

# 未来实验室： 利用云技术 实现集成的 研究实验室





全球领先的研究实验室中的科学家，都在努力解决存储、分析、集成和共享数据集等方面的难题。数据是由各种类型的设备和制造商创建的，在许多情况下，这些数据仅存储在本地服务器上。大多数这种实验室都不是为了优化研究人员的日常工作或简化重复耗时的过程而建立的。然而，借助创新思维和基于技术的解决方案的帮助，一些研究环境正在处理和解决这些问题，打造完全集成的“未来实验室”，该实验室由数据驱动，能够优化研究人员在科学实验上所花的时间。

建立数字化实验室有助于保护研究人员生成的宝贵数据，还可以利用机器学习等新技术，加快研究和发现的过程。未来实验室将能够自动完成数据的备份和归档、快速数据分析、人工智能的集成，以减少研究人员在繁琐任务上所花的时间，并确保与其他研究人员的内部和外部协作的顺利进行。

行业数据显示，从 2005 年开始的 10 多年时间里，各个制药公司之间建立了 9000 多个新的合作伙伴关系，旨在通过协作加快新药的创新。<sup>1</sup> 与外部合作伙伴无缝集成，同时保护组织的知识产权 (IP)，对这些合作伙伴关系的成功至关重要。Amazon Web Services (AWS) 通过自动存储旧有设备的数据、确保数据共享安全性以促进协作，以及实施标准化的数据格式，使集成研究实验室成为现实。通过将数据安全地存储在云中，无论是与全球研究团队的协作、利用按需高性能计算进行分析，还是实施人工智能或机器学习以加快获取可行见解的速度，AWS 都让它们变得更加简单。



<sup>1</sup> Reh, Greg 等等。“合作是制药行业研发成功的关键。” 医疗保健现状, 2017 年 1 月 17 日 [www2.deloitte.com/us/en/pages/life-sciences-and-health-care/articles/health-care-current-january17-2017.html](http://www2.deloitte.com/us/en/pages/life-sciences-and-health-care/articles/health-care-current-january17-2017.html).

## 获取和存储数据的挑战

由于缺乏技术整合，领先研究实验室中的科学家每天都要面临突发状况的挑战。即使在最先进的实验室中，宝贵的实验数据也只能存储在与研究设备相连的 PC 或服务上，很容易遭到破坏或被删除。缺乏保护可能是由于组织没有认识到其专有数据的价值。为了让研究实验室取得成功，数据应该被视为一种新形式的货币，它可以随着新数据流的开发而不断增值，还可以用作历史观点来测试新的假设。对于这些数据可以为您的研究和整个组织实现的未来价值而言，保护它们是一种投资。

将数据迁移到云端是一种显而易见的解决方案，但很少有设备是云原生的，这对希望轻松组合和集成数据的科学家来说是一个挑战。AWS DataSync 提供了一种解决方法，能确保将重要的实验数据从设备写入的文件系统中自动复制，然后在 AWS 云中安全存储和存档。借助云中的数据，研究人员可以在几分钟内开始分析他们的关键研究，甚至可以将工作流程自动化，只要数据一到达，就立即启动分析或归档任务。



将数据存储在云中可以创建自动备份，让研究人员无需再归档数据。

“没有 AWS 中的数据，我们就无法以最快的速度进行创新。AWS 的混合存储和传输服务产品组合可与我们现有实验室的计算环境和流程协同工作，帮助我们将不可替代的数据安全导入 AWS，在那里，科学家可以将其用于任何需要的地方：机器学习、数据分析或 HPC。” Celgene 研究计算主管 Lance Smith 说道。<sup>2</sup>

以这种方式存储数据，不仅能将数据更安全地存储在云中，还能防止由于设备故障而造成的意外丢失。设备或本地服务器上存储的数据可能无法长时间备份，并且，电源故障或这些系统的损坏可能会导致宝贵的数据丢失。将数据存储在云中可以创建自动备份，让研究人员无需再归档数据。

展望未来，市场上的新设备正在努力实现数据的无缝整合，这些设备包括：

- Illumina 的 BaseSpace Sequence Hub，它可以对从设备到基于 AWS 云的应用程序之间的数据流加密，让协作和数据处理变得更加便捷。<sup>3</sup>
- Thermo Fisher Connect，可连接实验室设备、自动执行数据备份、与实验室管理软件集成并提醒远程工作的科学家。<sup>4</sup>

研究人员面临的另一个常见挑战是整合来自不同制造商的设备的数据。各个工具生成的数据都采用了不同的格式，如果不做准备工作，将各个数据集统一成一种格式，就很难进行集成和分析。

2 Duso, Wayne, “通过 Snowball Edge 上的块存储扩展 AWS 混合云功能”。AWS Storage 博客, 2019 年 4 月 29 日。 <https://aws.amazon.com/blogs/storage/expanding-aws-hybrid-cloud-capabilities-with-block-storage-on-snowball-edge/>

3 Illumina BaseSpace Sequence Hub. 可在 <https://www.illumina.com/products/by-type/informatics-products/basespace-sequence-hub.html> 获取

4 Digital Science | 美国 Thermo Fisher Scientific. 可在 <https://www.thermofisher.com/us/en/home/digital-science.html> 获取

30%



通过合作伙伴关系进行创新的潜力已在业内广为人知，数据显示，大型制药公司目前正在开发的药品中有 30% 最初是由另一家公司开发的。

研究分析公司首席执行官 Savitra Sharma 说道：“大多数台式分析仪是由不同的供应商生产的，它们生成的测试结果数据都采用了不同的格式。”“为了合并数据，实验室技术人员必须手动处理这些数据，采用统一的格式。然后，需要将这些数据手动上传到实验室信息管理系统 (LIMS) 中。”

AWS 和 TetraScience 等 APN 合作伙伴可促进来自各种数据源的数据集成，并将数据标准化为与供应商无关且可随时进行分析的格式，消除了传统研究实验室中发现的数据孤岛。<sup>5</sup> 这可让研究人员更有效地专注于数据分析，而不是重新格式化从各种设备收集来的数据。

### 全球数据共享和高级分析

通过全球数据共享进行协作是推动研究发现的关键，制药公司和公共研究实体都在通过加大协作投入来推行这一理念。通过合作伙伴关系进行创新的潜力已在业内广为人知，数据显示，大型制药公司目前正在开发的药品中有 30% 最初是由另一家公司开发的。<sup>6</sup>

但是，作为一种新形式的货币，数据的价值越来越明显，因此，共享敏感的研究数据就要格外谨慎，以确保数据的安全并保护知识产权。使用闪存或硬盘的数据共享方法无法扩展到如今研究中很常见的大型数据集，在远距离协作中也不容易实现，还带来了数据安全风险。基于云的数据存储提供了一种经济高效、可扩展且安全的现代替代方案。一旦数据安全转移到云端，AWS 就可以通过允许主要组织为来自不同组织的研究人员创建单独访问权限来促进跨国合作。AWS 正在为许多大型研究财团提供技术支持，例如 CHARGE，<sup>7</sup> 在这个组织中，来自五个全球机构的 200 多名科学家协同工作，努力识别可能与心脏病发展有关的基因。CHARGE 项目中的数据量超过了 450TB，Baylor 首席程序员科学家 Narayanan Veeraraghavan 说道：“给这么多人发送硬盘将是后勤部门的噩梦。”“数据在任何时候都必须加密。如果有这么多的科学家处理那么多的硬盘，就会出现很多问题，因为不是每个人都能遵守安全准则。”

5 Amazon Web Services. “来自启用了云技术的实验室设备的数据集成。” YouTube, YouTube, 2019 年 2 月 6 日, [www.youtube.com/watch?v=F6lstTXFBjI](https://www.youtube.com/watch?v=F6lstTXFBjI).

6 Buvailo, Andrii. “医药研发外包正在兴起。” *BioPharmaTrend*, 2018 年 8 月 13 日. [www.biopharmatrend.com/post/30-pharma-rd-outsourcing-is-on-the-rise/](https://www.biopharmatrend.com/post/30-pharma-rd-outsourcing-is-on-the-rise/).

7 Baylor 案例研究—Amazon Web Services. 可在 <https://aws.amazon.com/solutions/case-studies/baylor/> 获取



基于云的合作还有一个额外的优势，那就是对资源有限的研究合作者来说，这是一种经济高效的合作方式。投资购买更好的计算机和支持性基础设施成本高昂，对于需要快速行动或资金有限的合作者和承包商来说，会限制他们做出重要的贡献。无论在哪里，每个贡献者都可以随时使用云的按需计算能力。

强大的云计算是一个关键组件，它让参与协作的研究人员能够并行运行数百甚至数千个程序、计算或模拟，并快速获得结果。Celgene、Bristol-Myers Squibb<sup>8</sup> 和 Fabric Genomics<sup>9</sup> 等公司已经在使用 AWS 云来实现高性能计算，帮助自己加快研究或临床分析的速度。

### 可推动行动的智能处理方式

随着集成式数据流水线成为全球高级研究实验室的标配，就能够更快地将原始数据转变为洞察信息并采取相应行动。通过在云中更安全地访问数据，我们可以结合使用机器学习等其他技术来补充科学家和研究人员的日常工作。如果采用能够分析数据、预测趋势、提醒研究人员注意有待进一步分析的问题，或能够执行简单但必要的日常任务的技术，就能提高实验室环境的效率。

Celgene 开发的 LabAlert<sup>10</sup> 之类的程序可以连续监控设备和应用程序的状态，让科学家能专注于其他增值任务。利用 Amazon Simple Notification Service (SNS) 和 AWS Lambda，研究人员可以在移动设备上收到设备是否出现故障，或离开后实验是否完成等通知，从而让数据的收集更加可靠并减少对人工干预的依赖。

8 “AWS 案例研究：Bristol-Myers Squibb。” Amazon, Amazon, <https://aws.amazon.com/solutions/case-studies/bristol-myers-squibb/>  
 9 “Fabric Genomics 案例研究 - Amazon Web Services (AWS)。” Amazon, Amazon, <https://aws.amazon.com/solutions/case-studies/fabricgenomics/>

10 Bookbinder, Maxine. “Celgene 的 LabAlert 系统专注于科学工作流程效率、用户体验。” *Bio-IT World*, 2018 年 6 月 18 日。 <http://www.bio-itworld.com/2018/06/18/celgenes-labalert-system-focuses-on-scientific-workflow-efficiency-user-experience.aspx>

同样，AstraZeneca<sup>11</sup> 在 AWS 上使用机器学习来帮助实现组织数据标记工作中一些最繁琐部分的自动化，从而帮助他们将研究人员花在样本分类上的时间减少了 50%，让科学家可以将更多时间集中在有价值的研究活动上。

### 未来该何去何从？

基于云的技术可以提供简化的方法来访问、标准化和共享设备数据，同时提供经济高效的强大计算和分析能力。随着人工智能和机器学习进一步融入现代研究实验室，科学家可以摆脱额外的单调任务，专注于增值研究活动。未来的创新可能包括：

- 使用机器学习来加快编目、分析或确定实验结果优先级的速度，并提出未来的行动方案。
- 机器人技术和自动化技术的广泛应用，让批量精准医疗生产成为可能。
- 实现整个组织范围内的数字化和数据集成，以连接从研究到生产再到商业化等典型孤立部门。

无缝集成研究、协作促进以及将机器学习纳入“未来实验室”的一部分，让药学研究人员能够每天加快科学和医学发展的步伐。对于需要全球合作来推动研究发展，以便在某一天能够解决重大人类健康问题的行业来说，改变数据的收集、存储和共享方式，是在这个行业中取得发展的明确途径。

11 Wired Insider. “机器学习正在改变 AstraZeneca 的药物研发。” *Wired*, 2019 年 7 月 2 日, [www.wired.com/brandlab/2019/07/machine-learning-is-transforming-drug-development/](http://www.wired.com/brandlab/2019/07/machine-learning-is-transforming-drug-development/).

十三年来，Amazon Web Services 一直是最全面、应用最广泛的云平台。AWS 在 22 个地理区域的 69 个可用区中提供了广泛的服务，包括计算、存储、数据库、分析、网络、机器学习、管理工具、IoT、安全性和企业应用程序。AWS 得到了全球数百万活跃客户的信任，包括发展最快的生命科学初创公司、大型企业和领先的政府机构，为其基础设施提供动力，帮助他们提高敏捷性并降低成本。

要了解有关 AWS 的更多信息，请访问 <https://aws.amazon.com/health/biotech-pharma/>