

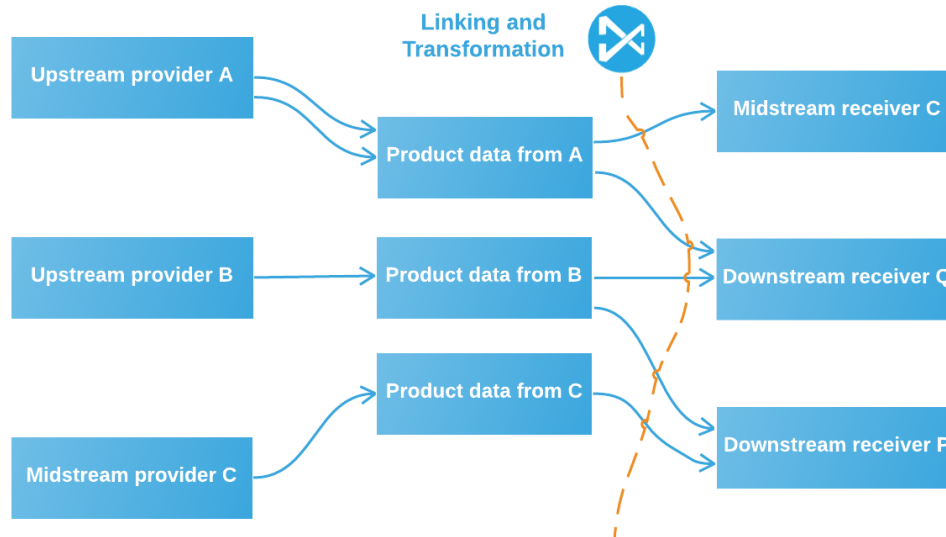
Due to the increasing use of e-commerce and other customer self-service sales approaches, the sharing of product data within business ecosystems of manufacturers, distributors, retailers and end users has grown dramatically. Most initiatives around handling product data has been focussed on internal processes and technology while **solutions that solves the problems in the exchange zones between trading partners has not been covered very well.**

The recent years has also seen an increased use of cloud based data management technology as well the rise of the term big data. A data lake is a concept with growing popularity related to big data. It focuses less on predefined standards, but makes variant use of data by the time needed possible. The data lake concept is, in its original form, a concept for mainly analytical use of big data within an enterprise. **The Product Data Lake takes this modern technology concept into the space of sharing product data between trading partners.**

By doing that, the **Product Data Lake solves major business issues:** It is a single trusted service where businesses can share consistent product data and tackle data quality challenges and share the burden of software licenses. All in all the Product Data Lake enables end customer self-service by **automating the chain of supplier self-service and B2B customer self-service for product data.**

## Solution brief

The service is a Software-as-a-Service offering where the Product Data Lake is a **cloud-based solution utilized by both the trading partner that provides product data and the trading partner that receives the product data.** The Product Data Lake will catch new product data updates from an upstream partner who subscribes to the system, link and transform the data tagging according to a set up and distribute the product data to those downstream partners who also subscribe to the service.



The Product Data Lake is built to ensure:

- **Completeness** of product information by enabling trading partners to exchange product data in a uniform way
- **Timeliness** of product information by connecting trading partners in a process driven way
- **Conformity** of product information by encompassing various international standards for product information
- **Consistency** of product information by allowing upstream trading partners and downstream trading partners to interact with in-house structure of product information
- **Accuracy** of product information by ensuring transparency of product information across the supply chain

## Conceptual model

The system handles the following entities related to sharing product information in business ecosystems:

**Parties** as manufacturers, distributors (wholesalers), retailers and large end users of product information as well as service providers within the data supply chain. All participating parties are registered with a legal entity identifier (the Duns number).

**Profiles** where a subscribing party can have one or several profiles representing different scenarios as an upstream or downstream role, different internal source systems, file formats and other characteristics.

**Products** also sometimes referred to as articles or materials. They can be treated as a flat model with only products or a hierarchy of products and their various stock keeping units (SKU/ item). Information can be exchanged between both models. The products common between trading partners can be linked either manually or using national or international product identification systems.

**Attributes** also known as features or properties. The attributes used by each partner in a trading partnership can be linked either manually or using references to national or international standards. Attributes can also be references to national or international product classification systems.

**Related products** as the Product Data Lake handles a range of associations between products and items, and transforms those using the product links set up between trading partners. Such relationships includes accessories, replacements, spare parts, bill of materials (BOM), cross sells and up sells.

**Digital assets** sometimes called rich media, mimes or blobs. This is product images, product sheets, videos and heaps of other documents. Their types can be linked between trading partners and exchanged as physical documents or references to those documents.

Subscribing parties and reservoirs makes partnerships in the Product Data Lake. Products, attributes and digital assets can be shared with all partners or with a specific partner.

## Participating parties

Three main kinds of parties participate in the Product Data Lake:

- **Subscribers**, who are the manufacturers, distributors, retailers and large end users of product information who also deals with the physical products
- **Ambassadors**, who are product information management (PIM) system vendors, system integrators and other professionals who supports the subscribers in managing product information
- **Reservoirs**, who are service providers that hosts collections of product data within a given industry and/or geography, or for a given purpose.

Subscribers can be:

- **Upstream**, typically a manufacturer or upstream distributor
- **Downstream**, typically a downstream distributor, retailer or end user
- **Midstream** (both upstream and downstream), typically a midstream distributor, or being upstream for some products as a manufacturer of direct finished products and downstream for raw materials and indirect goods

A reservoir can act with the functionality of a midstream subscriber and an ambassador.

## Enabled business processes

**For upstream subscribers** the Product Data Lake covers the following business processes:

- When you introduce new products to the market, you make the related product data and digital assets available to your downstream partners in one uniform way
- When you win a new downstream partner you have the means to immediately and professionally provide product data and digital assets for the agreed range
- When you add new products to an existing agreement with a downstream partner, you are able to provide product data and digital assets instantly and effortlessly
- When you update your product data and related digital assets, you have a fast and seamless way of pushing it to your downstream partners
- When you introduce a new product data attribute or digital asset type, you have a fast and seamless way of pushing it to your downstream partners
- An upstream provider may push product data and digital assets from several different internal sources

**For downstream subscribers** the Product Data Lake covers the following business processes:

- When you engage with a new upstream partner you have the means to fast and seamlessly link and transform product data and digital assets for the agreed range from the upstream partner
- When you add new products to an existing agreement with an upstream partner, you are able to link and transform product data and digital assets in a fast and seamless way
- When your upstream partners update their product data and related digital assets, you are able to receive the updated product data and digital assets instantly and effortlessly
- When you introduce a new product data attribute or digital asset type, you have a fast and seamless way of pulling it from your upstream partners
- If you have a backlog of product data and digital asset collection with your upstream partners, the Product Data Lake is the fastest and most cost effective approach to backfill the gap

## Data exchange

**Uploading data** to the Product Data Lake upstream is done by using hot folders on FTP sites, by importing files inside the Product Data Lake or by manually typing and twisting data in the Product Data Lake user interface.

**Downloading data** downstream is done by exporting files inside the Product Data Lake or by using a pull request. A pull request is an order to provide selected data on a FTP site at given time. The pull request can be repeated daily, weekly or monthly.

All attribute values can be uploaded and downloaded represented in a given culture, which is a combination of a country and a language. Different character sets for upload and download is available too.

## Technical information

As a participating party, you see the Product Data Lake as a cloud service where you should not worry about the technology behind.

However, the technical interested may want to know the following:

- The Product Data Lake is hosted on *Amazon Elastic Computing Cloud* using the *Amazon Linux* servers in Frankfurt, Germany
- The data is stored using *MongoDB*
- Other technologies used includes *Rails*, *Ruby*, *Golang*, *RabbitMQ* and *ElasticSearch*

## Levels and prices

For subscribers, the levels and related conditions are:

Level	Partners	Products (SKUs)	Digital assets	Users
S	1	1,000	2,000	2
M	10	5,000	10,000	4
L	100	25,0000	50,000	8
XL	1,000	125,000	250,000	15
XXL	10,000	500,000	1,000,000	30
XXXL	100,000	2,000,000	4,000,000	60

Yearly prices in selected currencies are:

Level	EUR	GBP	USD	DKK
S	400	300	450	3.000
M	1.300	1,000	1,500	10.000
L	4.000	3,000	4,500	30.000
XL	13.000	10,000	15,000	100.000
XXL	40.000	30,000	45,000	300.000
XXXL	130.000	100,000	150,000	1.000.000

A trial level with same threshold as the S (Small) level is available for two months.

## Vision and mission

Our vision is that the Product Data Lake will be the process driven key service for exchanging any sort of product information within business ecosystems all over the world, with the aim of optimally assist self-service purchase of every kind of product.

Our mission is to enable automated cross company data supply chains for product data by using technology that handles large volumes of product data and related digital assets, with the velocity needed for changing business environments and encompassing the variety of international, national, industry and other standards for product information.