大数据时代需要数据馆员

大数据环境下，数字图书馆拥有多种形态的数据，既有题录、书目、文摘等元数据，又有读者数据、借阅数据和日志数据，还有科研过程中产生的科学数据，以及全文数据和影像数据等。对这些数据进行收集、分类、编目、组织、存储、开发、创建和传播共享，全方位提供数据服务，激发数据价值和活力，需要图书馆员能力的转型和提升，大力培养数据馆员。

数据馆员，顾名思义，就是提供数据收集、组织、管理和服务的图书馆员。数据馆员主要负责管理、分析和利用图书馆所有类型的数据，涉及的工作内容主要包括数据采集、加工、管理与组织、数据存储、数据安全、数据开放与服务等，需要持之以恒地对各类数据进行监控，培育数据敏感力，发现数据涉及的事物发展规律和趋势。

第一，培养元数据馆员。通过研究数据间的关联关系，设计合理的元数据结构，选择适用的元数据字段，开发元数据模型，确立适合的元数据方案，制定元数据标准，协调数字化项目全过程活动；元数据馆员还需要对元数据进行加工、清洗、规范、去重和集成，保障现有的元数据能够支持数据管理和信息检索。第二，培养科学数据馆员。通过嵌入用户科研环境和科研过程，深入了解用户对科学数据的需求；开展科学数据管理培训，为科研团队制订系统全面的科学数据管理计划；建立科学数据服务平台，推进科学数据的发现、采集、保存、分析、导航、检索、出版和利用等，方便科学数据的共享。第三，培养数据分析馆员。数据分析馆员不仅要了解数据结构和标准，还要对科学数据、文本数据、用户数据、日志数据等进行收集、整理、注释、关联、分析、挖掘，熟悉数据分析工具；要作到既善于对数据进行描述分析，又敢于对数据作出预测分析，实现数据间的关联；帮助用户发现和获取相关数据，生成满足用户需求的数据分析报告，实现数据的增值和再利用。

《促进大数据发展行动纲要》提出要加强数字图书馆等公益设施建设，构建文化传播大数据综合服务平台，在此推动下，数据科学家、数据公民、数据馆员正在成为创新社会的新兴动力，图书馆需要界定数据的使用权利，合理合法进行数据获取，整理、分析有用数据，构建数据存储基地，形成数据服务新业态。由此可见，数据采集与整理、数据存储与应用、数据分析与解读、数据管理与服务、数据出版与发布等将成为图书馆员的必备技能。

■ 曾建勋