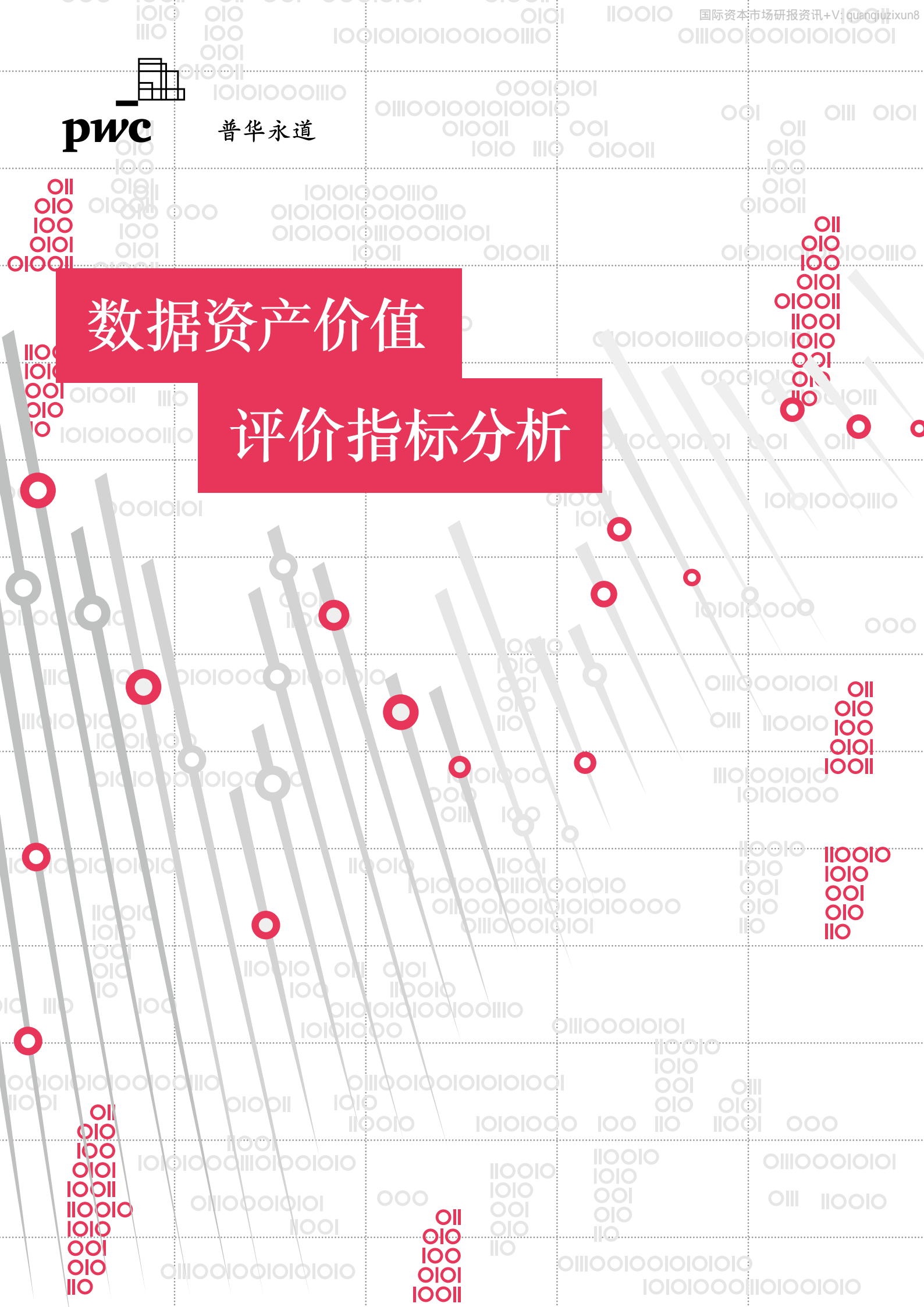


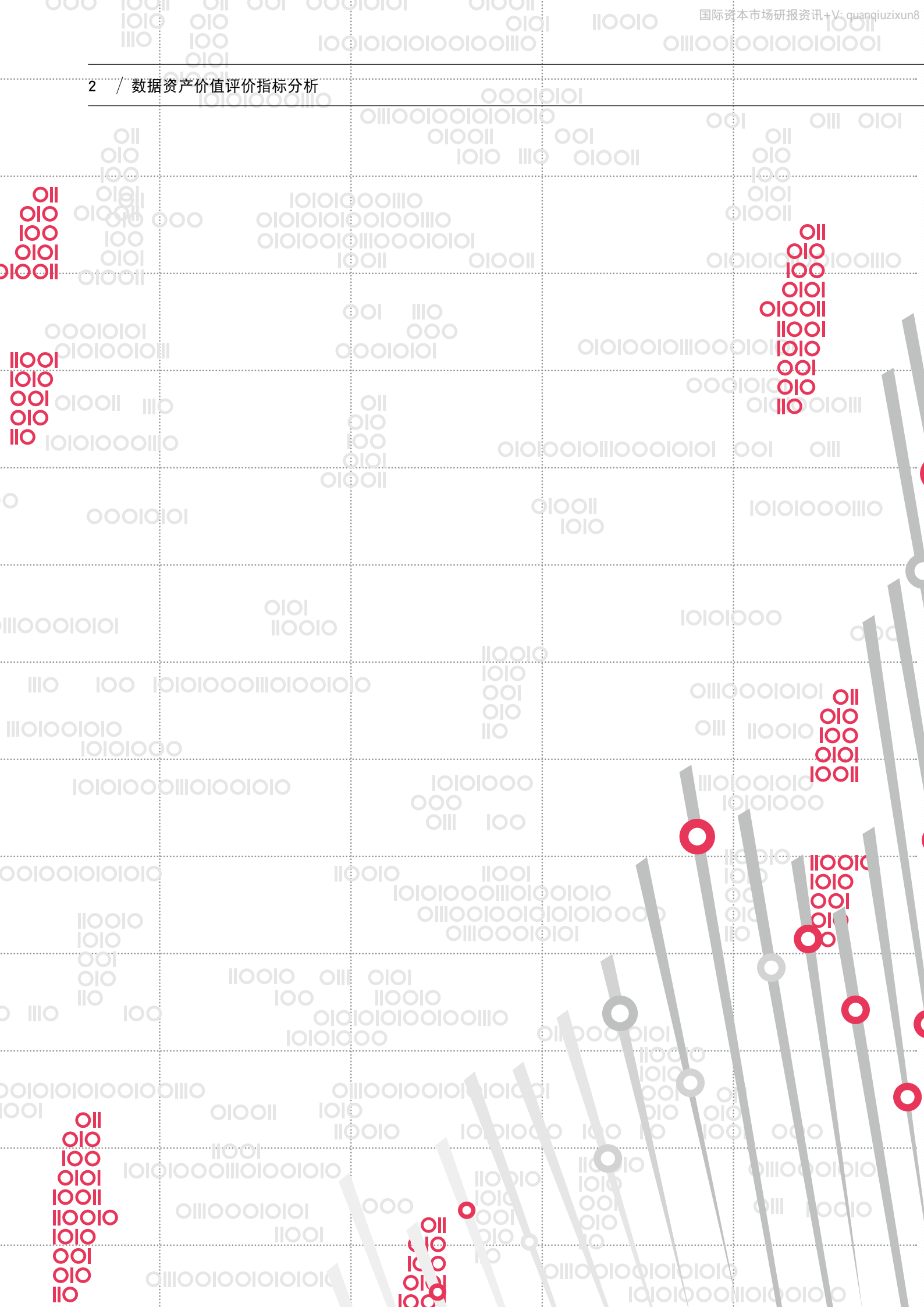
普华永道

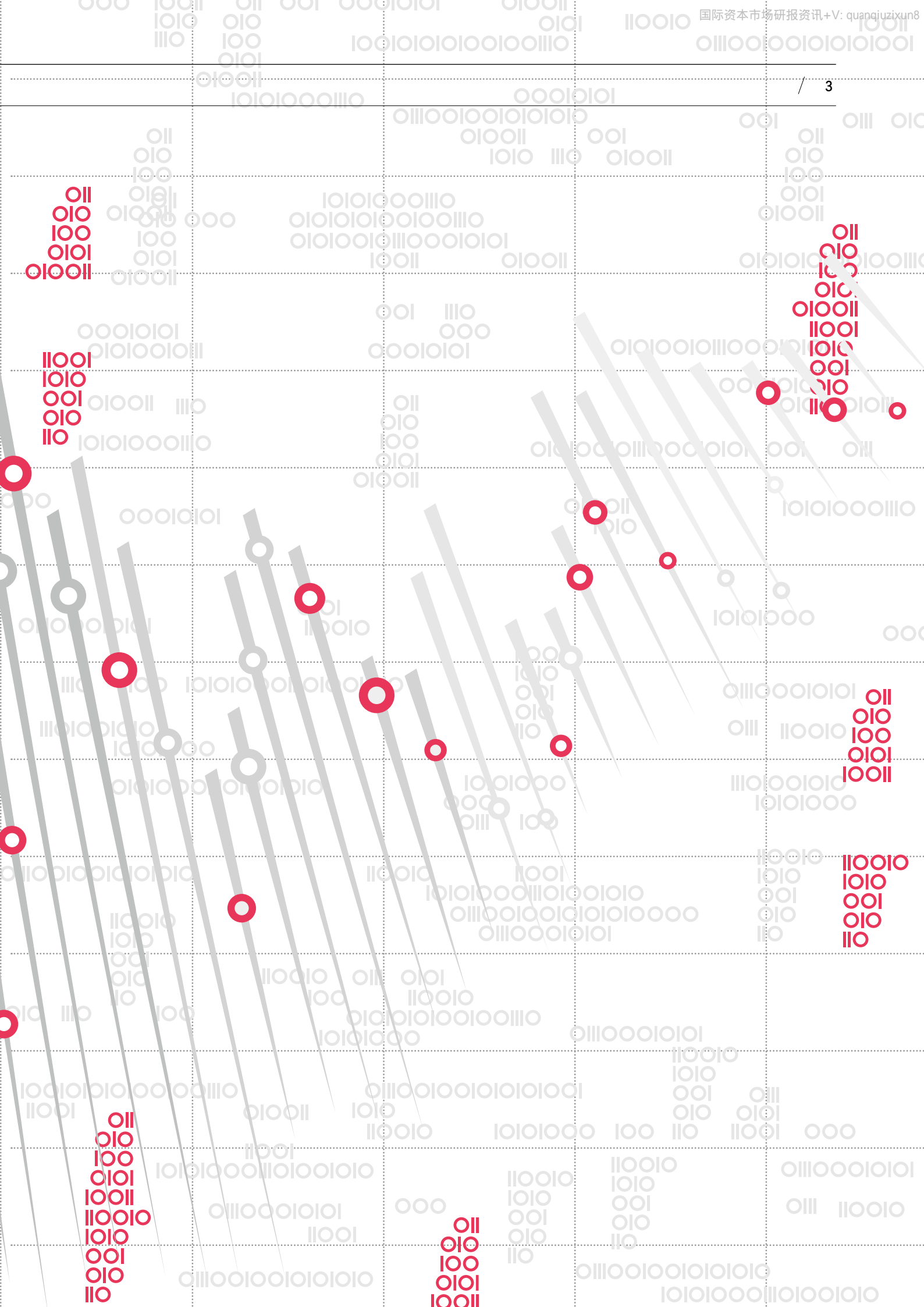
# 数据资产价值

## 评价指标分析



2 / 数据资产价值评价指标分析





# 前言

随着计算机相关软硬件技术的迅猛发展，大数据的相关应用已经深入到生产生活的各个角落。随着数据应用领域的扩展，对于数据的需求也随之水涨船高，利用数据进行分析、将分析成果应用于运营决策也已成为企业发展的新动力。党的十九届四中全会将数据增列为一种生产要素，要求建立健全由市场评价贡献、按贡献决定报酬的机制。在此背景下，数据作为一种新型生产要素，针对其进行价值评价将是商业化交易流通过程中必不可少的一环。本文立足于数据价值与应用场景相结合的理论基础，认为在不同应用场景下，数据所贡献的经济价值有所不同，提出开放式数据资产价值评价指标分析框架，初步建立评价思路、评价维度与指标，由数据市场的主要参与者“买卖双方”协商选择标的的数据资产的评价指标，并确定各指标的权重，尝试对数据资产价值影响因素标准化并进行计算，为数据资产的价值评价研究提供探索经验。



# 目录

· 数据资产分类与应用价值 .....	6
· 数据资产价值评价方法介绍 .....	8
· 开放式数据资产价值评价指标分析框架 .....	10
(一) 评价思路概述 .....	10
(二) 评价指标介绍 .....	12
成本维度评价指标 .....	12
质量维度评价指标 .....	13
应用维度评价指标 .....	14
风险维度评价指标 .....	14
(三) 评价结果处理 .....	16
(四) 评价运用实例 .....	17
· 数据资产价值评价应用 .....	22
· 数据资产价值评价展望 .....	23
(一) 数据权属合理拆分奠定数据流通基础 .....	23
(二) 隐私计算确保数据流转安全 .....	23
(三) 数据交易撮合平台助力数据流通 .....	24
· 结束语 .....	25
· 参考文献 .....	26
· 联系人 .....	27

# 数据资产分类与应用价值

未经清洗加工的原始数据，存在冗余、无序等方面的缺陷，导致其应用价值有限，而经过脱敏、分类、清洗、建模分析等数据加工处理后形成的可采、可见、标准、互通、可信的高质量数据，具备较高的应用价值。依照发展阶段，数据资产可以分为原始数据、粗加工后数据、精加工后数据、初探应用场景的数据、实现商业化的数据等。

数据的价值在于与应用场景的结合，不同应用场景下，数据所贡献的经济价值有所不同。数据资产按照应用领域不同，可划分为交通数据、医疗数据、金融数据、科研数据、社交数据、产业数据等。

表1数据资产分类示例

数据资产大类	数据资产子类	数据资产子类示例
交通数据	驾驶数据	踩刹车次数、驾驶稳定系数、里程数等
	位置数据	经度、纬度、车距等
	高速数据	高速车辆通行次数、高速运力指数、高速通行支付方式等
	车辆数据	车牌号、出厂日期、发动机型号等
	出行数据	出行时长、出行时间点、出行里程、出行起始点、行程停留情况等
医疗数据	疾病数据	诊断数据、病史、影像等
	健康数据	基因数据、体检数据、智能穿戴数据等
	物资数据	物资费用、物流费用、医药研发费用等
金融数据	基本面数据	监管文件数据、商业报告数据等
	市场数据	FIX数据、BWIC数据等
	另类数据	网络搜索、交易、天气等
教育数据	教学资源数据	教师个数、教师教育水平、教学设备费用等
	教育教学管理数据	网课网站个数、网课上线人数、寄宿学生个数等
	教育教学行为数据	学生出勤率、学生作业完成率、学生课数等
	教育教学评价数据	考试次数、学生成绩、教师考核成绩等
电商数据	总体运营数据	流量、订单、总体销售业绩等
	流量数据	流量规模、流量成本、会员数据等
	销售数据	购物车个数、下单个数、支付金额等
	商品类数据	商品总数、品牌存量、上架个数等
	市场营销活动数据	新增访问个数、新增注册个数、广告投资回报率等
	风控数据	评论卖家数、好评率、投诉率等
	市场竞争数据	市场占有率、交易额、用户份额等
房地产数据	土地数据	成交楼面价、土地出让金、土地供应面积等
	房企数据	拿地金额、销售面积、销售金额等
	住宅数据	成交面积、成交金额、成交套数等

数据资产的应用场景广泛，不仅可帮助企业进行精准营销、风险管控、商业决策等，也可应用于民生建设，促进社会发展和提高人民生活水平。例如，交通数据资产大类下的“出行数据”，可应用于移动出行的乘客风险监控。滴滴等出行服务商的数据分析师和算法开发人员，根据出行订单相关的“出行时长”、“出行时间点”、“出行里程”、“出行起始点”、“行程停留情况”等记录，判断乘客发生风险的可能性，以便随时监控风险和及时联系乘客确认乘车安全。

再如，医疗数据资产大类下的“疾病数据”，可应用于提高疾病治疗水平。医疗机构的科研人员，根据病患的“诊断数据”、“病史”、“影像”等记录，研究了解新冠肺炎的流行病学规律，优化治疗方案。同时，可基于医疗数据建立新冠肺炎流行病学模拟数理模型，预判疫情发展的速度和方向，帮助各级政府科学决策，让抗疫过程少走弯路。

表2数据资产应用示例

	数据资产	使用者	应用场景	应用价值
交通数据	驾驶数据	保险公司的数据分析师、算法开发人员等	保费计算	分析司机驾驶数据，判断事故风险概率
	出行数据	出行服务商的数据分析师、算法开发人员等	风险监控	监测乘客出行数据，判断风险发生的可能性
医疗数据	疾病数据	医疗机构的科研人员等	疾病研究	根据疾病数据进行实验，改进治疗水平
金融数据	分析数据	量化分析师	投资决策	根据各类分析数据得到研究报告，改善投资策略
教育数据	教育教学评价数据	学校教学团队	教学发展	通过教学评价数据，调整教学计划
电商数据	市场竞争数据	市场部门数据分析师、算法开发人员等	精准营销	通过市场数据分析可发现新的市场机会
房地产数据	住宅数据	房地产公司的数据分析师、算法开发人员等	客户挖掘	通过住宅数据分析识别和转化潜在客户群体

# 数据资产价值评价方法介绍

由上文我们了解到，数据资产具有应用价值，而不同应用场景下数据所贡献的价值不同，那么数据资产价值如何进行评价呢？比照无形资产价值评价方法，目前传统的数据资产价值评价方法主要为成本法、收益法和市场法三种基本方法。



成本法的原理是从产生数据资产所需花费的成本进行评价，在此基础上扣除各种贬值因素，并考虑数据资产的预期使用溢价，加入数据质量、数据基数、数据流通以及数据价值实现风险等数据资产价值影响因素进行修正，从而估算出标的的数据资产的价值。

## 计算公式

$$P=TC \times (1-L) \times (1+R) \times U$$

其中，P代表标的的数据资产价值，TC代表数据资产总成本，L代表数据资产贬值率，R代表数据资产成本投资回报率，U代表数据资产价值调整系数。

收益法的原理是对数据资产投入使用后的预期收益能力进行评价，考虑资金的时间价值，将未来各期收益进行加总，从而估算出标的的数据资产的价值。

## 计算公式

$$P = \sum_{t=1}^n F_t \frac{1}{(1+i)^t}$$

其中，P代表标的的数据资产价值， $F_t$ 代表数据资产未来第t个收益期的收益额，n代表剩余经济寿命期，t代表未来第t年，i代表折现率。

市场法的原理是基于相同或相似数据资产的可比市场交易案例进行评价，对数据资产的价值密度、交易期日、容量等数据资产的性质等相关因素进行修正，从而估算出标的的数据资产的价值。

## 计算公式

$$P=P_0 \times N$$

其中，P代表标的的数据资产价值， $P_0$ 代表可比案例的数据资产交易价格，N代表数据资产性质修正系数。



此外，业内学者在成本法、收益法和市场法三种基本方法的基础上，衍生出不同的数据资产评估公式，以及从不同的角度提出数据资产评估公式。

李永红等人基于市场法原理，提出衍生的市场法评价数据资产价值。

#### 计算公式

$$P = \frac{P_1 \times K_1 + P_2 \times k_2 + \dots + P_n \times K_n}{n}$$

其中，P代表标的的数据资产价值，n代表可比数据资产数目， $P_i$ 代表第i个可比数据资产的市场价值， $K_i$ 代表第i个可比数据资产根据各项指标确定的综合调整系数。

梁艳基于收益法原理，提出使用多期超额收益法评价数据资产价值。

#### 计算公式

$$P = \sum_{t=1}^n (E - E_w - E_f - E_i) \times (1+i)^{-t} \times K$$

其中，P代表标的的数据资产价值，E代表企业的自由现金流， $E_w$ 、 $E_f$ 、 $E_i$ 分别代表流动资产贡献值、固定资产贡献值、除数据资产外的其他无形资产贡献值，K代表数据资产价值调整系数，i代表折现率，n代表收益期限。

李希君认为数据的价值在于其减少不确定性的决策效用，提出以“信息熵”进行数据资产价值评价。

#### 计算公式

$$\text{Pr}(\cdot) = l(H(\cdot))$$

其中， $l(\cdot)$ 是一个非递减的联系函数，它应该满足如下条件：

$$\forall x_1 \geq x_2, l(x_1) \geq l(x_2),$$

$$\forall x_1, x_2 \geq 0, l(x_1 + x_2) \leq l(x_1) + l(x_2).$$

张弛提出基于深度学习方法进行数据资产价值评价：

$$P = (1+G) \times (1+D) \times (1+A) \times (1+S) \times (1+R)$$

其中，P代表标的的数据资产价值，G代表颗粒度，D代表多维度，A代表活性度，S代表规模度，R代表关联度， $G、D、A、S、R \in [0,10]$ 。

# 开放式数据资产价值评价指标分析框架

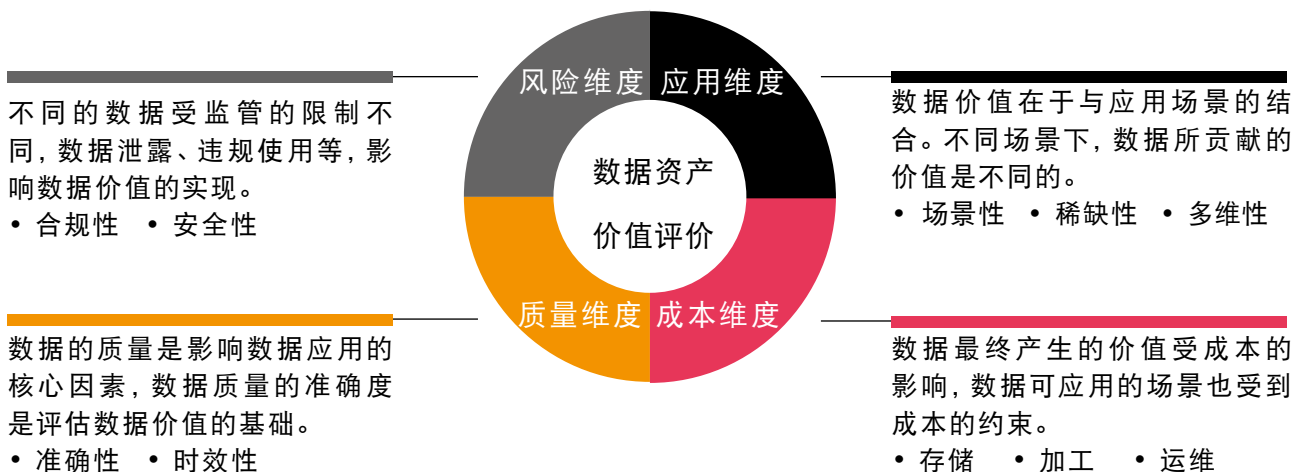
在数据资产交易逐渐成为社会各界关注重点的当下，设计和构建科学的数据资产价值评价体系对厘清数据资产属性、制定统一的定价机制、促进数据市场健康发展具有至关重要的意义。

本文提出开放式数据资产价值评价指标分析框架，初步建立评价思路、评价维度与指标，由数据市场的主要参与者“买卖双方”协商选择标的的数据资产的评价指标，并确定各指标的权重，尝试对数据资产价值影响因素标准化并进行计算，为数据资产的价值评价研究提供探索经验。

## （一）评价思路概述

党的十九届四中全会将数据增列为一种生产要素，要求建立健全由市场评价贡献、按贡献决定报酬的机制，因此本白皮书从数据交易市场的主要参与者“买卖双方”的关注因素出发，来探讨数据价值评价思路。通常情况下，卖方关注数据资产的交易价格能否覆盖数据投入的成本并实现溢价，买方关注购买数据资

产后带来多少应用收益，而数据质量是数据应用的基础，同时在数据应用过程中存在数据泄露、违规使用等风险，影响数据应用价值的实现，因此，我们认为影响数据价值的最基本因素包括数据成本、数据应用、数据质量和数据风险。



基于上述4项影响数据价值的基本因素，我们提出”数据资产价值指数“计算公式：

### 计算公式

$$P=C \times Q \times U \times (1-R) \times (1-E) \times F$$

其中，P代表标的的数据资产价值指数，C代表数据成本系数，Q代表数据质量系数，U代表数据应用系数，R代表数据风险系数，E代表其他负向影响系数，F代表其他正向影响系数。其他影响指数E、F需根据数据市场发展进行调

整。对于成本、质量、应用、风险这4项数据价值影响因素，可设置具体指标和权重进行量化评价，经标准化处理后，汇总得到百分制的评价分数。以数据质量系数Q为例，其计算公式为：

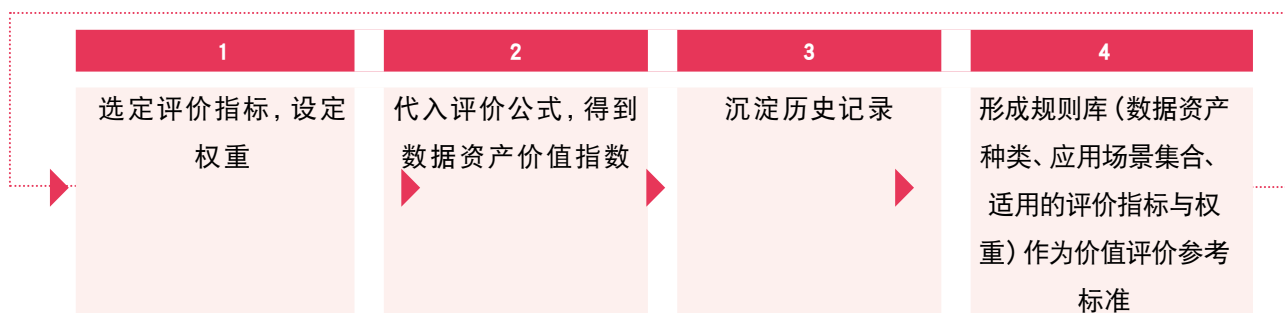
### 计算公式

$$Q = \sum_{j=1}^m Q'_j \times W_j \quad Q' = \sum_{i=1}^n p_i \times w_i$$

其中，Q'代表数据质量相关的一级指标评价价值，W<sub>j</sub>代表一级指标的权重，p<sub>i</sub>代表数据质量相关的二级指标评价价值，w<sub>i</sub>代表二级指标的权重。后文将在“评价指标介绍”章节，对成本、质量、应用、风险这4项数据价值影响因素的相关指标展开叙述。

遵循市场评价贡献、按贡献决定报酬的机制，对于不同应用场景下数据资产价值评价指标的选择与权重设定，由数据市场的主要参与者“买卖双方”协商确定。随着数据交易市场

的发展，大量的数据交易记录将沉淀出，各类数据资产在不同应用场景下的指标选择、权重设定记录。通过这些记录，我们可以获得各类数据资产的应用场景集合，并筛选出不同应用场景下针对不同数据资产最为适用的评价指标及对应权重，最终形成数据价值评价的规则库。未来，在积累了一定数据资产交易记录后，将历史交易信息与规则库结合，通过AI建模形成数据资产价值评价与定价模型，最终实现数据资产的自动化定价。



## （二）评价指标介绍

### 1. 成本维度评价指标

数据资产的取得成本需要根据创建数据资产生命的流程特点，分阶段进行统计。我们将数据资产的取得成本，划分为数据规划、数据获取、数据处理三个阶段的成本。

数据规划阶段的成本，主要包括项目经费、市场调研费、设计评审费、咨询费和数据规划相关的人工工资等。数据获取阶段的成本，与数据的获取方式有关，包括内部生产运营过程获取的数据、间接调研方式获取的数据，以及外购方式获取的数据。通过内部生产运营过程获取的数据，其成本主要包括数据存储相关的设备折旧费、场地租金、水费、电费、空调费、网络费和下包商运维费等；通过间接调研方式获取的数据，其成本主要包括人工工资、设备折旧、场地租金、打印费、数据采集相关的网络费和下包商采集调研费等；通过外购方式获取的数据，其成本主要包括数据采买相

关的人工工资、数据购买价款、税费、注册费和手续费等。数据处理阶段的成本，与数据处理流程有关，包括数据核验、数据预处理和数据分析挖掘相关的成本。数据核验成本，主要包括核验相关的人工工资、设备折旧和下包商核验费等；数据预处理成本，主要包括数据预处理相关的人工工资、设备折旧和下包商数据处理费等；数据分析挖掘成本，主要包括数据打标相关的人工工资、设备折旧和下包商数据处理费，以及数据挖掘相关的人工工资、设备折旧、下包商数据研发费和可容试错费等。

以上各成本项目的发生额，可通过数据拥有者的财务记录、数据购买合同等相关文档获取。在获取各成本项目发生额的基础上，对各项成本进行加总，得到数据资产的成本总额。成本总额越大，数据资产的价值越大。

## 2. 质量维度评价指标

我们以全国信息技术标准化技术委员会提出的数据质量评价指标（GB/T36344-2018 ICS 35.24.01）为依据，从规范性、完整性、时效性、准确性、一致性、可访问性六个方面，对数据质量进行评价。

表3数据质量评价指标示例

一级指标	二级指标	释义	示例	
			指标约束规则	指标评价结果
规范性	长度规范率	数据长度符合规定格式要求的数量占该字段总数据量的比率	字段“营业执照号”的数据值长度规定格式为15位数字	数据表中“营业执照号”的数据值有80%符合长度规范要求，则长度规范率为80%
	.....	.....	.....	.....
完整性	空值率	字段取值为空的数据量占总数据量的比率	客户信息表中的客户号不能为空	数据表中“客户号”有10%取值为空，则空值率为10%
	.....	.....	.....	.....
时效性	更新频率	实际信息发生变化后，间隔多久可正确同步至数据表	门店新增会员人数更新至客户信息表的时间间隔	门店新增会员人数T+1日更新至客户信息表，则更新频率为T+1
	.....	.....	.....	.....
准确性	精确度	满足精度要求的数据量占该字段总数据量的比率	利率数据需精确到小数点后4位	数据表“利率”有85%符合精确度要求，则精确度为85%
	.....	.....	.....	.....
一致性	逻辑一致率	字段间的数据值满足逻辑约束关系的数据量占相关字段总数据量的比率	“投保开始时间”应符合小于等于“投保结束时间”的逻辑约束关系	相关字段的数据值有95%符合逻辑约束要求，则逻辑一致率为95%
	.....	.....	.....	.....
可访问性	访问成功率	数据产品在约定时间长度内的可获取性	访问者在约定时间长度内通过数据接口对目标数据表发起访问请求且成功获取数据	访问者在1小时内通过数据接口对目标数据表发起10次访问请求，有9次请求成功获取所需数据，则访问成功率为90%
	.....	.....	.....	.....

以上内容仅为本文档的试下载部分，为可阅读页数的一半内容。如要下载或阅读全文，请访问：<https://d.book118.com/047004003134006025>