

资源聚合 积累数据势能

致谢

本研究是由普华永道团队、浙江大数据交易中心、全球数据资产理事会和海亮教育科技服务集团联合完成,在此对给予研究指导以及研究项目组成员致以感谢。

普华永道

詹睿
普华永道中国合伙人

侯建
普华永道中国高级经理

倪晨阳
普华永道中国高级顾问

浙江大数据交易中心

孔俊
浙江大数据交易中心总经理

刘梦迪
浙江大数据交易中心总经理助理

全球数据资产理事会

林建兴
全球数据资产理事会总干事

海亮教育科技服务集团

项灵刚
海亮教育CIO兼数字化中心总监

游军杰
海亮教育数字化中心副总监

谭立何
海亮教育数字化中心部长

郭乃菁
海亮教育数字化中心部长

梁宇
海亮教育数字化中心副部长

杨琳
海亮教育数字化中心副部长

贾竞源
海亮教育数字化中心资深经理

王莹
海亮教育数字化中心高级经理

汪文佳
海亮教育数字化中心高级经理

张北辰
海亮教育数字化中心中级经理

目录

前言	01
教育行业数智化概述	02
教育行业数字化现状	03
教育行业数字化面临的挑战	04
教育行业企业内部共享机制的构建	07
为什么要聚合数据	08
教育行业数智大富翁之聚合机制	10
基本概念介绍	10
交易机制	10
案例:海亮教育生涯规划大模型开发中“大富翁游戏”的应用	12
以收益/价值分配为目的的敏捷确权体系	13
数据确权的重要性	13
内部共享中确权的应用:特许经营权模式	14
案例:海亮教育生涯规划产品开发中确权的应用	15
高效的数据资源估值定价体系	16
探索统一的“标尺”	17
揭秘“标尺”的关键度量指标	18
“标准计量单位”的诞生	20
“海亮贝”的诞生	22
基于“市场规模”的定价	23
案例:海亮生涯规划产品在“大富翁游戏”下的估值以及定价	25
结语	26
参考文献	27
联系我们	28



前言

近年来，国家积极布局数据要素市场，一系列的数据要素政策不断出台，市场百花齐放，从体制、流通、标准、定价等方面开展研究和探索，掀起一波又一波的数据资产化浪潮。然而，对于非数字原生企业^[1]来说，顶层制度再灿烂，他们依然在数据实践中面临重重阻碍，陷入“从开始到暂停，从暂停到放弃”的恶性循环中。海亮教育作为教育行业非数字原生企业，近年来不断在数字化转型之路践行，以“提高市场分析能力，客户分析能力，优化提升产品服务水平，增强客户粘性，探索二次曲线的增长”为宗旨，探索并实践数据“聚类分析”的价值，从集团内实现数据聚合、到以“价值实现为导向”的数据资产建设，至“切实可行”的数据资产运营机制，逐步建立起一个健康、有序且平衡的数据生态。在此，我们也希望将我们的数据价值实现之路与业界进行分享，以期为在“数智化转型”过程中因为实践迷茫而踌躇不前的同仁们，提供一些我们的管理经验和劳动成果，贡献一份绵薄之力。

本次发布的白皮书为数据资产化系列白皮书之一，将以数据“聚合”为主题，通过分享海亮教育在建立数据内循环体系，以实现数据资源共享的实践经验，逐步解锁行业性的在数据资源聚合（内部共享）过程中如数据资源确权、定价、运营等一系列痛点。旨在深入探讨教育行业数据资源的聚合机制与策略，以及如何通过有效的数据资源聚合积累数据势能^[2]，同时为教育行业的数据资源聚合提供一套可操作性强、具有实践意义的指导方案。

[1]数字原生企业出自《华为数据管理之道》，解释为企业设立之初基本以物理世界为中心构建的，创建时围绕生产、流通、服务等具体的经济活动展开。

[2]数据势能是普华永道在《开放数据资产白皮书》中首先提出的概念，势能是指还未释放出来的能量，当势能积攒到一定的程度，通过适当的方式就可以释放或者转化为其他的能量，对于数据，这种能量就是价值。

教育行业 数智化概述



1 教育行业数字化现状

在教育总体战略的变迁和发展下,我国教育信息化已经走过20余年,大致经历了以下发展阶段:建设驱动阶段(2000—2010年)、应用驱动阶段(2011—2017年)、融合创新阶段(2018年至今),实现了跨越式发展和历史性突破,逐渐迈向教育数字化转型新阶段。2022年10月,党的二十大报告首次将“推进教育数字化”写进党代会报告,以及2024年全国教育工作会议强调开辟教育数字化新赛道,我国教育行业正式进入了数字化转型的新阶段。这一阶段不仅是对前期教育信息化成果的巩固与深化,更是对教育发展模式与治理体系的全面革新。

教育数字化转型基于教育信息化,又高于教育信息化。它不仅包含了教育信息化的所有内容,还进一步强调了数据的收集、分析、挖掘和利用,充分利用数字技术推动教育业务和教育发展从“物理世界”到“数字世界”的重塑和再造,实现教育全要素、全流程、全业务和全领域的数字化。

从教育信息化到教育数字化的转变是一个复杂而深刻的过程,它涉及到教育理念、教学模式、管理方式等多个方面的变革。在转变过程中,其特性体现在多个维度的深刻变迁(图1):



图1 教育信息化到教育数字化转变在多个维度的特性体现

2 教育行业数字化面临的挑战

为响应习近平总书记提出的教育数字化是我国开辟教育发展新赛道和塑造教育发展新优势的重要突破口这一指导思想，海亮教育致力于通过数字化、智能化等路径提供高质量的教育服务，成为全球优质教育服务的引领者。海亮教育的发展目标不仅限于经济效益和规模体量的增长，还包括品牌影响力的提

升、乡村教育的振兴、教育生态圈的构建以及科技在教育中的应用等多个方面。持续创新，全面拥抱数字化是海亮教育达成核心战略目标关键举措之一。但在通过数据要素的建设和应用推动教育数字化进程中，海亮教育仍然面临一些痛点：

1. 数据价值难以充分计量：

海亮教育拥有海量的教育数据资源，从学生的学习表现到教师的教学活动，再到学校的运营情况，可谓面面俱到。



数据分析
团队成员



高层管理
人员

是啊，这些数据听起来确实很有价值。但是，我们怎么确保这些数据能真正发挥作用呢？

这正是我们面临的一个挑战。虽然我们可以利用这些数据进行分析，比如预测学生的学习表现，但这种预测的准确性以及它如何改进教学策略，很难用具体的数字来衡量。



数据分析
团队成员



高层管理
人员

能举个例子吗？

当然可以。如果一个学生的历史成绩和作业完成情况显示他可能在数学上遇到困难，我们可以做出这样的预测。但是，问题在于，我们缺乏一个明确的量化指标来评估这种预测对后续教学策略调整的实际影响。



数据分析
团队成员



高层管理
人员

嗯，这确实是个问题。那教师的教学数据呢？它们是不是更容易量化一些？

其实也不尽然。比如，一位教师的教学方法在数据上显示出较高的学生满意度和成绩提升，但我们很难准确计算这种提升与教学方法之间的直接因果关系。



数据分析
团队成员



高层管理
人员

明白了。这么说来，我们目前对于数据价值的评估方法还不够完善，导致这些宝贵的资源难以被充分认识和利用。需要找到更有效方法，让数据价值最大化。

图2 数据价值难以充分计量在工作中的体现

2. 数据产品深度不足：

海亮教育在教育领域的数据积累确实很丰富，但我们在将数据转化为实际产品或服务方面，有什么可以提升的？



市场部门
负责人



数据产品
经理

目前市场上对个性化学习方案的需求日益增长，但我们的数据产品大多还停留在基础的数据分析和报告层面。

能具体说说吗？比如，学生在使用我们的产品时，他们期望得到什么样的内容？



市场部门
负责人



数据产品
经理

好的。虽然我们可以为学生提供一份包含其学习表现的报告，但报告中的建议往往较为笼统，缺乏针对个体特点的深度分析和个性化建议。这导致学生可能觉得报告内容不够实用，无法真正帮助他们提升学习效率。

那对于学校方面呢？他们有没有提出过类似的需求？



市场部门
负责人



数据产品
经理

当然。学校希望我们能够根据学生的学习数据来动态调整课程安排和教师分配，以优化教育资源配置和提升教学效率。但现有的数据产品在这方面提供的支持还不够详细和准确，无法满足他们的实际需求。

看来我们的数据产品在深度和广度上都还有待提升啊。这确实限制了数据资源的有效转化和应用，也影响了我们在市场上的竞争力。



市场部门
负责人

图3 数据产品深度不足在工作中的体现

3. 数据要素流通低效：



图4 数据要素流通低效在工作中的体现

基于以上痛点和不足，想要充分释放数据资源的价值，仅仅拥有海量的数据并不足以转化为实际的教育优势或业务增长动力。关键在于如何高效地挖掘、分析和利用这些数据。通过深

入挖掘数据的内在价值，提升数据产品的深度和广度，以及促进数据要素的流通和共享，我们有望充分释放数据资源的潜力，推动海亮教育在教育服务领域的持续创新和领先发展。

教育行业企业内部 共享机制的构建



1 为什么要聚合数据

数据资产化从运营角度看就是数据成为企业资产,为企业实现价值的过程,从实现路径来看,需要经过数据资源化,数据产品化等阶段。

为实现数据资产化,企业需要构建一套以价值为导向的数据资源运营体系(图5)。良好的数据资源运营组织体系、技术支撑以及数据治理体系构成数据资源运营的地基,但在这一

个基础上要真正的实现数据的价值,还需要进一步的根据企业的战略目标和业务需求挖掘价值场景,通过数据的聚合,数据产品建设,以及长效的评价机制开发和利用数据资源,才能让数据发挥真正的作用实现价值。这其中数据的聚合决定着数据产品建设和评价机制的深度和广度,是数据价值实现的关键一环。



图5 以价值为导向的数据资源运营体系(图片信息及资料来源:普华永道分析)

对于类似海亮教育这类集团型企业而言,集团内各个法人实体或者企业内各个部门在业务经营中积累了大量的数据资源,蕴藏着巨大的价值。然而由于部门间的信息系统不兼容,数据标准不统一,访问权限限制或者缺失,导致各个部门或者业务单元的数据资源无法进行有效的聚合,使得一些有价值的场景无法实现,阻碍着数据价值的挖掘。

然而所谓聚合并非易事,需要突破三个挑战:

1. **数据认知**,数据仍然被认为是业务流程的附带产物,用于事后记录,分析形成报告等作用,有时甚至认为整理和收集这些数据是在浪费时间;不认为数据可以提供更多的价值,对于数据的跨部门共享还是停留在口头沟通,一事一议,审批效率低下,甚至无法获取;
2. **利益分配**,业务部门通过投入成本和资源获得的数据,什么动力可以驱使他们把自己的数据资源提供出来呢,通过提供资源又可以得到什么?业务部门通过提供数据无法获得价值认同感和收益,如何向业务部门或者数据持有方提供驱动力,也是数据聚合所需要解决的一个难题。
3. **技术支撑**,即如何制定明确的共享机制和

标准,通过数据集成、数据交换平台、数据虚拟化、统一数据标准、中间件以及API管理等技术,实现数据的集中管理和跨系统的数据共享与流通。

所以,所谓数据汇聚并不是简单的把各个部门的数据物理性的放在一起,这样只是从技术层面解决了这个问题,更重要的是要同时建立一套体系可以解决各部门之间的数据资源的利益分配,同时通过聚合的资源创造价值反向的提升人们对于数据的认知。那么这种机制就可以充分聚合已有数据,挖掘更多数据应用场景,势必会提高市场分析能力,客户分析能力,优化提升产品服务水平,增强客户粘性,探索二次曲线增长。

为达到这个目的,海亮教育开创性借鉴**大富翁游戏**(图6),参考国家对于外部数据要素市场的相关管理制度同时结合企业内部数据资源特点及资产建设的蓝图,构建内部数据财富市场。在这个市场中,各个业务部门或者业务单元作为市场中的参与者(游戏玩家),在整个过程中,玩家逐步实现“数据致富”,而聚合是致富的基础环节,下面则通过介绍致富基础环节的聚合,开启海亮教育数智大富翁游戏之旅。

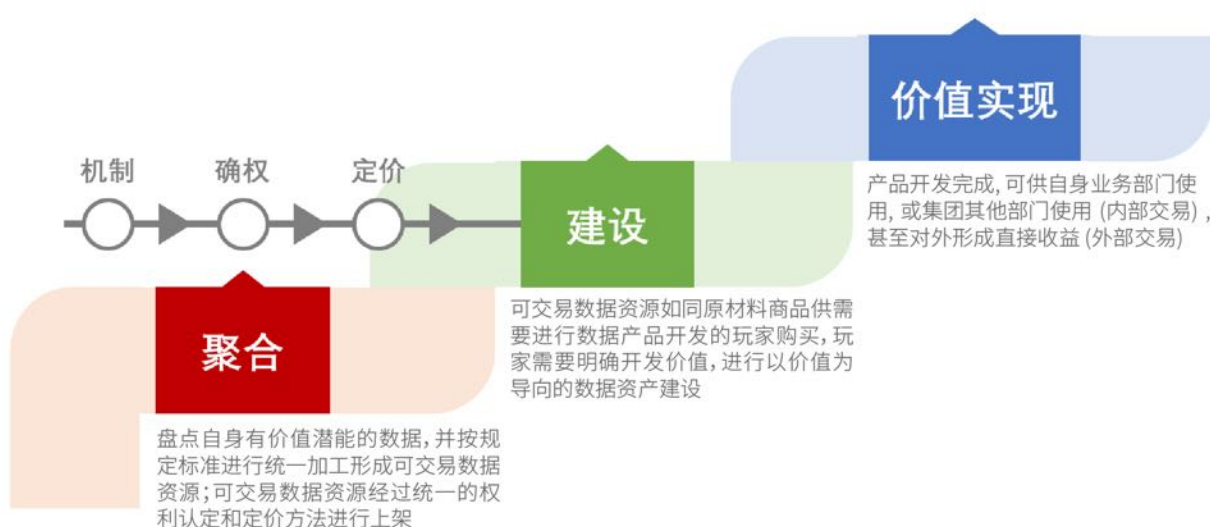


图6 数智大富翁游戏(图片信息及资料来源:普华永道分析)

2 教育行业数智大富翁之聚合机制

● 基本概念介绍

1. 数据需求方,提出数据资源需求的业务部门。
2. 数据提供方,提供数据资源的业务部门。
3. 数据加工方,有权利和能力对数据资源进行开发加工从而形成数据资产的部门,工厂的角色。以海亮教育为例,数据加工方为数字化中心。
4. 数据资源,数据资源是指经过初步加工和整理,对企业或组织具有潜在经济价值的各类数据记录,包括但不限于销售数据、客户反馈、学生数据和教师数据等,能够支持业务分析、决策制定以及潜在的商业机会探索。
5. 数据产品,数据产品是利用数据资源和分析技术,针对特定用户痛点,将业务数据转换成具有业务价值的数据资产的过程和结果。这个过程涉及数据的收集、清洗、整合、分析,以及通过技术手段将其转化为支持业务决策的格式和工具。如数据集合、报表(指标)、标签、画像、模型、分析报告等。
6. 交易标的:聚合阶段交易标的主要为数据资源,数据资源可以作为原材料被加工成为各类的数据产品,直接或者间接向企业提供价值,比如降本增效,带来新的业务收入等。
7. 虚拟账户,由于数据交易的初衷是为了促进数据的流动,往往一开始不会涉及直接的现金交易,故采用虚拟账户的形式记录各个参与方交易的情况,用于积分,晾晒或者利益分配的目的。
8. 数据交易平台,平台负责对数据资源交易的撮合,负责作为交易标的的数据资源的评估和定价,起到市场管理者的作用。同时还要管理虚拟账户以真实反应当期各

业务部门交易情况。

9. 数据管理方,企业内部数据资产最终的使用方和购买方,同时也负责企业所拥有的数据的授权,以及启动资金的发放。

● 交易机制

1. 设定虚拟账户期初余额,即各个部门的启动资金,启动资金的设定可参考企业整体所拥有的数据资源的估值水平,为公平起见,各部门的启动资金应一致。但后续周期可以考虑一定的奖惩机制。
2. 上架:对集团整体数据资源进行盘点,构成“资源产品清单”,通过确权和估值定价工作后完成“上架”,数据需求方可以在货架上找到所需资源。
3. 交易流程:首先数据需求方(买方)根据识别的业务场景产生了某一数据资源的需求,随后将购买需求提交至数据交易平台,由数据交易平台联系数据资源持有方(卖方)进行交易撮合,买卖双方达成初步意向后,根据定价情况进行交割,同时在各的虚拟账户上扣除和增加相应的金额,购买后,买方可将购买所得的数据资源用于相应业务场景。数据资源交易流程到此基本告一段落。需求方如有产品开发需求,则进入另一个环节可以向数据加工方提出,向数据加工方支付加工费,完成相关数据产品的开发。开发后的数据产品如果直接对外出售,则可根据开发阶段的交易情况作为利益分配因子,对相关收益进行分配,若相关数据产品用于内部使用,由数据管理方进行购买(产品阶段的交易和收益分配,我们将在后续系列白皮书详细介绍)。
4. 清零机制,根据考察周期确定每一轮游戏的时间,每一轮结束完成相应的评价后,余额清零,交易记录存档。同时根据评价情况对估值定价机制进行优化,上一轮的表现可以影响下一轮启动资金的分配。

随着公司数据资源整体的管理和运营水平的提升，“大富翁”游戏也可以引入更多的角色和规则以更好的帮助企业实现数据的聚合，比如如果标准化的交易需求较多，可以加入

数据银行或者数据市集，更便捷的进行交易；估值定价方面也可以引入更为复杂的议价机制(图7)。



图7 数据资源共享机制(图片信息及资料来源:普华永道分析)

整套共享机制的核心在于如下两点:

1. 以收益/价值分配为目的的敏捷确权体系,企业内部的数据确权不同于外部数据要素市场中的确权,在确定数据的归属的同时,更偏重于如何根据数据权力的归属进行相应的利益分配。在实践中,海亮教育集团根据实际业务情况,设计了一套统一的确权授权方案,为数据的流转提供了便利。相关内容会在下文做进一步的介绍。
2. 高效的数据资源估值定价体系,不同于外部市场估值,需要严格根据相关评估准则对评估标的的公允价值进行评估,目的在于基于可获得的信息获得一个公允的价值,而企业内部的估值定价注重于三个方面:
 - 1) 内部公认的估值方法,既然将公司内

部视为一个市场,那么所谓“公允价值”是要满足企业内部的需求,获得企业内部各个参与方的一致认可,不一定要达到外部市场公允价值的要求,所以可以在符合估值原理的前提下,在技术层面进行适当调整;

- 2) 快速敏捷的定价机制,以促进内部交易活跃度为出发点,那么定价机制在达到公认的基础上还要尽量简化,复杂的定价机制在数据资产建设初期并不适用,反而会阻碍数据的流通;
- 3) 内外汇率转换机制,虽然为了适应内部市场的情况在估值和定价方面做了简化处理,但是还是要建立内外汇率转换机制,为之后打通外部数据要素市场,乃至现金结算做准备。

相关具体内容会在下文做进一步的介绍。

案例

海亮教育生涯规划大模型开发中
“大富翁游戏”的应用

为了形象地将整个共享机制及相关两大体系（包括敏捷的确权体系和高效的估值定价体系）讲清楚，我们将会以海亮教育目前正在开发的生涯规划产品为例，在这个产品开发过程中，资源汇聚阶段就开启了大富翁游戏之路。

海亮生涯规划产品：依托海亮教育广泛的学校综合管理服务，独有的渠道和流量池，在学生成长的各个阶段，不断根据学生的个性化需求，为其推荐最合适的教育路径，确保学生在各个学习阶段都能获得量身定制的服务，真正实现“因材施教”的个性化教育理念。

第一阶段挑战：数据汇聚挑战：

1. 数据产品开发所需要的数据数据分散于不同系统之中，形成了信息孤岛，严重阻碍了数据的有效聚合和深入分析；
2. 不同系统间学生数据字段的维护标准不一，进一步增加了数据整合的复杂性等等，这些问题不仅影响了数据的准确性和可用性，也限制了海亮教育在提供个性化服务和精准教育方案方面的能力。

为了克服这些挑战，海亮教育在普华永道专家的协助下，致力于建立以价值为导向的数据资产运营体系，该体系复杂且庞大，在体系建立之初，我们重在建设数据治理体系及数据汇聚共享机制，以解决上述第一阶段挑战。具体地，海亮教育通过完善技术支撑及治理体系，统一数据标准；然后通过搭建内部交易市场，以“大富翁”游戏带动市场内各玩家积极参与其中；同时借助两大规则体系（敏捷的确权体系和高效的估值定价体系）维护游戏秩序，顺利突破在资源汇聚阶段的挑战。后续章节，我们将逐一介绍两大规则体系的详细内容，并结合该案例进行形象说明。

3 以收益/价值分配为目的的敏捷确权体系

● 数据确权的重要性

数据的确权是数据流通的前提，可充分保障数据流通各参与方的权益。

在大数据背景下，数据只有在流动和共享中才能发挥出最大的价值。然而由于数据权属不清，使得宝贵的数据不仅止步于企业内部完全无法实现对外赋能，甚至于企业内部的数据共享也问题重重，因此确权问题也成了

数据孤岛无法破除的关键障碍之一。

为了有效解决数据资源流通的根源问题，随着“数据二十条”对数据产权结构性分置制度(图8)的提出，海亮教育尝试先从集团内数据资源共享中开始突破，探索行之有效的解决方案，并为后续的外部流通奠定基础。

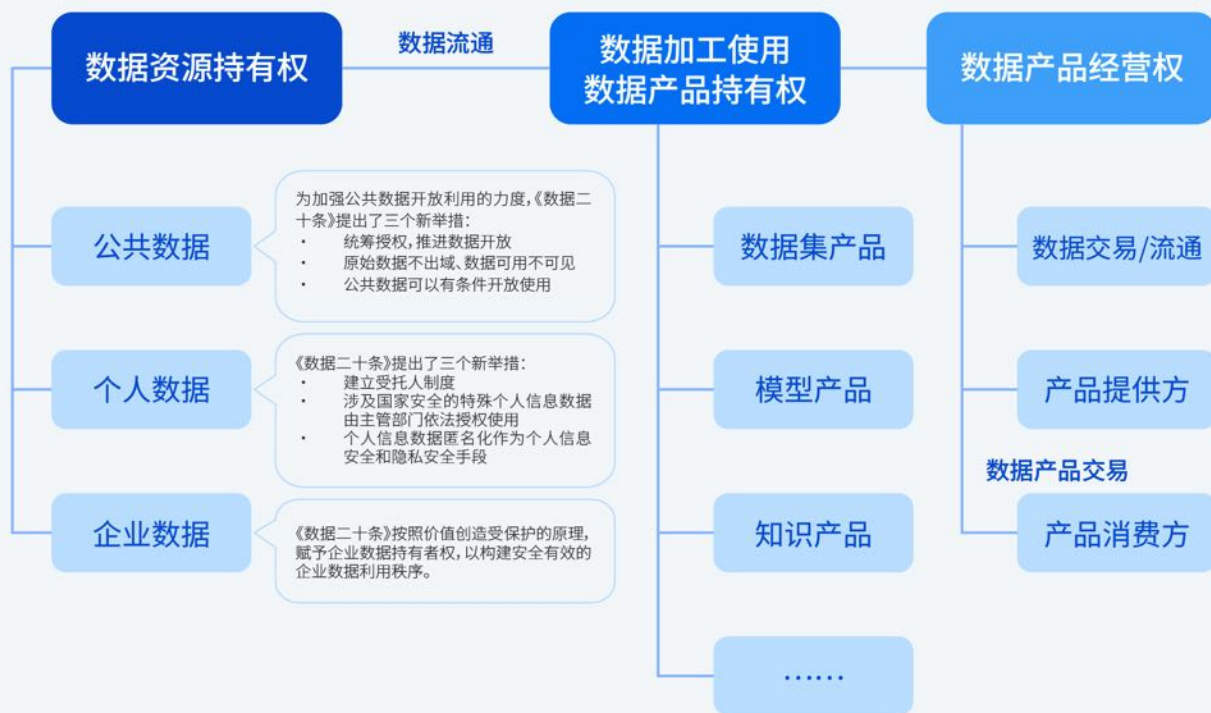


图8 数据产权结构性分置制度(图片信息及资料来源:普华永道分析)

● 内部共享中确权的应用:特许经营权模式

在企业数据内部聚合的过程中,还需要面临部门与部门之间,部门与管理层之间数据交互所产生的确权问题。和企业外部获取数据的确权工作侧重点不同,企业内部数据聚合需要解决的问题是如何确认权利以构建利益分配的基础。为了构建敏捷的内部交互体系,同时遵循《数据二十条》提出的三权分置原则,我们建议以特许经营权模式来运营企业内部数据,解决企业内部使用数据的确权问题。

商业特许经营权是指由商业企业(特许人)授予其他企业或个人(被特许人)使用其品牌、商标、经营模式等进行经营活动的权利。特许经营权具有如下特点:独占性,特许人在特定区域内通常具有独家的经营权力,意味着没有其他竞争对手可以从事相同的业

务;期限性,特许经营权一般具有期限规定,在固定期限内可进行相关经营活动。

那么假设在外部数据获取和使用不存在重大问题的前提下,将公司视为一个商业生态,除个别部门自有数据外,企业独立于各个业务部门拥有其余全部经外部授权的数据或业务数据。数据资源的持有权,数据产品的加工使用权和经营权,根据市场情况将各类权利以特许经营权的方式授权给特定业务部门运营、使用、加工;拥有特许权的业务部门可根据确定的业务场景运营数据或者加工数据,从而发挥数据的作用,同时数据的扭转链条因为确权工作而变的清晰,再结合内部交互机制,可以明确的确定参与方的利益分配因子。

特许经营模式确权的基本原则(以海亮教育为例)

1. 海亮教育科技服务集团(以下简称集团)拥有企业所持数据相关全部权利(包括数据资源的持有权,数据产品的加工权及数据产品的经营权等)(第4点所描述情况除外)。
2. 集团有权将上述相关权利以特许经营权的形式授予特定业务部门。
3. 集团有撤销或者变更授权的权利。
4. 业务部门对其基于部门职责获取,或者对其直接出资建设的业务系统获取的数据,默认拥有原始数据及数据资源的持有权、加工使用权及经营权。
5. **加工过程中**,加工方拥有加工数据产品所需的数据资源的加工使用权。
6. **加工完成后**,在数据产品维护期间,数据产品持有方可授予维护方数据产品及所需数据资源的加工使用权。

实际实践中,可基于基本原则,同时根据情况制定数据的加工方,数据维护方以及数据经营方的权利,以及相关数据流转管理规则。

案例

海亮教育生涯规划产品开发中确权的应用

生涯规划产品开发过程中,首先需要基于各类学生数据,通过大数据和AI模型等技术为学生生成个性化的生涯规划报告,帮助他们识别最适合的未来职业方向。鉴于海亮教育之前完成了信息化建设,各个学生相关的信息散落在各个系统和部门如:招生服务中心的学生系统里面包含了大量的学生相关的基础信息,拔尖人才培养学校的星迹平台有学生成绩相关信息,后勤服务中心有和学生相关的成长报告相关信息等。学生相关信息是伴随学生在校全生命活动周期而产生,其三权应归企业所有,对应到海亮教育,就是总校。

故在“生涯规划”这个数据产品的开发过程中,招生服务中心、拔尖人才学习、后勤服务中心等部门通过总校的特许经营权授权享有学生基础数据的持有权,加工使用权和经营权,而生涯规划中心通过内部共享平台,可以从以上各部门获得学生相关数据,提供给海亮教育的数字化中心进行加工,用于“生涯规划”产品的开发。



4 高效的数据资源估值定价体系

数据资源在企业内部循环,并非必须要定价。然而,随着内循环实践我们发现,如果没有估值定价机制,似乎无法调动各业务部门的积极性,而数据资源很有可能如一盘散沙,没有得到合理的利用。同时,一套良好的估值定价机制,可以从落实确权后解决各方收益分配的问题。

然而,过去几年里,数据的估值一直是困扰业界的难题,虽然理论研究层出不穷,中国资产评估协会也于2023年推出《数据资产评估指导意见》,就三大经典方法,即收益法、成本法和市场法在数据资产的应用提出了对应的估值思路,但实操中仍面临不小挑战,这其中除了估值方法使用的效率、信息获取难度、假设不确定性等挑战外(表1),还有估值效率的问题。从数据资源内循环实践看,如若企业集团内部大量的数据资源共享中完整采用上述方法中的任何一种,从效率上都很难实现“高

效”的效果,反而给企业带来沉重的评价工作。为推进数据资源内循环机制,海亮教育在普华永道专家的专业指导下,通过多次实践磨合,找到了一条高效且可量化的估值定价机制,希望给业界提供一定的参考。

本章节专注于深入探讨数据资源共享过程中所涉及的估值定价机制。该机制的核心原则为侧重于公平有效记录交易和各业务部门的“贡献”,找到一个统一的标尺进行高效的计量,从而保障各方的权益,减少可能出现的争议与分歧,推动数据共享活动能够更加高效、更有成效地进行,进一步释放数据共享在促进企业发展,创造更大价值等方面的巨大潜力。同时,普华永道专家也在此提示,该估值体系只适用于企业/集团内部标准化程度较高的数据资源共享交易,对于外部市场估值以及定制化数据产品的估值,本体系并不适用。

经典估值方法	企业内部市场数据资源估值中的挑战
成本法	一般直接成本最常见的参数(工时)是相对容易得到的,然而,对于集团或企业已形成初步清洗分析后的数据资源,如需做回溯估值,则要求企业有非常完整的工单记录,而多数企业需要耗费较大的精力去挖掘这些信息。因此对于数量较大,且标准化程度较高的数据资源,逐一挖掘直接成本(工时)信息并非易事。成本组成中的间接成本分摊复杂,效率低,无法满足企业内部高效敏捷的需求
收益法	数据资源、产品尚未形成,预测未来的收益的不确定性较大,且难以形成企业统一的度量标尺
收益法	内部市场尚在建立,无法形成可靠的交易参考

表1 经典估值方法在企业内部市场估值中的挑战(资料来源:普华永道分析)



● 探索统一的“标尺”

何为统一的标尺,用估值的角度来说,其本质是形成标的资产价值的主要驱动因素,即我们需要回答一个问题,价值是如何产生的。早在18世纪,亚当·斯密提出的劳动价值论,认为商品自然交换取决于生产该商品所需的劳动量。因此,从最朴素的角度理解价值,我们可以找到一个放之四海而皆准的统一标尺——劳动量。进一步看本文提及的计量标的数据资源,同理数据资源的产生,伴随着数据分析师们以及数据开发团队等的劳动投入。而这些劳动投入最为直观的表现形式,也即是参与生产该数据的广大数据工作者们的工时投入。

我们在日常生活中经常所遇到的数据,例如关乎我们个人重要信息的个人档案,它详细记录了我们的诸多基本情况、学习经历、工作经历等,凝聚着相关工作人员在信息收集、整

理、录入等环节的辛勤付出;再如城市的精密地图,其制作过程更是复杂繁琐,需要专业的测绘人员进行实地勘测、数据采集与处理,以及绘图人员精心绘制和校对等,这其中每一个步骤都饱含着无数相关数据工作者的智慧与汗水。

同样,从另外的角度去审视,对于一个已然比较成熟的数据,当一个企业期望另起炉灶、从头开始去重新生产出该数据时,也必然需要花费相同的劳动才能够获得。换言之,选择支付之前已付出的劳动工时的报酬,亦或是重新付出相同的劳动工时,是共享方获得该数据的唯一途径。而这两个角度无疑都将数据的价值明确指向了劳动工时投入,深刻揭示了劳动工时在数据价值形成过程中的核心地位和关键作用,强调了数据价值与劳动投入之间不可分割的紧密联系。

● 揭秘“标尺”的关键度量指标

知晓劳动工时投入对于评价数据价值而言是一个非常重要的维度，参与共享的各方如何才能清晰的了解一个数据所需要付出的必要劳动工时的数量呢？一个最常见的方式即是回溯历史产生该数据的工时，具体包含了哪些工作人员的参与，这些工作人员分别花费了多长的时间在该数据的开发项目中。

经典方法中的成本法，计算直接成本时既是这种思路，毫无疑问，该方法非常精确。然而，数据的世界是丰富且庞大的，一个企业所包含的数据天罗万象，涉及到人的数据、业务过程的数据、供应商的数据、财务收款的数据等方面，如果企业希望对历史所有的数据

相关劳动投入进行盘查、检索、枚举，这将会是一个非常庞大的工作量，耗时耗力，并不符合企业对于一个高速运转的共享机制的诉求。

数据的共享需求是旺盛的，他依赖高效的数据流通，及时的需求反馈，因此，我们迫切的需要一个简单且有效的途径去评价数据产生所需要的劳动投入，更深的窥探不同数据产生所需不同劳动投入时背后最核心、最关键的影响因素是什么。在我们总结了教育行业不同种类的千千万万的数据之后，我们得到了结论，即**数据资源中所包含的信息量、相关数据源信息的复杂程度以及质量评价**（图9）。

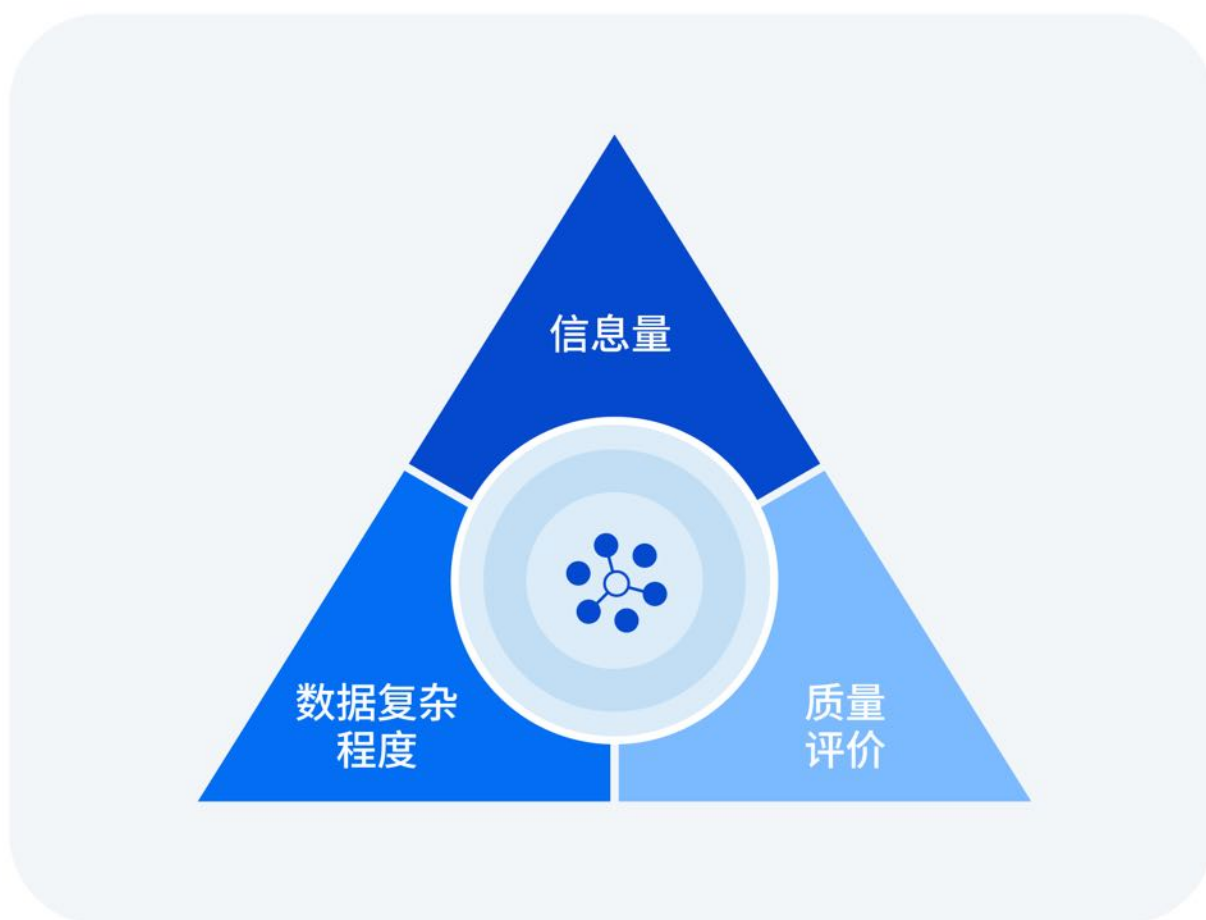


图9 数字价值的关键度量指标 (图片信息及资料来源: 普华永道分析)

信息量

数据所包含的信息量,在通常情况下,往往是较为直观地呈现出来的。一个数据究竟是只包含了一个对象的信息,还是涵盖了多个对象的信息;是仅仅包含了对象一个维度的信息,还是涉及到同个对象的多个维度的信息;其信息是零散的依稀存在几个,还是非常充分的面面俱到、无一遗漏的,这些方面都在很大程度上对一个数据的价值产生深远的影响,以上所提及的各方面的内容,也可以总结概括为数据的记录数量、字段数量以及数据的信息密度。其中字段数量是最为关键的要点。

相较于数据的记录数量,也即一个数据中包含多少条数据,数据的字段数量往往包含更多的劳动投入。企业对数据资源进行日常清洗分析过程中,一般需要针对数据源本身建立一个特定的规则,之后将该规则应用至所有可以触达到的对象并伴随着业务活动的开展,从而累计形成源源不断、日渐丰富的数据,也即是数据集中的一项项记录。而这些规则的建立便成为了数据的一个字段维度,随着更多的规则的建立,数据的字段数量也渐渐丰富。因此,工时往往投入在规则建立中,而显然规则建立所耗费的时间和字段数息息相关,字段数越多,则所花费的工时也越多。

数据复杂程度

数据所包含的信息量在通常情况下是由该数据的字段数量所决定的,然而,不同维度(字段)的信息,其蕴含的信息获取的难易程度是存在显著差异的,这些差异也影响着工时的投入。有些信息能够直接获取,比如用户的登记信息等结构化的数据获取相对较为便捷。而另一些非结构化的信息,如以音频、图片等形态存在的数据,则需要花费更多的工时投入。这些非结构化的数据或者是半结构化的数据,在我们需要对其进行处理之前,往往需要对其进行结构化的操作或者更多的,例如编解码、数字化等操作,这些操作形成较结构化数据更多的工时投入。

质量评价

另一个同样需要重点关注的特征,则是数据的质量评价。数据中的信息是否是准确的,信息是否是及时有效的,甚至包括信息是否完整,是否有空的信息等等,也大幅度的影响了数据的价值,即一个质量优秀的数据应当是包含了更多的劳动工时投入以针对数据原本的种种问题进行治理。在教育集团中,数据的治理通常是一个较为完整的大工程,其中包含了数据治理到入数据湖等一系列的操作。当一个企业关注数据的价值时,企业通常已经站在了数据治理入湖的基础上进行,一般经历了入湖清洗后的数据之间通常没有明显的质量差异。因此,数据质量评价这一关键点,在本文中暂不进行更多的展开。

综上所述,决定数据资源形成过程中工时投入的两大关键度量指标为数据资源中所包含的信息量(以字段数为代表)和相关数据源信息的复杂程度。同时,这两个指标在我们评价数据资源标的是时是较容易获取的,即可通过直接观测获取。

● “标准计量单位”的诞生

数据作为一种具有商品属性的特殊存在，其估值定价与其他商品存在着诸多极为相似的地方。数据是纷繁复杂且多种多样的，倘若我们试图从所有全量的数据入手展开研究，往往会遭遇诸多棘手的困难。然而，换一种思考角度与方式，我们可以通过选取一个典型的样本的具体情况去推测全量样本的整体状况。例如，当我们想要确切知晓一个城市的商场消费情

况时，我们能够通过对几个具有代表性的商场的表现进行全面且深入的分析，从而由小样本的情况合理引申至全体的情况。我们可以通过对数据中最为典型、最为常见的特殊案例展开深入分析，以此来助力我们更为全面地了解数据世界的整体面貌，正所谓窥一斑而知全豹（图10）。



图10 以最典型的数据作为标准计量单位 (图片信息及资料来源: 普华永道分析)

在这其中，对于最为典型、最为常见的数据集/单元，经过与普华永道专家的反复讨论，决定赋予其一个特殊的称谓——“数据标准计量单位”。通过对数据世界的深入探寻与研究，我们从统计学的专业角度寻觅到最为常见的

一种类型的数据集/单元，这种数据其所包含的字段数量处于适中的水平，其字段中所包含的信息全部源自最为常见的结构化的数据(图11)。

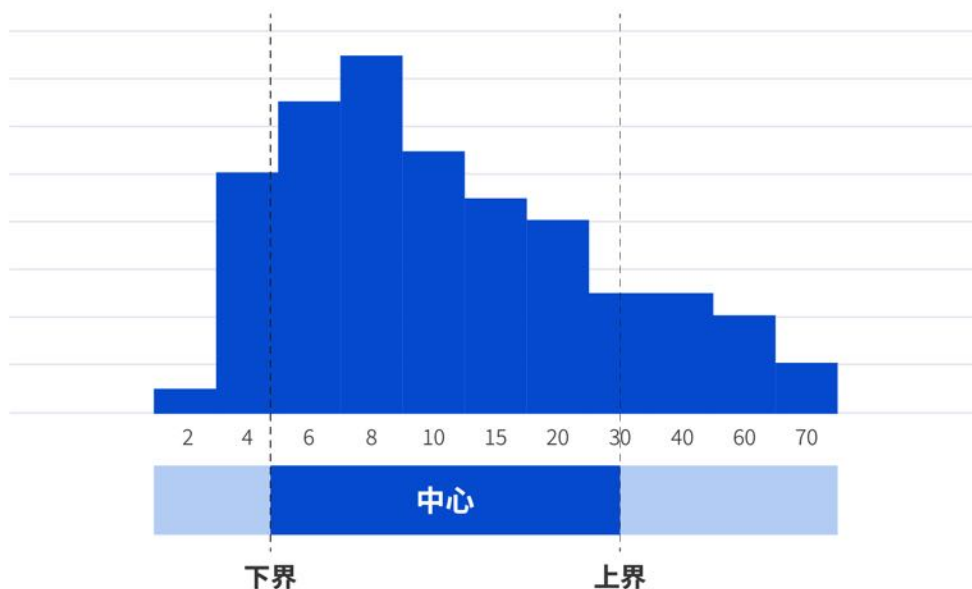


图11 以数据字段数量为例的数据统计分布情况(图片信息及资料来源:普华永道分析)

我们会发现这样各方面属性都较为标准的数据，也就是和“数据标准计量单位”相似的数据，产生该数据所需要的劳动工时投入是较为相近的。因而，在我们对数据进行估值、定价时，对于符合标准计量单位的数据，我们以生成一个数据标准计量单位的劳动工时投

入，作为其估值、定价的重要参考依据。而对于那些与标准计量单位存在差异的数据，我们依据数据在两个核心维度上的差异对其进行合理调整，进而得出不同数据各自准确的估值(图12)。



图12 从数据标准计量单位调整到任一数据价值(图片信息及资料来源:普华永道分析)

对存在差异的商品的价格进行调整的思路广泛应用于市场的各个领域与层面，此处我们以商业房地产的价值评价作为一个具体实例，随着商业房屋所处地块的不同、面积的大小、不同种类数量的配套设施等附属属性的差异，房屋的价格通常会存在大大小小的个体差异，越好的地块、越大的面积往往伴随着更高的价格，同时，一个在市场中各方面属性都比较居中的房屋通常其价格也会集中在一个相对固定的区间范围内。因此，我们将这一常见且实用的思路引入数据估值定价的科学方法之中。

回到数据世界中，当我们需要评价一组数据资源集合，我们可以直接获取到的信息，及这一集合的标准字段数和结构复杂程度。当该

集团的标准计量单元对应的工时输出值为A时，那么我们评价的这一组资源集合就可以快速通过其与标准计量单元之间的字段数及结构复杂程度的差异调整得出待评价数据资源集所需花费的工时。这种方式便捷且所需信息较少，非常适合在标准程度较高的资源层面的估值。

普华永道专家提示，标准计量单元设计及调整系数设计，初期是需要累计大量样本量，通过一系列的统计学分析得到，每个集团或企业都会有不同的标准计量单元及调整系数，但一旦设计完成，则会大大提高后续内部市场运行过程中的估值效率。鉴于统计学模型的复杂性及技术保密的原因，本白皮书将不再详细对此进行说明。

“海亮贝”的诞生

细心的朋友可能发现，根据上一章节推出，对评价标的的评价的输出结果似乎仅是“投入工时”，那么怎么联系到价值呢？劳动投入对于数据价值的评价是一种基于企业内部流转的需求，从最重要的关键因素入手的一种快速敏捷的评价方法。此外劳动工时投入本质上是可以与现实货币相联系的，每一工时根据涉及到人员成本的不同参考不同的报价。而聚焦在当前阶段，服务于企业的内部共享流转，为了区分于真实的货币，引入了“海亮贝”作为参考数据价值的衡量单位，采用“贝”来计价，也曾是古代流行的计量传统。以虚拟的“贝”作为计量，对内可以“海亮贝”的数量作为内部的收益利润分配的因子，对外实现获利也可以直接通过汇率转化为实际的现金收益。

● 基于“市场规模”的定价

由于数据具有可多次重复访问的特点，显区别于一般的常见商品。通常情况下，一般商品的价格往往与其自身价值呈现较为直接的关联，然而数据却并非如此，数据的价格与其价值之间存在着一定程度的差异。正如在前文中对于数据权属所进行的详细说明那样，数据的共享并非是将数据所拥有的所有权力一次性地、毫无保留地全部让渡给另一方，而是仅仅共享其中的部分权力给使用方，并且在这一过程中，数据所有者自身所原本拥有的权利并不会因此而受到任何影响。

因此，从本质上来说，数据的共享过程实际上

是一个部分权利的共享过程，其价格自然而然地应当会受到愿意共享该数据的客户数量的变化而产生相应的改变。对于一个有着热门共享需求的数据而言，大家可以通过一种平分的形式共同来支付该数据的价值，这里所说的数据价值也就是其生产过程中所必须投入的必要劳动工时。在市场中，通常一个物品的总价值与其对外售卖的总价格是相等的，即物品的价值 V 等于其售卖的价格 P 与售卖的市场规模 N 的乘积，在价值相同的情况下，市场规模越大，物品则可以以相对便宜的价格出售给市场中的每一个参与者（图13）。

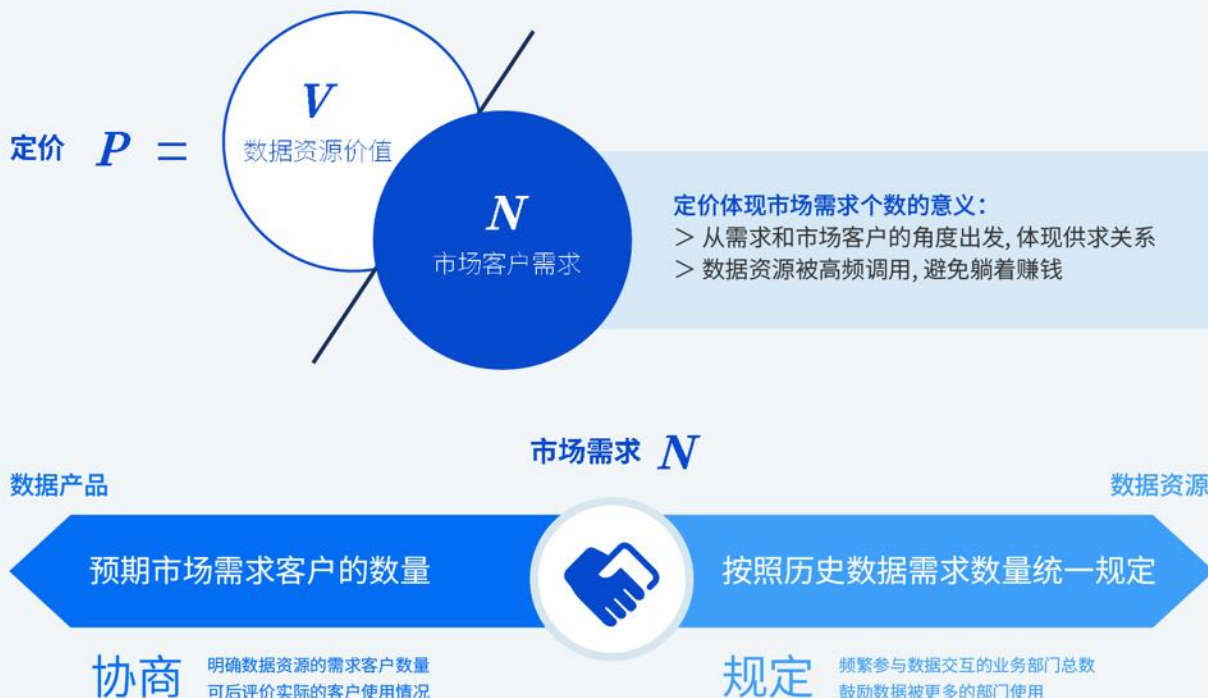


图13 数据资源的定价模式(图片信息及资料来源:普华永道分析)

以教育行业中的教师辅导为例，一位教师在特定的固定时间内，其劳动工时的投入是相对固定的。倘若他针对一个阶梯教室中几百个学生进行大班课教学，那么每一个学生所需支付的上课费用便会相应地被摊薄。反之，若学生期望能与这位老师进行一对一的补习，显然，这种一对一补习的价格必然会远远高于大班课。在这种情况下，其中每一个共享方所需要支付的价格则会因为参与共享的人数众多而进行一定程度的稀释；与之相反，当面对一个冷门共享需求的数据时，由于不存在更多的共享方参与其中，此时的需求方则需要支付相对更高的价格。

为了确定数据的价格，我们可以参考以上的公式。其中，数据的价值，在前文的估值部分已经较为清晰地阐述过。而其潜在的市场需求数量如何确认？则仍是需着重关注的问题。继续以前文中的授课教师为例，作为一名教师，应当如何了解倘若自己开设大班课，预期会有多少学生愿意前来报名。

一个较为直观的方法便是，通过调查历史上与自身情况相似的老师开展大班课教学时的学生数量，以此来推测可能会来上自己课程的学生数量，进而确定每个学生的报名费用。

倘若实际报名的学生数量比教师预期的要多，那么教师便会在学生的教学过程中额外获得相应的报酬；反之，如果教师的预估过于乐观，实际报名情况不及预期，那么对应的教师则会承受开班上课所带来的损失。

数据亦是如此，对于在历史上极为热门的数据，我们可以参考以往市场的需求状况；另一方面，对于那些市场需求不确定的数据，我们在进行开发之前应当率先明确潜在的客户，清晰地界定共享的需求。

这样的，针对数据的估值定价形成了一条较为清晰的路径，这一方法论的确定便可广泛应用在企业的数据交互中了。进一步的，回顾上文中劳动投入决定了数据价值，劳动投入是数据产生过程中最为核心最为关键的投入，但是，我们也需要认识到，劳动投入并不是产生数据的全部成本投入。

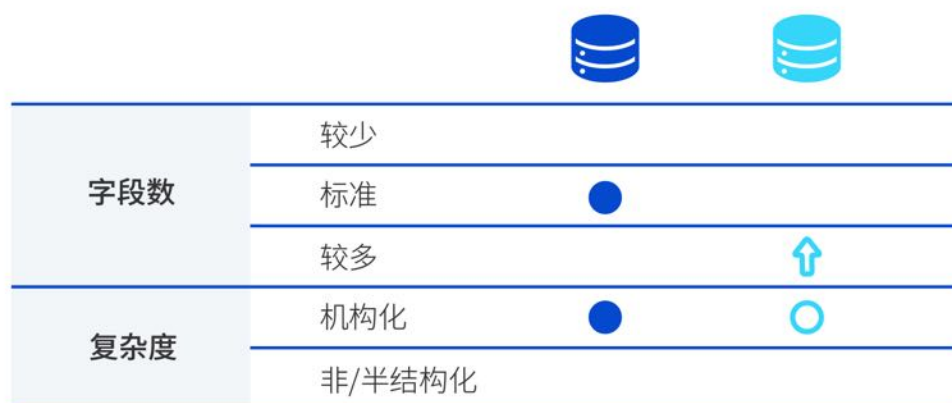
服务器、存储系统等的运行维护，数据安全性防护等也都在时刻伴随着数据源源不断的形成成本投入。只是当我们考虑在企业内部进行数据交互流转时，这些投入可以视作是企业总体层面上的共同承担的费用，各部门之间的交互则可以视情况忽略这些共同的间接成本。

案例

海亮生涯规划产品在“大富翁游戏”下的估值以及定价

我们继续以前文中的生涯规划大模型产品为例，产品的开发过程中迅速的产生了大量的基础数据的需求，其中广泛的包含跨着部门的数据。如何对这些跨部门的数据交互进行快速的估值定价，上述的方法论便可以很好的助力企业实现这一数据交互的需求。

首先，聚焦于生涯规划产品开发所依赖的基础数据，例如学生基础信息（来自招生部门）、学生成绩信息（来自教管部门）等方方面面的基础数据。而后，从两个关键的影响因素出发，依赖的基础数据因其字段数量的不同，我们可以划分为三个类别：与标准计量单位相似的数据，以及字段数量明显多于或少于标准计量单位的数据；同时，生涯规划产品所依赖的数据为结构化数据，与标准计量单位是相同的。我们以依赖的其中一个基础数据为例，该基础数据的字段数量相较于标准字段数量是较多的，同时该基础数据依赖的原始数据是结构化的数据（图14）。



字段数	较少		
	标准	●	
	较多		↑
复杂度	机构化	●	○
	非/半结构化		

图14 数据与标准计量单位的差异调整（图片信息及资料来源：普华永道分析）

- 我们规定一个标准计量单位的价值为1个海亮贝；
- 通过复杂统计模型得到，一个字段数较多的数据集应当在标准计量单位的基础上乘以1.2的倍数^[3]，即为1.2海亮贝；
- 该海亮贝的权属部门在基础数据的共享方面考虑有10个数据需求方，因而设定该基础数据的定价为1.2海亮贝除以10个预期数据需求方，定价0.12海亮贝的价格；
- 整个生涯规划项目约参考了百余个类似的大小各异的基础数据，因而整体生涯规划参考的基础数据价值在12个海亮贝左右。

从而，我们根据上文中确定的不同调整下的关键参数，确定了每一种数据的估值。另外的，基于估值，以及每一个不同部门的数据的交互需求的数量，即得出了生涯规划产品需要依赖这些基础数据所需要支付的成本，也即是数据的定价。

[3] 此处的1.2倍即为章节“标准计量单位”的诞生”中描述的，字段数量偏高因素对于数据劳动工时投入的调整系数。该系数通过对集团历史数据信息的拟合得到。



结语

我们身处信息爆炸的时代，数据不仅仅是数字的简单堆砌，它们可能尚未被挖掘，但对企业而言毫无疑问是无价之宝，等待着被转化为全面提升企业价值，推动行业前进的强大动力。现阶段，海亮教育作为教育行业在数据资产化的践行者，通过数据资源共享汇聚机制，将这些散落的数据聚合起来，就像汇聚溪流成为江河，积累起巨大的势能。这不仅是数据资产化过程中的坚实基础，更是我们教育行业数据生态建设的壮丽起点。

但是，这仅仅是开始。我们仍在数据资产化长河中不断探索，后续我们将沿着我们实践的

脚步，推出《教育行业数据资产化实践系列白皮书》之后续系列，包括以价值为导向的数据资产建设，激发数据动能的系列之二以及深入探讨数据资产运营过程中难点及解决方案，建立一个健康、有序且平衡的数据生态的系列之三。

我们坚信，通过不懈的努力和持续的探索，教育行业的数据生态将日益完善，成为推动行业高质量发展的不竭动力。让我们携手同行，共同推动教育行业的数据生态建设，为教育的明天注入活力，为未来的教育贡献我们的智慧和力量！

参考文献

《中国教育数字化的现实基础、实然困境与改革方向》教育部评估中心，郑永和 刘士玉 王一岩

《中小学信息技术教育》2023年第11期 国家信息中心信息化和产业发展部主任 单志广

《数据资产评估指导意见》，财政部，中国资产评估协会，2023年

《企业数据资源相关会计处理暂行规定》，财政部，2023年

《资产评估专家指引第9号-数据资产评估》，财政部，中国资产评估协会，2019年

《数据资产化前瞻性研究白皮书》普华永道 2021年

《华为数据之道》，华为公司数据管理部，2020年

《开放数据资产估值白皮书》，普华永道，2021年

《数据资产价值与数据产品定价新思考》普华永道，2022年

联系我们

海亮教育科技服务集团



项灵刚

CIO兼数字化中心总监

电子邮件

xianglinggang@hailiang.com

普华永道



詹睿

普华永道中国合伙人

电子邮件

kate.zhan@cn.pwc.com

浙江大数据交易中心



孔俊

总经理

电子邮件

kongj@zjdex.com

全球数据资产理事会



林建兴

全球数据资产理事会总干事

电子邮件

lin@dacglobal.cn



本报告版权属于海亮教育科技服务集团及著作人(詹睿、侯建、倪晨阳)所有, 未经事先书面授权许可, 任何人不得修改、复制、转载、摘编或以其它任何方式使用本报告的全部或部分内容。本通讯中所含内容乃一般性信息, 任何海亮教育科技服务集团及著作人(詹睿、侯建、倪晨阳)并不因此构成提供任何专业建议或服务。在作出任何可能影响您的财务或业务的决策或采取任何相关行动前, 您应咨询合资格的专业顾问。我们并未对本通讯所含信息的准确性或完整性作出任何(明示或暗示)陈述、保证或承诺。海亮教育科技服务集团及著作人(詹睿、侯建、倪晨阳)不对任何方因使用本通讯而直接或间接导致的任何损失或损害承担责任。