

● 刘家真

馆藏文献数字化工作模式的选择^{*}

摘 要 进行文献数字化有3种工作模式:本单位自主开展,业务外包,综合方式。3种方式的选择,原因不同,各有不同的特点和要求。参考文献2。

关键词 图书馆 文献数字化 工作模式 选择

分类号 G253

ABSTRACT In this paper, the author discusses three patterns for the digitization of library collections, i.e. the independent development within the library, outsourcing, and comprehensive pattern. We can use different patterns with different characteristics and requirements for different purposes. 2 refs.

KEY WORDS Library. Document digitization. Work pattern. Selection.

CLASS NUMBER G253

馆藏文献的数字化,要求将传统的文献通过模数转换、图像优化、压缩归档等数字化技术,加工成电子化的信息数据,并存入光盘或其他媒介中,或保存在镜像服务器大容量硬盘中,进行高质量的保存和管理,并提供网上高速的存取服务。在这一过程中,所需技术力量与专业人才大致有:(1)数字化工程管理者:设计项目目标与本单位数字化的预期值,确定人员与设备的费用,协调与调整相关部门间的工作。(2)计算机与技术人员:负责安装与维护计算机网络、设计存储系统等。(3)质量监督人员:提出与维护图像质量的控制标准与产品目标。(4)执行模数转换的技术人员:对原件或其替代品(如缩微片)进行数字采集和编辑。(5)后处理人员:待数字图像采集完毕,将图像处理到归档存储媒介上,并为它们的短期或长期使用作准备^[1]。

所需的主要设备有:(1)用来采集原件或其替代品的数字采集设备,如扫描仪、数码摄像机以及相关的配套设备。(2)供数字采集、后期处理和编辑的大容量图像处理计算机;为工作室内外配置网络的服务器。(3)合适的照明设备。(4)电压控制器:对扫描过程中灯光与电力引起的电压变化进行控制。(5)保持工作室室内温度与湿度恒定与洁净的设备,如空调、空气滤过设备、吸尘设备等。(6)色彩管理控制:用来控制色彩管理的工具。(7)后处理设备:包括磁带驱动器、CD刻录机等。(8)数据存储设备:保存数字图像的设备,如磁盘、磁带与近线网络存储器等。(9)传输

媒介:内网、外网等。

数字化工程一旦启动,以上人力、物力需求就会立即提到议事日程。数字化工程开始运行,还要投入大量后续资金。对于这样一项巨大的系统工程,必须确保其成本效益和工作效率。前者与数字产品潜在用户及用途有关,可通过合理的优先数字化方案加以保证,而后者取决于数字化的工作模式选择,是本单位自主开展、通过业务外包来执行还是采用综合方式来完成。其实,数字化的工作模式选择也与成本效益有关,但这是一个十分复杂与有争议的问题,不在本文讨论范围之内。

1 数字化工作模式

1.1 本单位自主开展

本单位自主开展数字化,包括由本单位职工利用本单位设备来进行文献数字化工作,并对图像进行处理、存储与归档管理。它至少需要在以下几个方面进行投资:(1)购买或租借运行数字化工程所需的一切设备。(2)雇佣技术人员进行数字化系统整合,设计保存系统以及解决网络与图像传输等问题。(3)职员培训。机构内的职员,包括具有计算机一般专业技能的人员,都需要进行培训,以掌握色彩管理、技术系统整合和产品质量控制等数字化工程的基本技能。

由本单位自主开展数字化工程,从长远来看是有益的。在数字化工程运行中,培养与锻炼了一批专

^{*} 本文系教育部人文社会科学基金项目(01JA870010)的成果。

业队伍,他们成为本单位新的资本,不仅可以服务于当前,也可服务于今后,还可以与其他机构共享技能与经验,为本单位今后的工作带来新的活力。

但我们也必须意识到,在某一个方面投资过高,花费太多的时间和资金必然会影响到该工程在其他方面的进展。例如,如果将有限的资金过多地投入购买设备与人才培养,就可能没有足够的资金来运行该工程等。必须在经济实力雄厚、人力资源充沛的单位才有条件考虑采用这种模式。

1.2 业务外包

业务外包是指一个单位通过签定合同把整个项目或项目的一部分包给专门从事这种工作的公司来完成。数字化工程业务外包通常由专业的系统集成企业提供专门的数字化工程解决方案并进行施工,完成采集数字图像并将它们处理到储存介质上。这些服务也包括一些相关的建议,如数字化和生产策略、颜色校正、数字元数据的管理等。采用业务外包的模式运行数字化工程,可为单位带来以下效益:(1)避免购买昂贵的设备。文献资源管理部门长期以来就存在资金不足问题。业务外包可使资金不充裕的部门用有限的钱解决数字化工程中必须解决的重要问题,这对于经费不多的小单位尤为重要。(2)提供了本单位欠缺的技能与经验,使数字化工程可能运行。(3)使本单位可以腾出更多精力从事数字化工程的前期准备工作与后期管理工作,以保证数字化工程的效益与质量。(4)使本单位在短期内完成数字化工程成为可能。专门化的集团设备齐全、技术人员完备,一旦合同签定立即就可以投入工作,缩短了自主开展数字化工程的前期准备工作所需要的时间。(5)通过合同可以严格控制与保证数字化的质量。只要合同中含有严格的质量要求,承包者必须履行合同。在后期管理中,只要本单位执行严格的质量检查与监督,就可以保证数字化工程高质量完成。

鉴于业务外包的以上优势,近些年来IT外包服务发展很快。IBM的郭士纳在2000年12月的一篇讲话中曾指出:“我并不怀疑,在今后长时间内,很多客户还将继续拥有服务器、中间件、应用软件、存储设备和管理所有IT人员。但是越来越多的企业会发现,自己拥有所有的设施是不经济的。越来越多的客户已经开始购买IT服务。服务提供商们已经为适应发展需要建立了巨型数据中心——巨大的‘服务器农场’,提供最先进的基础设施。我相信IT外包服务将会从根本上改变信息技术行业的市场运作模式。”

确实,在我国馆藏文献数字化的实践中,许多单位和部门深感靠自己的力量来进行数字化工程,周期长、耗资多、费人力、技术要求高,信息化进程受到影响,因而转向了业务外包。在北京地区,仅就我们所知的就有北京市文物局、文化部档案处、中共中央对外联络部档案处、外交部档案馆、教育部档案处等多个部门选择了数字化工程业务外包。

业务外包也有不少弊端,如:

(1)可能失去开展这方面工作的技术基础。长期外包,使本单位不得不依靠外界的服务,无法使本单位的职工在数字化专门技术方面得到锻炼与提高。长此缺乏相关的知识与技能,就会失去对承包商的控制与监督能力,如无法判断承包商在某项花费上是否合理、难以制定切实可行的合同、无法评价承包方的技术与产品质量等,合同费用也会因此不合理地增长。(2)如果在业务外包中,不保持直接参与而是简单地移交一切,重要的敏感资料、机密信息有可能丢失或失真,珍贵文献有可能损坏。即使合同中有相应的惩罚措施,但其损失是无法挽回的。

为保证业务外包顺利进行,减少各类风险,必须制定严格的、全方位考虑的合同以保护本单位利益,必须有专人直接参与业务外包的管理,以减少和控制风险的发生。

1.3 综合方式

综合方式是本单位与IT外包服务商以协作方式,整合内部力量与外部资源,共同开展数字化工作。它一般有以下几种形式:(1)由内部职员来数字化源文件,但是采用外界服务把数字图像处理到光盘上和进行磁盘备份等来达到归档和存取的目的,它缩减了与数字化相关的某些管理工作。如果机构内部人员有数字化方面的专长或某些文献不宜采用外界人员处理,可以采用这种方式。(2)外界服务人员同内部职员在同一地方共同完成数字化原件的工作。在共同的工作中,机构内的工作人员可获得相关的技能与经验,同时可监督数字化质量。(3)外部采购整个数字化工程,但是用内部资源进行后期的处理工作。这种方式适用于数字化的原件不是珍贵的原始材料,或是原件的胶片拷贝。这种协作方式可以使机构把精力主要集中在数字图像的存取和应用项目的处理中。

2 数字化工作模式的选择

数字化工作模式的选择必须考虑工程的性质、

机构实力以及被数字化的文献性质。

2.1 工程性质

数字化工程可以分为两大类:一类属于特定用途工程,另一类为多重用途工程。特定用途工程是为满足单一目标,为特指用户或特定用途服务,它仅限于提高原文献的利用率而不包含管理和保存目标。通常只是为了满足数字图像的短期使用或其他目的,如设立在网页上供用户临时浏览等,因而特定用途的数字化工程并不需要高质量的图像。文献数字化后并不能满足其他用途,除非对原件按要求再数字化。

多重用途的数字化工程,文献数字化后可达到最高质量要求,既可用于储存目的(如网站或文件管理系统),也能满足多种需求(如长期保存数字图像或按需进行格式转换等)和普遍应用。对于这类数字化工程,文献只需数字化一次,便可多目的使用,可避免原件多次数字化而带来的损失。该类数字化工程要求高质量的图像采集,在颜色、校正、锐化、分割等方面对图像作最小的编辑,当需要某个数字图像用于特定用途时,再对它进行最优化来满足该用途。

以上两类数字化工程,均可采用业务外包与自主开发的方式进行,但对于多重用途的数字化工程,如果本机构没有足够的资金、技术力量和管理人员,是很难达到工程的最终目的的。

2.2 机构实力

要自主开展数字化工程,除具备足够的经济实力,还必须具有足够的人力资源保障。人力资源包括两个方面:技术人才与管理人才。技术人才是进行数字化的技术保证,而管理人才是数字化建设的组织保证与质量保证。数字化工程的费用是巨大的,如果一个机构不能承担预算的费用,切不可降低数字化质量要求来硬撑,这将带来无穷的后患。以上3种数字化运行模式,哪种方式最经济,这是一个极难回答的问题,它取决于该工程的性质、该机构的人员知识结构等。如果机构内部人员知识结构多元化,具有高素质的管理人才与技术人才,就可以降低人力资源外聘的费用,为自主开展数字化工程或综合开展数字化工程创造条件。

一般说来,没有足够的资源来支持数字化工程的机构大多选择业务外包方式。如果投入设备上的资金总额过大且缺乏技术人员与管理人员,是可以考虑采用业务外包方式的。业务外包的优点很多,但也很容易失败,即达不到预计目的或适得其反,关

键在于合同的制定与数字化后期的质量监督。

如果机构具有一定实力,并希望凭借运行数字化工程的契机发展内部的技术力量,采用综合模式是较好选择。

2.3 文献性质

文献的珍贵程度与内容的敏感性、保密要求也影响着数字化工程运行模式的选择。珍贵与敏感的文獻有损坏、失真、丢失、泄密等风险。本单位自主开展数字化工作,可减少风险发生的机率;若本单位资源缺乏,也可以考虑采用综合法实行文献数字化,台湾国史馆档案数字化工程就是采用后种方式进行的。他们聘请计算机公司来馆辟专室架设数字化设备,雇专业人员进行图像采集,而质量审查、验收以及文献运行的全程管理由馆内专人承担。以下列出该项目分工,以供参考。

(1)每天由馆内专人向库房管理人员提调档案,交付厂商专职点收人员点收签名,并标记时间。

(2)归档监督:厂商驻馆实作,馆内专人在扫描现场监督厂商操作作业,以避免馆藏档案的毁损及遗失。

(3)回收复原:厂商每扫完一卷文件,馆内专人立即收回,并检查顺序有无错置与遗失。

(4)还卷归档:馆内专人按日将厂商提调的档案清点无误后,交还库房管理人员归档。

(5)每天厂商将影像数据传输到本馆计算机,由本馆人员校对审查,如发现有问題,应登记于记录表,交厂商限期修订。

馆内人员的严格监督,不仅有利于保证数字化质量,而且最大程度地降低了珍贵文献损坏、丢失的风险。

3 外包合同

信息化时代,第三产业迅速发展,自给自足模式结束,外包成为整个社会范围内合作与分工的产物,很多单位的数字化工程选择了业务外包的方式。他们借助外包服务供应商来获得专业性的服务与质量,但很少有客户与外包服务供应商的关系能十分融洽。据相关数据,只有8%的客户表示他们的外包服务超过了预期,而近20%的客户认为他们的外包服务并没有达到预期目标^[2]。

惠普公司在分析客户对服务满意度不高的问題时说:“这与那些要求外包服务的单位最初的念头有关,他们认为可以移交一切。”惠普中国区专业与支持服务集团市场战略及联盟事业部总经理孙秀芳女

士认为,目前,国内某些企业在IT外包上混淆了概念,真正的外包服务,外包的不是责任而是执行。所谓的执行,就是执行双方签定的合同条款。因而业务外包的合同内容成为双方责权的凭据,彼此沟通的基础。

3.1 合同应当明确指定双方责权

合同必须以可操作方式明确指出双方的责权,以便协作与沟通,一旦问题出现,也便追究责任。以下是相关实例:

(1)部分业务外包的台湾“国史馆”档案数字化的责权。馆文献处负责图像文件制作的规划、监督、提调、校对与修缮等工作;承包商提供设备与技术人员,并按“国史馆”提供的文献类别分别提出技术建议,选择最适于转换的存储格式进行数字化操作,该建议经本单位同意后方可执行。

(2)以业务全外包方式运行的台湾地区文献数字化工程。《台湾地方文献数字化工作手册》规定,职能部门指导各单位在签订外包合同时,应当分工明确:各单位对文献进行分类,并执行对最终结果的质量检查;承包商应按本单位提供的分类文献,根据《地方文献数字化工作手册》所提出的数字化参考格式,提出最适于转换的数字存储格式,经本单位同意后再进行资料转换,进行资料转换前厂商需依本单位所提供的资料分类清册逐项进行储存格式确认,确认格式的清册被作为验收依据。

(3)应将质量要求与控制写入合同。质量是数字化工程的关键,必须以可执行的方式写入合同。以下是几份合同的内容,供参考。

例1:台湾“国史馆”档案数字化项目对数字产品验收的要求:①备份光盘2份(TIFF〈非压缩〉及JPEG〈压缩50%〉各1份),依厂商所提供清单,核对光盘之数量是否正确。②检查光盘盒及每片光盘是否加注文件名称、光盘编号、文件号起讫。③检查同一卷之影像资料是否拷在下一片或编号相邻之光盘内。④检查是否有前一大类的影像资料与后一大类的影像资料拷录在同一片光盘中。⑤发现有前述错误时,登记于登录表后,交予厂商修正或重新拷录再提交品检。

例2:台湾“国史馆”档案数字化项目对质量控制的要求有:①每天由厂商将影像数据传输至本馆10部计算机进行品检。②检查项目:影像资料歪斜或歪曲度是否超过5度,是否有内、外折,是否有漏页,影像是否模糊不清,档案分辨率及文件名是否有误?

如发现前述情形,应登记于校对记录表,交予厂商限期修正。

例3:台湾地方文献数字化工程中,对质量控制的要求有:①转换之数字资料本单位得进行检验,资料不良率高于1/4者,每笔错误得依待标价之3/4进行罚款。资料辨识程度良窳,由签约前厂商与本单位共同协商,双方同意后并入合约执行。②应将降低文献损坏的风险的措施写入合同。承包商追求工作的高效率,可能使用不适当的设备来完成数字化工作,或由于工作疏忽等原因带来文献的损坏,均应考虑到这些因素,并写入合同。例如《台湾地方文献数字化工作手册》中要求各单位在合同中必须有以下条款以减少风险发生:第一,资料运送和转换过程中,不得损坏本单位提交的资料原件,若有损坏发生,依本单位规定的财产价格3倍赔偿,情形重大者得由本单位决定中止合同,不得有异。第二,进行资料转换时,若因使用的设备造成原件损坏,依本单位规定的价格3倍赔偿,本单位应立即中止转换作业,情形重大者,由本单位决定中止合同,不得有异。第三,厂商使用转换资料的高速扫描设备及数码相机等相关设备型号及作业方式,需经本单位同意,本单位有权选择设备使用方式及流程控制方式。

业务外包时阐明自身要求、目的与条件十分重要,但这些问题的提出都与IT技术的掌握程度以及对数字转换技术的了解关系密切。我国很多用户正好缺乏这方面的知识,例如在外包中图像格式的提出与敲定,基本由商家确定。用户认为只看数字化后的效果就可以了,但效果与质量有时是不统一的。直观效果不能表明特定数字图像的用途,更不能说明这就是该数字化的最佳选择,因而用户在起草外包合同时必须多找专家咨询,特别应多借鉴国内外已成功案例。我国图书馆、档案馆在缩微工作方面有相当多的经验可以借鉴,使业务外包的潜在优势得到更大程度的发挥,将不利因素限制到最低范围。

参考文献

- 1 Planning an Imaging Project: <http://www.rlg.org/visguides/visguidel.html>
- 2 阿甘·惠普:对待外包,态度不能含糊.中国计算机用户, 2003-01-02

刘家真 武汉大学信息管理学院教授,博士生导师。
通信地址:武汉。邮编430072。(来稿时间:2004-07-08)