TDWI South Florida Chapter

tdwi

Forging a Data Strategy

A Score Card for Establishing Goals

Evan Levy

March 2, 2022

EvanLevy@IntegralDataLLC.com
Evan@EvanJLevy.com

@EvanJayLevy, #DataStrategy , #TDWI

Agenda

- **Strategy Fundamentals**
  - Needs, Pains, & Drivers
  - The Data Strategy Development Effort
- **The Core Components**
  - Definitions and Details
  - Scorecard Attributes
  - Real World Scorecard Examples

1630-1715 –Segment #1
1715-1720 - Break
1720-1810 – Segment #2

Data Challenges…

"We need to be able to add data quickly. A new data source should take a few days"

"We use internal and external data. Do we even know what we have?"

"With the new privacy laws (CA & NY), are we able to protect and track data usage?"

"You want me to fix the data? I don't own the data; I can't fix it."

"Everyone agrees that the business should own the data. What does that mean?"

"I have to beg, borrow, and steal data to do my job. Why isn't data sharing an obligation?"

"We have to approach data differently. Reuse and economies-of-scale is the goal."

"A tool or a new platform won't fix inaccurate data. Until we have governance, we won't have the mechanisms to fix and maintain the data"

3

What is Strategy?

A method or plan chosen to bring about a desired future, such as achievement of a goal or solution to a problem.

BusinessDictionary.com

a careful plan or method for achieving a particular goal usually over a long period of time

Merriam-Webster

Strategy is different from vision, mission, goals, priorities, and plans. It is *the result of choices on where to play and how to win, to maximize long-term value*

strategy+business

….as there is always an element of uncertainty about the future, strategy is more about a set of options ("strategic choices") than a fixed plan
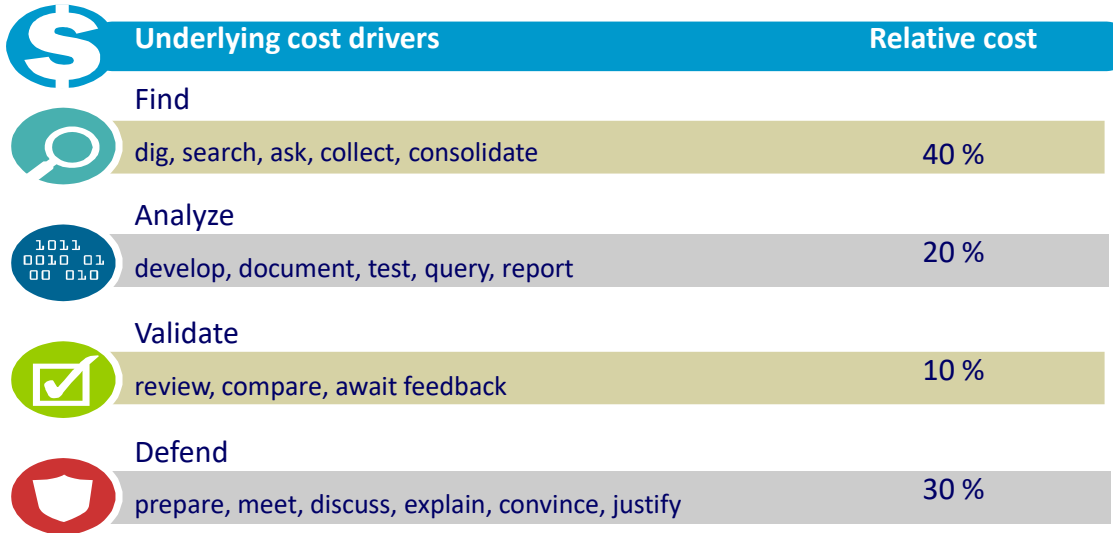
WIKIPEDIA

4

Data Sharing
The Business Impact

	Underlying cost drivers	Relative cost
	Find	
	dig, search, ask, collect, consolidate	40 %
	Analyze	
	develop, document, test, query, report	20 %
	Validate	
	review, compare, await feedback	10 %
	Defend	
	prepare, meet, discuss, explain, convince, justify	30 %

5

The Multitude of "Data" Initiatives

Numerous and disparate data integration efforts

Self service data, self service analytics

The explosion in data storage

Migration of applications (and data) to external platforms

Analytics Sand Boxes

6

Data Strategy Drivers

A plan designed to improve all of the ways you acquire, store, manage, share, and use data

Sharing (and publishing) internal, 3rd party, and external data

Tracking data access/usage across systems and 3rd parties

Data storage and data retention

Data privacy, data protection, individual's consent, etc.

Enterprise naming standards and value conventions

Supporting different usage needs (reporting, sand boxes, data science, applications, etc..)

7

Strategy Outputs

Today's discussion

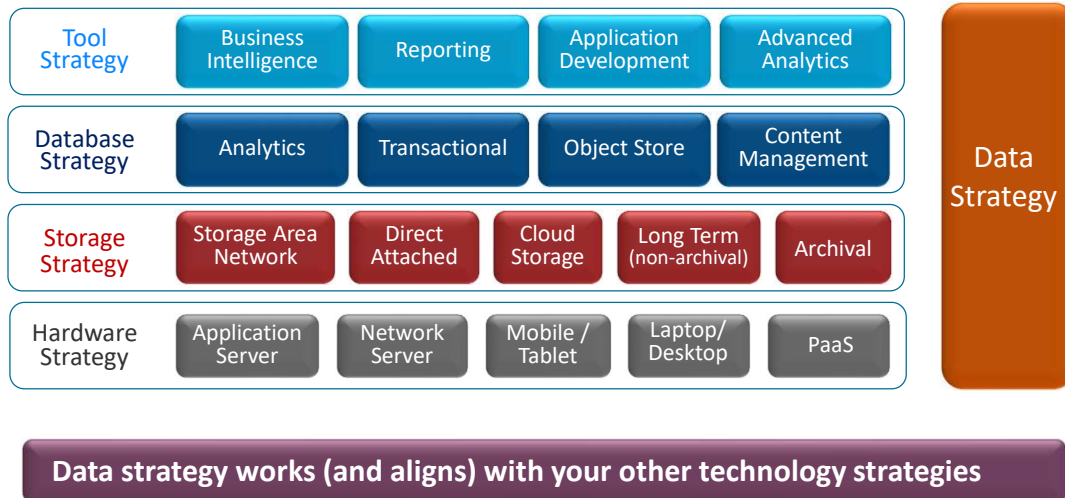| Needs & Challenges	Goals & Objectives	Methods and Artifacts	Implementation Roadmap
• Activities that can't be analyzed • Data that exists but isn't available • Decisions that occur w/o data • Misunderstood and misused data • New staff learning curve • Decisions that can't be made	• New business activity KPIs • New business processes analysis • Supporting self service analysis • Access to new systems (or data) • New/Improved data tools • Improved data	• New processes – Data governance – Data mgmt – Development – Stewardship – User engagement • Deployment of new tools • Creation of new data artifacts	• Milestone accomplishments • Phased/ordered deliverables – 3-6 month phases – Illustrates dependencies – Allows for update/review • Categorized across disciplines

Analysis

Development

8

Companies Have Multiple Technology Strategies

Tool Strategy	Business Intelligence	Reporting	Application Development	Advanced Analytics	
Database Strategy	Analytics	Transactional	Object Store	Content Management	Data Strategy
Storage Strategy	Storage Area Network	Direct Attached	Cloud Storage	Long Term (non-archival)	Archival
Hardware Strategy	Application Server	Network Server	Mobile / Tablet	Laptop/ Desktop	PaaS

Data strategy works (and aligns) with your other technology strategies

9

The Value of a Data Strategy…

- Establishes common goals and objectives across projects
- Involves a cross-functional set of stakeholders from IT and business areas
- Allows for the creation of common methods and practices across development projects
- Provides a method for approving (or dismissing) new data-centric project requests
- Aligns developers and end users as data consumers

A Data Strategy serves as a common thread across initiatives

10

The Core Components

@EvanJayLevy, #DataStrategy , #TDWI

11

Data Strategy
The Core Components

12

The Core Components Defined

	Identify	The ability to identify data and understand its meaning regardless of its structure, origin, or location
	Provision	Enabling data to be packaged and made available while respecting all rules and access guidelines.
	Store	Persisting data in a structure and location that supports access and processing across the enterprise
	Assemble	Cleansing, standardizing, combining, and moving data residing in multiple locations and producing a unified view
	Govern	Establishing, communicating and monitoring information practices to ensure effective data sharing, usage, and protection

13

Data Strategy
Data Strategy is not "One Size Fits All"

Organization ——————————————— **Enterprise**

Identify
Application or Team defined ←————→ Business-defined

Provision
System Centric ←————→ Self-Service Oriented

Store
Host owned & Aligned ←————→ Online/bulk w/change control

Assemble
Source/Appl Defined ←————→ Multi-Org/App Support

Govern
Application Specific ←————→ Cross-Functional Teams & Process
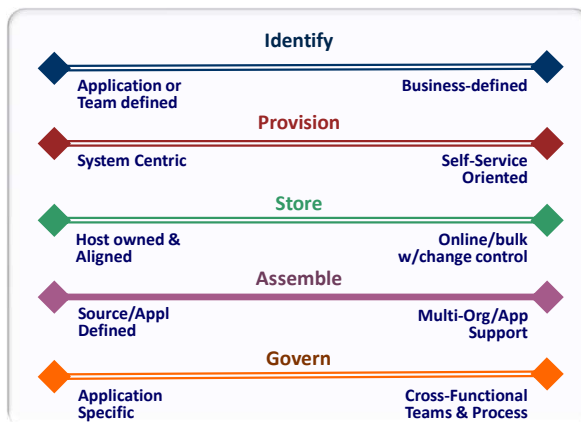
- Each component is independent and can evolve as needed
- Each component is specific to an individual set of skills and capabilities
- Not all organizations need an "enterprise-class" data strategy
- Complexity (and maturity) increase with organizational scope
- The Data Strategy should establish goals for each component
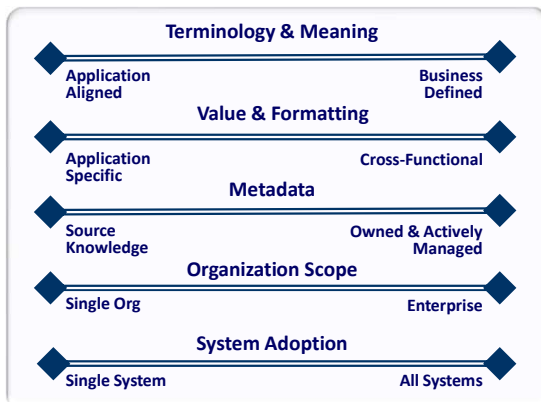
14

Data Strategy
Multiple Component Capabilities

Identify | Provision | Store | Assemble | Govern
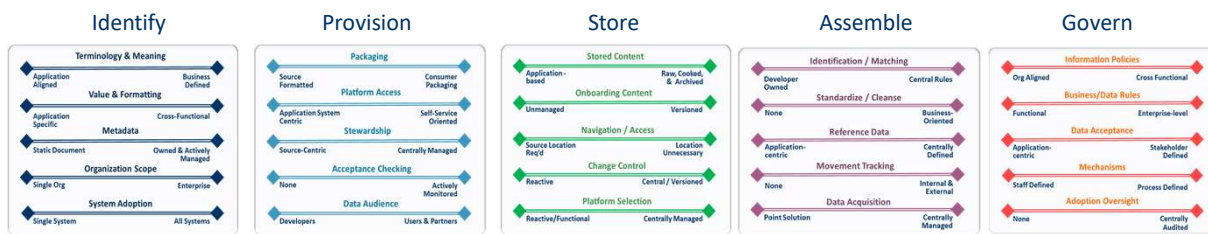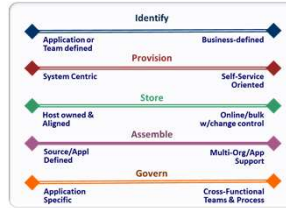
15

The Core Components
Identify

The ability to identify data and understand its meaning regardless of its structure, origin, or location


- Who defines terms and their meanings?
- How is the data represented? How is it determined?
- Does everyone use different terminology? Are there reports with conflicting terms?
- Is there any documentation or a tool that contains the reference details?
- What's the scope (or breadth) of your data management standards? Are they voluntary or mandatory?

16

Slide 17

The Scope of Your Strategy
Identify
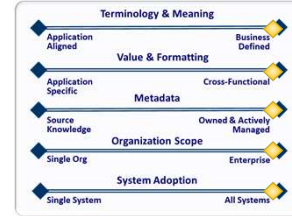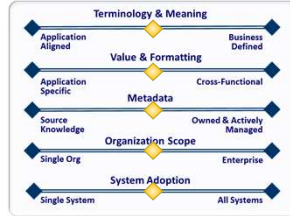
Developer Domain
- Terminology is based on the developer or the source system
- Data representation is based on the source application
- Data knowledge is tribal (or developer known)
- Data usage is based on the requester; there's no thought to sharing or reuse
- Every request is a custom effort – there is no data management

Report and Data Sharing
- Reports contain standard terms, values, & meanings
- Data is shared across systems; the format is usually based on the source
- Some data is defined and documented
- There are standards for customer data; everything varies across tools and organization
- Reporting systems usually use consistent terms; there's no consistency across operations

Business Centric Terms/Data
- There are established business terminology and value standards
- There's a focus on data sharing and self-service
- A centralized data dictionary and source catalog is maintained
- Org- or process-centric terms are supported
- New systems and reports adopt existing standards (where practical)

17

Slide 18

The Scope of Your Strategy
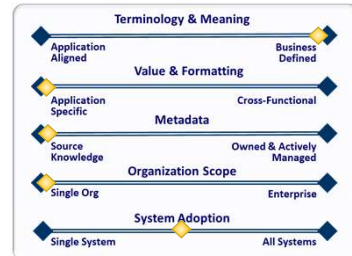# Identify – Real World

Most environments are NOT "aligned" to a single capability level
- Staff skills are inconsistent across teams
- Some DBAs are good with naming conventions and standards
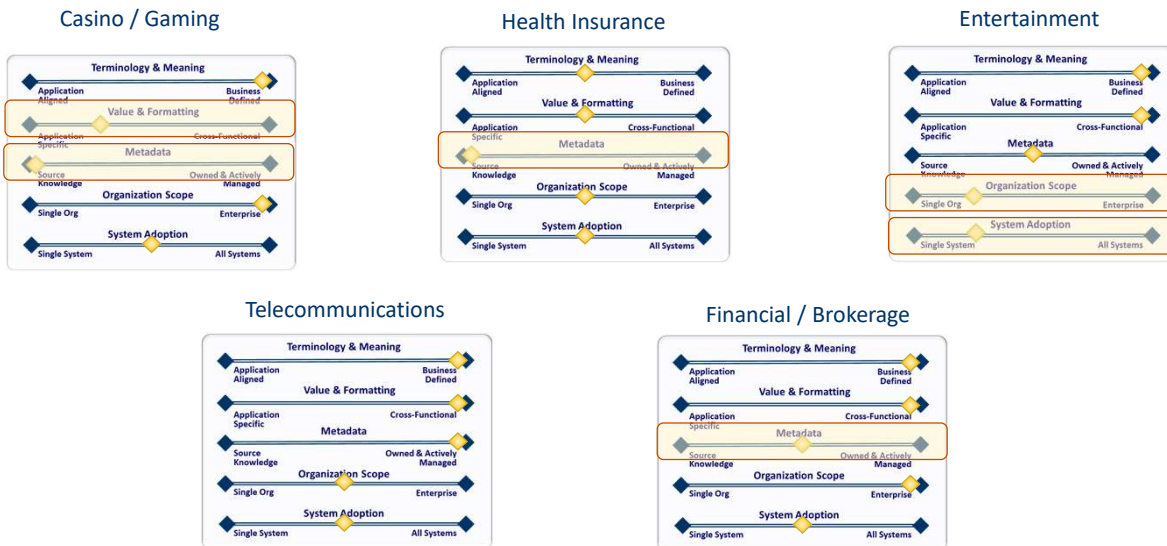- Few have any user-oriented metadata

One client that I worked with…
- The DW had good naming standards
- …but value standards were non-existent
- Data marts were autonomous but typically followed the DW

18

Slide 19

Sample Clients and Their Candidate Goals

Casino / Gaming — Health Insurance — Entertainment — Telecommunications — Financial / Brokerage

19

Slide 20

The Core Components
Provision

Enabling data to be packaged and made available while respecting all rules and access guidelines

Packaging
Source Formatted — Usage-Oriented

Platform Access
Application System Centric — Self-Service Oriented

Stewardship
Source-Centric — Centrally Managed

Acceptance Checking
None — Actively Monitored

**Data Audience**
Developers — Users & Partners
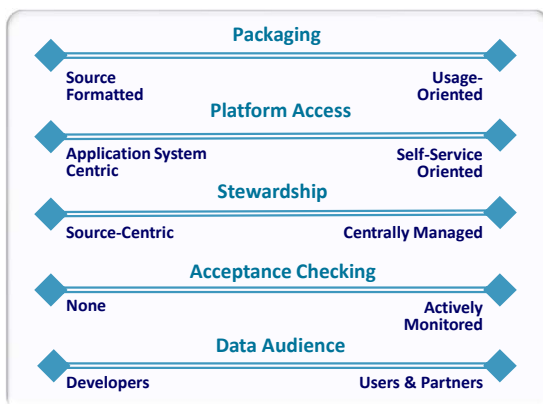
- How is data packaged for distribution to other systems and users? (is it even packaged?)
- Is data sharing (and source data access) a courtesy or responsibility?
- Is there someone responsible for supporting source data access?
- Who typically asks for data (developers or users)?
- Is data content measured or accepted prior to delivery?

20

The Scope of Your Strategy
Provision
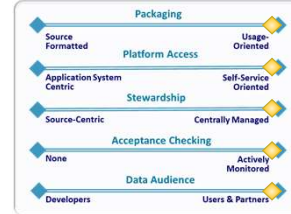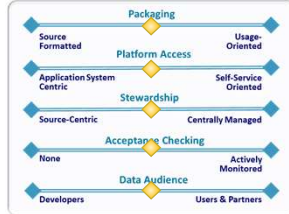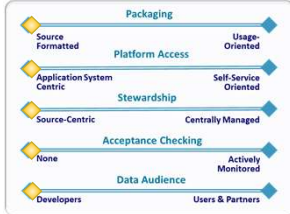
Custom Source Creation
- There are no packaging standards.
- The data developer must query and retrieve the data
- There is no source data steward
- Data isn't checked to conform to standards or accuracy
- Access requires source knowledge and programming skills is required

Source Data Distribution
- Data layouts are defined (and access requires developer skills)
- Data extracts scheduled and stored in defined location
- Someone is identified as source data steward – they generate the extracts
- Data has standard formats & values. Detection of data issues relies on usage
- Access requires ETL skills and data domain knowledge

User Data Self-Service
- Data formats reflect business defined standards
- Data is delivered to a central repository (mart, data lake, etc.)
- The data steward generates extracts and tracks changes
- Data content is checked for completeness and quality
- There are various tools available to developers and users (based on skills & needs)

21

Example #1: Data Provisioning

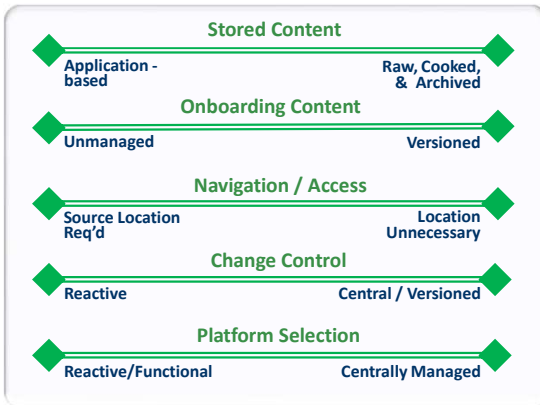## We have a DW and several data marts. We receive data from systems nightly.

- Source extracts are created using a vendor utility. The files change at least once a year (due to SW upgrades)
- The DW is supported by a central team; each DM is supported differently
- We receive external data from a few providers. Everyone manages it individually
- If users want new data, they are dependent on developers for the DM and DW.
- Lots of users load data onto their own systems and analyze with Excel (and other tools).
- If we have problems with data – it's usually found by users after it's been loaded (or not loaded)

22

Slide 23
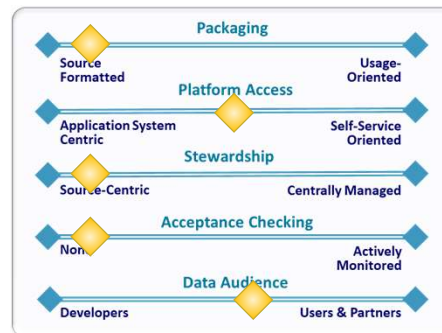
The Core Components

Store

Persisting data in a structure and location that supports access and processing across the enterprise

Stored Content
Application - based — Raw, Cooked, & Archived

Onboarding Content
Unmanaged — Versioned

Navigation / Access
Source Location Req'd — Location Unnecessary

Change Control
Reactive — Central / Versioned

Platform Selection
Reactive/Functional — Centrally Managed

- Is there a central copy of data to share, or does each system maintain its own copy?
- How do you determine available data assets (how does someone find it)?
- How do you share or store an internally created data set?
- How do you hear if a source generated data set changes (layout, format, etc.) ?
- Who decides where the data is stored (last month? Last year? Last 5 years)?

1341729

23

Slide 24

The Scope of Your Strategy

Store
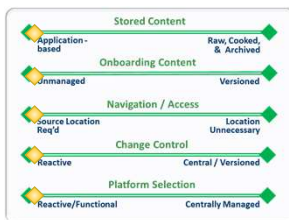
Application-Centric
- Data formats and layouts reflect source application. Changes are common (and unexpected)
- There's no documentation or detail of acquired data
- Data access requires developer skills, permission, and knowledge
- Source data changes are common, frequent, and a surprise
- Data storage is based on developer discretion.
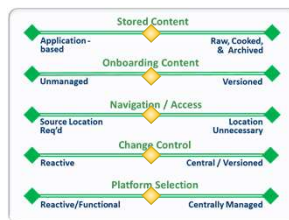
Application/System Sharing
- Data layout is based on source; the format is consistent
- External data exists but it's available via word-of-mouth
- Source data is available via extracts (with developer skills and location knowledge)
- Source changes are common but production profiling jobs identify those issues
- Historical data exists in a DW; source content is managed by individual systems
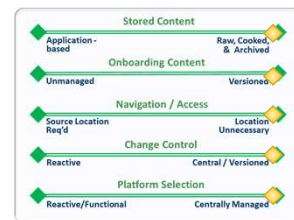
Usage Oriented
- Data is stored in raw and ready-to-use format and documented
- External data is documented like all sources and made available
- Source data is available via a data lake or extract that is documented
- Data change control is a production process
- Source and business data is documented and centrally managed

1341729

24

Example #2: Storage

We have a diverse set of analytical capabilities and a large number of servers / data marts supporting these diverse needs.

- If developers want data – they request an extract. We have hundreds of custom source extracts
- Each analytics team is responsible for their own budget. Some have premise systems, others use the cloud. Everyone will be in cloud over the next 18-24 months
- We have production change control – but it doesn't include impact to extracts or data
- Because of our budgeting process, all analytical apps are silos. IT manages HW/SW platforms and cloud, but there is not resource sharing

25
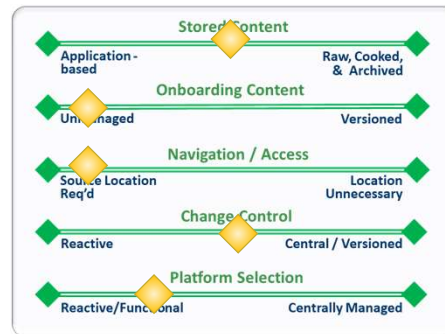
## The Core Components
Assemble

Cleansing, standardizing, combining, and moving data residing in multiple locations and producing a unified view
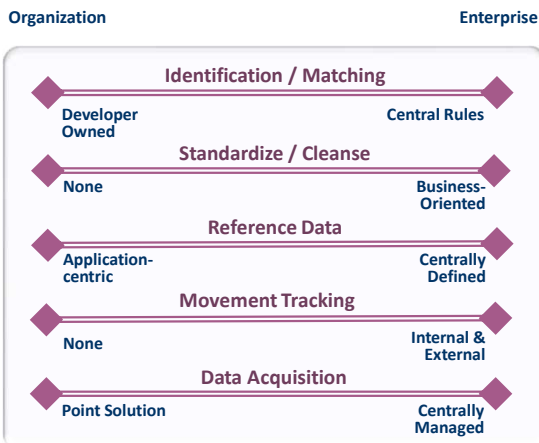

- How do developers determine the match logic across different data sources?
- Which elements should be retained from which source after integration?
- Will the data be standardized and/or cleansed? If so, how?
- What is the source of reference data?
- Is there any tracking of data source movement within your company? Outside your company?

26

Slide 27

The Scope of Your Strategy
Assemble
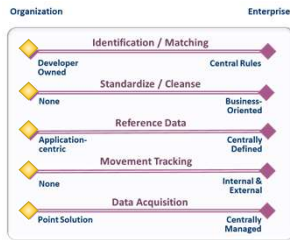
Custom Programming
- Developer must write code to identify/match records
- There's no record of data location or propagation details
- Data/Business rules are unique for each system
- Each development team handles their own integration and movement activities
- There no defined system of record – sourcing based on developer discretion
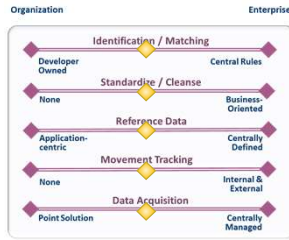
Developer Logic Sharing
- Matching is based on data steward rules/details
- There's lineage tracked for data warehouse data loads
- Data rules are documented in a the data dictionary
- There's a centralized team that handles data movement and processing for multiple systems
- There is a designated system of record for major subject areas
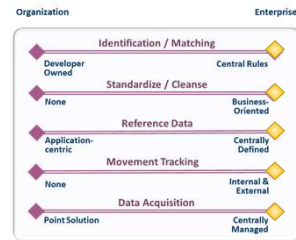
Managed/Published Rules
- MDM hub exists; a central cross-reference provides matching map
- Data lineage and movement tracking is logged centrally
- There's a rules engine/dictionary that is documented and accessible
- Systems (analytic, transactional) share integration & access code
- MDM manages/owns system of record functionality

27

Slide 28

The Core Components
Govern

Establishing, communicating and monitoring information practices to ensure effective data sharing, usage, and protection
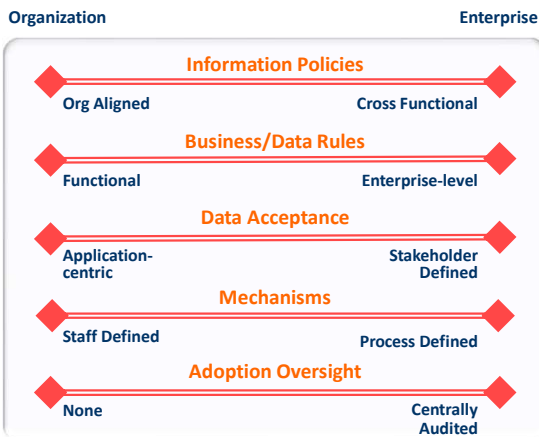
- Who decides and defines business information policy?
- How are policy decisions socialized?
- How business data rules defined and published
- Is data acceptance measured, tracked, and published? (which data)?
- What are the mechanisms and process for implementing new policies and rules?
- Are data governance decisions tracked and audited for adoption?

28

14

Slide 29

The Scope of Your Strategy

Govern
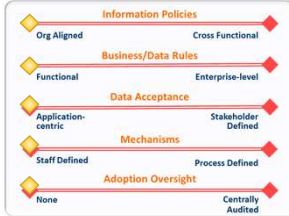
Developer-Owned
- Data policies (access, usage rights) are defined by the system-of-creation
- Data rules reflect the transactional application
- Data accuracy reflects the rigor of the hosting application
- There is no data management oversight; decision driven by developers
- Change management and decision mechanisms are at development discretion

Platform-Oriented
- Policies include stakeholders but are specific to a platform
- Rules reflect data usage context (report, publish, etc.)
- Data quality is based on usage context
- Data management standards are specific to a team or system
- Changes and decision authority or defined based on application or system needs

Business Collaboration
- Policies are aligned to subject areas and managed by a centralized body
- Data rules reflect business-centric criteria that are cross-functional
- Data is measured and accepted prior to delivery
- Data standard adherence is centrally monitored and published
- Change processes align to cross-functional usage needs

29

Slide 30

Example #3: Governance

We have a data governance program. It's a "bottom-up" program that's focused on data standards and data quality.

- Our DG team has built a data glossary and established terminology and value recommendations
- The DW follows DG recommendations. No IT system is required to follow the standards
- We have a lot of consumer data (incl. non-US consumers. We're in the midst of implementing a new data privacy and protection initiative
- We don't have a single/central information policy authority. There are lots of rules and they are managed by individual orgs (legal, HR, etc.)
- There's nothing in place to measure the adoption and use of standards or rules

30

Data Strategy
The Core Components

31
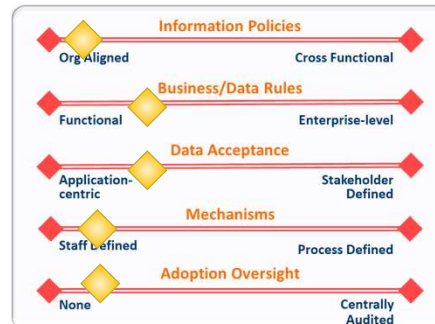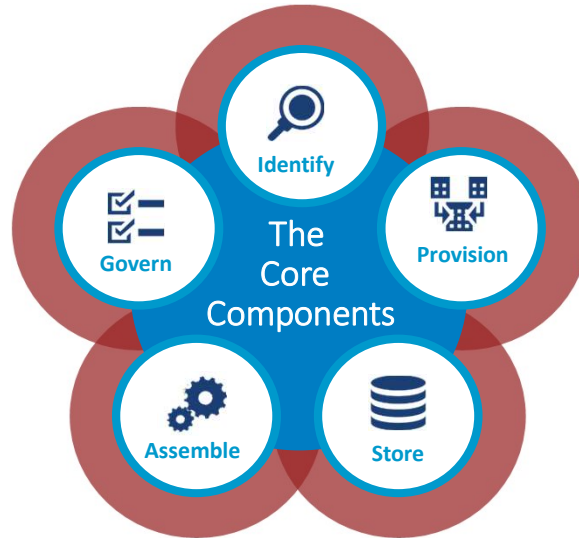
What We're Telling Folks…

A Data Strategy isn't a "once and done" activity
- It's an an ongoing program that identifies and refines deliverables on a regular basis
- Must include goals (with a business benefit) that are measurable
- Addressing changing business needs is the norm

· The Scorecard is used to target short and long term investment
- It's not a pass/fail grading tool; it helps identifies strengths and needs
- The goal is not perfection. The goal is improved productivity

· A Initial Data Strategy shouldn't include everything
- "A plan designed to improve all of the ways you acquire, store, manage, share, and use data"
- Focus on the components where investment is practical

32

Thank You!

@EvanJayLevy

EvanLevy@IntegralDataLLC.com

www.EvanJLevy.com

33