

# 丰富人文研究多重证据参照体系的数据基础设施建设——从文化记忆的角度

夏翠娟

上海图书馆

发源于“集体记忆”的“社会记忆”理论，系统性地将记忆从历史中剥离开来，不仅对社会学、人类学、文化学、民族学产生了巨大的影响，也为史学研究提供了全新的视角，在“二重证据法”、“三重证据法”、“四重证据法”的基础上，发展出了“多重证据法”。表现在除了各种史学典籍之外，重视民间文献和田野调查，且奏章黄册、方志家谱、口述历史、私人档案、私家笔记、诗词歌赋、绘画艺术、金石瓷器、传说歌谣、戏曲戏剧、节日仪式、遗址遗迹等文化记忆资源，甚至气候灾害、树轮冰川等自然科学研究的成果，无不可成为参照和互证的证据。而这些文化记忆资源，大多保存在图书馆、档案馆、博物馆、美术馆（GLAM）等不同文化记忆机构中。

数智（大数据和机器智能）时代，内容和载体的分离，缩小了 GLAM 之间因资源载体不同而造成的差距，数据、事实和知识成了“文化记忆”的最小单位，不仅 GLAM 的文化记忆资源都应成为多重证据参照体系中的一部分，人、地、时、事、物等实体的知识图谱，以及基于大规模、长时间、多维度、细粒度的数据所生成的量化分析数据、可视化图表等，也构成了多重证据参体系中不可忽视、愈加重要的另一重证据。

通过对文化记忆资源的长期保存、载体管理、知识整序和传承传播工作，GLAM 都成为了文化记忆的基础设施，都在参与“构建未来的文化遗产”的工作。而在数智时代，文化记忆机构经过数字化转型，建设以“数据”为基本的组成单位的“数据基础设施”，成为支撑人文研究传承文化记忆的重要工作。“数据”最初的定义是“可被计算机传输和存储的信息”，对于人文研究来说，“数据”可理解为可被机器处理的信息单元，如文献或实物资源对象、概念、人物、机构、团体或其结构化的描述信息（包括变量、数值、文字符号或事实等）。

“数据基础设施”，除了具备“基础设施”的开放性、公共性和可持续性，还应充分体现数据规模大，覆盖时间长，地域范围广，描述粒度小、维度多等特点，以支持全网域（Webb-scale）的数据调度、融合和自动化分析、统计，和数据可视化。同时还应独立于具体应用开

发和特定领域研究之外，遵循通用的数据组织规范和开放共享规范，成为介于信息基础设施“后台”和特定领域研究“前台”之间的“数据中台”

本报告试图通过理论建构和实践探索，阐述文化记忆机构如何通过数据基础设施建设来丰富数智时代人文研究的多重证据参照体系。

***Building a Data Infrastructure to Enrich the Multiple Sources of Evidence for Humanities Studies:  
From the Perspective of Cultural Memory***

***Xia Cuijuan***

***Shanghai Library***

The theory of "Social Memory", which originated from "collective memory", systematically separates memory from history. It not only has a great influence on sociology, anthropology, culture and ethnology, but also provides a new perspective for historical research. On the basis of "double sources of evidence method", "triple sources of evidence method" and "quadruple evidence method", it develops "multiple sources of evidence method". In addition to all kinds of historical books and records, it attaches great importance to folk literature and field investigation, as well as cultural memory resources such as memorials, local chronicles, family genealogies, oral history, private archives, private notes, poems and odes, painting art, gold and stone porcelain, legends and ballads, operas and dramas, festival ceremonies, site relics, and even natural science research achievements such as climate disasters, tree rings and glaciers, All of them can be the sources of evidence and mutual confirmation. Most of these cultural memory resources are stored in libraries, archives, museums, art galleries (GLAM) and other cultural memory institutions.

In the era of Data & Intelligence (big data and machine intelligence), the separation of content and carrier has narrowed the gap between GLAM due to different resource carriers. Data, facts and knowledge have become the smallest unit of "cultural memory". Not only GLAM's cultural memory resources should become parts of the multiple sources of evidence reference system, but also the knowledge graph of people, places, times, events, things and other entities, as well as the quantitative analysis data and visual charts generated from large-scale, long-term, multi-dimensional and fine-grained data also constitute another important source of evidence that can not be ignored and is becoming more and more important in the multiple sources of evidence reference system.

Through the long-term preservation of cultural memory resources, carrier management, knowledge organization, inheritance and dissemination for culture, GLAMs have become the infrastructure of cultural memory and is participating in the work of "building future cultural heritage". In the era of Data & Intelligence, cultural memory institutions have undergone digital transformation, and the construction of "data infrastructure" with "data" as the basic unit has become an important work to support the humanities research and inherit cultural memory. The original definition of "data" is "information that can be transmitted and stored by computer". For

humanities research, "data" can be understood as information units that can be processed by machine, such as literature or physical resources, objects, concepts, people, institutions, groups or their structured descriptive information (including variables, values, text symbols or facts, etc.). In addition to the openness, publicity and sustainability of the "infrastructure", the "data infrastructure" should also fully reflect the characteristics of large-scale data, long time coverage, wide geographical scope, fine-grained and multi-dimensional, so as to support the data request, fusion, automatic analysis, statistics and data visualization in a web scale. At the same time, it should be independent of specific application development and specific field research, follow the general data standards and open sharing specification, and become a "data center" between the "back-end" of information infrastructure and the "front-end" of specific field research.

Through theoretical construction and practical exploration, this presentation attempts to elaborate how cultural memory institutions enrich the multiple sources of evidence reference system for humanistic research in the Data & Intelligence age through the construction of data infrastructure.