同 按照片内容分:学照/政要照/大型团体照/同事照 |

| | | | |
|----------|---------------------|--|----------------------------------|
| | | 本馆采访照/社会活动照/学术活动照/生活照/旅游观光照/出访照/采访照/工作照/会议照/展览会照/学位照/证件照/演员造型照/艺术照/生日照/结婚照/悼念照/历史事件照/舞台设计照/剧照/学术资料照/建筑照/模拟照 按是否原件分:原照/翻拍照 | |
| 证件文献类型 | typeOfPaper | 出入证/医疗证/工作证/代表证/选民证/挂号卡/名片/其他证件 | |
| 证书文献类型 | typeOfCertificate | 获奖证书/聘书/委任书/毕业肄业证书/其他证书 | |
| 责任方式 | role | 著/编/译/书/画/篆刻 | |
| 实物文献类型 | typeOfPracticality | 第一层分类: 奖章/其他实物 | 第二层分类: 徽章按 照类别: 奖章/纪念章 /徽章 |
| 签名本文献类型 | typeOfSignatureFile | 签名本/捐赠本 | |
| 技法 | skill | 国画/版画/油画/素描速写/水粉水彩画/其他 | |
| 异名类别 | typeOfNickname | 原名/笔名/字/号/曾用名/初名/又名/室名/昵称/绰号/简称/谥号/化名/外文名 | |
| 与个人标目之关系 | relationOfPerson | 友/父/母/兄/弟/姐/妹/子/女/夫/妻/媳/师/一般亲戚/ | |
| 国别 | nation | 国别 scheme 表 | |
| 地区 | region | 地区 scheme 表 | |
| 语种 | language | 汉语,英语,日语,德语,法语,俄语,朝语 | |
| 民族 | race | 汉,蒙,满,回,藏 | |
| ISO8601 | ISO8601 | 类似 2001-08-07 或 20010807 | |
| 国际标准书号 | ISBN | 国际标准书号说明 | |

名人手稿馆元数据著录系统

名人手稿馆元数据著录系统项目于 2004 年 5 月开始启动, 一边研究一边开发, 数据加工也同期开展, 而确定元数据方案和元数据加工管理的功能需求是该系统开发的核心工作。基于此系统的重点在对馆藏内容的揭示, 元数据方案的广度和深度方面有较高要求, 因此元数据元素及其修饰词和编码体系的设定和配置是方案的难点, 项目的设计、开发和使用者如果不能就这个方面尽快达成一致, 必然会影响到系统的开发进度。事实证明在整个开发过程中研究开发与应用实施是互相牵制而又互相促进的。

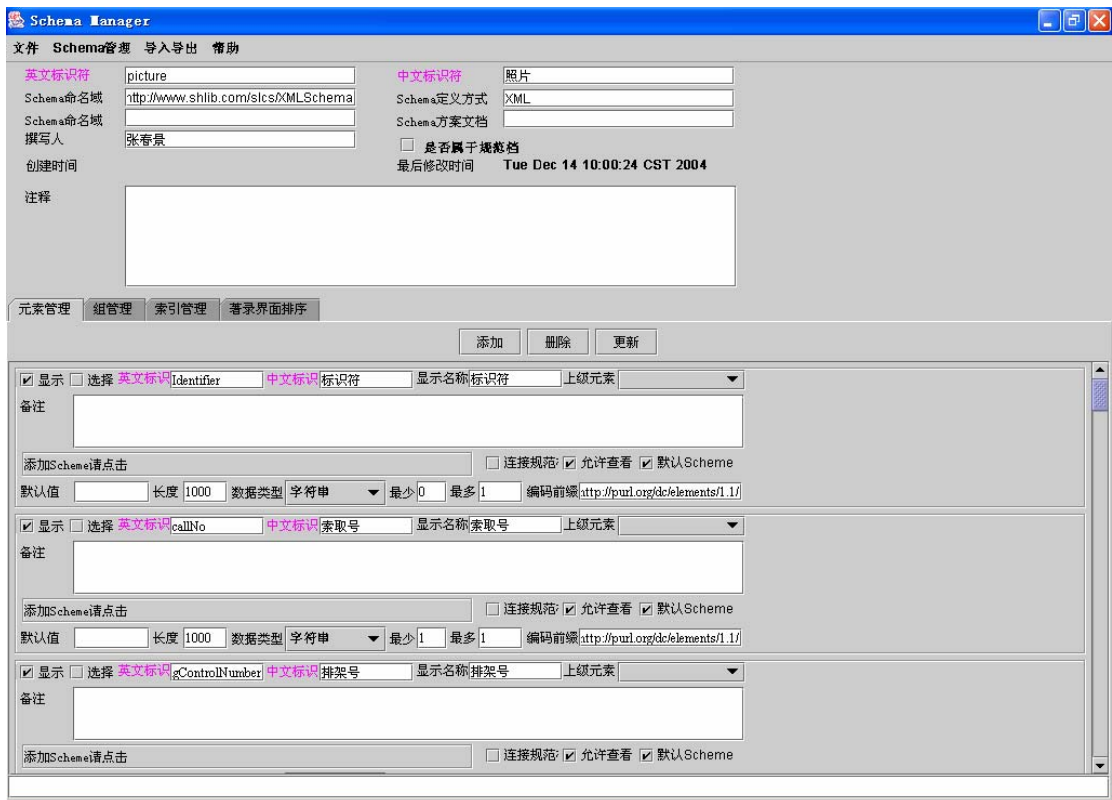
名人手稿馆著录系统的整个系统架构分著录流程和管理流程两大功能。著录系统的著录过程支持 TCP/IP 及 HTTP 协议的因特网 WEB 环境, 即著录为 WEB 方式, 属浏览器/服务器结构 (B/S); 系统后台采用 MS SQL 或功能相当的关系型数据库为管理系统, 前端采用 WEB 服务器加 JSP 方式的应用模式, 同时辅之以采用 JAVA 等开发的系统工具程序, 构

成一个完整的支持多种元数据应用规范的联机元数据著录系统。著录系统的客户端为标准的浏览器，主要为微软的 IE 浏览器。但是由于多种元数据应用规范维护上的复杂性，对元数据维护，采用专用的客户端软件形式。

由于此项目的定位是不仅要满足对名人手稿馆现有十三种资源的著录，而且可以满足将来上海图书馆所有资源的著录。基于以上目标，名人手稿馆著录系统中有一个重要的模块：元数据模式(Metadata Schema, 指经过 XML 或 RDF 等形式化语言编码的元数据方案)的管理工具，此工具可以支持任意多种元数据方案，支持对任意多种资源的著录。因此使用此工具不仅可以任意增加元素和修饰词，任意配置不同元素组成的多个组，任意添加索引字段，还能支持元素和组之间的任意排序。而针对每个元素可以规定其上级元素，可以添加元素的默认值，可以规定元素属于字符串还是长文本，可以规定元素是否必备，规定元素在某个数量区域内重复添加，并且可以指定该元素是否连接人名规范档，是否给读者浏览的权限，等等。

利用此工具还可以实现元数据模式的导入导出，通过元数据模式导入功能把一个 XML/RDF Schema 导入到能够进行 Schema 转换的功能模块中，以便与元数据著录、存储进行映射；对导入的元数据模式进行系统设定，在导入元数据模式的基础上设定其他参数，并转换成系统内部元数据存储、元数据著录等操作相关的数据结构及其他内容；还可以进行系统中元数据模式的格式转换，并且可选择同时导出与 XML schema 相关联的数据。

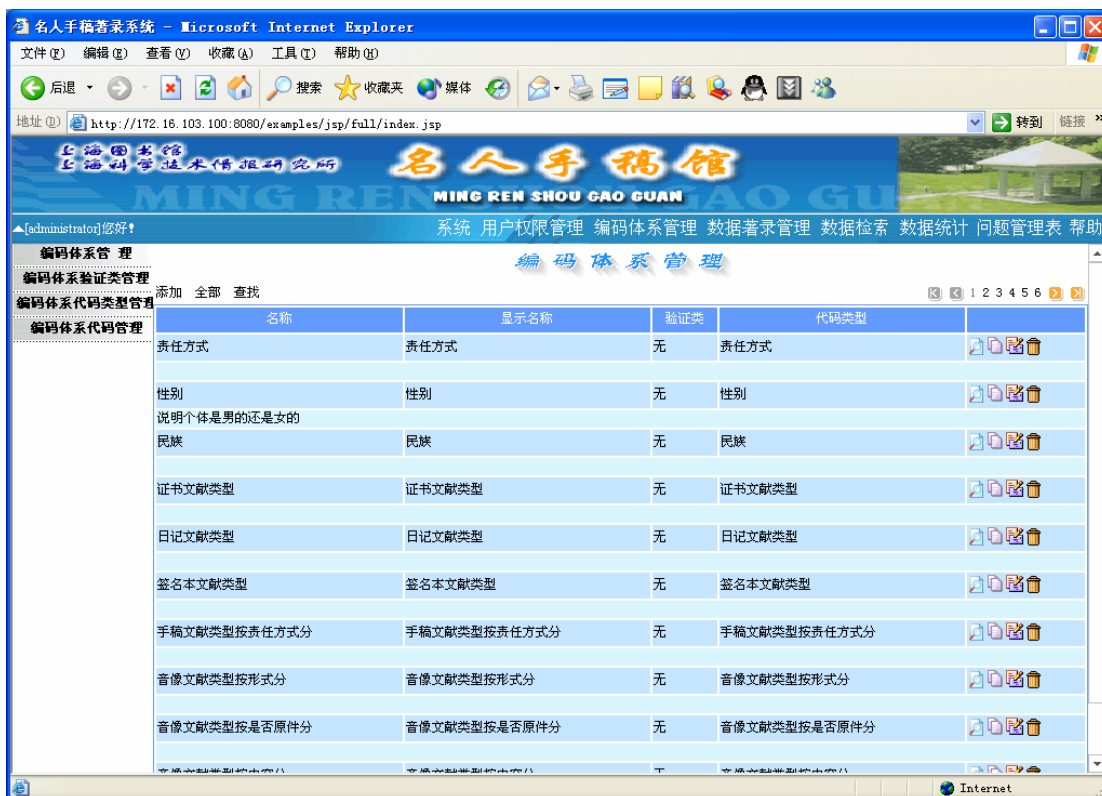
图三即为名人手稿馆著录系统的 schema 管理工具的配置界面。



图三：schema 管理工具配置界面

对 schema 管理工具对某种资源的元数据方案进行配置以后，还需要对元数据方案中所涉及到的编码体系进行配置。在本项目中，为了便于管理，将值列表和编码体系用统一的模式进行管理。由于系统支持任意多个 schema 的配置，从而导致 scheme 和值列表必定是不断增加的，并且对应于一个元素，可能有多种编码体系和值列表对其进行修饰，每一种编码体系和值列表都有自己的数据验证方式，因此名人手稿馆著录系统的编码体系管理可以实现对 scheme 和值列表的手工添加、修改、删除、查找等功能。

下图即为编码体系管理的系统界面



图四：编码体系管理界面

在 schema 管理工具和编码体系管理中对某种资源的元数据方案进行配置以后，就可以利用著录界面进行著录了，在系统的整个著录过程中，除了提供主要进行著录的著录界面功能以外，还提供了“编目树”功能，如果要全屏显示著录界面，可以将该编目树关闭。

在著录界面中，用红色星号显示了必须著录的必备字段，否则无法保存，并且还提供任何字段的无限重复添加，字段的删除，字段的成组添加，成组删除功能，对于著录人名的字段，可以直接将其著录内容链接到人名规范档。著录工作结束后，系统提供保存和临时保存功能，点击临时保存，则该条记录存入临时库，经校对人员校对后正式保存入库。系统在此还提供工作单的载入功能和对整条记录的密级管理的功能。系统中的编目树功能，可以提供点击编目树的任何一个字段，系统就将此字段添加到著录界面中的相应的元素下方，而且该编目树还可以反应该资源的元数据方案元素和修饰词，元素和编码体系修饰词之间的层级关系。



图五：著录界面

名人手稿馆著录系统作为一个专门管理各种名人资源而开发的系统，具有其独特性，名人和名人之间，名人和资源之间，具有复杂的关系。因此需要提供一种在名人与资源间的浏览功能，即名人导航功能。名人导航界面主要分三大块，分别是：“当前人物”，“关系人”和“与当前人物相关的资源”。系统在此提供两大功能，第一：系统首先在界面的导航栏显示名人手稿馆中的重要名人，点击任何一个名人姓名后，该名人成为“当前人物”，并显示和其有关系的其他名人及和相关资源。当前人物除了能浏览界面提供的规范名，职衔，小传，点击当前人物的规范名可以对整条规范档数据进行浏览，并提供当前人物的大头照。也可以点击与该名人有关系的其他名人，并以被选择的名人作为一个基准点进行下一步的信息显示；点击相关资源的题名时，则显示资源中的详细内容。第二：界面提供用于人名检索的检索框，进行检索后，页面显示该名人及其与该名人有关的其他名人和资源的关系图。此后的操作与第一项功能相同。这样可以层层叠套，将名人手稿馆的整个资源囊括和串联起来，形成名人手稿馆信息系统的一个特点。

存在的问题和未来发展

名人手稿馆元数据方案的实现有许多独特的难点，一些已经得到解决，一些还没有很好的方案。例如同样的元素对于不同的资源具有不同的名称（如：书信的责任者为“下款名”），而系统内部必须默认为一致，我们的解决方案是为每一个元素设立三个名称，即中文标识，英文标识和显示名称，中文标识保证了元数据方案的统一，英文标识有利于与国外元数据方案的互操作，显示名称又有利于专门领域工作人员的标引和用户对资源的利用。另一个相关的比较棘手的问题是不同的资源类型对同一个元素具有不同的元素修饰词与编码体系修饰词，目前我们只能同时支持此元素下所有的元素修饰词和编码体系，否则只能将元数据方案按照资源类型拆开，这与我们的设计思想是相悖的。

目前名人手稿馆的元数据方案主要考虑资源系统内部的互操作（即跨库检索），主要依靠核心元素来实现，而核心元素的互操作主要通过其支持和复用 DC 等元数据标准来实现，在名人手稿馆著录系统的各个资源类型子库的基础上对各子库中的名人信息进行勾连，即名人手稿馆著录系统中已经实现的“名人导航”功能，此外还可以与上图的馆藏书目数据库进行链接，更大范围之外可以与档案馆、陈列馆、文化馆、博物馆、纪念馆和其他图书馆的相关书目进行链接，例如国家图书馆的名人手稿馆，收藏有 300 多位名人的约 5000 多种手稿，书目记录采用 MARC 格式。目前名人手稿馆著录系统已经实现了元数据方案与 MARC 的映射，因此通过 Z39.50 网关与上海图书馆、国家图书馆乃至全国其它图书馆进行链接没有任何问题，将来通过开发 OAI 接口与互联网上其它的手稿资源库进行链接也是有可能的。但要实现丰富、灵活的外部勾连还必须考虑更多的互操作协议和技术标准，关注这些标准规范的发展动态。

完整的名人手稿数字图书馆系统是一个包含数字化加工制作、元数据生成加工处理、元数据存储转换检索发布管理、数字资源对象管理等各个部分，如下图所示。最近上海图书馆初步完成了“数字资源服务平台”项目，整合了上海图书馆内大部分数字资源和数据库系统，能够为用户提供一个统一的资源整合平台，实现了统一用户登录。使用此系统，用户只需要一次登录便可以随便查询里面整合的数据库，逻辑上实现了统一检索的功能。名人手稿数字图书馆系统也将作为这个数字图书馆的组成部分，集成到这个数字图书馆门户中去。名人手稿数字图书馆系统在元数据加工处理方面为上图数字图书馆提供了一个样板，即核心元数据采用上图资源集合元数据方案，但每种资源都保留自己相对独立的元数据方案。目前在解决多种元数据模式的管理问题时采用一张融合大表的方式，可以看成将多个元数据方案的 XML Schema 合并成一个大的 XML Schema，而没有采用多模式映射或多模式表层级管理的方式，主要考虑到现有软件工具和技术条件下实现起来较为简单。

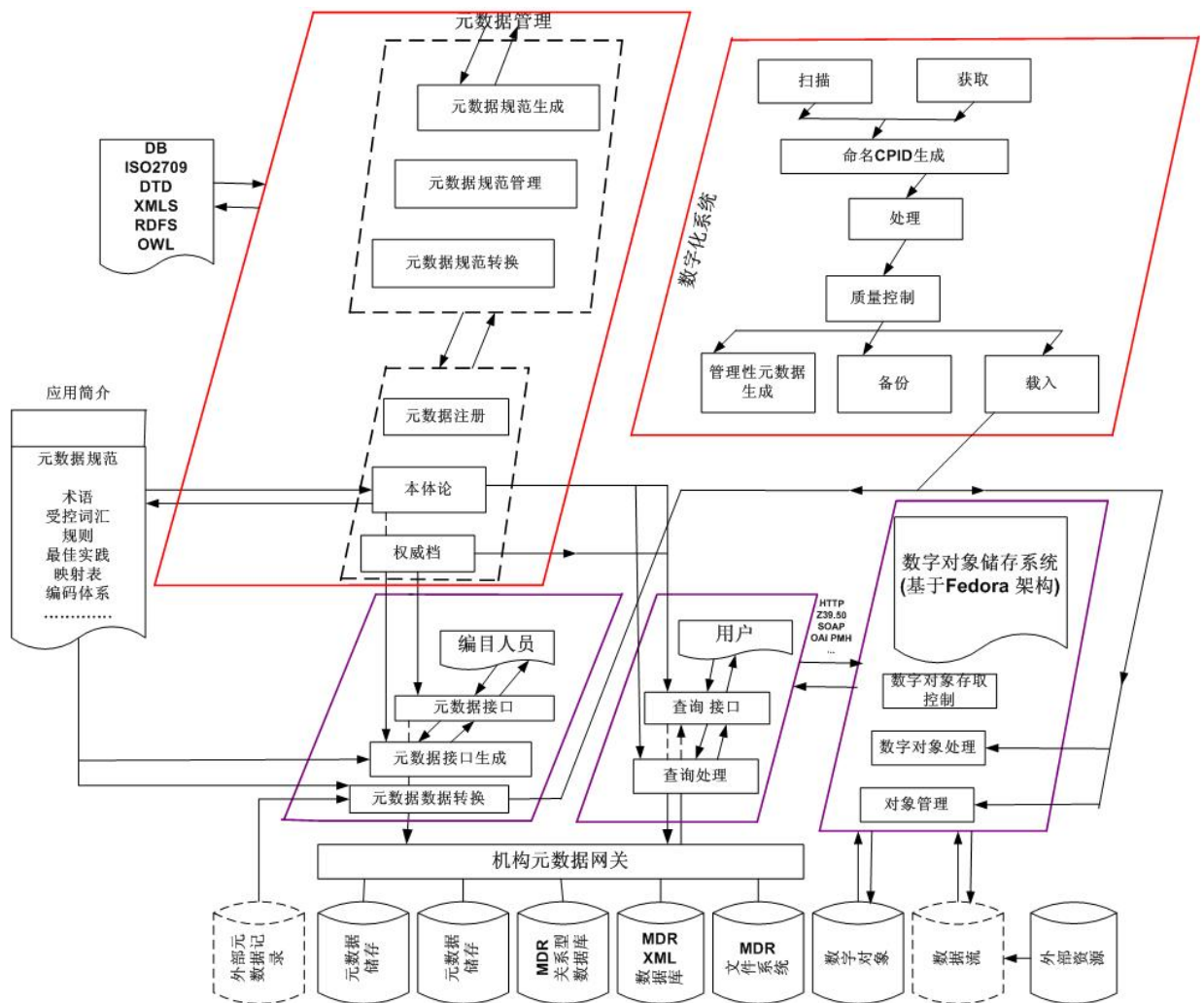


图 3：上海图书馆名人手稿数字图书馆系统示意图

名人手稿馆的资源是上海图书馆丰富馆藏的一部分，名人手稿馆元数据方案是上海图书馆元数据方案的一个实例和具体应用。上海图书馆元数据方案也还有许多不尽完善的地方，通过名人手稿馆的元数据实践能够加以补充和完善。但是如前文所述，名人手稿馆的元数据方案目前还缺少管理型元数据、保存元数据和技术性元数据等其他部分，这些元数据需要在名人手稿馆数字图书馆的实现过程中加以提炼和总结。并且上图的各类元数据方案从整体上还缺乏注册登记管理方法和手段，自己建立注册系统也涉及到长期维护和长效管理问题，这些问题都将在今后的工作中进行探索和解决。

参考文献：

1. “外滩源”保护开发全面启动. 王蔚. 文汇报, 2003-01-10
2. MALVINE:Manuscripts and Letters via Intergrated Networks in Europe . URL : <http://www.malvine.org/>
3. Reference Model for an Open Archival Information System (OAIS). Consultative Committee for Space Data Systems (CCSDS) . URL: <http://www.classic.ccsds.org/documents/pdf/CCSDS-650.0-B-1.pdf>(检索日期: 2004-2-1)
4. 中文元数据方案(征求意见稿). 国家图书馆. 内部资料, 2001年6月

5. Functional Requirements for Bibliographic Records (FRBR). IFLA Study Group on the Functional Requirements for Bibliographic Records. URL:
<http://www.ifla.org/VII/s13/frbr/frbr.pdf>(检索日期: 2004-2-1)
6. Dublin Core Abstract Model. Andy Powell. URL:
<http://dublincore.org/documents/abstract-model/>(检索日期: 2004-1-14)
7. RDF Vocabulary Description Language 1.0: RDF Schema. W3C Proposed Recommendation. URL: <http://www.w3.org/TR/rdf-schema/>(检索日期: 2004-2-1)
8. Library Application Profile. Rebecca Guenther. URL:
<http://dublincore.org/documents/library-application-profile/>(检索日期: 2004-1-14)
9. Guidance on the Structure, Content, and Application of Metadata Records for Digital Resources and Collections. URL: <http://www.ifla.org/VII/s13/guide/metaguide03.pdf>(检索日期: 2004-1-14)
10. Metadata Encoding and Transmission Standard (METS). URL:
<http://www.loc.gov/standards/mets/>(检索日期: 2004-1-14)
11. Metadata Object Description Schema (MODS). URL:
<http://www.loc.gov/standards/mods/>(检索日期: 2004-2-1)
12. Web Services Description Language (WSDL) 1.1. W3C Note. URL:
<http://www.w3.org/TR/wsdl>(检索日期: 2004-2-1)
13. 《专门数字对象描述元数据规范研制工作手册(修订稿)》. 专门数字对象描述元数据规范子项目组. 内部资料, 2003年7月
14. Categories for the Description of Works of Art(CDWA). URL:
http://www.getty.edu/research/conducting_research/standards/cdwa/(检索日期: 2004-2-1)
15. RLG REACH Element Set for Shared Description of Museum Objects . URL:
<http://www.rlg.org/reach.elements.html>(检索日期: 2004-2-1)
16. 张晓林. 元数据研究与应用. 北京: 北京图书馆出版社, 2002-05
17. 陈雪华, 陈昭珍, 陈光华. 数位图书馆 XML/Metadata 管理系统. 台北: 文华图书馆管理资讯公司, 民国 90[2001]