IDC's Worldwide Data-as-a-Service Taxonomy, 2018

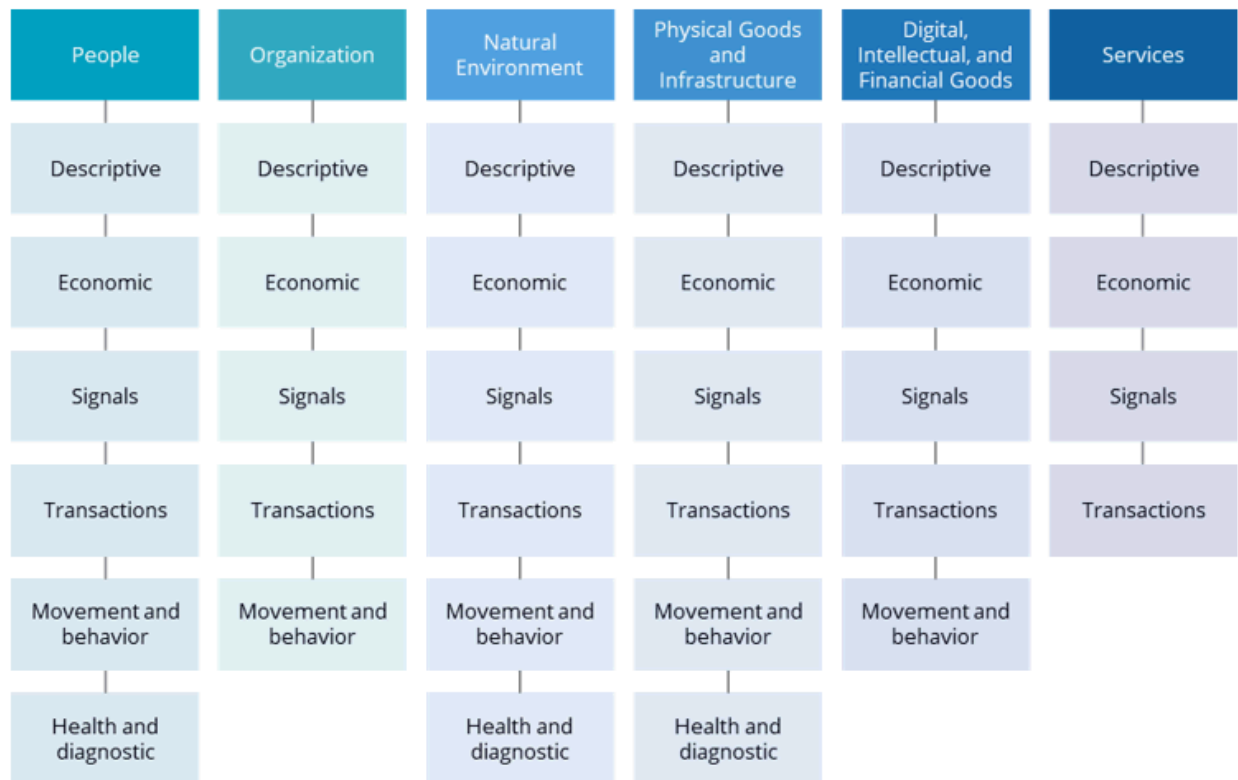Lynne Schneider   Dan Vesset

IDC'S WORLDWIDE DATA-AS-A-SERVICE TAXONOMY

FIGURE 1

Data-as-a-Service Primary Segments

Source: IDC, 2018

DATA-AS-A-SERVICE TAXONOMY CHANGES FOR 2018

This is the first taxonomy for the data-as-a-service (DaaS) market.

TAXONOMY OVERVIEW

This IDC study provides a detailed description of IDC's data-as-a-service market taxonomy. IDC organizes the DaaS taxonomy based upon the primary domain or subject of the data being provided, and further segments the market into multiple categories of DaaS within each domain.

This IDC taxonomy will serve as the basis for our ongoing sizing of the DaaS market. Such sizing will be based on the total revenue garnered by both data providers (owners) and intermediaries, such as marketplaces, that might facilitate the sale of DaaS. This does not include the software, hardware infrastructure, or professional services needed to transmit or otherwise maintain the data.

DEFINITION

Data as a Service

Data as a service (DaaS) is data delivery and access services that result in the trade in machine-readable data. These data can include raw data or value-added information derived from the raw data and/or from aggregated third-party data. The requirement for machine readability expressly excludes long-form text, pictures or photos, video, and sheet music or musical performances. However, IDC's definition of DaaS includes metadata from human consumable content such as books, magazines, newspapers, movies, TV, and recorded music but not that content itself.

The DaaS market is where an enterprise can monetize its data assets to external parties (or participate as a member of a data-sharing consortium). DaaS can be provided as a one-time sale or batch or real-time updates. It may be provided by the original "owner" of the data or by a third-party aggregator or marketplace.

DaaS providers are companies that are generating, aggregating, curating, analyzing, and adding value to information for transmission and licensing or sale to others. These providers include dedicated DaaS vendors as well as any entity across any industry that sells DaaS to external parties.

Some of the many use cases employ the data to augment organizations' internal data sets to improve business processes and decision making, train cognitive/AI solutions, and wrap around other product or services.

Market Boundaries

IDC's DaaS definition does not include the software, IT infrastructure, professional services, and internal labor costs to develop and distribute DaaS.

This market also excludes vendor services that create custom, original research for specific clients; for example, polling for a specific political candidate, employee or customer opinion polling, outsourced scientific or laboratory reports, and bespoke market analysis and sizing are all excluded. The results or outputs of these services could be part of the DaaS market, but only when packaged for sale to other audiences.

Logical Groupings

The goal of this taxonomy is to describe the breadth and variety of a market in a way that is meaningful for organizing its measurement and identification of solutions being provided and consumed. IDC defined six DaaS domains, each of which contains multiple DaaS segments. We also define categories as collections of segments across domains.

Domains

DaaS offerings are highly varied, and in the DaaS taxonomy, the domains are defined by what the data itself is about – its primary subject. For example, data in the physical goods and infrastructure domain is about the good itself, that is, what kinds of breads there are, how much bread is being purchased, price of bread sold, and who is buying bread. The people domain is about people who buy bread – and could include information about other things they buy when they buy bread.

The DaaS market domains are not segmented by the method with which the data is initially captured or the way in which it is provided as a service or the technical type of data (e.g., structured versus unstructured).

The domain definitions are agnostic to the industry of the entity that does the gathering. In this view, what matters most is the subject of the data set. The six domains of the DaaS market are shown in Table 1.

TABLE 1

DaaS Domains

Domain	Description
People	Data sets about people as individuals; the data may include personally identifiable information (PII) or it may be anonymized
Organization	Data sets about socially constructed organizations, such as businesses, governments, trade organizations, schools, or sports teams; organization also includes cities/towns, counties, and countries
Natural environment	Data sets about things in the natural world, including land and water, animals, plants, and weather
Physical goods and infrastructure	Data sets about physical goods, structures, and facilities (e.g., consumer goods, buildings, roads, capital equipment, and power supplies); this domain also includes all developed real estate properties, whether residential or commercial
Digital, intellectual, and, financial goods	Data sets about nonphysical goods such as literary works, images, patents, music, financial assets, laws, and regulations; this includes metadata about intellectual property but does not include the content of the intellectual property, such as music, text or patent drawings, and descriptions
Services	Data sets about services (as generally defined in economic literature), including business and personal services such as accounting, childcare, healthcare, and public utilities

Source: IDC, 2018

Categories

In addition to domains, IDC defines categories of DaaS that span multiple domains. When referring back to Figure 1, this would be providing DaaS that spanned across one or more rows horizontally rather than focusing on a particular domain (column). For example, a DaaS provider may include descriptive category data across both the people and organization domains.

There are some common elements to category definitions, shown in Table 2. Specific definitions of the segments (intersection of a domain with a category) are in the Detailed Domain Definitions section of this document.

TABLE 2

DaaS Categories

Category	Description
Descriptive	▪ Attributes of the data in that domain such as name, address, age, basic physical characteristics; includes basic relationships between data within the category (family relationships, corporate relationships, brands associated with a company, sensors within an array, etc.); also includes data about the creation or production of the item, particularly applicable to the goods and services domains ▪ Attitudinal data, including responses to surveys, polls, and other opinion-gathering methods; provides data about sentiments by asking the subjects rather than inferring attitude based on other characteristics (i.e., answers to political opinion polls rather than rating likelihood to vote a particular way based on demographic or behavioral factors)
Economic	▪ Financial characteristics of the data in that domain, such as income, net worth, credit score, and monetary value
Signals	▪ Signals (news, conference attendance, website mentions) about individuals or groups within the data set; examples include both traditional and social media mentions, product or service reviews, ranking lists, and other websites and promotional materials; the signal can be about connections between the individuals/groups, activities the individuals/groups undertake, evaluation of services/features, and so forth
Transactions	▪ Exchanges of intellectual and financial property, goods and services, or other property when organized by the data domain; may include an array of data about the transaction, including temporal, spatial, financial, and other factors; includes summary data of individual purchasing or selling behavior ▪ Consumption or use of natural resources, intellectual and financial information, goods, services, and installed industrial, infrastructure, or real estate property assets; examples include reduction of oil deposits, erosion, nonfinancial information about volume of uploads and downloads of content, food eaten, bus trips taken, and volume of concrete used
Movement and behavior	▪ Physical movement, tracked by any means that makes it directly attributable to the person, natural element, or good ▪ Also includes activities undertaken by people, organizations, or elements within the natural environment, such as physical exercise, eating, web searching, and smoking; also includes social interactions, whether digital or nondigital; includes the nature of the activity and time spent and may include location or other characteristics
Health and diagnostic	▪ Health and medical information; includes diagnoses, treatments, and outcomes; includes the environment within the natural environment domain ▪ Data would be captured by personal devices, electronic medical records, physician/hospital treatment and invoicing data, and sensors and remote monitoring devices ▪ Also applicable to the operations and health of machinery, electronics, or other goods with sensing and/or self-diagnostic capability

Source: IDC, 2018

Detailed Domain Definitions

This section of the taxonomy provides further definitions of DaaS domains and the specific segments within each domain. A segment is a category within a domain.

For each DaaS segment, IDC has provided a few sample commercial, governmental, or nonprofit DaaS examples. These are meant to be only illustrative of the specific segment within the domain. Longer lists of DaaS providers will be available from future IDC research reports and data products.

People

The people domain consists of data sets that are about people as individuals. It includes descriptive information about them, their transactions, movements, and other data that can be organized around the person (or groups of people) as the subject. The data may include personally identifiable information (PII), or it may be anonymized. The data is considered within the people domain if the data can be reasonably attributed to an individual (i.e., tracked by a cell phone or wearable device) that is not commonly shared. If there is some ambiguity (i.e., tracked by a vehicle, which could have different occupants), the appropriate data domain is that of the object doing the tracking.

This domain does not include people's performance as a member of an organization (i.e., an employee's skills or review ratings within a company or an athlete's statistics with a sports team), as these are included in the organization domain.

Table 3 provides a detailed description of each DaaS segment within this domain.

TABLE 3

People Segments

Segment	Description
Descriptive	- Attributes of a person such as name, address, age, basic physical characteristics, gender, and marital status - Also includes attitudinal data, such as responses to surveys, polls, and other opinion-gathering methods; provides data about sentiments by asking the subjects rather than inferring attitude based on other characteristics (i.e., answers to political opinion polls rather than rating likelihood to vote a particular way based on demographic or behavioral factors)
Economic	Data about financial characteristics of a person, such as salary, net worth, and credit scores
Signals	Signals (news, conference attendance, website mentions) about individuals or groups within the data set; examples include both traditional and social media mentions; the signal can be about connections between the individuals/groups, activities the individuals/groups undertake, and so forth; examples include current employment and employment history, school attendance, and alumni membership
Transactions	Data about exchanges of intellectual and financial property, goods and services, or other property when organized by person; may include an array of data about the transaction, including temporal, spatial, financial, and other factors; includes summary data of individual purchasing or selling behavior
Movement and behavior	- Data about physical movement of people, tracked by any means; includes people's movement within means of transit but not the movement of the vehicle or system - Data about activities undertaken by people, such as physical exercise, eating, web searching, smoking, social interactions (whether digital or nondigital), and criminal activity; includes the nature of the activity and time spent and may include location or other characteristics
Health and diagnostics	Health and medical information about people; includes diagnoses, treatments, and outcomes; data would be captured by personal devices, electronic medical records, physician/hospital treatment and invoicing data, and sensors and remote monitoring devices

Source: IDC, 2018

Example providers in this space include:

- TransUnion
- Kantar
- Gravy Analytics
- UnitedHealth Group
- Acxiom

Organization

The organization domain consists of data sets that are about socially constructed organizations, such as businesses, governments, trade organizations, schools, healthcare institutions, or sports teams. Organization also includes cities/towns, counties, and countries as social constructions.

Table 4 provides a detailed description of each DaaS segment within this domain.

Organization Segments

Segment	Description
Descriptive	▪ Attributes of an organization such as name, address, number of members/inhabitants, year of organization, and key individuals associated with the organization ▪ Also includes responses to surveys, polls, and other opinion-gathering methods; provides data about sentiments by asking the subjects rather than inferring attitude based on other characteristics (i.e., answers to economic optimism rather than using purchasing or other proxies) when answered by an individual on behalf of an organization
Economic	▪ Data about financial characteristics of an organization, such as revenue and credit scores
Signals	▪ Signals (news, conference attendance, website mentions) about individuals or groups within the data set; examples include both traditional and social media mentions, reviews, ranking lists, and other websites and promotional materials; the signal can be about connections between the individuals/groups, activities the individuals/groups undertake, and so forth examples include announcements of new corporate officers, hiring/layoffs in a geography, and ownership of equipment ▪ Reviews could be regarding the organization overall — the organization as an employer, reputation/value of an organizational brand or particular employees of the organization, and so forth; product or service reviews would be in the appropriate goods or services domain
Transactions	▪ Data about exchanges of intellectual and financial property, goods and services, or other property when organized by organization; may include an array of data about the transaction, including temporal, spatial, financial, and other factors
Movement and behavior	▪ Data about activities undertaken by organizations, such as hiring and layoffs, registration of patents, and adoption/implementation of new tools and technologies; for government organizations, includes the creation and enforcement of legislation and regulation; includes the nature of the activity and time spent and may include financial or other characteristics ▪ Includes crime statistics when organized by governing entity or other group; individual criminal history is classified in the people domain

Source: IDC, 2018

Example providers in this space include:

- Dow Jones DNA
- Dun & Bradstreet
- Government agencies
- Thinknum
- Onvia

Natural Environment

The natural environment domain consists of data sets that are about things in the natural world, including land and water, animals, plants, and weather. Geospatial and mapping data is also within this domain. The data sets in this segment include descriptive information about items in the natural environment and their transactions, movements, and other data that can be organized around the element of the natural environment (or groups of elements) as the subject.

Table 5 provides a detailed description of each DaaS segment within this domain.

TABLE 5

Natural Environment Segments

Segment	Domain
Descriptive	▪ Attributes of a natural element, such as type of element, location, volume, age, and ownership
Economic	▪ Data about values of natural elements, including animals, plants, arable land, water, solar energy, and mineral deposits
Signals	▪ Signals (news, conference attendance, website mentions) about individuals or groups within the data set; examples include both traditional and social media mentions; the signal can be about connections between the individuals/groups, activities the individuals/groups undertake, and so forth; examples include reporting on weather phenomena, natural disasters, wild animal population, and sea levels
Transactions	▪ Data about transactions of natural elements in exchange for payment or other goods/services when organized by natural element; this includes any commodities that are transacted in their raw form (i.e., amount of ore mined, sale of undeveloped land from one party to another) but not the processing of those items or their sale in a processed form (i.e., iron smelting, sale of ore, sale of a commercial property); may include an array of data about the transaction, including temporal, spatial, financial, and other factors ▪ This segment also includes the use or depletion of natural resources; examples include reduction of oil deposits, erosion, and animals herd reductions
Movement and behavior	▪ Data about the physical movement of animals, water, wind, and other natural elements that change their location ▪ Activities undertaken by animals, such as physical exercise, eating, mating, and migration; includes the nature of the activity and time spent and may include location or other characteristics ▪ This segment also includes weather data
Health and diagnostics	▪ Health and medical information about plants or animals; air, water, and other environmental quality data; and geological readings ▪ Plant and animal data includes vital statistics, test results, and treatment efficacy; data could be captured by (but is not limited to) devices attached to/implanted, electronic medical records, and treatment and invoicing data

Source: IDC, 2018

Example providers in this space include:

- IBM's The Weather Company
- Orbital Insights
- World Organization for Animal Health
- Global Forest Watch

Physical Goods and Infrastructure

The physical goods and infrastructure domain consists of data sets that are about any physical goods, things, physical assets, structures, and facilities (e.g., consumer goods, buildings, roads, capital equipment, and power supplies). The domain includes descriptive information about these elements and their transactions, movements, and other data that can be organized with this type of item (or group of items) as the subject.

The range of goods covered in this domain is broad and includes soft goods and hard goods, consumables, and capital equipment.

Table 6 provides a detailed description of each DaaS segment within this domain.

TABLE 6

Physical Goods and Infrastructure Segments

Segment	Description
Descriptive	▪ Attributes of a physical good, piece of installed industrial equipment, infrastructure or developed real estate property such as name, producer/developer, owner, description, age, or other identifying information ▪ Data about the production of the good or its installation; examples include inputs, volume produced, and time to produce using different equipment
Economic	▪ Data about financial characteristics of a good or a piece of industrial, infrastructure, or real estate property such as cost of production or economic value
Signals	▪ Signals (news, website mentions) about a specific good or class of goods; examples include both traditional and social media mentions; the signal can be about connections between the individuals/groups, new offerings, product support, recalls, and so forth ▪ Also includes product reviews, rankings, and directories
Transactions	▪ Data about exchanges of goods and services when organized about that element; may include an array of data about the transaction, including buyers and sellers (owners), temporal, spatial, financial, and other factors ▪ Also real estate transactions, construction contracts, and other exchanges of capital goods, industrial and infrastructure property when organized about that property ▪ Consumption or depletion of goods; examples include food eaten, volume of concrete used, and number of automobile crashes; also use or depletion of the installed industrial, infrastructure, or real estate property assets that is nonfinancial in nature (not accounting treatment of depreciation)
Movement	▪ Data about physical movement of goods as work in process or finished goods; includes the element's movement within means of transit; movement of a vehicle or transit system is also included but based on the vehicle or transit system as the organizing element ▪ Also physical movement of any of the other segments of data, when organized by the industrial, infrastructure, or developed real estate item; examples include volume of trains stopping at a transit station, foot traffic to a particular piece of real estate, and number of vehicles traversing a toll road
Health and diagnostics	▪ Data from process and machine sensors relating to the status of the good, production and usage processes, and the good's own diagnostic data; examples include flow meter readings and variance reporting ▪ This does not include information that is used to track the health and diagnostics of people or elements of natural environment; those are considered within their respective data domains

Source: IDC, 2018

Example providers in this space include:

▪ IRI
▪ INRIX

- Uber Movement
- CoStar
- IDC

Digital, Intellectual, and Financial Goods

The digital, intellectual, and financial goods domain consists of data sets that are about nonphysical goods such as literary works, images, patents, and music. It also includes financial instruments such as currencies, stocks, and bonds. Laws and regulations are also part of this domain. The domain includes descriptive information about these elements and their transactions, movements, and other data that can be organized with this type of item (or group of items) as the subject.

This includes metadata about intellectual property but does not include the content of the intellectual property, such as music, text, or patent drawings and descriptions. In other words, IDC's DaaS market sizing is not intended to count the box office value of movies or revenue derived by streaming content services. It would, however, include the sale or licensing of data about box office sales, and so forth.

Example types of subdomains include:

- Legislation and regulation
- Financial
- Media and entertainment content
- Intellectual property
- Scientific research

Table 7 provides a detailed description of each segment within this domain.

TABLE 7

Digital, Intellectual, and Financial Goods Segments

Segment	Description
Demographic and descriptive	▪ Data about attributes of a piece of intellectual property or financial instrument such as name, author/vendor, description, age, or other identifying information; includes metadata or descriptors accompanying written or spoken works, imagery, video, or music (e.g., labeled content) ▪ Information about the creation of intellectual and financial items, including the inputs, time expended, and human and machine resources
Economic	▪ Data about financial characteristics of a piece of intellectual property or financial instrument such as cost of production or economic value; includes the value of an item relative to another, such as a currency exchange rate
Signals	▪ Signals (news, conference attendance, website mentions) about individuals or groups within the data set; examples include both traditional and social media mentions; the signal can be about connections between the individuals/groups, activities the individuals/groups undertake, and so forth; examples include announcements of currency devaluations, adoption of textbooks, and new regulation ▪ Also includes reviews and rankings of items within this domain
Transactions	▪ Data about exchanges of intellectual and financial property when organized about that property; may include an array of data about the transaction, including buyers and sellers (owners), temporal, spatial, financial, and other factors ▪ Includes registrations, such as patent applications and trademark registration ▪ Use or depletion of intellectual or financial items; examples include nonfinancial information about volume of uploads and downloads of content ▪ Also includes data from scientific experiments and clinical trials
Movement and behavior	▪ Data about physical movement of intellectual and financial property, in terms of storage and transmission

Source: IDC, 2018

Example providers in this space include:

- ▪ Netra
- ▪ Data Simply
- ▪ App Annie
- ▪ BuzzAngle Music

Services

The services domain consists of data sets that are about a broad array of services including business function services, childcare and education services, construction, legal services, entertainment, transportation, telecommunications, public utilities, medical, hospitality, and retail and food service.

The domain includes descriptive information about these elements and their transactions, movements, production, and consumption information that can be organized with this type of item (or group of items) as the subject.

Table 8 provides a detailed description of each segment within this domain.

Services Segments

Segment	Description
Descriptive	▪ Data about attributes of a service such as name, producer, description, varieties, or other identifying information
Economic	▪ Data about financial characteristics of a service such as cost of production or economic value
Signals	▪ Signals (news, website mentions) about a particular service or group of services; examples include both traditional and social media mentions; the signal can be about connections between the individuals/groups, activities the individuals/groups undertake, and so forth; examples include announcements of new service launches, new pricing arrangements, or awards ▪ Also includes reviews and rankings of services within this domain
Transactions	▪ Data about exchanges of services when organized about that element; may include an array of data about the transaction, including buyers and sellers, temporal, spatial, financial, and other factors ▪ Also includes data sets about the volume of services consumed/used

Source: IDC, 2018

Example providers in this space include:

- Wharton Research Data Services
- Staffing Industry Analysts
- WorldACD
- Bureau of Labor Statistics

Alternative DaaS Market Groupings

This study also suggests a list of alternative ways to look at the DaaS market. For sales purposes, some vendors in this space choose to market their offerings by the use cases rather than focusing on a specific domain, category, or market segment. This is not IDC's taxonomy structure because we believe that the rapid development of this market will lead to ongoing change with these use cases and applications for DaaS.

Industry

The DaaS market can be segmented by industry — an alternative view that combines whole or fractions of segments associated with a particular industry. An example would be a view of the automotive

DaaS market that might contain data about cars sold, raw materials purchased and consumed, patents filed, and so forth.

Use Cases

Use cases are another alternative lens to examine the market. Use cases are combinations of the whole or fractions of the domains and categories that reflect such market dynamics as the business process that is being augmented or the type of AI system that is being trained. Use cases can be industry specific (vertical) or cross-industry functional (horizontal). This is also an area where buyers will seek to make unique combinations that provide competitive advantage for them.

Examples of Industry and Use Case Markets

- **Business-to-customer marketing** – Providing mailing lists, audience segmentation, or other customer preference markets that have long been delivered by data management platforms (DMPs) and used by advertising agencies
- **Financial and energy trading desks** – Combinations of economic, production, and emerging use of news and social media platforms to inform transactions
- **Tourism and travel** – Adding weather and additional travel service availability and ratings to flight booking applications

Market Roles

Market roles are an alternative sorting mechanism for examining the DaaS marketplace, which categorizes the types of service providers. IDC defines three market roles, and those are based on the role of the vendor in the DaaS marketplace. Table 9 provides an overview of the market roles and segment definitions. The revenue from contributor roles are not included in market sizing.

TABLE 9

DaaS Market Roles

Market Role	Definition
Provider	▪ Producer or distributor of data that sells or licenses to paying users; providers may be ingesting, organizing, cleansing, and tagging data from others to provide DaaS offerings for sale or license; the data may be raw or enriched in some way, but the term *provider* only applies to the entity that is selling the data ▪ It is possible for a provider to sell or license only via data marketplaces and/or brokers; these are most likely to be enterprises engaged in a noninformation industry as their primary industry
Marketplace/broker	▪ A medium that allows buyers and sellers of data to interact in order to facilitate exchange; a marketplace is any medium through which two or more parties can engage in an economic transaction, even those that do not necessarily need to involve money; these marketplaces and brokers facilitate the exchange and may set prices or other rules ▪ IDC's definition only includes the legal trade of data
Contributor	▪ Provides service, support, software, and/or hardware to facilitate provision or ingestion of DaaS; the contributors provide necessary tools and services for the DaaS market to function but are not selling or licensing data ▪ Contributors' revenues are not included in market sizing for DaaS research

Source: IDC, 2018

LEARN MORE

Synopsis

This IDC study provides a detailed view of IDC's data-as-a-service (DaaS) market taxonomy. IDC's taxonomy incorporates six major DaaS domains, each with several categories of data. This taxonomy describes each of the DaaS domains, categories, and segments with the goal of providing DaaS providers and consumers with a structure that supports DaaS market suppliers' and consumers' decision making about participation in this market.

"IDC's Worldwide Data-as-a-Service Taxonomy provides a comprehensive framework of the data-as-a-service market, used to support a wide variety of industries and use cases," said Lynne Schneider, research director, IDC's Data as a Service. "As enterprises continue their drive for more intelligent decision making and greater knowledge, the data-as-a-service market provides an opportunity for data monetization and data consumption."

About IDC

International Data Corporation (IDC) is the premier global provider of market intelligence, advisory services, and events for the information technology, telecommunications and consumer technology markets. IDC helps IT professionals, business executives, and the investment community make fact-based decisions on technology purchases and business strategy. More than 1,100 IDC analysts provide global, regional, and local expertise on technology and industry opportunities and trends in over 110 countries worldwide. For 50 years, IDC has provided strategic insights to help our clients achieve their key business objectives. IDC is a subsidiary of IDG, the world's leading technology media, research, and events company.

Global Headquarters

5 Speen Street
Framingham, MA 01701
USA
508.872.8200
Twitter: @IDC
idc-community.com
www.idc.com