

## 第四届中国数据馆员培训班顺利举办

2017年3月20日，第四届中国数据馆员培训班在中国科学院文献情报中心的院士厅正式开班。本次培训班以提高专业性知识和技能为目的，课程设置分为基础班（3月20日-24日）和进阶班（3月27日-31日），学员们根据自身知识背景自主报名相应班级，本届基础班培训学员49名，进阶班培训学员52名，共101人次，经过为期1~2周的集中培训，合格学员获得了由中国图书馆学会专业图书馆分会颁发的证书。

基础班培训期间，首先由中科院文献情报中心、中国科学院大学经济与管理学院的顾立平老师为大家介绍了目前有关数据治理的发展趋势和背景，强调了数据科学的重要影响，并提出了数据馆员自我实现的可行方案。中科院计算机网络信息中心孔丽华老师讲授了科研数据在出版、发布和引用过程中需要注意规范和问题。中科院文献情报中心罗娇老师介绍了开放数据的权益问题与管理方式。中科院文献情报中心赵昆华老师讲解了数字资源的开放内容管理机制。培训班第二天上午，中国农业科学院农业信息研究所的杜若鹏老师为大家详细讲解并演示了DSpace的安装、调试、测试、配置及相关注意事项。下午，来自北京大学图书馆信息化与数据中心的罗鹏程老师为大家讲授了Dataverse的安装、调试、测试、配置及相关注意事项。在接下来三天的培训中，学员们还学习了有关数据处理、数据分析和机器学习的课程，包含R语言的统计处理和可视化，以及丰富的项目案例分析探讨。

中国科学院文献情报中心副主任刘细文在进阶班开班致辞中讲到，“这是专业图书馆学会与高校图书馆学会合作的第一场培训，希望在此基础上，能够推出以数据挖掘探索知识服务的一系列开放政策、分析方法和信息技术。”进阶班阶段，来自超人学院的吴超老师利用两天时间讲授了大数据工程、存储、计算、管理、分析、采集、处理等需要学员深度学习的内容，包含大数据工程基本概论与工具、Hadoop原理与结构演示、MapReduce概论和爬虫与反爬虫的基础知识等，并指导学员学习Hadoop和Python的安装、调试、测试，HDFS数据导入导出，Hive数据查询和报表导出。去哪儿网的裴振老师和文献情报中心顾立平老师在之后讲解和示范Python的数据处理、网页爬虫的法律问题和案例实现，学员们还学习了Tensorflow安装和基本操作，体验深度学习的几个案例。

本届培训班由中国科学院文献情报中心知识技术研发中心、中国图书馆学会专业图书馆分会、中国图书馆学会高校图书馆分会联合举办，旨在推动图书馆员在开放科学的信息环境下向数据馆员发展转型。有学员表示，通过本次学习，不仅了解了国内外大数据的最新发展现状、信息政策前沿的著作权议题、开源软件、开放数据、开放获取等的权益机制，还能掌握R、Python、Hadoop、Tensorflow的安装与操作，对今后的研究视角和工作创新都有很大帮助。

即将由中国科学院文献情报中心和中国图书馆学会专业图书馆分会、中国图书馆学会高校图书馆分会联合主办的第二届中国合理使用周也即将举办，欢迎大家参加。



(赵昆华、顾立平供稿)