

浦江创新论坛 研究报告

(2020 年第 7 期，总第 126 期)

上海浦江创新论坛中心

2020 年 12 月 8 日

2020 浦江创新论坛专题简报之七

把握云时代战略机遇 助推科技创新与发展

编者按：2020 年浦江创新论坛——新兴技术论坛 2 上，来自国内外的知名专家学者以“云时代的科学第四范式”为主题，围绕如何通过云计算等先进技术架构推动科学数据的开放和共享展开深入研讨，为云时代下的科技创新、企业发展提供全新的解决方案。本期基于新兴技术论坛 2 嘉宾¹报告整理而成，供参考。

¹ 与会嘉宾包括：国家科技基础条件平台中心主任苏靖，上海市科学技术委员会副主任傅国庆，中国科学院上海巴斯德研究所首席科学家、法国科学院院士、美国科学院外籍院士、德国科学院院士、瑞典工程院院士、英国皇家学会会士 Philippe Sansonetti，中国科学院院士、复旦大学常务副校长金力，中国科学院院士、中科院上海营养健康研究所首席科学家赵国屏，上海市研发公共服务平台管理中心副主任朱悦，上海科技创新资源数据中心 CTO 张嘉锐，北京青云科技股份有限公司 CEO 黄允松。

2020 浦江创新论坛专题简报之七

把握云时代战略机遇 助推科技创新与发展

科研 4.0 时代下，科学数据已成为国家科技创新和经济社会发展的重要基础性战略资源，科学数据的开放共享也上升为国家战略。云计算为科学数据的汇集和互通提供了良好平台，欧美发达国家已经把云计算作为国家科学基础设施，推动科学数据的共享和重复利用。与会嘉宾一致认为，科技创新进入了数据驱动的第四范式时代，要充分利用新一代信息技术，更好地挖掘、汇集和利用科学数据，推动科技进步和创新发展。

一、科技创新进入数据驱动的第四范式时代

一是数据密集型科学研究是大势所趋。国家科技基础条件平台中心主任苏靖指出，海量的科学数据把科学研究快速推进到一个前所未有的大数据时代，数据密集型科学研究第四种科研范式快速发展。上海市科学技术委员会副主任傅国庆指出，伴随以互联网、云计算和大数据为代表的数字经济的迅速发展，科学数据已成为传播速度最快、影响面最广、开发应用前景最广阔的科技资源，科技创新进入了数据驱动的第四范式时代。上海市研发公共服务平台管理中心副主任朱悦认为，科学数据作为科技创新的基石，已成为国家科技创新和经济社会发展的重要战略性资源。如何更好地挖掘、汇集和利用科学数据已经成为国家科技创新和经济社会发展的首要任务。上海科技创新资源数据中心 CTO 张嘉锐表示，数字化、网络化、智能化深入发展并正在改变所有人的生产和生活方式，科研领域同样如此。

二是数据驱动的第四范式呈现三大特征。特征一：海量数据。苏

靖指出，随着大科学装置的建设，重大科学实验的实施和无处不在的科学传感器、传感网络的广泛应用，源源不断产生海量的科学数据，科学数据已经成为科学研究的战略性、基础性资源。中国科学院上海巴斯德研究所首席科学家 **Philippe Sansonetti** 指出，所有的研究过程中都会生成大量数据，随着科技的发展和变革，获得数据的方式越来越便捷和多样化，数据处理方法也更为强大，使得每一次研究都发生质的飞跃。中国科学院院士、复旦大学常务副校长金力认为，第四范式下科学研究的第一步要通过“测量”获得大量数据，然后才能进行分析。中国科学院院士、中科院上海营养健康研究所首席科学家赵国屏指出，生物学经历了三次革命，伴随着每次革命，数据量在不断增长。21 世纪初的第三次革命，使生物学进入了生命科学、物理学和工程学的会聚时期，数据量已达到 TB 级。

特征二：关联共享。 **Philippe Sansonetti** 认为，数据交流互通十分重要，在生物学的传染疾病研究过程中，微生物整个发展的过程包括遗传学、化学、基因组等各方面的相互作用，所有的研究过程均会产生大量数据，需要相互之间的协同。北京青云科技股份有限公司 CEO 黄允松认为，数据如果不能共享、没有交叉，研究则没有意义，上海有很多团队从事微生物、病毒传染病研究，这类研究都是项目交叉的，此时无论公有云还是行业科研云的形态都变得至关重要。

特征三：高效发现。 金力认为，传统的科学研究是问题导向的，而大数据时代第四范式下的科学研究是产生问题的，核心是通过产生大量数据而产生问题，这两种范式能够相互补充，提高科研效率。黄允松指出，云计算提高了数据处理效率，研究者能更快速地从数据中发掘逻辑和关联性。

二、科学第四范式在生物医学领域中的体现

一方面，大数据有效支撑生物学领域前沿研究。**Philippe Sansonetti**指出，所有人体中都有微生物，且是微生物群落，存在于人体皮肤的各个地方，新的测序和新的生物信息方式可以帮助科研工作者了解整个微生物群落的情况，相关数据采集量越来越大，总体上的数据复杂度也越来越高。**赵国屏**指出，流行病学是研究人群中疾病以及健康相关状况及其影响与决定因素，并研究防治疾病及促进健康的策略和措施的科学。在流行病研究过程中的每个层次上都会产生大量数据，这些数据能够有效支撑研究工作。

另一方面，数据获取分析正全面渗透生命健康产业创新链。**金力**指出，新范式下的生命健康产业创新链主要包括：（1）以精密测量为导向，突破一批生命健康与临床诊疗关键技术，推动健康管理与医疗设备的研发；（2）通过对健康人群和疾病人群精密测量，发现全新生物表现标志物，研发新型诊断试剂和产品；（3）解析健康与疾病表型的遗传基础及致病机制，获得新型新药靶点及组合，为原创新药提供源头支撑；（4）挖掘表型组图谱大数据，发掘全表型组基因关联分析方法，形成系列健康管理数据产品和智能化辅助诊断系统。

三、新科学范式下生物学领域研究的相关建议

一要全力做好数据积累工作。从考虑应用目标和追求基本认识两个维度，**赵国屏**将生物医学研究划分为四大象限：（1）巴斯德象限的应用基础研究，既寻求扩展知识的边界，又受到应用目的影响的研究活动；（2）爱迪生象限的应用研究，单纯的应用研究，不寻求对某一学科的全面认识；（3）NCBI象限的数据/样本积累活动；（4）波尔象限的基础研究，只受认识需求推动，不受实际应用引导的研究活动。其中，数据/样本积累活动是基础性工作，真正需要国家支持才能做好。**金力**指出，在国家支持下推行国际大科学计划可以有效解决相关

数据收集和数据安全等问题。

二要重视新范式下的传染病学研究。 Philippe Sansonetti 指出，今年新冠疫情的爆发，提醒研究者们要更加注重微生物群落之间的作用，研究应对各类细菌和病毒。如果在感染发生时对病菌有充分了解并加以干预，效果会比较好。赵国屏指出，传染病防治体系中应当纳入传染病相关研究。在疾控、医疗、研究教育机构相关单元之间应建立起经常性的联系和有效的科研合作，在安全可控及绩效共赢的基础上实现信息与资源的交流共享。

三要避免“数据孤岛”现象。 Philippe Sansonetti 认为，在生物学还有传染疾病的研究过程当中生成大量数据，但大量数据因为各种原因丢失，这需要大家去共同支持、合作，也需要一个开放的科学与分享。金力指出，要建立数据标准，大数据的数据质量非常重要。表型大数据的获取、使用和共享涉及到很多问题，包括个人隐私、数据权益、人类遗传资源管理、网络安全等问题。通过国际大科学计划等国际间合作方式能够在一定程度上解决这些问题。黄允松指出，数字孪生技术可帮助研究者进行病毒传播过程的有效还原，而该技术的基础架构需要全球合作，基础架构一定是全球一体化的。

编辑：傅翠晓、姚景怡、毛朝梁