Capability: Designing and Managing Data Products

Description

Data product defined

A **data product** is a rational and governed collection of data and associated data artifacts managed together under the *'Data Product Criteria'* section.

An additional important concept is that of '**data as a product'**. This refers to the importance and priority in the way that an organization considers data and its use and value. Organizations that treat data as a product as part of the organizational culture that includes data as a first class citizen and product management concepts are typically more successful in driving data product value. Consequently, organizations that focus on objectives and resources (e.g. organization structure, planning, operational processes etc.) to support that primary concern, achieve the optimum outcome more often.

Data Product vs. Data Asset

Data product is a data asset but not all data assets are a data product. Data asset - collection of data owned by an org that has intrinsic value [CDMC].

Data Product Criteria

- Collection of data and associated data artifacts
- A value on its own and can be consumed (easily used, easily applied) to solve business problem(s)
- Define Purpose: Primary usage - bounded (value propositions - specific use cases), explainable (easily understood), discoverable (easy to find), usability (governance, documentation, value prop), accessible
- Data Product Lifecycle - details below
    - Ideation
    - Design
    - Build
    - Deploy
    - Consume
    - Retire
- Consistently managed and governed
- Terms of use/categories of purpose – may include what you can do with the product (rights) - assembly/application instructions, access to the data product, functionally defined, and restrictions of use according to purpose.
- Clear ownership, accountabilities, responsibilities - provider (inputs, outputs and distribution) and consumers
    - Data consumer
        - Consumer of the data product/data domains

        - A process, application or stakeholder that receives or uses data from a data producer. (EDMC Business Glossary)
    - Data producer/provide - includes business and technical roles
        - Producer of data components that feed the data product/domain
        - A process, application or stakeholder that provisions data to one or more data consumers (EDMC Business Glossary)
    - Data product team
        - Product owner, engineers, analytics engineers, et al
- Value defined - Addresses a specific need. Data has value if it can be measured. Additionally, data has value when it has been consumed and an outcome or benefit has been identified.
    - Value defined examples include:
        - Commercial Value: increases revenue or profit
        - Efficiency Value: reduces costs (e.g., money or time)
        - Market Value: attracts more users or customers
        - Customer Value: increases likelihood that a customer continues to use a product
        - Risk-Reduction Value: mitigates reputational and/or financial risk
        - Reputational Value: increase brand, customer loyalty/trust
- How to manage different types of products - end delivered data product owner - touched it last principle who owns the responsibility
    - Any transformation of existing data products may be defined as a new product and therefore ownership is for the person that transformed the existing data product - can be a new product or obtain an agreement of the new data that improved the existing data product to take ownership.
    - Data products can be componentized - e.g. one or more data products combined to create new data product
    - Data product components are focused on the output of the data product
    - Data products can be created from data resources. It is up to the data product owner to determine the input data is fit for purpose (data quality, context, value). If Data is not fit for purpose, the data product owner is responsible to work with the input data owner to improve the data to meet the requirements
    - Data products are classified into data components by use case.
        - Example: Dun and Bradstreet
            - Internal to Dun and Bradstreet can be an aggregated data product
            - External to Dun and Bradstreet can be a referenced data product to the company that purchased that service.

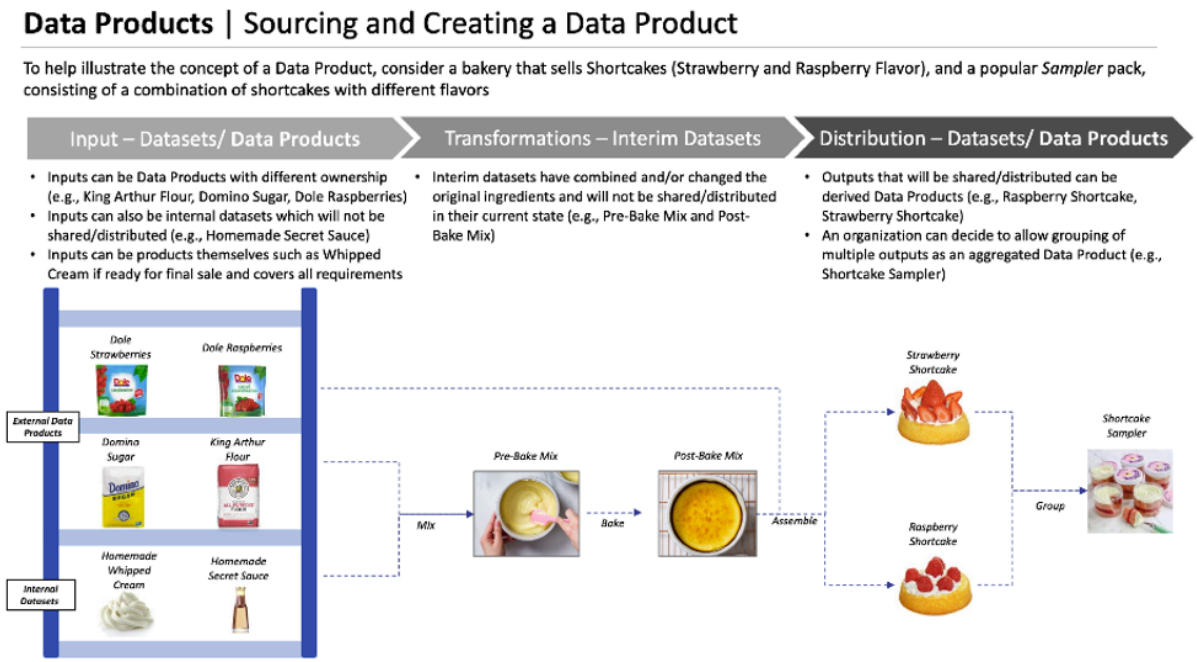
Data Product Examples

The intent of this section is to provide a visual representation of a data product to show examples of data products development.

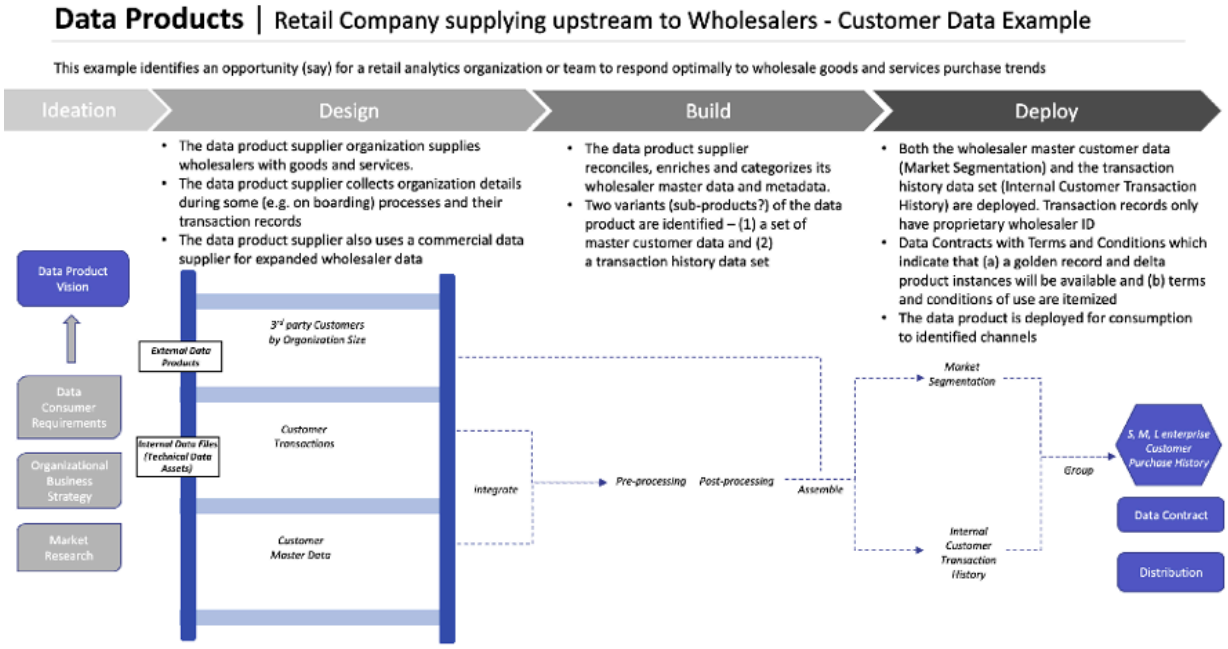
Bakery Example

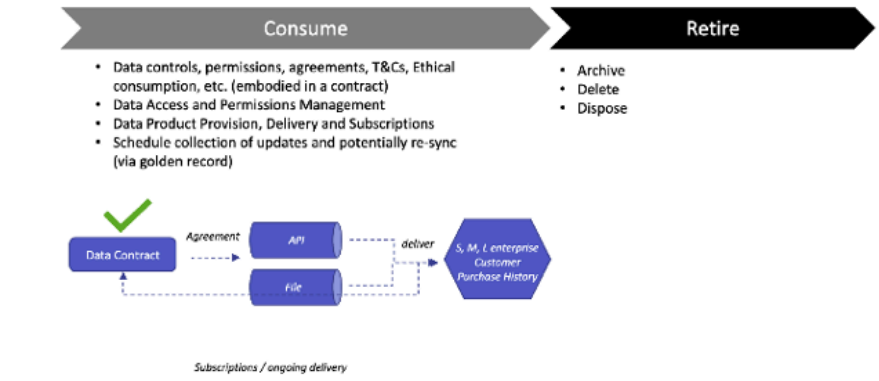
To simplify a complex topic of a data product, using a bakery analogy provides an overview of how a bakery product relates to a data product.

Customer Data Product Example



The following visualization provides a lifecycle for a retail company that supplies upstream to wholesalers





Data Product Objectives

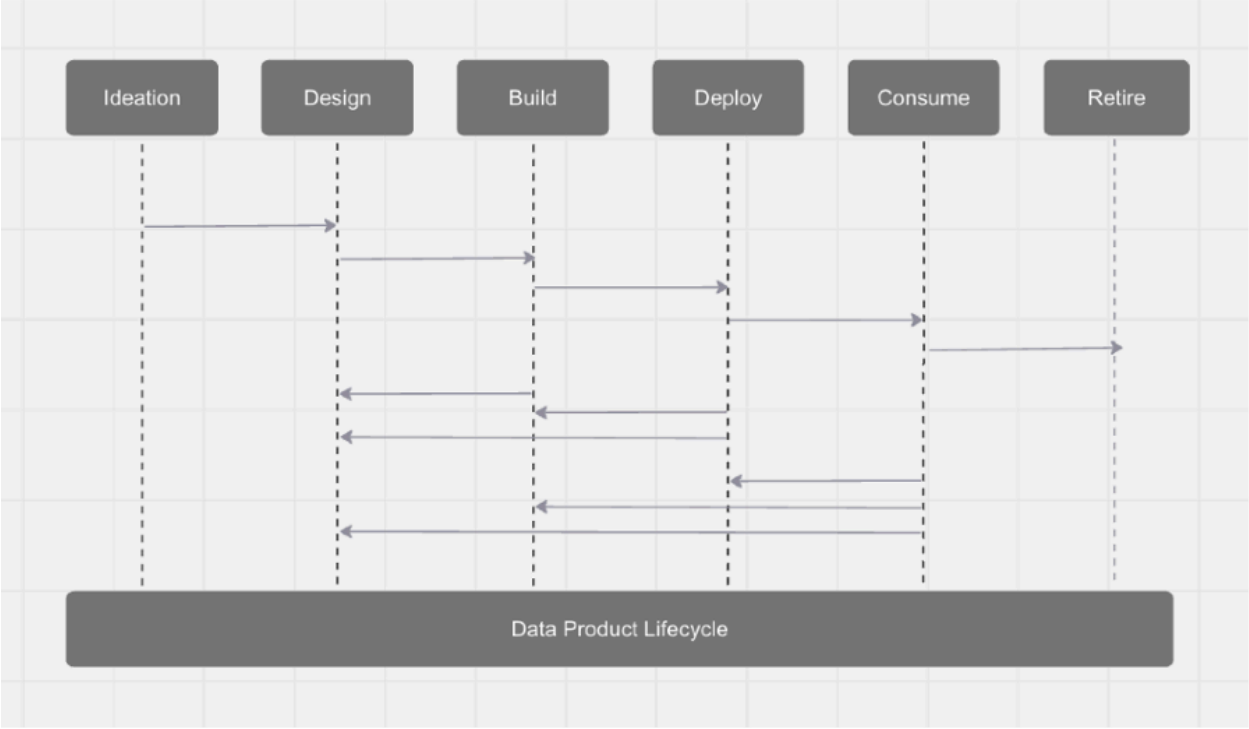
Data products become a first class corporate citizen that results in leadership focus, proper investment, and cross functional collaboration to drive value

- Data products become a first-class citizen
- Instill a sense of product thinking for data practitioners
- Reduce compliance risk, cost and carbon footprint by identifying and remediating ROT (redundant, obsolete, trivial) data
- Manage observability of the data product
- Ensure regulatory and corporate compliance requirements are met.
- Increase Data Products usage by enabling supportability (validate and communicate quality), access, understandability, findability, discovery to data and streamlining the Data Product lifecycle process including operationalize data sharing and delivery
- Enable business leaders to make data driven informed decisions boosting efficiency/decreasing costs, increasing corporate profitability by monetizing/valuating data products to drive revenue/create new revenue streams and maintaining corporate relevance.
- Increase or maintain the value of the data product by developing and measuring Objective and Key Results (OKRs)

Data Product Life Cycle

The Data Product Lifecycle is an iterative process and isn't always linear. It has the potential of going back to a preceding step in the life cycle if needed, even moving back to the Ideation stage.

Part of the data product lifecycle is to promote data strategy as part of the corporate strategy planning vs. being an afterthought. Data products can be requested from data consumers as well as being a part of the corporate strategy. Data products can support or drive corporate strategies.



Ideation

Identify a business need and conceptualize an idea.

- Data Product vision
    - Define data product's purpose and business value: This should be continuously evaluated.
        - The intended purpose is drafted
        - Conceptual understanding of the data product consumer
        - Estimate data product's value proposition (not restricted to a financial measure)
        - Data product type (model, visualization, dashboard, other)
        - Explicit to the need of the data product vs. data product definition – what is the problem that it is trying to solve versus what it is
        - Data may or may not be existing to ideate a data product
        - Target marketplace considerations
        - Viability/feasibility of creating and offering the data product
        - Incorporating user research and testing, as well as monitoring performance and usage metrics.
        - Incorporate insights from data product consumption (should be defined within the usage terms with the consumer) on how the product is being used (i.e. composability, etc.), which speaks to measures of ROI and intended use (purpose).

- Operational Business Strategy
    - Data product team organization considerations
        - Operating model (building and supporting the product)
        - Governance (when is a product considered complete)
    - Review process for validating data product viability - variable approval process depending upon total cost of operating (TCO), complexity and regulatory requirements
    - Key Performance Indicators (KPIs) or Objective and Key Results (OKRs)
    - Return on investment (ROI)
        - Pricing strategy or charge back model
        - Total cost of ownership including cost of creating, managing and retiring a data product.
            - Hardware, software, licensing, data acquisition cost, legal/contractual cost, resource costs (technical, business)
- Market research
    - Identifying data consumers, buyers, and other stakeholders

Design

Design the data product to align with the business need identified in Ideation

- Product Architecture, Design and prototyping
    - Data product owner collaborates with data product technical lead, designers, developers, and other stakeholders to build and iterate on the product
- o Specification
    - Sourcing and curation
        - Data discovery and exploration
            - Search and use of existing data products or data assets including registration of data products to inform consumers. Creation of data marketplace, data catalogs, and other methodologies of being able to “shop and search” for data products. The sourcing and curation should include metadata that identifies a data product explicitly.
        - Ensure data used in data products will be fit for purpose, relevant, and up-to-date (data observability).
            - The notion of ‘fit-for-purpose' implies:
                - Data quality - relevant to the data itself as well as the data product.
                - The statement of, monitoring of, reporting of data quality must be adequate for the correct procurement, use and support of data products.
                - Data product quality refers to the offering, supply and support of data products

        - Quality of data itself is both quantitative against some set of rules as well as qualitative through some understanding of the benefit of consumption of the data or data product.
            - Context - semantic description to understand the various uses of the data product and what it actually is
        - Value - understanding usage and constant measurement to ensure the value of the data product - will the product work for the consumer?
    - Regulatory compliance, data privacy, data security, and other business rules defined
    - Outline ethical usage and potential inherent bias'
- Feedback model
    - Data product owner goes to the data producer(s)/data product technical lead when the design is not fit for purpose
- Data product meta-model - basic structure of a data product
    - Appropriate metadata of the data product to be captured and made transparent and available to consumers.
    - Metadata collected aims to achieve trust, understanding and confidence in a data product.
- Consumption mechanism
    - Multiple modes of delivery/consumption can be established for a given data product
- Setting measurable metrics on usage of the data product to determine ROI (consumer and producer) and identify product improvement opportunities
- Incorporate insights from data product consumption on how the product is being used (i.e. composability, etc.).
- Data product owners should attest and be accountable to internal lineage management including authoritative sources that are used and applicable
- Develop of retention policies for the data product where applicable
- Retirement of data product process for replacement of the data product itself or for authoritative sources that may change.

- Confirmation and potential usage with other use cases validating other business values including ROI.

Build

Build must cover each deliverable for a data product to be consumed and cover all aspects of the design phase.

- Build
    - Development of data products can be developed from an IT or a business perspective
        - IT perspective

        - Via CI/CD (continuous Improvement / continuous development) means that includes supporting metadata to manage the data product lifecycle.
        - IT may have their own products, and a marketplace that is separate from the business
    - Business perspective
        - Interoperability - ownership of data product implementation should work cross functionally to validate data product implementation
    - Data quality measurements that are fit for purpose as defined by the data producer and accepted by the data consumer. Data Quality must be transparent and have a defined feedback loop that aligns with the data usage agreement. Data quality must be at the data product level
        - Monitoring of data product drift (changes over time) and outlier validation (data that can impact the data product due to an anomaly) as part of the data being fit for purpose
        - Data quality includes observability (is the data refreshed, volume of the data, change in schema, etc) aspects as well
        - Data quality issue management - go/no go decisions by data product owner are logged and communicated prior to putting into production (marketplace), as well as future ongoing data quality management of the product
    - Business and Technical Lineage (inclusive of data mapping and traceability)
        - Data lineage must be transparent and have a defined feedback loop that aligns with the data usage agreement but has a different level of visibility and granularity with data lineage (Provenance)
        - Data and business processes
            - Level of lineage may be at different levels of visibility identifying source, transformations, and accessible landing zone for consumer accessibility
            - Lineage detail level may depend on the consumer of the data product
                - IT/Data Science personas may require technical data lineage for root cause analysis to resolve DQ and context of data issues
                - Business may require only the data product level back to the data or process level
        - Automation is necessary to ensure accuracy, up-to-date lineage
    - Cataloging, publication and transparent accessibility of Data products
    - Explicit metadata identification of sensitive data
    - Explicit metadata identification that a data asset is a data product by the data owner
    - Construct/develop the data product per finalized requirements including collecting/harmonizing needed data and any development work that may be required

        - Monitoring and usage of data products that includes validation of the data contract on how to use the data product. Data consumers could report back on the data usage and any marketplace changes that may affect the validity of the data product. Incorporate insights from data product consumption on how the product is being used (i.e. composability, etc.)
        - Data products must adhere to internal data protection and external regulatory policies
    - Quality assurance ensures that the product functions as intended

Deploy

Data Product producer's focused requirements

- Data product is published in Data Marketplace
    - A place to shop or share for data products providing
        - List/publish data product to the directory/registry to allow for self-service, discovery, fulfillment. The data product must also have metadata describing it
        - Provide means of registration to obtain access to those data products
- Define and associated Data Sharing Agreement/Data Contracts - see Governing Data Sharing
    - Sharing data products internally or externally - define regulatory and internal policy compliance for data consumers [Reference Data Sharing CDMC+]
    - Data retention policy of the data product should be part of the agreement/contract
- Feedback mechanism for the consumer to ensure the data product is delivering the requirements under the data sharing agreement/data contracts
    - Issue management process for closed loop from issue creation to issue resolution
- Data product is deployed for consumption to identified channels
- External consumer cost - price to use the product or revenue stream for data producer
- Internal cost for data producer
    - Total Cost of Ownership (TCO) of the data product including data creation, data production and data maintenance
- Develop best practice - publish data product guides that include objectives of the data product, how to use the data product, regulatory considerations, access controls, identify critical assets, monitored critical assets, etc.
    - Data Product owners are accountable for ensuring data product guides are up-to-date
    - Change management process to enable communication on data products

Consume

Consumer focused requirements where the consumer finds and access the data product

- Data discovery mechanism via "Data Marketplace"
- Data product provision, delivery and subscriptions

    - Schedule collection of updates
- Data Sharing Agreements/Data Contracts
    - Data controls and permissions, agreements, ethical consumption
    - Terms & Conditions, Service Level Agreements
    - Data access and permissions management
    - Permitted usage of the data product to comply with the intent and purpose defined of the data product
- Feedback mechanism for the consumer to provide if the data product is delivering the requirements under the data sharing agreement/data contracts
    - Issue management process for closed loop from issue creation to issue resolution
- Price or consumer cost to use the product
- Measure return on Investment (ROI) of the data product to ensure it meets business objectives
- Leverage/consume/follow best practice - publish data product guides that include objectives of the data product, how to use the data product, regulatory considerations, access controls, etc.
- Identification of potential alternative usage of data products as well as changes to the data products to enhance the data product. Feedback mechanism from ongoing use of the data product
- Insights on how many users, how much, how often they are using, how they are using it. Stickiness metrics: Daily active vs monthly active
- Consume best practice - review data product guides that include objectives of the data product, how to use the data product, regulatory considerations, access controls, identify critical assets, monitored critical assets, etc.

Retire

Data product retirement components

- Adherence to retention policies for the data product
    - Data consumer and data producers may have separate retention policies for data products
- Remediate redundant, obsolete, and trivial (ROT) data products
    - Archive - copy of data that is put into long-term storage. The original data may or may not be deleted from the source system after the archive copy is made and stored, thought it is common for the archive to be the only copy of the data [CDMC 6.0 Data & Technical Architecture]
    - Delete - remove, erase or overwrite the data
    - Dispose/Destroyed - removed from operational use and is not recoverable by any forensic means
- Changes of authoritative sources for data products that are no longer accessible should be communicated to data consumers and to associated products.

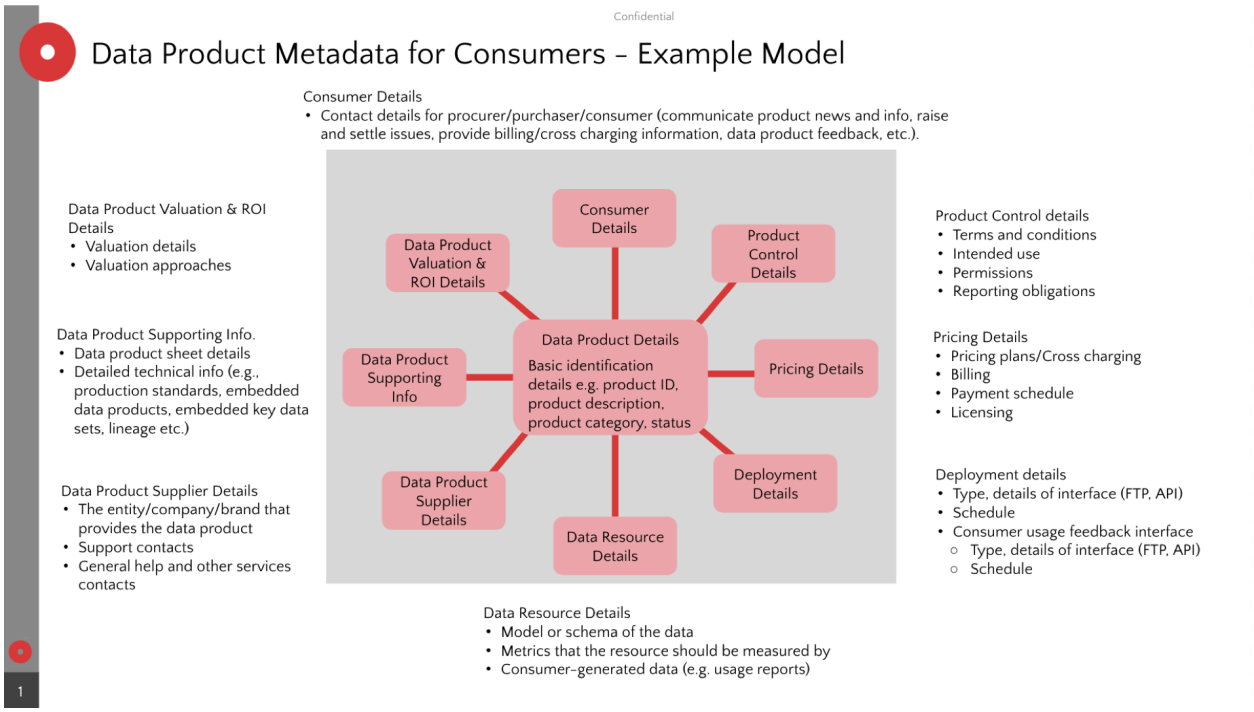
Data Product Metadata for Consumer - Example

Metadata is critical to managing data products in order to ensure the data product is identifiable, fit for purpose, used as intended, and provide support for the consumer. The visual below highlights key categories for data product metadata models and is not meant to be a comprehensive list of metadata required. The value of the metadata model is to highlight the different categories to managing a data product to be used as a high level overview of different requirements. Design phase should address each component and may differ by data product.

Different personas will use different metadata across technical, business, and operational metadata. Metadata accessibility may be set by persona to drive their requirements and simplify for the persona consumption as well as secure metadata.

The data product details are part of the data product itself vs. other metadata components that support the data product.

Meta model below starts once the data product is selected.



Advice for Data product practitioners

- Integrate data product into data management framework, business strategy and processes
- Educate stakeholders on product lifecycle or get a practitioner
- Recognize org maturity around data literacy, product management, change management, analytics and governance programs
    - If programs exist, us existing programs
    - If programs do not exist, then develop specific requirements to ensure data product outcomes
        - Identify data product development and take data product inventory that exists across the org. Align those functions with the data product framework developed in this document
        - Align data product stakeholders to manage data product outcomes
        - Socialize data products
        - Develop escalation path and approval framework defined that aligns to existing analytics and governance frameworks
        - Identify influencers and supporters to meet their needs first; not get to 100% alignment
            - Define what is good enough
            - Start simple to deliver a data product that is adding defined value - deliver to the last mile
                - Minimal Viable Product product ensuring deliverable aligns with business value
                - Have clear path from MVP to productizing @ scale
            - Break the deliverables down into smaller parts to show milestone achievement during the data product development process
- Accountability on a data product and ensure succession plan for sponsors and data product owners
- Lessons learned from past experiences
    - Why is now different from past failures?

Advice for cloud and technology providers

- Support and engage with data product concept & standards
- Data product roles and responsibilities are clearly defined and transparent for data product producers, consumers, and ‘brokers’ including production, marketplace, consumption and retirement of the data product.
- Data product capability support for data producers and consumers including:
    - Managing Metadata
        - Schema and other metadata for the characteristics of the data resources

        - Maintaining classifications including sensitivity of the data product and data assets
    - Providing access, storage, distribution and usage to data products (e.g. marketplace)
    - Monitor and report on the use of data products including detection of misuse of data product
    - Transparency and accessibility to Data product usage agreements
    - Enforcement mechanism of data usage and resource support agreements (e.g. through contracts, SLAs)
    - Data product lifecycle management process execution (Ideation to Retire)
- Procurement process including external billing, internal cross-charge, and cost of the data product.

Questions

- Has data product been defined and aligned across the enterprise?
- Has a data product strategy been defined, published, or communicated to stakeholders?
    - Alternatively, does an existing data strategy provide direction that is supportive of the introduction and development of use cases supported by Data Products?
- Has a data product strategy been implemented?
- Do we have a product development process and platform defined?
- Does it include certification, versioning, governance, ownership, and lifecycle aspects?
- Has a data product lifecycle been identified and communicated throughout the organization?
- Has a data product governance framework been established that provides support & assurance of data products and supporting processes throughout the lifecycle?
- Do supporting good practice documents or statements such as policies (e.g. retention, access management), standards & procedures for the realization of data products and their utilization?
- Has technology been identified or planned to support the development & production of data products?
- Has a supplier framework been established to ensure sufficient assurance and due diligence (validation data sharing agreement) has been applied to the acquisition and management of input data and metadata resources that will support quality data products, their development and consumption from the data suppliers?

- Do we have the required metadata attributes (for the data product) specified and clarity about who is providing and maintaining them?

Artifacts

- Mechanisms to publish and discover and use of data products
- Data products
    - Strategy
    - Governance framework
    - Lifecycle framework
    - Supplier framework