

档案领域元宇宙基础研究

■ 赵生辉 / 西藏民族大学管理学院

■ 胡莹 廖月妮 / 云南大学历史与档案学院

摘要: 从档案学视角解读元宇宙概念与元数据、本体、档案等术语之间的联系, 分析基于档案资源的元宇宙建设技术路线, 探索档案部门助力元宇宙建设的实践路径, 展望档案领域元宇宙研究的若干理论问题和技术方向。元宇宙是建立在现实世界映射与交互基础上的虚拟世界和数字生活空间。档案是构建社会系统本体模型的重要基础材料, 在元宇宙数据基础设施建设中发挥着不可或缺的作用。基于档案资源的元宇宙基础设施建设应采用“由实及虚”路径, 通过对大量档案文献内容信息的数据抽取、本体建模和场景再构驱动虚拟和现实资源深度融合。档案部门可以通过“历史时空回溯”和“数字孪生世界”两种实践路径赋能元宇宙建设, 面临众多有待深入探讨的理论问题和需要系统解决的技术问题。

关键词: 元宇宙; 虚拟现实; 本体; 数字孪生; 档案数据化

分类号: G270.7

The Primary Discussion on Metaverse Related Fundamental Researches in Archive Studies / Zhao Shenghui Hu Ying Liao Yueni

Abstract: Based on the interpreting of relationship between metadata, ontology, archive and Metaverse from perspective of archival science, the technical route of Metaverse construction depending on archival resources is analyzed in this paper. Then, two practical pathways of Metaverse construction empowered by archival agencies are explored, some theoretical issues and technological directions of Metaverse researching in the archive studies are proposed. Metaverse is a virtual world and digital living space based on the mapping and interaction of the real world. Archive is an kind of crucial basic material for constructing ontology model of social system and plays an indispensable role in construction of Metaverse data infrastructure. The practice of Metaverse infrastructure construction based on archival resources should adopt the path of "from real to virtual", which drives the deep integration of virtual and real resources through data extraction, ontology modeling and scene reconstruction of a large number of archival literature content information. Archives management agencies can empower construction of Metaverse through two practical pathways of "historical time-space backtracking" and "digital twin world" which facing

many theoretical issues to be deep-going explored and technical problems to be solved systematically.

Keywords: Metaverse; Virtual Reality; Ontology; Digital Twin; Archives Data Transformation

元宇宙 (Metaverse) 是2021年开始在全球信息产业界广泛传播并受到全社会高度关注的新兴技术概念。到目前为止, 人们对元宇宙的理解还处于初级阶段, 对很多基础问题的讨论尚未取得共识, 但是其所代表的“虚实深度融合”“时空交错再构”等理念还是得到了众多企业的追捧, 被视为沉浸式移动互联网的雏形和人类“两栖”生活的起点, 2021年还被称为“元宇宙元年”^[1]。从构词法角度理解, Metaverse就是“universe about universe”, 即致力于映射和重构现实宇宙资源的数字化虚拟生活空间。北京大学陈刚教授认为, 元宇宙是“利用科技手段链接和创造的, 与现实世界映射与交互的虚拟世界, 具备新型社会体系的数字生活空间”^[2]。元宇宙是在虚拟现实 (visual reality, VR) 技术基础上发展起来的一种以沉浸式仿真交互为特征的技术愿景, 目前支撑其实现的各项关键技术尚未成熟, 产业发展总体上处于早期战略布局和基础技术酝酿阶段。20世纪下半叶至今, 信息技术的每一次重大变迁都会使人类信息资源管理和利用模式发生显著变化, 催生众多的新兴理论和技术方法。元宇宙技术潮流的兴起在给信息资源管理学科提出挑战的同时也必然会创造出新的发展机遇, 这一趋势已经得到图书情报领域部分学者的关注, 相关研究成果呈现出逐渐增多的趋势^[3-5]。档案学界对元宇宙相关问题的探索尚处于概念启蒙阶段, 对基础理论、实践模式和技术方法缺乏相对系统清晰的认识, 亟待开展一场以“元宇宙与档案学科发展”为主题的学术讨论。

1 档案学视域中的元宇宙理论解读

2020年12月, 腾讯科技董事局主席马化腾在公司内部刊物《三观》上撰文指出: “移动互联网的十年发展, 即将迎来下一波升级, 我们称之为全真互联网。”^[6]“全真互联网”是马化腾从另外一种视角对元宇宙进行的表述, “全真”即“深度沉浸式立体仿真交互”, 既表示用户在互联网虚拟场景中的真实性体验, 也反映了元宇宙与现实世界各类真实存在事物之间的关联与映射关系。以“真实性”作为核

心理理念的学科领域,档案学理论和方法将在元宇宙基础设施建设与运营过程中发挥不可或缺的作用。

1.1 元数据与元宇宙

元数据(metadata)是与元宇宙(metaverse)构词法较为接近的术语。“metadata”的经典解释为“data about data”,后者是被描述的资源对象,前者是用来描述资源对象属性特征的数据。人们要对资源对象的属性特征进行数据描述的主要原因在于其存在形态不便于计算机处理。元数据创建之后,用户就能以元数据为工具间接地获取资源对象的属性特征而不用逐一感知和验证。因此,元数据本质上是资源对象在网络空间的信息模型。与此同理,“metaverse”可以理解为“universe about universe”,后者是现实存在的宇宙,前者则是现实宇宙在虚拟空间当中的信息模型。如果人类社会的万事万物及其互动关系都可以在元宇宙当中建立数据模型,人们就可以通过与元宇宙的交互和处理完成对现实宇宙运行规律的认知和把握,从而为管理和决策提供全景式的基础信息支持。当现实世界的各类运行场景都在元宇宙中建立了数字化模型,之前分别位于不同时间、空间的社会活动就有了一致性的逻辑基础,就可以跨越时间、空间的阻隔进行“混搭”,从而创造出多种类型之前在现实世界难以实现且用户能够体验到真实感、沉浸感的新型应用场景。

1.2 本体与元宇宙

如果把元宇宙的“元”理解为哲学意义上的“本原”,那么元宇宙就是指客观世界原本的运行状态,其内涵与“本体”非常接近。本体(ontology)一词来源于哲学,意为“存在”,即事物在客观世界存在的本原形态。计算机科学、图书情报与档案管理等学科的学者在研究中引入本体概念的目的在于为了解决文献信息符号多变性与事物本体形态稳定性之间的矛盾。哲学意义上的时空本体、社会本体都是只能体验而无法由计算机直接处理的资源,研究人员通过建立与时空本体、社会本体具有类似逻辑结构和属性特征的数字模型并使其作为数字代体(digital surrogate)参与计算和分析,就成为信息学领域的本体。“(信息)本体”是指特定领域明确的、详尽的、可共享的形式化概念体系,早期表现为分类表、叙词表等人工检索语言,随着知识图谱(Knowledge Graph, KG)、建筑信息模型(Building Information Model, BIM)等技术的发展演变成为可以对社会系统万事万物复杂关系进行建模的工具。“(信息)本体”是独立于多变的符号层逻辑而存在的,是模拟现实世界事物属性关系及其互动规律的数据模型^[7]。“(信息)本体”通常局限在特定领域,而元宇宙可以跨越多个领域实现本体的关联与融合。“(信息)本体”采用便于计算机分析和处理的结构化模型,对于现实世界的模拟侧重于逻辑关系,而元宇宙是一种可供用户体验的虚拟生活空间,在逻辑关系模拟的基础上,还需要考虑真实社会场景的建模和用户体验的模拟问题。

1.3 档案与元宇宙

档案是机构和个人在社会实践过程中直接形成的历史记录,其内容反映了特定时空之下机构、个人、实物、事件等多个维度的属性特征,是现实世界本体建模最可靠的信息依据和基础资源。元宇宙是基于现实宇宙的虚拟空间,无论是哪种类型的本体模型,其建模过程都不能凭空进行,而是要以反映空间或社会系统特征的原始档案为依据,确保元宇宙与现实世界在逻辑层面上的一致性。元宇宙的实现需要依托三种本体模型,即模拟地球表面空间特征的地理信息本体模型、模拟现实世界建筑及空间关系的建筑信息模型、模拟社会系统属性关系的社会本体模型。在实践当中,与地理信息本体模型和建筑本体建模相关的档案主要表现为城建档案,是由地质、测绘、水文、气象、建设、房产等多个职能部门在其工作当中采集和生成地对区域空间信息的原始记录。与社会系统本体建模相关的档案主要表现为机构、个人在社会活动中形成的各类文书、图表或数据,其内容记录和反映了特定时空范围内机构、个人、实物之间的属性关系及其互动过程,呈现的是人类社会活动的多维度信息。本体模型遵循语义层逻辑,与现实世界各类机构、个人等社会实体的唯一性存在方式类似,在本体模型当中特定的地名、机构、个人都被赋予唯一不变的标识符URI(Universal Resource Identifier),从而使其可以跨越档案文件的边界和语言文字的差异性,在逻辑上与描述同一对象属性特征的所有档案联结在一起。在本体建模过程中,档案内容信息被抽取和表达为采用资源描述框架RDF(Resource Description Frame)格式保存的语义数据,可以脱离载体和信息组织单元的束缚,按照与现实世界一致的逻辑集成与融合,从而将大量“信息碎片”拼接为“全局信息”,成为反映社会系统属性规律、可以在管理和决策活动中发挥重大作用的“大数据集”。反映空间场景、建筑样貌、个人身体特征等方面的档案文献还可以为元宇宙中的场景建模和个人数字分身建模提供依据。由此可见,档案是构建社会系统本体模型的重要基础材料,在元宇宙数据基础设施建设中发挥着不可或缺的作用。脱离了各类档案资源的支持,元宇宙与现实宇宙之间的映射与关联关系就会被割裂,而完全基于虚构场景、虚拟人物、虚拟事件构建的元宇宙与饱受诟病的网络游戏并没有本质区别。

2 档案资源赋能的元宇宙实践路径

档案学是贯通人类社会系统几乎所有子领域、所有方面的“横断学科”。档案载体形式的原始性、内容与社会活动的高度相关性、涉及领域的广泛性等独有的特征,决定了档案管理领域可以为元宇宙的社会系统本体建模提供丰富、可靠的基础资源,也使得档案学成为支撑元宇宙大厦的基础性学科之一,档案学也必将在参与元宇宙建设、发展与演进的过程中获得新的发展机遇。

2.1 基于档案资源的元宇宙建设技术路线

与当前全球部分科技公司基于网络游戏平台、社交媒体平台驱动初级版本元宇宙建设的“由虚及实”路径不同,基于档案资源的元宇宙建设采用与之相反的“由实及虚”路径,主要通过对大量档案文献内容信息的数据抽取、本体建模和场景再构来驱动元宇宙的建设。

现实世界(universe)当中的机构和个人在社会实践当中生成各类档案文献,这些文献的内容信息是对社会系统特定场景(scene)的直接或间接记录或描述。场景是元宇宙的重要概念,主要由所处时间段或时间点、空间范围、地表形态、建筑信息、人物属性关系、实物属性关系、事件演化过程、事件最终结果等方面的信息构成,涉及区域城建档案、机构文书档案、新媒体档案等多种类型。如果将档案文献的内容信息抽取为RDF数据并基于空间本体GIS、建筑本体BIM和社会本体ADO实现聚合之后,就可以完成特定领域社会系统的整体性数据建模。基于社会系统的本体模型,对其中与用户需求相关的部分进行重新组织,就可以构建为数众多的社会场景数据集。其中,部分场景是根据档案文献内容进行推断之后形成的场景回溯,部分场景是对现实场景的精准再现,而另外一些场景则是对本体库当中涉及的时间、空间、机构、个人、实物等信息进行重组之后形成的交融型场景。这些场景需要通过虚拟现实VR、增强现实AR或混合现实MR设备进行访问,可以使用户产生与现实世界场景类似的真实感与沉浸感。用户的场景访问活动在以统一身份管理URI、资产权益管理NFR、虚拟货币交易NFT等机制作为保障的情况下,就构建起了与现实世界类似的一种可持续的虚拟生活环境,即建成了元宇宙(metaverse)。元宇宙与现实宇宙是一种模拟、再现和重构的关系,用户可以通过VR/AR/MR设备实现在现实宇宙和元宇宙之间自由切换,跨越地理位置的阻隔参与千里之外其它空间的社会活动,获得与在现实世界类似的真实性体验。元宇宙是建立在现实世界信息本体模型基础上的场景化虚拟生活空间,以档案大数据为基础支撑,运行和管理都带有人工智能社会的特征。如,生成与用户需求高度契合的个性化场景,在虚拟社交场景中提示陌生人的背景信息,决策可以基于全局数据的精准分析并推荐优化方案等。

2.2 档案部门助力元宇宙建设的实践路径

元宇宙作为与现实宇宙平行存在的虚拟生活空间,在时间维度上同样呈现出线性延续特征,而特定时间段元宇宙建设的完备程度与留存的档案资源的丰富程度具有密切的联系。根据元宇宙对应档案资源时间段的不同,档案部门主导或参与的元宇宙建设实践路径可以分为以下两种。

2.2.1 历史元宇宙:以档案为依据的历史时空回溯

历史时空回溯是指档案部门以留存的档案文献的内容信息为依据,通过语义数据反向推导与关联融合,在网络空间重构历史空间格局,模拟实体属性关系和社会活动过程并通过多种形式予以呈现的过程^[8]。近年来,以历史时空回溯为

主题数字化项目在数字人文领域得到广泛关注,其中影响最大的是由瑞士洛桑联邦理工学院数字人文中心与意大利威尼斯大学、威尼斯国家档案馆等机构联合推动建设的“威尼斯时光机”(Venice Time Machine)项目。该项目通过数字化手段重建了中世纪时威尼斯亚里托地区的空间格局与建筑风貌,并通过对同时期历史档案的数据挖掘,重建了中世纪时期威尼斯人口的社会关系网络。“威尼斯时光机”项目追求的虚拟现实空间体验与元宇宙的建设目标是一致的,即在虚拟网络空间重现特定历史时期特定地理区域内现实世界的真实风貌,可以视为“历史元宇宙”建设的成功范例。继“威尼斯时光机”项目之后,阿姆斯特丹、巴黎、布达佩斯、耶路撒冷等城市的时光机项目相继启动,旨在连通各类城市时光机项目的“欧洲时光机(Time Machine Europe)”项目被欧盟列入未来十年重点资助的旗舰型研究计划。我国各大城市的档案管理部门,可以联合图书馆、博物馆等公共文化机构,通过“城市逆向记忆工程”^[9]在网络空间再现所在城市的建筑格局和生活风貌。

2.2.2 现实元宇宙:档案赋能的数字孪生世界构建

数字孪生(digital twin)又称数字克隆、数字镜像^[10]等,是指利用信息感知、虚拟建模等技术在虚拟空间构建与现实实体完全对应的数字实体并依据历史数据进行参数设定,使其可以与现实实体在生命周期内同步变化的信息技术。数字孪生技术目前主要运用于工业控制和建筑工程领域。随着信息化建设的推进,数字孪生技术呈现出向社会领域扩展的趋势,政府、企业、个人都可以在网络空间拥有与实体具有类似属性的数字分身(digital avatar),使其成为社会实体各类属性特征在虚拟世界的载体。元宇宙当中的任何虚拟实体构建过程中,无论是外部形象还是内部属性特征的建模都不应该凭空完成。只有以现实实体的档案信息为依据,数字分身才能尽可能模拟现实实体的特征,在虚拟网络空间当中以符合现实实体行为特征的方式参与社会活动。在数字孪生体创建之后,需要采集和保存实体参与社会活动的原始记录数据,以保持实体与数字孪生体之间的平行映射关系。当现实世界的地理实体、建筑实体、机构实体、个人实体、实物实体等在虚拟世界建立了各自的数字孪生体且可以相互关联和互动,平行于现实世界的元宇宙虚拟空间就可以建成。数字孪生世界本质上是现实世界的“平行宇宙”,需要档案部门与多种社会机构基于协作机制共同完成。

3 档案领域元宇宙相关研究展望

元宇宙目前在档案学界还属于新鲜事物,存在大量有待深入探索的理论问题和有待解决的实践问题,现通过列举部分代表性研究领域和技术方向,期望达到抛砖引玉、促进领域学术讨论的目的。

3.1 档案领域元宇宙相关理论问题

第一,全息档案理论。我国学者罗润泽早在1989年就

提出了全息档案学和宇宙全息系统网络理论^[1],受制于当时的社会信息化环境和数据资源基础,该理论并未受到足够重视。全息档案学认为,档案全宗本质上是个全息系统,与其所在的社会系统之间存在同构现象,通过分析档案全宗就可以获取社会系统的全方位信息,还原社会系统运行的过程。元宇宙构建过程中,档案资源与社会系统之间的全息映射关系正是档案本体数据集建模的理论依据。全息档案理论的内涵和实践形式有待学界进一步探索。

第二,实体数据全宗理论。全宗是纸质档案管理时代非常有效的实体档案资源组织单位。然而在档案数据本体当中,档案文件之间的分隔不复存在,数据与数据之间的边界日益模糊,只有与特定实体属性特征高度相关的所有档案数据才在逻辑上被视为同一整体。实体数据全宗是指与特定社会实体属性特征相关的所有档案数据的集合,是元宇宙当中档案数据资源组织与检索的基本逻辑参照单元。

第三,数字孪生体档案理论。传统档案管理以文件及其集合为基本单位,有关同一实体对象的档案信息分散于多个来源的档案文献当中。领域本体建模和数字孪生体的构建实现了上述信息的集成,如果将数字孪生体视作档案资源的一种特殊表现形式,则在元宇宙虚拟空间当中,档案工作的重心就由管理档案文献转向维护社会实体与孪生体属性关系的一致性。在两者保持一致性的情况下,用户需要了解特定社会实体相关信息时,可以通过与数字孪生体的互动获取而不用查阅作为其来源的各类档案文献。

第四,元宇宙档案馆理论。档案学是专门研究人类社会记忆留存和利用问题的学科,如果人类社会活动的边界从现实空间拓展到了元宇宙,如何记录和保存个人或机构在虚拟空间当中的社会记忆,就成为“元宇宙档案馆”研究需要探索解决的问题。元宇宙档案馆并不是实体档案馆在元宇宙当中的数字孪生体,而是记录与描述元宇宙虚拟空间当中人类社会活动的数字资源及相关软件资源的总和。

3.2 档案领域元宇宙相关技术问题

第一,面向元宇宙的场景时空档案数据资源采集技术。

基于档案资源的元宇宙建设本质是一种以档案为中介的技术路线,通过对档案内容信息的发掘与分析,利用数字化手段部分还原社会生活场景的原貌。如果应用三维数字扫描技术对现实生活中的场景进行时空数据的连续性采集,则可以减少因档案数据抽取造成的信息损耗,从源头上构建立体化的时空数据集,为基于场景数据的时空再构奠定资源基础。

第二,面向元宇宙的测绘、城建、社会档案数据资源深度融合技术。元宇宙是一种贯通各类领域本体的社会系统通用本体集成架构,需要将测绘部门生成的地理信息档案、城建部门生成的建筑档案和档案管理部门收藏的社会档案的内容信息融合于同一的数据空间,面临大量有待解决的技术难题。

第三,基于机器学习的档案文献数据抽取技术方法和流

程规范。档案领域数据本体的真实性是由档案文献资源传递的一种衍生属性^[12],只有确保作为来源档案资源的真实性并确保档案数据抽取的方法、流程符合规范,本体的整体真实性才能得到确认,因而档案文献数据抽取的技术方法和标准规范是亟待研究的问题。鉴于档案文献数量的海量性特征,引入机器学习等人工智能领域的技术方法是非常必要的。

第四,元宇宙空间虚拟实体唯一性标识与解析技术。元宇宙当中的各类实体的“数字分身”需要有与现实世界对应的身份来参与虚拟化的社会活动,身份的认证与管理可以依托本体模型的资源统一标识符URI管理机制,为地名、机构、个人等各类实体赋予唯一的编码,使其在虚拟场景当中可以相互区分。

第五,基于沉浸式虚拟现实技术的数字档案馆服务空间重构。沉浸式虚拟现实技术可以为用户创造类似现实生活场景的真实性体验,从而使数字档案馆超越传统的信息系统模式,以参与式虚拟信息空间方式为用户提供服务,用户通过手势控制等方式在虚拟空间调阅档案文献并完成点选批阅等加工操作,还可以通过与虚拟形态的智能助理进行自然语言交互,直接获取与其需求高度匹配的精准档案数据。

本文系国家社会科学基金项目“多民族语言数字资源语义互联框架研究(项目号:19BTQ04)”的阶段性研究成果。

注释与参考文献:

- [1] 刘连松,黄宣,韩爱青,等.元宇宙:开启人类“两栖”生活[N].天津日报,2021-11-24(10).
- [2] 丁建英.元宇宙:或重新定义人类文明4.0[EB/OL].(2022-01-16)[2022-03-19].<https://www.163.com/dy/article/GTS453V20511831M.html>.
- [3] 杨新涯,钱国富,唱婷婷,等.元宇宙是图书馆的未来吗?[J].图书馆论坛,2021(12):35-44.
- [4] 赵星,陆绮雯.元宇宙之治:未来数智世界的敏捷治理前瞻[J].中国图书馆学报,2022(1):52-61.
- [5] 陈定权,尚洁,汪庆怡,等.在虚与实之间想象元宇宙中图书馆的模样[J].图书馆论坛,2022(1):62-68.
- [6] 腾讯网.案例解码“全真互联网”[EB/OL].(2020-12-03)[2022-03-19].<https://new.qq.com/omn/20201203/20201203A0DI8Q00.html>.
- [7] 赵生辉,胡莹.拥有整体性记忆:档案领域数据本体管理论纲[J].山西档案,2020(6):17-27.
- [8] 赵生辉,胡莹,黄依涵.打造“时光机器”:城市逆向记忆工程理论与实践研究[J].档案学研究,2021(6):114-121.
- [9] Time Machine Project Website. Time machine organization[EB/OL]. [2022-03-19]. <https://www.timemachine.eu/time-machine-organisation/>.
- [10] 郭炯.数字孪生在城市治理中的应用研究[J].新型工业化,2021(11):35-37.
- [11] 罗润泽.全息档案学与宇宙全息系统网络理论[J].档案与建设,1989(3):7-9.
- [12] 赵生辉,胡莹.档案领域本体数据集衍生证据价值实现机理探析[J].浙江档案,2021(1):19-21.