



数据与决策

□ 文 青

2022 年地平线技术报告上周新鲜出炉了，命名为数据与分析版，即重点预测大学里的数据和分析。报告提出六种关键技术和实践：

- 1) 数据管理与治理；
- 2) 统一数据来源；
- 3) 现代数据架构；
- 4) 开展数据素养培训；
- 5) 多样化、公平、包容的数据和分析；
- 6) 评价和改进大学的数据分析能力。

在我看来，报告提出的六种关键技术和实践，核心是第一种——数据管理与治理，其他五种是为实现这个目标服务的。

比如，要实现数据管理与治理，学校将面临两大挑战：一是全校各种数据的整合；二是各种数据结构需要统一。这就是第二、三种技术和实践的任务。大学数据虽然众多，但大都分散在各部门，像新生入学、学生资助、学生事务、师资力量等。有的可能还不存储在计算机系统中。这些都可能限制基于数据的分析和决策，都需要把它们整合起来。另外，大学使用的软件众多，每种软件的数据结构不同，有的可能已过时，很难满足当下机器学习、自然语言处理等要求，需要把他们更新，采用现代数据架构。没有足够数量、可以改动和灵活的数据结构，将很难对数据进行分析。

第四种显然是出于全面采集数据的需要，因为原有的数据标准大都根据多数人的特点建立的，这就会忽视某些人的数据，造成数据不全。重新思考谁选择数据，选择哪些数据，如何采集数据，做什么用，其中隐含什么偏见，是这一部分的任务。

基于数据决策不能仅针对领导层，还需要覆盖到广大老师和员工。这就需要开展员工数据素养培训——也就是第五种任务。只有广大教职员能从大数据里获得洞见，理解数据的涵义以及如何解释和使用数据，数据治理才是可行的。事实上，高校采集的数据数量和类型逐年增多，但是许多高校没有能力解释和使用这些数据的发现。

第六种是大学持续发展的需要。大学要提高和改进高校数据分析能力，获得高质量、有影响的洞见。

那么，大学为什么要推进数据管理与治理，原因正如报告指出的，一是基于数据决策是社会发展的趋势，二是技术发展为此提供了可能。总的目标是希望通过数据发现平时难以发现的洞见，使得决策更加科学有效。规划、创新、测试、运作和监控形成良性循环，推进大学持续改进和优化。

当然，这六种关键技术和实践仅仅是预测。要把这项庞大的系统工程从可能变成现实，显然需要大学领导层的远见，需要他们的魄力。它可能使得大学决策变得更加精准，也会挑战大学传统的决策惯习和文化。