

# 李鹏辉：在海量数据中找到相关关系，就能产生价值

◆ 本刊记者 杨璐

清华大学研究生院专硕办副主任李鹏辉老师与数据院渊源颇深。从校领导决定建院之初，校长及四位副校长齐聚工字厅开会探讨，到如今教学规模达一千多名学生，李老师是一步一步看着，支持着数据院走到今天。他印象最深的是2018年初的学生RONG奖学金答辩，他说他看着各个院系的学生们，包括美术学院的、经管学院的等等，能用大数据思维去跨学科学习，他突然觉得这些年自己的工作特别有成就感。归根到底，教书育人这件事还是使命感的先导。

清华大学数据科学研究院成立于2014年。当时正值大数据热潮，时任研究生院常务副院长（现任清华大学副校长）的杨斌敏锐地捕捉到大数据发展的趋势，认为做大数据人才培养正值其时。时任研究生院培养办主任刘惠琴也认为，国

外已有商务分析大数据的相关尝试，清华也应发挥自己的工科优势开展大数据有关项目。二人一拍即合，当时在培养办工作的李鹏辉也加入筹备工作。

数据院的筹备与成立得到了校领导的高度重视。2014年1月2日上午，在清华大学工字厅的东厅，杨斌教授就“大数据行动在清华”作了主题发言，就数据科学的影响、国内外行动态势、清华现有成果以及未来建设等内容进行了论述。李鹏辉老师也回忆道：“那天的会议一共去了五位校长，这几乎是史无前例的，可以看得出来学校对这件事的重视。包括经管学院的陈国青老师、信息学院的刘云浩老师、交叉信息研究院的徐葳老师、公管学院的孟庆国老师等近十个院系的老师都参与到其中，也是在那一天学校定下了指导数据院前进及发展的24字方针。”“学校统筹、问题引导、



清华大学研究生院专硕办副主任李鹏辉

社科突破、商科优势、工科整合、业界联盟”这24个字就此确立，而首四字“学校统筹”说明了数据院的战略地位。

谈起加入筹备数据院的初衷，李鹏辉老师认为这既是与自己工作的结合，又是未来的大势所趋。从工作岗位上说，李老师在培养办负责专业学位（包括MBA、工程硕士等），专业学位本身就要服务国家经济社会发展的需求，而社会上大数据发展迅速，“我们肯定要跟上人才培养的目标”。而从大数据的时代意义上说，大数据不是学科的改变，而是学科的颠覆。传统学科因为有了大数据，已经不适用原来的一些研究方法。此外，大数据是一种思维方式的颠覆性变化，相比于因果性，大数据强调的是相关性，在海量数据中找到相关关系，就能产生价值。“所以我觉得建立数据院确实是挺好的一件事，自己也觉得挺愿意干这些事的。”

经过初步了解，清华校内大数据相关研究都散落在各个院系，急需整合资源。于是在数据院成立的同年9月，研究生院牵头调研了各院系情况，联合信息学院、社科学院、公管学院和经管学院等院系分设几个方向，筹备并开设《大数据分析》《大数据系统基础》《大数据算法基础》《大数据平台核心技术》等一系列课程，面向清华各院系在读研究生招生，完成学习并考核合格的学生可获得清华大学大数据能力提升项目证书。身为大数据人才培养计划的主要起笔者，李鹏辉老师指出，第一个目标就是“要在全校范围内普及大数据的思维方式”，这种大数据素养主要通过几个模块的课程学习实现。第一期的项目由10个学分的课程组成；从2015年起，加入了实践环节，学生需要通过所学的方法、手段去解决实际存在的问题，这也正符合了大数据“问题引导”的特点。

在不到四年的时间，数据院已聚集了一千多名学生，覆盖了全校所有院系，大数据能力提升项目迅速成为全校最有影响的、学生受益面最宽的能力提升项目之一。虽然受制于当下的学位体

系，第二个目标“培养大数据硕士、博士”目前无法实现，但学生最后的就业证明了大数据能力提升项目的社会认可度。当下，中国大数据IT市场蕴含强烈市场需求与良好的发展趋势。因此，既具备行业背景，又拥有大数据技术能力及思维方式的交叉学科人才，成为了企业急需的“宝马良驹”。正是在此背景之下，具有“跨学科”背景的学生在企业中变得“炙手可热”。

“从学生自己的一些体会、感受，我觉得他们真的是通过这个大项目受益了。”数据院也通过酒会等活动为学生、老师、企业搭建一个交流的平台，李老师也由此接触了不少学生。他大致把学生们分为三类。“一种是把课程学得很好的，把大数据的手段、方法学得很好的人。还有一种是传统专业学科的学生，学完大数据以后发现原来自己的专业还可以干这么多月的事，这就是一开始说的学科颠覆。”李老师用自己原本的专业水利举了例子，水利上有海量的数据，这些数据是有序、归类的，但由于仪器、时间等原因是非结构化的，原本无法分析，而大数据可以把海量的数据拿到一起进行分析，得出对一个大坝甚至整个流域整体的评价。“第三类就是新学科，它更是这样，大数据激发了很多活力。所以说学生们的视野得到了很大的开阔。”大数据的思维方式也给学生们打开了另一扇门。在大数据思维中，重要的在于通过相关性找到有用的价值。李老师引用了彭凯平（现任清华大学社科学院院长、心理系主任）当年的一句话：“大数据不是方法的导向，是思维的导向；不是管理的导向，是问题的导向。”

从当初一纸方案，多方奔走，到如今有广泛影响力的能力提升项目，数据院的人才培养并不会就此止步。“下一步，一定要这个项目走向国际化，”李老师说，“数据科学肩负着提高全球胜任力的使命，它是没有国界的。并且在清华走向世界一流大学的过程中，我们的项目可以大有作为。”