

725.4T Token:首个“词元”计量数据产权登记落地

中经记者 许璐 李晖 北京报道

当人工智能产业开始以Token计算调用量、成本和商业回报时,数据资产登记环节也开始引入

新的计量口径。

近日,青岛华大基因研究院“青岛华大千种海洋生物基因组测序数据资源集”在青岛数据集团所属青岛数据资产登记评价中心完成

数据产权登记。该数据资产主体为青岛华大基因研究院,计量单位为“词元”,词元规模为725.4T Token。

对于词元计量的作用,青

岛华大基因研究院副院长徐梦阳在接受《中国经营报》记者采访时表示,词元计量剔除了无效的“噪音”,真正反映了数据中蕴含的“语义有效性”和知识

密度。

“我们能够清晰评估出7000余万元的市值,关键在于在计量中提供了‘词元’这一颗粒度极细的核算基础。可以准确计算

每一个基因特征从测序、组装到功能注释的全流程投入,并结合其在生物医药靶点、合成生物学元件等应用场景中的预期收益贡献进行折现。”徐梦阳说。

从“条数”“容量”到“词元”

以词元为计量单位,代替传统的条数、字段数、存储容量等概念,直接体现基因序列数据的信息密度。

在传统数据资产登记、流通和交易中,结构化数据多以条数为计量单位,文本、图像等非结构化数据则以文件大小,即MB、GB等存储容量为计量单位。

“对基因测序数据之类的非结构化数据而言,传统计量方式难以反映其作为AI‘燃料’的真实颗粒度。”徐梦阳指出。

词元是大模型处理信息的基本单位。国家数据局局长刘烈宏在2026年3月份的“中国发展高层论坛2026年年会”中明确“Token”的中文名为“词元”,具有可计量、可定价、可交易等特征。不仅是智能时代的价值锚点,更是连接技术供给与商业需求的“结算单位”,为商业模式的落地提供了可量化的可能。

从产业数据看,词元调用规模正在快速增长。公开数据显示,2024年年初,我国日均Token调用量为1000亿;到2025年年底,跃升至100万亿;到2026年3月,已超过140万亿,两年增长超千倍。自2026年1月底以来,有的模型企业创下20天收入超越2025年全年总收入的业绩纪录。

在此次登记中,基因序列与词元之间建立了具体对应关系。对于此次以“词元”为核心计量单位进行登记的计量方法,青岛数据集团首席数据官赵传启表示:“在本次华大基因的登记中,以词元为计量单位,代替传统的条数、字段数、存储容量等概念,以1个碱基对对应1个词元(1bp=1Token),直接体现该数据的信息密



视觉中国/图

度,衡量出的总词元规模为725.4T Token。”

在徐梦阳看来,基因序列中的A、T、C、G本质上是大自然写下的代码。A、T、C、G分别对应DNA遗传序列中的腺嘌呤、胸腺嘧啶、胞嘧啶、鸟嘌呤四种基本核苷酸。

“我们将1个碱基对直接映射为1个词元,相当于把青岛华大全球最大海洋基因库里的数据,直接转化为AI模型可以‘阅读’和‘理解’的标准化语料。例如,本次入表的大生物基因组,就包含了约1600亿个词元。选择‘词元’作为单位,是因为这能让我们的海洋生物数据无缝对接全球顶级的人工智能计算范式。”徐梦阳表示。

词元计量进入登记环节后,也对评价模型提出了适配要求。赵传启表示,为适配词元和基因数据,相关评价指标和算法进行了优

化,引入了BUSCO完整度、QV值、ContigN50等指标衡量基因组数据质量,并以词元计量基因数据的信息熵和各物种在总样本中的权重。

对于此次数据登记完成的意义,赵传启表示,词元计量方法,能够解决不同格式、不同行业的数据各有各的计量标准的问题,为结构化数据、文本、影像视频、生物基因等非结构化数据,多模态数据提供统一的度量衡,并适配大模型、智能体的计量方法,为数据赋能大模型应用进一步铺平路。

在一些业内人士看来,基因数据是非结构化数据中专业门槛较高、信息密度较大的类型之一。此次案例如果能够在确权登记、计量评价、授权使用和收益分配等环节形成可复制路径,将为医疗、教育、传媒、工业、交通等领域的非结构化数据资产化提供参考。

词元计量进入数据要素流通链条

“按量授权、按词元计费”的微服务生态,大幅降低了中小型创新企业和科研团队获取顶尖海洋数据的门槛。

从数据流通链条看,产权登记解决的是“数据是谁的、能否登记、如何描述”的问题;计量评价解决的是“数据有多少、如何衡量”的问题;交易运营解决的是“数据如何使用、如何定价”的问题;收益分配解决的是“使用后如何分账”的问题。此次案例中,词元计量被用于产权登记、评价模型、贡献度核算和收益分配等环节。

公开信息显示,此次登记的数据资源源自青岛自贸片区“千种海洋生物基因组测序”项目。2023年9月,青岛华大基因研究院承接该项目后,依托自有高通量测序平台完成本土物种测序,同时整合NCBI、EBI、JGI等国际公开数据库资源,经过数据清洗、组装、标注和全流程质检,最终形成覆盖上千种海洋生物基因组信息的数据资源集。

对于该数据资源集从原始测序到最终形成的过程,赵传启表示,在数据采集处理环节,主要涉及样本采集与DNA提取得到原始的DNA序列,然后进行数据清洗、组装、标注和校验,形成最终的基因组。

“在权属上,华大基因研究院自主开展加工处理,形成最终的基因组数据,并对此拥有持有权、使用权和经营权。同时,委托第三方律师事务所对公司主体、数据来源、数据内容、加工处理过程进行全面的合规审查,并确认其在此过程中实质性投入了大量资

金、技术、人力、算力、场地等资源。”赵传启进一步表示。

除登记计量外,词元也被用于科研协作中的贡献度核算。“从科研创新的角度看,过去科学家之间的合作,数据的提供方和使用方很难精准界定各自的贡献比例,这也是制约数据开放共享的一大痛点。我们首创了‘数据量占40%+稀缺性占30%+应用价值占30%’的贡献度评分模型,其中数据量维度的基础就是词元数量。这样一来,无论是与国内机构联合开发新酶,还是在国际海洋基因组学联盟框架下协作,每一份智力投入都能被量化,收益分配变得透明公平。”徐梦阳说。

对于该数据资源集后续是否进入数据交易、授权运营或产业应用环节,赵传启表示,该项数据可应用于蛋白质预测、工业菌株基因编辑靶点挖掘、合成生物学元件设计、濒危物种遗传多样性评估和海洋生态保护等多类场景,涉及数据交易和产业应用,也可以由华大基因授权第三方开展运营。

从流通方式看,徐梦阳认为,词元计量可以驱动“按需调用”的数据流通,孕育海洋科研的“微服务”生态。

“在流通使用上,词元计量带来了前所未有的灵活性。一位寻找新型抗菌肽的研究者,完全可以精准调用、只为他所需要的那部分代谢通路基因词元付费,无须购买其他冗余信息。这种‘按量授

权、按词元计费’的微服务生态,大幅降低了中小型创新企业和科研团队获取顶尖海洋数据的门槛。”徐梦阳说。

此次登记完成后,生物制造、海洋药物和生态保护等都是海洋生物基因数据的后续使用场景。

对于海洋生物基因组数据的产业价值体现,徐梦阳表示,在生物制造领域,已有耐高温测序酶、高活性漆酶等直接应用于环保和工业生产。在海洋药物方面,数据集提供了大量抗癌、抗衰老的先导化合物线索,可以帮助合作方缩短新药筛选周期。在生态保护上,它提供了一套高分辨率的基线,可以精准监测江豚等濒危物种的动态。

围绕高价值遗传数据流通,数据安全也是后续产业应用中的重要环节。徐梦阳表示,青岛华大基因研究院设计了一套安全的输出机制:不交易原始数据,而是交易基于数据训练出的AI模型能力。

“在资产化的第一步,我们已经剥离了物种的地理位置、生物属性等标注信息,将数据彻底转化为纯粹的生命语言片段,这在技术上是完全不可逆转、不可追溯的。对外合作时,青岛华大基因研究院交付的是像‘ACCESS蛋白功能预测模型’或‘ThermoMod酶最适反应温度预测模型’这样的AI模型,而不是原始序列。大模型本身的‘黑盒化’特征,构成了一道天然的安全护城河。”徐梦阳说。

一线调研

百亿信贷稳产增效 金融“守护”海南橡胶产业

中经记者 杨井鑫 海口报道

天然橡胶、钢铁、煤炭、石油并称世界四大工业原料。其中,

天然橡胶广泛用于交通、医疗、军工等领域,是重要的战略物资。

我国是全球最大的橡胶消费国和进口国。多年以来,国内

1800万亩的橡胶种植区一直发挥着“保安全”的作用。而海南作为橡胶的主要产区,种植面积高达近755万亩,更是在保供上起到了

关键作用。

然而,天然橡胶普遍8年才进入旺产期,漫长的种植周期与大规模的种植都需要金融的持续支

持。同时,海南部分东北部地区不时会有台风侵袭,危害橡胶产业发展。

目前,海南橡胶产业种植情

况如何?金融又是怎样支持海南橡胶发展的?《中国经营报》记者赴海南橡胶种植基地进行了实地调研。

产业面临周期低谷

在海南省农垦投资控股集团(以下简称“海垦集团”)的金鸡岭种植基地,橡胶林中胶工们正忙碌着割胶。胶刀划过树皮,洁白的胶乳缓缓渗出,沿着胶舌滴入胶杯。

“一株胶树每四天能割一次胶,大概四个小时能收集到满满一盅。”金鸡岭种植基地的一名胶工表示。

她向记者表示,一名胶工负责2000株胶树的割胶工作。“每天凌晨2点就要赶到胶林工作,挣的是一份辛苦钱。”

据介绍,胶工是橡胶生产的核心力量,但收入长期偏低。如今,橡胶市场价格正处于行业周期的上升期,橡胶产量和胶工收入都十分可观。

在海南东北部,台风的时常侵袭让当地橡胶产业发展面临挑战。前述胶工告诉记者:“部分胶树被摧毁了。以前一亩胶林产胶的胶树有33株。上次台风过后,一亩地里产胶的胶树锐减到12株。”

为了保障橡胶的产能,金鸡岭种植基地需要对橡胶林进行补种,而这需要一大笔资金。金鸡岭基地分公司负责人表示:“海垦集团旗下的海南橡胶主营橡胶种植,有18家基地分公司。金鸡岭基地分公司所辖范围有三个橡胶种植的片区,土地面积共计13万亩。此前台风过

后,金鸡岭片区的橡胶林一直在休割和重新育苗,目前产能正逐渐恢复中。”

“在海南,中西部地区受自然灾害影响要小,橡胶种植整体在向中西部转移。但海南中西部的土地资源有限,产能进一步增长也受到限制。”上述负责人对记者表示,海南东北部发展橡胶产业并不占优势,但是海垦集团金鸡岭基地分公司的主责主业在此,同时担负着国家橡胶产业保供的重任。“难,但还是要继续发展橡胶产业。”

据了解,目前海垦集团金鸡岭基地分公司的橡胶干胶年产量总体保持稳定。

“目前,橡胶干胶一吨的价格大概在1.7万元左右,而2005年至2008年橡胶干胶的市场价格超过4万元一吨。”实际上,在2024年台风过后,2025年海南全岛天然橡胶产量大幅下滑。但是,金鸡岭基地分公司攻坚克难,保持了经营成果的稳健。

对于当前橡胶产业的发展模式,该负责人介绍称,橡胶种植由海垦集团金鸡岭分公司负责。在产胶阶段,胶工割胶后将橡胶原料交给分公司。金鸡岭基地分公司主要负责种植管理,在收购橡胶后运往集团的加工企业进一步加工。

百亿资金“驰援”产业

“在海垦集团橡胶种植业务受到台风侵袭之后,农业银行(601288.SH)紧急投放了7亿元资金,支持企业补种橡胶苗。”农业银行海口海秀支行彭经理表示。

他向记者介绍,作为海南省省属最大的国企,海垦集团自1952年成立之初就与农业银行建立了紧密合作关系,合作时间长达70多年。

据了解,海垦集团是全国第三大农垦,仅次于新疆生产建设兵团农垦和黑龙江农垦,公司业务涉及八大板块,涵盖了天然橡胶种植、加工与贸易、南繁育种、热作农林、畜禽海产、冷链物流等,并拥有进出口经营权。海垦集团的农业板块收入占集团总收入的86.1%,是集团的核心支柱。其中,天然橡胶产业是集团的核心产业,拥有A股唯一橡胶上市公司——海南橡胶(601118.SH)。海南橡胶经营管理土地490万亩,境内外天然橡胶生产基地30家,胶园土地380万亩,约占全球的2%。2025年,其天然橡胶年加工量131万吨,年销售贸易量337万吨。

农业银行海南分行相关负责人表示,在海垦集团改革发展的

70年进程中,农业银行始终相伴左右。仅2004年海垦集团启动改企、社企分离改制一项,农业银行就主动与改制企业协商,盘活了30家企业8.4亿元不良贷款,助力海垦集团产业和市场化改革。此后,农业银行又全力推动海南橡胶成功上市,协助海垦集团财务公司搭建资金归集平台,一步步陪伴企业走出改革深水区。

据介绍,2012年、2016年、2025年,农业银行海南分行与海垦集团三次签订全面战略合作协议,意向性信用额度从150亿元提升到300亿元,合作领域也从传统信贷扩展到投行、国际业务等新赛道。近5年来,农业银行海南分行累计向海垦集团授信超200亿元。

值得一提的是,农业银行海南分行还积极帮助海垦集团“走出去”。2023年,农业银行海南分行落地海南省首笔“FT+”自贸账户项下跨境并购业务,投放5亿元助力海南橡胶开展海外并购;2025年,农业银行海南分行成为海南橡胶海南自贸港跨境资金集中运营中心的合作银行。在债券市场,农业银行海南分行牵头为海垦集团及海南橡胶注册了85亿



图为金鸡岭基地分公司一名胶工正在割胶工作。

本报资料室/图

元中期票据和60亿元超短融,主承销发行33.5亿元,创下全国首单乡村振兴数字人民币中票、全国农垦首单科创债等多个“首单”纪录。农业银行海南分行相关负责人表示,2026年计划进一步扩大对海垦集团的授信,并加深与该集团各个子公司的合作。

“天然橡胶是国家重要战略资源,产业向‘高质高效’转型过程中,在培育良种、提升产能、人才培养以及国际化布局上都离不开金融支持。”该负责人表示,在金融“活水”浇灌之下,这片承载着战略使命的橡胶林,必将焕发更强生机。