

盒装数据：一种基于数据盒的数据产品形态

叶雅珍^{1,2}, 朱扬勇^{1,2}

1. 复旦大学计算机科学技术学院, 上海 200438;
2. 上海市数据科学重点实验室, 上海 200438

摘要

与通常的商品市场一样,数据市场中流通的数据商品也分为标准产品和非标准产品两大类。目前,市场中有效流通的数据产品是音乐、图片、电影等标准数据产品,而一般意义上的、大规模的大数据流通则遇到了诸多障碍,其中一个障碍就是数据产品的计量计价问题,需要设计一个可计量的标准数据产品形态。基于数据盒模型,设计了一种数据产品的基础标准形态,称之为盒装数据。一个盒装数据产品包括盒内数据、盒外包装两部分,其中,盒内数据是指“时间+空间+内容”三维度的数据立方体组织,一般包括图像、图形、视频、音频、文本、结构化数据等多种数据类型;盒外包装包括产品登记证书,以及产品说明书、质量证书、合规证书等内容。盒装数据旨在为数据要素市场提供一个可计量计价的标准数据产品形态。

关键词

数据产品;盒装数据;产品形态;数据市场

中图分类号:TP311,F49

文献标志码:A

doi: 10.11959/j.issn.2096-0271.2022030

BoxedData: a data product form based on databox

YE Yazhen^{1,2}, ZHU Yangyong^{1,2}

1. School of Computer Science, Fudan University, Shanghai 200438, China
2. Shanghai Key Laboratory of Data Science, Shanghai 200438, China

Abstract

Same as those usual product markets, data products circulating in a data market can be categorized into standard products and non-standard products. Currently, standardized data products such as music, images and video clips are effectively circulating in the market, while large-scale big data product in a broad sense is facing numerous circulation obstacles. One such obstacle is the measurement and valuation of data products, which requires a measurable standard data product form to solve. On the basis of the data box model, BoxedData was proposed, as a standard form of data products. A data product that adopts the form of BoxedData consists of two parts, which are inbox data and packing materials. Inbox data refers to the cubic data structure with three dimensions of time, space and content, which generally includes images, shapes, videos, sound, text, structured data and other types of data. Packing materials includes product registration certificate, product

instructions, product quality certificate, product compliance certificate and other documents. BoxedData aims to provide data factor market with a standard data product form which is measurable and evaluable.

Key words

data product, BoxedData, product form, data market

0 引言

数据是数字经济的关键要素,数据要素市场建设是发展数字经济的基础性工作,数据产品在数据市场上的有效流通是数字经济持续健康发展的重要标志。当前,音乐、图片、电影等标准化的数据产品已在数据市场上有效流通^[1-2],但一般意义上的、更大规模的大数据流通却面临困境^[3]。造成这个局面的主要原因是这类产品主要是非标准化的、产品形态难以界定,不能准确地计量,因而也难以计价。近年来,国内数据交易机构探索通过数据交易方式实现大数据流通的机制和可行路径,但在未能明确数据产品形态的情况下,数据的计量计价机制难以形成,数据交易面临难以开展数据交易业务的困境,这也是数据交易机构面临的困境之一。

数据产品的界定尚在研究阶段,还未形成统一的标准定义。Loukides M^[4]将运用数据生产更多新数据并由此实现数据价值的应用程序视为数据产品,他认为数据科学的研究为数据产品的创建创造了条件。Hazen B T等人^[5]认为数据产品与制造业生产产品类似,是数据生产的产出品。Huang G Y等人^[6]把从原始数据中通过移除噪声与无用信息而获得的关键信息所形成的一种新的数据格式,称为特定领域的数据产品。Bengfort B等人^[7]认为数据产品是数据与用于推理或预测的统计算法的组合。Cao L B^[8]将数据产品定义为来自数

据或由数据支持、驱动的可交付产品,其可以是某种发现、预测、服务、决策、模型、范式、系统等形式,数据产品的最终价值将体现在知识、智能及决策等方面。Pei J^[9]给出的数据产品定义是从数据集派生出的用于产品和信息服务的数据集。作为一个需要在市场上流通的产品,数据产品应该满足一定的标准规范要求,是可以独立使用、交易流通的数据集。

数据产品形态是数据产品得以计量、计价的基础,如果能够设计出一个数据产品形态,那么数据产品的生产、流通、监管都将变得更有效率。数据产品到底应该具备怎样的形态才可以被可靠地计量、计价,并得以在数据市场上有效流通呢? Storch T等人^[10]对某领域的科学数据产品标准进行了研究, Bing C等人^[11]对某应用领域的数据产品类型进行了划分。在数据产品相关实践方面,上海数据交易所积极探索数据产品交易环节和步骤,开展数据产品登记凭证、数据产品说明书等实践工作;山东数据交易有限公司联合中国信息通信研究院制定《数据(产品)登记管理办法》,探索数据产品的登记流程。上述工作仅涉及数据产品的一些原则性规范要求,尚未考虑数据产品的形态。朱扬勇等人^[12-13]设计了一个数据盒模型,将结构化和非结构化的电子数据汇聚在一起,形成外部可见可用的数据容器,该模型具有数据产品形态雏形;叶雅珍等人^[12,14]提炼了音乐、图片、电影等单一类型数据产品的形态。值得注意的是,图书将文字、图片、图形等非电子数据汇聚在一起形成标准化产品,这种做法

对于非标准数据产品的标准化有很好的借鉴作用。基于图书的基础页码数量为49页,以及外形要求、版权规则等标准规范,将文字、图片、图形等非电子数据汇聚在一起就形成了一种标准化的非电子数据产品——图书。本文针对数据产品形态问题,参考图书形态设计,基于数据盒模型设计了一种数据产品标准形态——盒装数据。盒装数据主要由“时间+空间+内容”三维度的数据立方体组织——盒内数据,以及包括产品登记证书、产品说明书、质量证书、合规证书等内容的盒外包装组成。

1 数据产品问题描述

作为一类新型产品,数据产品与其他产品一样,要想实现在数据市场中有效交易和流通,就需要能对其开展可靠的计量,而数据产品计量要有可用可见的数据产品形态。

1.1 数据产品的计量问题

数据的多样性和复杂性加大了数据产品计量的难度。数据产品由数据组成,是电子化的非实物产品。网络空间中的数据通过键盘、摄影机、手机、对地观测卫星、正负电子对撞机、DNA测序仪等电子仪器设备不停地产生,随着时间的推移,所积累的数据类别、形式越来越多样,复杂度越来越高^[15]。因而,对不同行业、不同格式、不同类别、在互联网中/不在互联网中、公开/非公开的数据进行统一计量存在困难,对由多种数据组成、类型多样、规模庞大的大数据集进行可靠计量也存在巨大挑战^[2]。另外,数据的信息属性^[16]使得数据产品的价值因人而异、存

在相对性,这给数据产品计价带来很大的技术挑战。

数据产品计量方面已经有一些探索和实践,有些简单的数据产品计量方法是相对可行的,使得相应数据产品能在数据市场上有效流通。例如,当前在市场上流通的音乐数据产品是一首曲子、一首歌等,以一个完整的音乐作品为计量单位进行计量;类似地,图片数据产品是以一幅图为计量单位,影视数据产品是以一部电影、一集电视剧等为计量单位。这类数据产品的计量单位是早期非电子化情况下就已经设计了标准产品形态的结果,是这些数据产品能在市场上有效流通的基础和前提^[1]。

1.2 数据产品的形态问题

数据产品具有可用可见的形态是其能被可靠计量和计价的前提。

数据产品的生产方式主要有两种:一种是通过数字化实物产品形成数据产品,另一种是直接加工有关数据形成数据产品^[2]。组成数据产品的数据的形式、类型是多种多样的,可以是整数、小数等数值型数据,可以是符号、字符、日期等数据,也可以是文本、音频、图像、照片和视频等类别的数据^[15,17]。这些数据难以有统一的标准,这对设计统一的数据产品形态造成了极大困难,包括形式统一困难和规模统一困难。

● 数据产品形式统一困难。数据产品形式可以由单一类别数据组成的形式,亦可是由多种不同类别的数据构成的形式。例如,一个音乐数据产品可以由统一的数据形式组成;再如,一个电子病历的数据产品包含多种类别形式的检疫检验数据和医嘱、治疗方案数据;等等。如何将这些不同类型的数据统一管理是一个难题,这个难题在数据库领域已经存在多年。

● 数据产品规模统一困难。数据产品的数据类别形式多样且复杂,使得数据产品规模难以有固定大小的基本量。例如,一个音乐数据产品的规模通常在MB级;而一个用于某类疾病大数据研究的电子病历数据产品,一般要有数万份该类疾病的电子病历才能达到研究的要求,其数据规模通常会在GB级甚至以上;等等。那么,使用多大规模作为数据产品的衡量标准呢?这也是一个极具挑战的难题。

由于上述两个难题,数据产品的形式和规模至今未能有一个统一的计量方法,从而导致组成数据产品的形式、类型以及规模都是非标准的。

2 图书的形态与数据盒

图书汇聚了文字、图表等多种类型的数据,图书以“本/册”为计量单位,图书的计价不以图书的内容为基准,而以图书编撰、印刷、流通过程的成本为基准。对于多类型数据汇聚这种做法,图书与数据产品非常相似,甚至可以说是一致的。因此,可以借鉴图书的形态,设计数据产品的形态^[2,14]。那么,如何将多种类型的数据包装成“一本书”呢?数据盒是一个很好的选择,用数据盒包装多种类型的数据,形成一个数据产品的标准形态——盒装数据。

2.1 图书的形态

文字、图片、图形等数据汇聚通过图书的形态,从原来的非标准化产品变为标准化产品。图书的形态主要包括正文主体内容和相关配套内容两大部分。文字产品按一定的排版格式分章节形成图书的正文主体内容,从而满足一定的可阅读性,这部分内容是图书的内容主体,具有独立完整

性,是一个有机整体。为了让文字、图片、图形等数据汇聚能合规出版发行,并成为标准产品得以在市场上流通交易,图书形态配有封面、扉页、版权页、前言(或序)、后记(或跋)、目录、插图及其文字说明等内容^[18]。其中,版权页又称版权记录页^[19],由版权说明、图书在版编目数据和版本记录等内容构成,将纸质图书的书名、作者、国际标准书号(international standard book number, ISBN)、出版责任人、出版发行者、规模大小(开本、字数、印张等)、印刷发行记录等信息都囊括在内^[20]。图书形态的相关配套内容对于保护版权人的相关权益和读者使用引导等方面具有非常重要的作用。

作为一类记录和传播数据的载体产品,图书经过长时间的发展和实践,已形成非常成熟和固定的产品形态。图书的形态能很好地解决文字、图片、图形等数据汇聚在数据类型、数据规模等方面不统一的问题,使文字、图片、图形等数据汇聚产品成为标准化产品,并在市场上有效流通。

2.2 数据盒内部的数据组织

数据盒的基本形态如图1所示。数据盒自带自主程序单元,封装在数据盒中的数据只能通过该单元接口进行受控访问,以实现在数据流通过程中,数据盒中的数据在在保证方便使用的前提下能很好地维护数据拥有方的权益,即数据盒外部可见、可理解、可编程,内部可控、可跟踪、可撤销^[13,17]。每个数据盒设有唯一的字符串类型的标识——数据盒标牌。

数据盒是数据集标准化的一个框架模型。如何设计组织数据盒内部的数据呢?先看一个简单的数据产品的例子。

例1:“某证券交易所2020年挂牌股票

的行情数据集”数据产品。该数据产品的数据集如图2所示。可以看出，单一类型数据产品（如表格、点集、照片集等）很容易用内容维度、时间维度和空间维度来表达。

- 内容维度：是数据集中每个数据对象的内容，即数据对象有哪些属性，这些属性描述了数据对象的完整内容，使得数据对象作为实体独立存在，属性可以是一个或多个。从图2可以看到，内容维度有开盘价、收盘价、最高价、最低价。

- 时间维度：指每个数据对象的时间覆盖范围，即数据对象在不同时间上的值。很多数据产品描述了一段时间的事物或行为的变化，因此数据产品会用一个时间维度来描述每个数据对象在不同时间上的具体值。数据产品需要描述其数据对象具体的时间段和时间粒度。从图2可以看到，时间维度为每个交易日。

- 空间维度：指符合数据产品描述的数据对象的空间覆盖范围，即满足数据产品描述的数据对象全体。需要可区分、可甄别、可检验地描述出数据产品所含数据对象的范围。对于一个数据集，应当明确指明其对象空间应覆盖的范围。从图2可以看到，空间维度为在某证券交易所挂牌的所有股票。

一些数据产品可能没有时间维度（如一个照片数据集就没有时间维度），但内容维度和空间维度是必须要有的。内容维度是对数据对象的描述，空间维度涉及数据对象全体。

针对数据集的三维特征，设计数据盒内部的数据组织为一个包含内容维度、时间维度、空间维度的三维立方体结构的数据形态，简称盒内数据。将图2的数据集装入数据盒后的盒装数据产品如图3所示。没有时间维度的图片盒装数据产品如图4所示。

对于多类型的数据产品，可以用数据

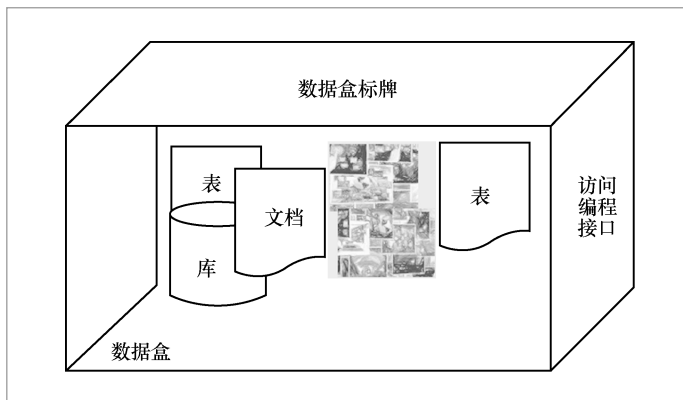


图1 数据盒的基本形态

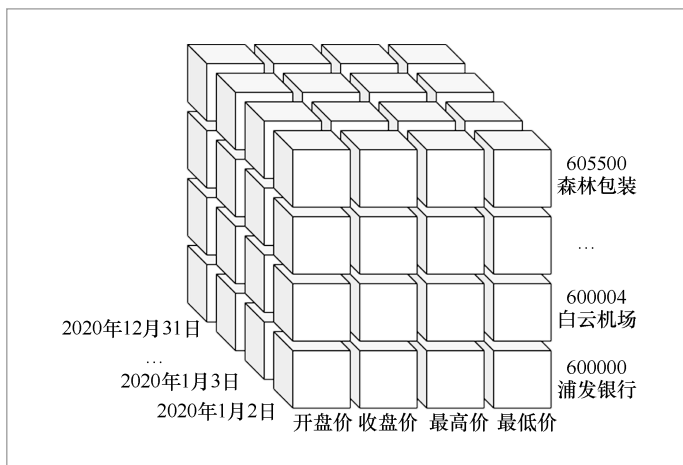


图2 “某证券交易所2020年挂牌股票的行情数据集”数据集

盒的组合形式来表示，即将多个数据盒装入一个大的数据盒中，形成复合型盒装数据产品，如图5所示。这很类似于用零部件组装成一个大的部件，然后还可以用大部件继续组装，直到形成需要的产品。

2.3 盒装数据的计量

像图书容纳文字那样，数据盒可被看作一种能够容纳数据的容器。规定了49页的图书基础规模后，形成了图书的计量单位，并且设计了ISBN等。这样，图书就能方便地在市场上流通。参照图书的做法，可以设计一个标准

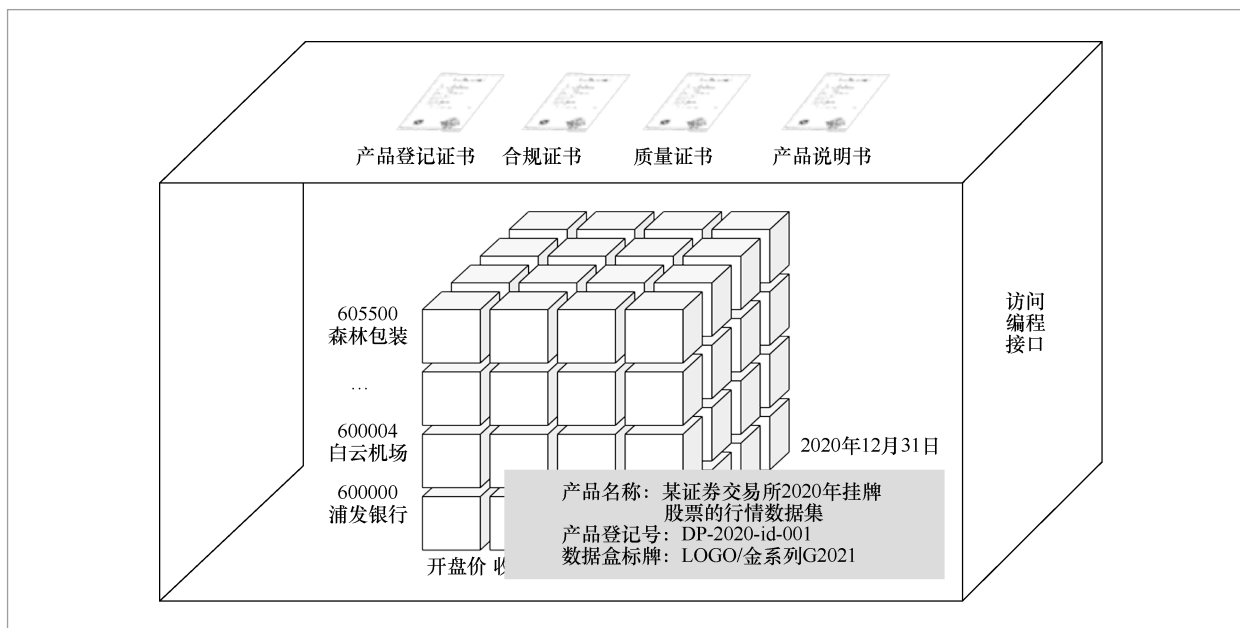


图3 “某证券交易所2020年挂牌股票的行情数据集”盒装数据产品

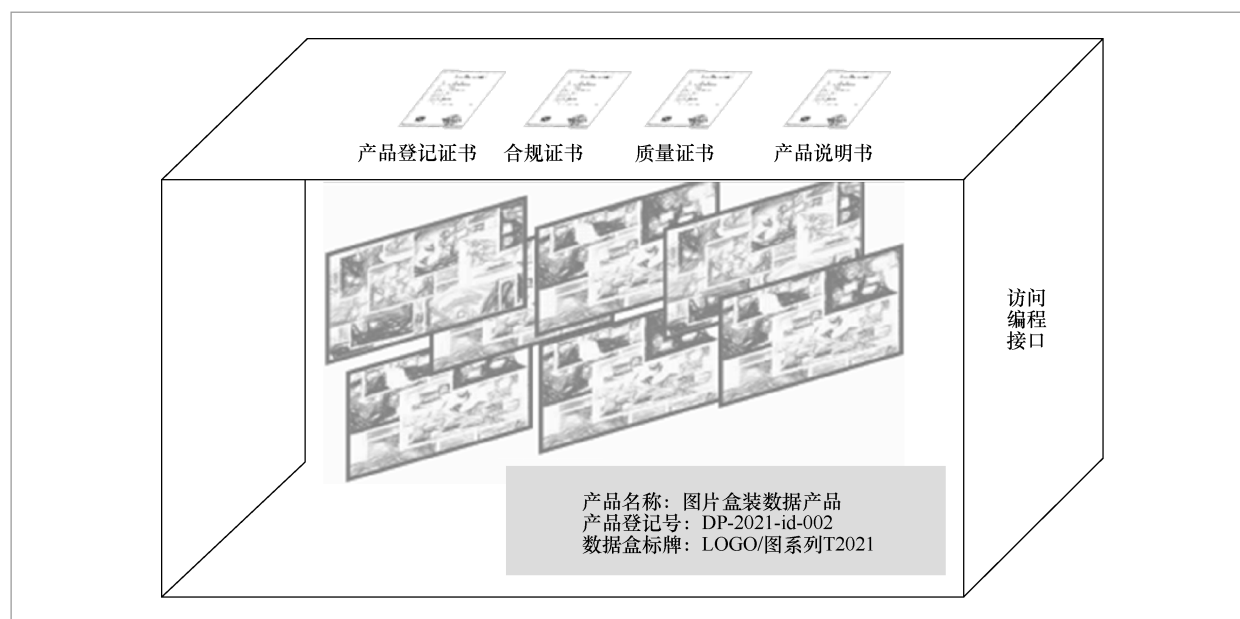


图4 没有时间维度的图片盒装数据产品

化的盒装数据产品的基础规模和内容。在规定盒装数据的基础规模后，盒装数据就可以作为标准化的计量单位。

盒装数据的基础规模和内容的设定要体现市场需求和监管需求。

首先是市场需求。显然，数据流通的基本单位不是一条条数据，也不是一个简单的数据集。数据资源是有含义的数据集结到一定规模后形成的^[16,21]。“一定规模”是数据资源的要求，没有“一定规

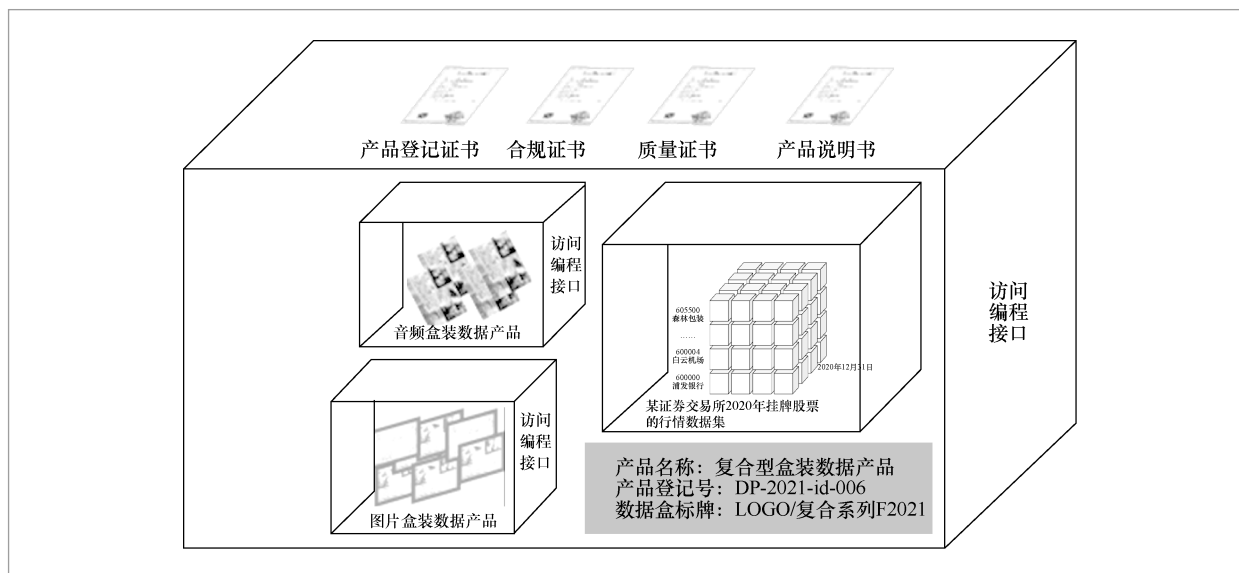


图5 复合型盒装数据产品

模”不能被称为数据资源。以个人数据为例，一个人的身份数据不能被称为数据资源，但是一个城市所有居民的身份数据是很重要的数据资源。在第三届数据科学家大会上向110位参会的数据科学家和数据从业人员发起的大数据问卷调查显示，有96.3%的受访者认为一个大数据产品至少应该达到GB级别以上的数据规模，其内容应至少包括图像、图形、音频、视频、结构化数据、文本等两种以上数据类型。

其次是监管需求。盒装数据作为标准产品在市场上流通，就需要政府监管。对于数据市场，政府监管有两个目的：一个是维护市场的公平、公开、公正；另一个是保护国家数据安全和公民个人信息。前者是所有商品市场都需要的，主要是市场法规的建设；而后者则是数据市场特有的。那么，多大的数据规模有利于监管的实施呢？以公民个人信息保护为例，我国法律中针对侵犯公民个人信息有专门的量刑标准，非法获取、出售或者提供公民个人信息5 000条以上的，应当被认定为

《中华人民共和国刑法》所规定的“情节严重”，已构成犯罪事实，要接受法律的惩处。我国《最高人民法院、最高人民检察院关于办理侵犯公民个人信息刑事案件适用法律若干问题的解释》指出，公民个人信息主要包括姓名、身份证件号码、联系方式、住址、账号密码、财产状况、行踪轨迹等内容。一条公民个人数据涉及十几个数据项（1 KB），加上公民的一张证件照片（200 KB）、一个指纹（10 KB）等内容，其数据规模约211 KB，5 000条公民个人数据大概为1 GB。

鉴于上述两类需求，将盒装数据的最小规模设定为1 GB，这作为其基础规模的数据量是合适的。盒装数据采用1 GB的基础规模作为计量单位，这基本参照了调查结果，也基本符合5 000条个人数据的规模，兼顾了市场需求和监管需求。与“达到49页的规模才能构成一本图书”类似，“达到1 GB的数据规模才能构成一个盒装数据”也是一个认定的规模数值，不是一个科学的界定，具体的标准规模还有待未来实践总结和完善。

对于小于1 GB数据规模的数据集是否需要作为数据产品的问题,从生产、登记、管理和流通成本多个方面来看,目前暂时不予考虑,以免使情况过于复杂,成本过高且难以被监管。例如,人们在朋友圈中传播的照片等,如果将这些都作为数据产品进行登记的话,行政成本就会过高。就像不能把几页的文字材料以图书的方式进行出版一样,在数据产品登记初期,先不考虑将较小数据规模的数据集作为数据产品。另外,单一照片、图书、音乐等在传统电子音像出版领域已经有一套完整的产品体系,因此并不需要考虑新的产品形态。

3 盒装数据外部形态

盒装数据外部形态(又称盒外包装)包括产品登记证书、产品说明书、质量证书、合规证书等。

3.1 产品登记证书

盒装数据产品登记证书是拥有者对相关数据产品权属的声明,是对盒内数据的

概述性介绍,由专门的数据产品登记主管部门审核发放。只有登记后的盒装数据产品才具有合法性,才被允许在数据市场上进行流通和交易,并受法律保护。

盒装数据产品登记证书主要包含产品名称、产品登记号(编号、条码或二维码)、数据盒标牌、生产商/著作权人、数据量、访问编程接口、权利等方面的内容,还包括产品说明书、质量证书、合规证书等附件,具体见表1。

其中,产品登记号(编号、条码或二维码)是保护知识产权和确认数据权属的一种工具,便于盒装数据发行、流通过程中的维权和授权。这类版权标识需经过专门的管理机构登记给出,具有唯一性,可为盒装数据后续的维权、授权等方面的追踪提供便利。数据盒标牌是一种数据商标,由生产商根据一定的标码规则融入品牌元素给出。此外,在互联网上盒装数据还需要给定专门的访问标识,以便人们使用相关系统在互联网上快速查询和访问盒装数据。

3.2 产品说明书

盒装数据产品可以让购买者在购买前看

表1 盒装数据产品登记证书

登记项目	体例
产品名称	某证券交易所2020年挂牌股票的行情数据集
产品登记号	DP-2021-id-001
数据盒标牌	LOGO/金系列G2021
生产商/著作权人	某证券交易所
数据量	2 GB
访问编程接口	NoSQL/Open
权利	所有权:原始所有权/著作所有权(数据经加工后形成的权利)
附件	产品说明书:包括内容说明、生产方式/著作方式说明(被加工数据来源的合法性证明)、使用说明等 质量证书:由数据质量检测机构出具 合规证书:由律师事务所出具

到数据是什么、怎么用,所谓“可见可用、先看后买”。做法是提供详细的产品说明书,包括数据产品内容说明、生产方式/著作方式说明(被加工数据来源的合法性证明)和使用说明等。其中使用说明详细介绍了使用环境、使用接口、使用举例、接口代码等内容。

- 使用环境:包括对硬件(计算机集群、内外存、CPU、集群总线等)和计算机软件(Windows、macOS和Linux等操作系统,Hadoop、Spark等软件)的要求。

- 使用接口:主要为外部软件受控访问、使用开发盒内数据提供接口,包括数据访问接口和数据使用软件应用编程接口等。其中,数据访问接口主要用于用户访问盒内数据,主要分为元数据查询接口、原始数据查询接口、数据统计接口;数据使用软件应用编程接口主要用于用户开发数据访问程序,主要包括开发态数据访问接口和运行态数据访问接口^[17,22]。

- 使用举例:开展用户使用盒装数据的引导工作,主要以图文形式、通过案例来指导用户使用盒装数据的方法步骤。

- 接口代码:主要提供访问、使用、开发盒内数据的接口代码。

3.3 质量证书

盒装数据只有达到一定的数据质量要求才允许被登记,才能进入数据市场流通。盒装数据的质量证书就是盒装数据中的数据集达到相应质量标准和要求的证明性文件,是其开展交易流通的重要凭证。质量证书的相关内容可以作为判断盒装数据价值高低的依据。

盒装数据的质量证书内容主要涵盖数据集3个维度的质量保障,用完整性来表达,即时间完整性、空间完整性、内容完整性。质量证书由数据质量检测机构出具。

3.4 合规证书

盒内数据必须合法合规,即符合国家相关法律规定,这需要律师事务所出具合规证书。

盒装数据的合规证书主要用于承诺盒内数据符合《中华人民共和国数据安全法》《中华人民共和国网络安全法》《中华人民共和国个人信息保护法》等国家有关法律要求,是由律师事务所提供的相关证书。

4 结束语

数据流通需要一个可计量的产品形态,但如何定义数据产品形态仍面临诸多困难和挑战。参照图书,基于数据盒模型设计了一种数据产品标准形态——盒装数据,一个标准盒装数据的数据规模为1 GB。本文设计的盒装数据是针对公开数据市场数据流通的一种探索,鉴于数据要素市场建设的迫切性,先行将阶段性研究成果公开。后续将基于标准盒装数据产品形态,进行数据产品定价和估值等相关研究工作,并开发相应的技术平台,开展数据产品的交易流通实践工作。

参考文献:

- [1] 叶雅珍,刘国华,朱扬勇.数据资产化框架初探[J].大数据,2020,6(3):3-12.
YE Y Z, LIU G H, ZHU Y Y. An initial exploration on framework of data assetization[J]. Big Data Research, 2020, 6(3): 3-12.
- [2] 叶雅珍,朱扬勇.数据资产[M].北京:人民邮电出版社,2021.
YE Y Z, ZHU Y Y. Data assets[M]. Beijing: Posts & Telecom Press, 2021.
- [3] 何培育,王潇睿.我国大数据交易平台的现实

- 困境及对策研究[J]. 现代情报, 2017, 37(8): 98-105, 153.
- HE P Y, WANG X R. Predicament and countermeasure research about big data trading platform in China[J]. Journal of Modern Information, 2017, 37(8): 98-105, 153.
- [4] LOUKIDES M. What is data science?[M]. Sebastopol: O'Reilly Media, Inc, 2011.
- [5] HAZEN B T, BOONE C A, EZELL J D, et al. Data quality for data science, predictive analytics, and big data in supply chain management: an introduction to the problem and suggestions for research and applications[J]. International Journal of Production Economics, 2014, 154: 72-80.
- [6] HUANG G Y, HE J, CHI C H, et al. A data as a product model for future consumption of big stream data in clouds[C]// Proceedings of 2015 IEEE International Conference on Services Computing. Piscataway: IEEE Press, 2015: 256-263.
- [7] BENGFORT B, KIM J. Data analytics with Hadoop[M]. Sebastopol: O'Reilly Media, Inc, 2016.
- [8] CAO L B. Data science[J]. ACM Computing Surveys, 2018, 50(3): 1-42.
- [9] PEI J. Data pricing: from economics to data science[C]//Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. New York: ACM Press, 2020: 3553-3554.
- [10] STORCH T, BACHMANN M, HONOLD H P, et al. ENMAP data product standards[C]//Proceedings of 2014 IEEE Geoscience and Remote Sensing Symposium. Piscataway: IEEE Press, 2014: 2586-2589.
- [11] BING C, PENG C. Construction of power data products based on STP theory[J]. IOP Conference Series: Earth and Environmental Science, 2021, 680(1): 012035.
- [12] 朱扬勇, 熊贲, 廖志成, 等. 数据自治开放模式[J]. 大数据, 2018, 4(2): 3-13.
- ZHU Y Y, XIONG Y, LIAO Z C, et al. Self-governing openness of data[J]. Big Data Research, 2018, 4(2): 3-13.
- [13] 熊贲, 朱扬勇. 面向数据自治开放的数据盒模型[J]. 大数据, 2018, 4(2): 21-30.
- XIONG Y, ZHU Y Y. Data box: a novel data model for self-governing openness of data[J]. Big Data Research, 2018, 4(2): 21-30.
- [14] 叶雅珍, 刘国华, 朱扬勇. 数据产品流通的两阶段授权模式[J]. 计算机科学, 2021, 48(1): 119-124.
- YE Y Z, LIU G H, ZHU Y Y. Two-step authorization pattern of data product circulation[J]. Computer Science, 2021, 48(1): 119-124.
- [15] 朱扬勇. 旖旎数据: 100分钟读懂大数据[M]. 上海: 上海科学技术出版社, 2018.
- ZHU Y Y. Charming data-to understand big data in 100 minutes[M]. Shanghai: Shanghai Scientific & Technical Publishers, 2018.
- [16] 朱扬勇, 熊贲. 数据学[M]. 上海: 复旦大学出版社, 2009.
- ZHU Y Y, XIONG Y. Dataology and data science[M]. Shanghai: Fudan University Press, 2009.
- [17] 朱扬勇. 数据自治[M]. 北京: 人民邮电出版社, 2020.
- ZHU Y Y. Data autonomy[M]. Beijing: Posts & Telecom Press, 2020.
- [18] 国家新闻出版总署. 图书质量管理规定[Z]. 2005.
- National Press and Publication Administration. Provisions on the book quality management[Z]. 2005.
- [19] 牛慧兰. 谈图书版权页的著作权问题[J]. 中国出版, 1999(7): 48-49.
- NIU H L. On copyright issues on copyright pages of books[J]. China Publishing Update, 1999(7): 48-49.
- [20] 中华人民共和国国家质量监督检验检疫总局. 图书书名页: GB/T 12450-2001[S]. 2004.
- General Administration of Quality Supervision, Inspection and Quarantine of

- the People's Republic of China. Title leaves of books: GB/T 12450—2001[S]. 2004.
- [21] 朱扬勇, 熊贇. 数据资源保护与开发利用[M]//专家论城市信息化. 上海: 上海科技文献出版社, 2008: 133-137.
ZHU Y Y, XIONG Y. Protection and utilization of data resources[M]//Expert forum on urban informatization. Shanghai: Shanghai Scientific and Technological Literature Press, 2008: 133-137.
- [22] 吴毅坚, 陈士壮, 葛佳丽, 等. 数据自治开放的软件开发和运行环境[J]. 大数据, 2018, 4(2): 31-41.
WU Y J, CHEN S Z, GE J L, et al. Software development and runtime environment for self-governing openness of data[J]. Big Data Research, 2018, 4(2): 31-41.

作者简介



叶雅珍 (1985-), 女, 博士, 复旦大学计算机科学技术学院教师, 上海市数据科学重点实验室数据资产研究室主任, 复旦大学数据产业研究中心主任助理。专著《数据资产》作者, 在数据资产研究与实践方面提出“数据资产框架”“数据产品运营的两阶段授权模式”等创新观点。主要研究方向为数据确权、数据资产和数据流通。



朱扬勇 (1963-), 男, 博士, 复旦大学计算机科学技术学院教授、上海市数据科学重点实验室主任, 复旦大学数据产业研究中心副主任。《大数据》期刊编委会副主任, 农业大数据产业技术战略联盟副理事长兼首席科学家, 大数据协同安全国家工程实验室副理事长, 中国自动化学会国防大数据分会副主任。国际数据科学倡导者, 提出数据界、数据学、数据身、数据自治、数据财政等概念和体系。发表学术论文200多篇, 出版《数据学》《旖旎数据》《特异群组挖掘》《数据自治》等专著, 并任《大数据技术与应用丛书》(22册)主编、《大数据资源》主编。主要研究方向为数据科学和数字经济, 近期研究重点为数字化转型、数据财政、数据资产、数据自治与数据跨境等。

收稿日期: 2021-12-27