Data Architecture

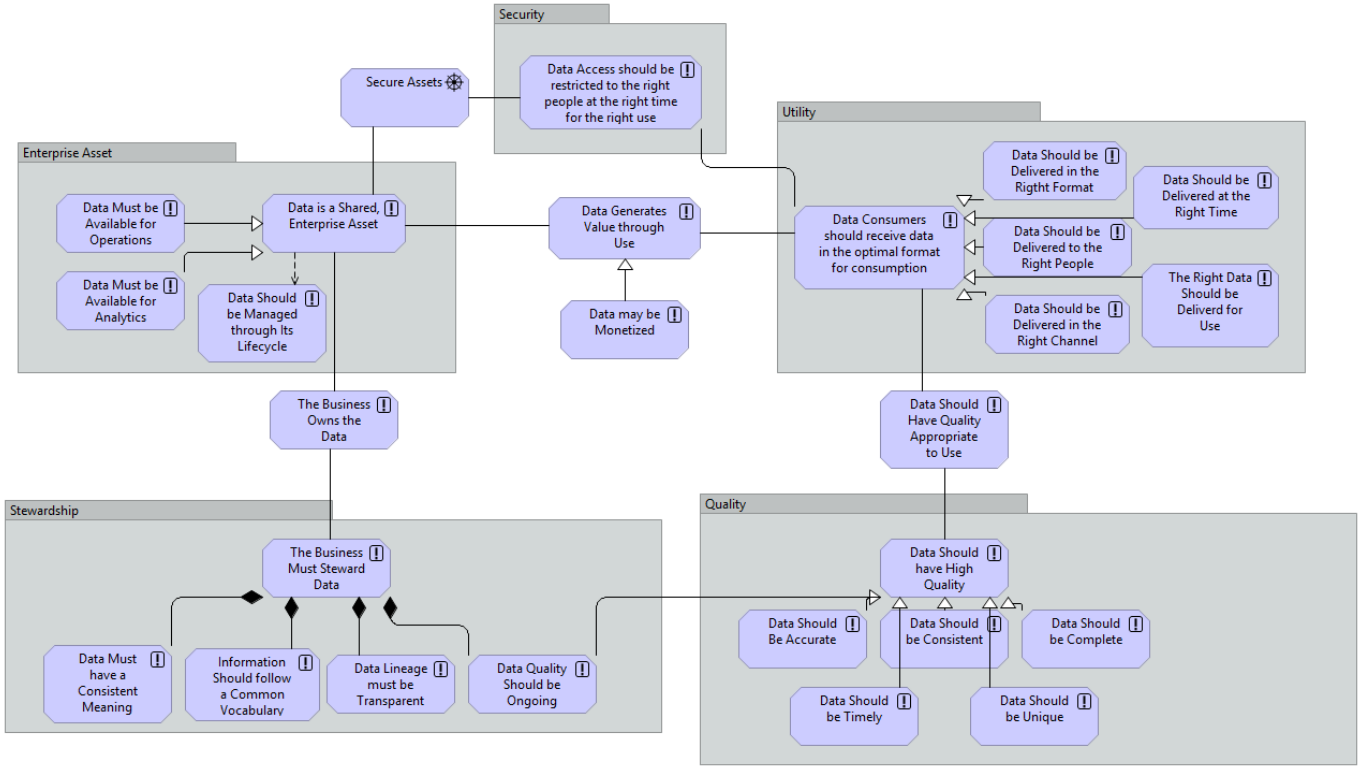
8 Feb 2018 14:47:56

Purpose

Views

Data Architecture Principles

Motivation viewpoint



Elements

Element	Type
Data Access should be restricted to the right people at the right time for the right use	Principle
Data Consumers should receive data in the optimal format for consumption	Principle
Data Generates Value through Use	Principle
Data is a Shared, Enterprise Asset	Principle
Data Lineage must be Transparent	Principle
Data may be Monetized	Principle
Data Must be Available for Analytics	Principle
Data Must be Available for Operations	Principle
Data Must have a Consistent Meaning	Principle
Data Quality Should be Ongoing	Principle
Data Should Be Accurate	Principle
Data Should be Complete	Principle
Data Should be Consistent	Principle
Data Should be Delivered at the Right Time	Principle
Data Should be Delivered in the Right Channel	Principle
Data Should be Delivered in the Rigtht Format	Principle
Data Should be Delivered to the Right People	Principle

Element	Type
Data Should be Managed through Its Lifecycle	Principle
Data Should be Timely	Principle
Data Should be Unique	Principle
Data Should have High Quality	Principle
Data Should Have Quality Appropriate to Use	Principle
Information Should follow a Common Vocabulary	Principle
Secure Assets	Driver
The Business Must Steward Data	Principle
The Business Owns the Data	Principle
The Right Data Should be Deliverd for Use	Principle

Motivation

Data Access should be restricted to the right people at the right time for the right use

Type	Principle
Statement or Description	Data access must be aligned with orgnaizational outcomes, limiting the use and access of data assets to only those users (system and human) who require access for a business use, and only when that data access is required.
Rationale	Deriving from the priniciple that data is an asset, and the driver to secure assets, data access should allow and enable legitimate uses, but should not be enabled for non-legitimate uses or users. When data is not being used for a legimiate use, restricting access ensures the data is secure.
Implications	Systems and solution designs must take into account access controls that incorporate who, what, when, and why. Technical controls will be required to enforce access controls. Monitoring and auditing tools will be required to validate the proper functioning of access controls.

Data Consumers should receive data in the optimal format for consumption

Type	Principle
Statement or Description	Consumers of data should expect that required data be available in the format required for consumption, and that efforts be made to deliver the data in that format. This principle encapsulates the concept of data quality.
Rationale	There are 2 primary rationales: one for the consumer, one for the provider. For the consumer, setting the expecation that data is in the format required streamlines consumption and use. It removes the requirement to perform data quality functions. And, it helps define an interface boundary between consumers and producesrs. For the producer, providing optimally formatted data provides a control point through which other concerns, such as data security, can be applied.
Implications	Knowledge of data consumption needs is required by the data provider to ensure that data is delivered in the optimal format. Data consumers have an additional burden to effectively define the consumption format requirements.

Data Generates Value through Use

Type	Principle
Statement or Description	Latent value in data "as an asset" is only realized when the data is used.
Rationale	While data as an asset suggests that data has value, only though using the data can that value be realized. Without use, managing the asset, data, incurs only costs, and not value is generated to offset that cost.
Implications	Data that is not used does not generate value. Data that may be used may generate value, and the prospective value should be considered. Data's value can conceivably be calculated through an understanding of the value derived in use. In some cases, data value can be monetized.

Data is a Shared, Enterprise Asset

Type	Principle

Statement or Description	Data is an asset that has utility across the enterprise, and across groups and use cases.
Rationale	Business activities often have implications across the value chain. Data from business activities, when shared, can enable capabilites, improve outcomes and inform decisions around business activities far from the data source.
Implications	Data generated from activities or sourced externally should be considered when contemplating business solutions. Effort should be made to make data available beyond local data producers and consumers. A local view of data will miss opportunities to leverage enterprise data resources.

Data Lineage must be Transparent

Type	Principle
Statement or Description	For any data element, it must be clear and available the history of the data, any transformations, filters, summarizations, or other changes to the data, to effectively use the data.
Rationale	For there to be trust in the outcomes of business activties, there must be a complete and transparent accounting of the supporting data's history. Without this, the outcome of the business activity will suffer from trust and utilty.
Implications	Systems and tools will be required to capture, maintain, and make transparently available data lineage for all data assets put into use.

Data may be Monetized

Type	Principle
Statement or Description	Data may have explicit monetary value to external consumers.
Rationale	Data produced by business activities may be valuable to partners in our ecosystem. In some cases, this value may best be represented as a monetary exchange.
Implications	We may have an opportunity to generate revenue, or offset costs, by requiring payment for access to data. Data security concerns must be addressed. Additional legal considerations may come into play when data is traded.

Data Must be Available for Analytics

Type	Principle
Statement or Description	A special use case of data as a shared asset is data as an asset used for anlaytics.
Rationale	Data's utility for operational activties is assumed. However, analytic uses have historically been secondary. This principle states clearly that analytics uses should be treated equally (in principle) to other shared enterprise uses.
Implications	Operational data sources must be externalized and made available to analytics uses (as demand requires).

Data Must be Available for Operations

Type	Principle

Statement or Description	Operational systems must have the access to the data necessary to support their functions.
Rationale	Systems supporting operations will deliver higher quality outcomes to the extent they have access to data from across the enterprise.
Implications	Additional intergration may be required for operational systems to have the fuller context represented in enterprise data sets.

Data Must have a Consistent Meaning

Type	Principle
Statement or Description	The meaning of data must be consistent across contexts of use.
Rationale	Mis-applying data that means one thing in a context where data means another results in an erosion of trust in the business outcomes.
Implications	Business context requirements need to drive the data requiements, and data that doesn't have the precise required meaning must not be used in those circumstances. When the business terminology across business contexts is the same, but the meaning is different, there will need to be some couching of the naming and terms to avoid confusion. New data sets will be required in some contexts where data meaning needs do not meet the data meanings of the data in use.

Data Quality Should be Ongoing

Type	Principle
Statement or Description	Data Quality is a process, not just a characteristic or attribute of data.
Rationale	Data quality declines over time, so resources need to be applied to maintain quality.
Implications	Ensuring that data quality is reviewed adds to the requirements of data stewardship. Tools to evaluate and maintain data quality need to be available to validate that data quality standards are being met. Resources need to be planned to perform data quality activities.

Data Should Be Accurate

Type	Principle
Statement or Description	Data representations accurately represent the entities they model.
Rationale	The fundamental tenant of data quality, and hence trust, is that data accurately represents what is intended. Inaccurate data erodes trust in the outcomes of dependent business activities.
Implications	Systems, tools, and resources are required to ensure that data accuracy is maintained. Processes are reqiured to capture, report, validate, and correct data inaccuracices. Systems and solutions must accommodate receipt or managment of accuracies. To the extent that inaccuracies cannot be corrected, business activities must accommodate a corresponding level of inaccuracy.

Data Should be Complete

Type	Principle

Statement or Description	Data should be complete, in that expected values are present, related records are present, and the related historical record is complete.
Rationale	Complete data supports data quality, in that decions made within the business context can be seen to be based on full knowledge of the context.
Implications	Having complete data implies that systems are in place to ensure that data is both internally consistent, ie. all known data is captured and present, and externally consistent, ie. all expected facets of the data are present. In cases where completeness cannot be guaranteed, systems must support processes that accommondate incompleteness.

Data Should be Consistent

Type	Principle
Statement or Description	Data should be consistent, from use to use, from system to system, and across time.
Rationale	Trust is eroded when data values are not consistent across usages, across systems, or across time. The results of business activities should be expected to be consistent in time, and across systems.
Implications	Processes are required to ensure consistency across the information landscape. Where data is used, transformed, or processed, technology is required to ensure consistency, and to monitor consistency. If there cannot be an expetation of consistency, systems and business ativities must accommodate.

Data Should be Delivered at the Right Time

Type	Principle
Statement or Description	Data consumers should expect required data when it is required.
Rationale	Data consumers should not be burdened with time or process shifting to accommodate timeliness in delivery of data.
Implications	Technology will be required to ensure that data is available at the SLA required. This means that both data sources are available, and distinct data elements are available. In some cases, this may mean that queueing, eventing, or alerting solutions are reuqired. Timeliness may be unduly expensive, so the concerns of timeliness need to be balanced against other concerns.

Data Should be Delivered in the Right Channel

Type	Principle
Statement or Description	Data consumers should expect that data is made available in the right channel, through the necessary technical channels.
Rationale	Data consumers should not be burdened with extending channels to access data reuqired for consumption. The need to create value through use requires that data be presented where it can be used.
Implications	Channel capabilities are a requirement of the data mangement function, so that the data itself is made available as needed. Channel flexibility technology (such as an ESB) is required. Channel management costs need to be considered in the costs of overall solutions

Data Should be Delivered in the Rigtht Format

Type	Principle

Statement or Description	Data should be delivered in the format required by the consumer.
Rationale	Data consumers should expect that translations into structure or form have been addressed, and any related concerns regarding meaning or content reconciled, prior to accessing data.
Implications	Technical solutions are necessary to provide format conversions, including date type, format, language, etc. Solutions are required to manage representational structure formats, such as XML vs. JSON, file structures, etc. Crosswalks and conversion serviecs may be required, and if so, they must also include governance and management capablities.

Data Should be Delivered to the Right People

Type	Principle
Statement or Description	Data consumers should expect to receive the data that is relevant to their needs, both in terms or roles and person, and in terms of business context.
Rationale	Effective business outcomes occur when the individual context is met. Security concerns are met when data is provided only to the right consumers.
Implications	The Right People includes system as well as human actors and their proxies. Identity, role, and context must be accounted for as part of all data interactions. Ideally, the consumers identity, role and context are made available to the data provider, and only that data necessary given the consumer's context is made available.

Data Should be Managed through Its Lifecycle

Type	Principle
Statement or Description	As an asset, data should be managed throughout its lifecycle.
Rationale	As an asset, the business and IT can only effectively manage data if it's existence and qualilty are known and available. If data is not managed, it may be lost, misplaced, misused, or otherwise not provide value, while still creating costs on the organization.
Implications	Tools, processes, and organizations must exist to manage data throughout its lifecycle. Catalogs of data, it's location, origin, meaning, and uses exist and be maintained. Onboarding activities must be defined. When data is no longer useful, processes must be in place to retire and/or destroy that data.

Data Should be Timely

Type	Principle
Statement or Description	Consumers should expect data that does not suffer undue latency, and that is timely in the context of the business activity.
Rationale	Trust is eroded if business outcomes are determined based on information that is not current, or after the business need has passed.
Implications	Data access and synchronization must be provided as part of the data management activities. Timeliness requirements of the data consumer must be understood to ensure SLAs are maintained. In situations where timeliness cannot be expeceted, solutions must account for the lack of timeliness in the solutions. Timeliness can be expensive, so the contextual need must be balanced against value of timely delivery.

Data Should be Unique

Type	Principle

Statement or Description	Entities should be represented in the proper count, without duplicates.
Rationale	Trust is eroded when entities are over-represented, or represented multiple times. Systems records become incomplete and inconsistent when associated data is distributed across multiple records.
Implications	Technical measures of duplication, and processes for consolidation, are required to ensure uniqueness is maintained. Where uniqueness cannot be ensured, systems must accommodate potential duplicates.

Data Should have High Quality

Type	Principle
Statement or Description	Data consumers should expect data to have the quality characteristics necessary for users to be able to trust the business outcomes.
Rationale	Business outcomes will be informed based on the data presented. If the data does not exhibit the characteristics of quality, the business outcomes may be wrong, or may not be trusted.
Implications	Data quality services must be employed to address the realities of data quality issues. Technical solutions to assure, and measure, data qualilty are required. Shortcomings in data quality must be accommodated by data consumers. The costs of data quaility cannot be avoided, but their nature changes as data quality adaptations propagate to consumers. In extreme cases, data quality issues can undermine internal and external acceptance and reliability of buiness activities, or the business itself.

Data Should Have Quality Appropriate to Use

Type	Principle

Information Should follow a Common Vocabulary

Type	Principle
Statement or Description	Our data vocabulary should be applied consistent to data that shares a common meaning.
Rationale	Common language helps bridge undestanding across business contexts, and helps ensure that data consumers use data that means what is intended.
Implications	Business terms are often reused in different contexts. The vocabulary of our must balance the advantages of the business vocabulary with the advantages of unambiguous identifiers for data sets, particularly when the use cases are similar but not identical.

Secure Assets

Type	Driver
Statement or Description	Assets should be secured
Rationale	Assets represent an investment and a potential value through use. To ensure that value is realized, the assets must be protected.
Implications	Data security is an essential requirement for data management.

The Business Must Steward Data

Type	Principle

Statement or Description	The business, those who are responsible for delivering business services, must provide governance of the data that supports and informs those services.
Rationale	Business users are uniquely aware of the business context, and the data that represents the business context. Only the business users have the knowledge and authority to determine meaning, vocabulary, and the required quality of data.
Implications	Business users need tools, processes, and resources to perform stewardship functions. The business must work with technology custodians to realize stewardship outcomes, including data quality processes.

The Business Owns the Data

Type	Principle
Statement or Description	The utiltimate owner of data is the business.
Rationale	Data exists and has meaning only insofar as it supports business outcomes, and in the context of business activities.
Implications	Data's costs and benefits accrue to the business and business functions the data supports. Appropriate use, quality, mangaement, and disposition of data is ultimately determined by the business. If there is a vacuum of ownership of data, that data's disposition will be undetermined.

The Right Data Should be Deliverd for Use

Type	Principle
Statement or Description	Data consumers should expect that data is delivered appropriate to the needs of the data consumer.
Rationale	Data consumers should expect to receive data appropriate to the intended use. Consumers should not be expected to repurpose data intended for other uses.
Implications	There must be a means for data consumers to communicate context to data providers. There is value in generating re-usable data services, though that must be tempered against the need to provide the data necessary to the use case.

Relations

Specialization relation

Type	Specialization relation
Source	Data Should Be Accurate
Target	Data Should have High Quality

Specialization relation

Type	Specialization relation
Source	Data Should be Timely
Target	Data Should have High Quality

Specialization relation

Type	Specialization relation
Source	Data Should be Complete
Target	Data Should have High Quality

Specialization relation

Type	Specialization relation
Source	Data Should be Consistent
Target	Data Should have High Quality

Specialization relation

Type	Specialization relation
Source	Data Should be Unique
Target	Data Should have High Quality

Association relation

Type	Association relation
Source	Data Access should be restricted to the right people at the right time for the right use
	Secure Assets

Specialization relation

Type	Specialization relation
Source	Data Should be Delivered in the Right Channel
	Data Consumers should receive data in the optimal format for consumption

Specialization relation

Type	Specialization relation
Source	Data Should be Delivered in the Rigtht Format
	Data Consumers should receive data in the optimal format for consumption

Specialization relation

Type	Specialization relation
Source	Data Should be Delivered at the Right Time
	Data Consumers should receive data in the optimal format for consumption

Association relation

Type	Association relation
Source	Data Should be Delivered to the Right People
	Data Access should be restricted to the right people at the right time for the right use

Specialization relation

Type	Specialization relation
Source	Data Should be Delivered to the Right People
	Data Consumers should receive data in the optimal format for consumption

Specialization relation

Type	Specialization relation
Source	The Right Data Should be Deliverd for Use
	Data Consumers should receive data in the optimal format for consumption

Association relation

Type	Association relation
Source	The Business Must Steward Data
Target	Data is a Shared, Enterprise Asset

Association relation

Type	Association relation
Source	The Business Must Steward Data
Target	Data Generates Value through Use

Association relation

Type	Association relation
Source	Data may be Monetized
Target	The Business Must Steward Data

Association relation

Type	Association relation
Source	Data Generates Value through Use
Target	Data is a Shared, Enterprise Asset

Specialization relation

Type	Specialization relation
Source	Data may be Monetized
Target	Data Generates Value through Use

Composition relation

Type	Composition relation
Source	The Business Must Steward Data
Target	Information Should follow a Common Vocabulary

Specialization relation

Type	Specialization relation
Source	Data Must be Available for Analytics
Target	Data is a Shared, Enterprise Asset

Composition relation

Type	Composition relation
Source	The Business Must Steward Data
Target	Data Quality Should be Ongoing

Composition relation

Type	Composition relation
Source	The Business Must Steward Data
Target	Data Lineage must be Transparent

Composition relation

Type	Composition relation
Source	The Business Must Steward Data
Target	Data Must have a Consistent Meaning

Association relation

Type	Association relation
Source	Data Should have High Quality
	Data Consumers should receive data in the optimal format for consumption

Association relation

Type	Association relation
Source	Data Access should be restricted to the right people at the right time for the right use
	Data Consumers should receive data in the optimal format for consumption

Association relation

Type	Association relation

Source	Data Quality Should be Ongoing
Target	Data Should have High Quality

Association relation

Type	Association relation
Source	Secure Assets
Target	Data is a Shared, Enterprise Asset

Association relation

Type	Association relation
Source	Data Consumers should receive data in the optimal format for consumption
	Data Generates Value through Use

Specialization relation

Type	Specialization relation
Source	Data Must be Available for Operations
Target	Data is a Shared, Enterprise Asset

Association relation

Type	Association relation
Source	The Business Owns the Data
Target	The Business Must Steward Data

Association relation

Type	Association relation
Source	The Business Owns the Data
Target	Data is a Shared, Enterprise Asset

Association relation

Type	Association relation
Source	Data Should have High Quality
Target	Data Should Have Quality Appropriate to Use

Association relation

Type	Association relation
Source	Data Should Have Quality Appropriate to Use
	Data Consumers should receive data in the optimal format for consumption

Influence relation

Type	Influence relation
Source	Data is a Shared, Enterprise Asset
Target	Data Should be Managed through Its Lifecycle