



European Space Agency (ESA) + InterSystems

## 处理数据，绘制银河系地图



Eduardo Anglada,  
来自欧洲航天局 (ESA)  
和欧洲太空天文学中心

来自欧洲航天局 (ESA) 和欧洲太空天文学中心 (ESAC) 的 Eduardo Anglada 目前正效力于“盖亚”任务，以绘制银河系中的 10 亿个天体。Anglada 向来自 InterSystems 的 Jordi Calvera Sagué 解释了对 10 亿个天体进行分类所涉及的数据量之大，以及支持这项地外任务对技术的要求。

欧洲航天局 (ESA) 的任务是探索宇宙，它有一支由 10 艘飞船组成的舰队在太空中完成这项任务。“盖亚”太空任务始于 1993 年，航天器于 2013 年发射。“盖亚”的目标是为 10% 的银河系绘制精确的三维地图，在此过程中，对多达 10 亿天体进行“定性”。

Eduardo Anglada 是欧洲航天局 (ESA) 的计算机分析师和网格工程师，他解释道，所谓“定性”，是指他和他的团队要绘制这些天体的位置、速度、物理参数和温度。

他们还希望能够回答与天体的金属度和年龄有关的问题，以及它们是否是双星系统（两颗恒星靠得很近）。这一过程将揭示星系的构成、形成和演化。

正如人们可以想象的那样，这些数字是惊人的。这颗耗资近 7 亿欧元的卫星距离地球 70 万公里 ~150 万公里。赌注很高，因为硬件、软件和数据的安全性都必须无懈可击，因为这样的距离意味着无法对卫星进行维修或补充燃料。从根本上来说，从这颗卫星上下载的数据是不可替代的。

我们已经分析了超过一万亿次的观测结果。这个相当大的数字要归功于 Caché。

*Eduardo Anglada,*  
来自欧洲航天局（ESA）和  
欧洲太空天文学中心

飞船上相机的像素是 9.38 亿，有 106 个电荷耦合器件（CCD），这是大多数数码相机使用的图像传感器类型。“这是有史以来最大的相机之一。” Anglada 说。这颗卫星绕其轴旋转一周需要 6 个小时，每天发送 45 到 100 千兆字节的数据，拍摄约 7000 万张照片。“数据一旦传回地球，我们就会在 24 小时内进行分析。”

来自相机的数据每天都会被下载下来，并由位于西班牙、澳大利亚和阿根廷的三个天线进行接收。这些不同的天线，令欧洲航天局（ESA）具备了一种“跟着太阳走”的工作方式，这样一来，卫星始终能够被其中一个国家的卫星持续跟踪到。“如果我们失去了它，那将是一场灾难。仅需几个小时，它就可能会远离它应该在的位置很多很多公里，而且很难再次追踪并找到它。” Anglada 解释道。

数据从这些位置被发送到德国的任务运营中心，工作人员在这里检查卫星是否有问题，如果有问题就修复。然后，压缩后的数据被发送到位于马德里的欧洲航天局（ESA），观测数据会被解压缩、校准，然后被存储在缓存数据库中。

在此之后，数据会被发送到数据专家中心，进行模拟和对象、光度和光谱处理等等任务。然后再返回到位于马德里的中央数据库。

该系统以轮辐模式建立，以便满足所有研究人员访问数据的需求，以达成其目的。辐条分布在欧洲各地，包括剑桥、都灵和图卢兹。

数据处理分为四个阶段，包括：日常和周期操作，主数据库，校准活动和有效载荷指挥，最后是开发。

Anglada 说，现在每天的数据库都相当大，“差不多有 30TB。这还只是一个单独的实例。我们的大型服务器有 1.5TB 的 RAM 和 7TB 的固态硬盘。”

周期操作每 4~6 个月完成一次。在这段时间内，数据中心必须完成数据处理，并开始将数据发送到其他数据中心，以便改进结果。

InterSystems Caché 是主数据库和日常操作的一部分。航天局与 InterSystems 区域总经理 Jordi Calvera Sagué 取得联系，称其面临着在短时间内插入和处理大量数据的挑战。航天局向 InterSystems 提供了一些虚拟数据，软件工程师在三天内对数据库缓存进行了配置。他说：“在概念验证后，我们改变了架构，我们仅做了配置，但它已经变得更快了。”

除了 InterSystems Caché 之外，欧洲航天局（ESA）还使用了 Aspera 在数据处理中心之间进行分发，并使用 Atlassian Jira 进行漏洞跟踪。欧洲航天局（ESA）的团队由 400 名来自欧洲的人员和 26 名来自欧洲太空天文学中心（ESAC）的人员组成，他们具备从校准到管理再到日常下载等不同领域的专业知识。

那么，InterSystems 的吸引力是什么？Anglada 认为，它非常可靠，而且 InterSystems 能够提供全面的支持。“在过去的 11 个月里，我们没有遇到任何问题。我是一个非常精通技术的人，也是日常操作团队的一员，InterSystems 能够为我们提供迄今为止得到的最好的支持。”Anglada 补充道，他能够非常容易地联系到客户经理并了解该工具是否能够切实满足新要求。“项目一旦成熟之后，一家公司能够与客户继续保持这样的合作非常罕见。”他补充说，Caché 非常健壮，日常操作检查可以在几分钟内就完成。

如前所述，这个项目涉及到一些大数字，对 Anglada 来说，其中一个数字尤其令人印象深刻。“我们已经分析了超过一万亿次的观测结果。这个相当大的数字要归功于 Caché。它可以毫无问题地处理 30TB 的数据库。”InterSystems 甚至延长了这一项目。Calvera Sagué 说，这颗卫星经过优化后，任务可以比预期延长两到三年。

这项任务于 2016 年首次发布目录，并绘制了 300 光年的地图，这只是银河系的一部分。在 2018 年 4 月的第二次发布中，绘制了大约 8000 光年的地图。“盖亚”团队记录了 17 亿颗恒星的亮度，以及和 14 亿颗恒星的表面温度。

这种详细程度，意味着有一些令人惊讶的发现。“拥有速度意味着我们可以研究星系的运动学。例如，银河系在许多年前与另一个星系发生冲突，我们已经看到了这两个星系合并的最后阶段。这些多年来已经为人所知，但到目前为止还没有人能测量到它。有了这些数据，这就成为可能。”

此外，去年，人类探测到第一颗星际彗星奥陌陌（Oumuamua）。“这是在太阳系中探测到的第一个不属于太阳系的天体。这颗彗星在太空中运行了大约 6000 年，它进入太阳系的速度如此之快，以至于太阳无法捕获它，但太阳的引力改变了它的轨道。”

第三次发布可能会在 2021 年上半年发布，但最终发布的日期尚未确定，最终发布将包括完整的天体测量、光度和径向速度目录。Anglada 说，到目前为止，天文学界对这项任务的结果非常满意，平均每天有三到四篇文章会引用此目录，“这在科学界来说是惊人的成就。”

“现在一切都很物质化。”安格拉达说，“每当科学家意识到我们可以测量一些新的东西时，与之相应的发展就带来新的想法。有了这个目录，科学界将能够在未来 30~40 年内研究他们想要的这些恒星的物理特性。”

