

数字图书馆的“大数据”

2012年3月奥巴马政府宣布投资2亿美元启动“大数据研究与发展计划”，并将之定义为“未来的新石油”，标志着大数据时代的到来。

大数据实际上是指针对海量数据的处理方法。当数据量和规模巨大到无法利用目前主流软件工具处理时，如何在合理时间内摄取、管理、处理并整理分析成为有用的资讯，是大数据要解决的问题。大数据对人类的数据驾驭能力提出了新的挑战，也为人们获得更为深刻、全面的洞察力提供了前所未有的空间与潜力。

数字图书馆作为人类知识的储存空间、社会文化的传承渠道，积累着大量数据，如各类型书目数据、文摘数据、全文数据、事实型数据，等等。这些数据中蕴含着大量的知识及其关联，对各种知识主题、实体对象和载体形式，以及科研要素、科学文献、科技项目、事件活动、专家学者、产品技术、组织机构和呈现之间的关系进行揭示和挖掘，形成知识网络平台，进行相关计量分析和监测评估，以形成新的科学知识和实事对策，可以丰富数字图书馆的服务内涵。

随着互联网与人们生活的不断融合，人们的每一行为都会成为数据，不知不觉中人们的信息行为和消费习惯都会被记录。数字图书馆拥有丰富的用户数据资源，如用户注册信息、用户利用信息（如浏览、检索、下载等）、用户交互信息等各类日志等，通过分析这些用户数据，能够对用户的阅览习惯、资源利用行为、网络使用痕迹进行细粒度的分析，进而拓展数字图书馆的服务形式。

大数据的发展，提供了一种看待世界的全新方法，更多地基于事实与数据作出判断，使社会也经历思维变革、商业变革和管理变革；大数据也必将促进数字图书馆数据管理、数据分析、数据使用及数据服务的深层次变革，既提出针对数据本身的处理和服务要求，也提出针对图书馆员的职责和意识要求。这更多地需要图书馆员适应大数据时代，改进自身的数据处理观念并提高数据挖掘能力。