



Context Awareness Services as a Means to Enable Information Discovery and Access

Bryan Aucoin
Alanthus Associates, Inc.
bryan.aucoin@alanthus.net
703.307.3191 (c)
703.684.9888 (o)
www.alanthus.net



Agenda

- The Basic Concepts from the Federal Enterprise Architecture (FEA) Data Reference Model (DRM)
- What the DRM Says about Data Context
- Information Discovery and Access in Large Federated Environments



The Basic Concepts from the DRM

- The DRM is a complex document, but the underlying concepts are fairly simple.
- The DRM is intended to:
 - Provide a roadmap for Communities of Interest (COIs) to share the information needed to perform their mission/business.
 - Provide some high-level, basic guidance to information technologists and architects on the services that are required to enable the information sharing.



Why are Communities of Interest at the Focal Point?

Because:

- Information Sharing is Increasingly:
 - Intra-organizational
 - Cross-functional
 - Interdisciplinary
- We need to start thinking about information sharing within the context of an “extended enterprise”
 - Federal Government Organizations
 - State, Local and Tribal Partners
 - Private Sector Entities
 - Academia

Increasingly, we solve problems by forming coalitions...

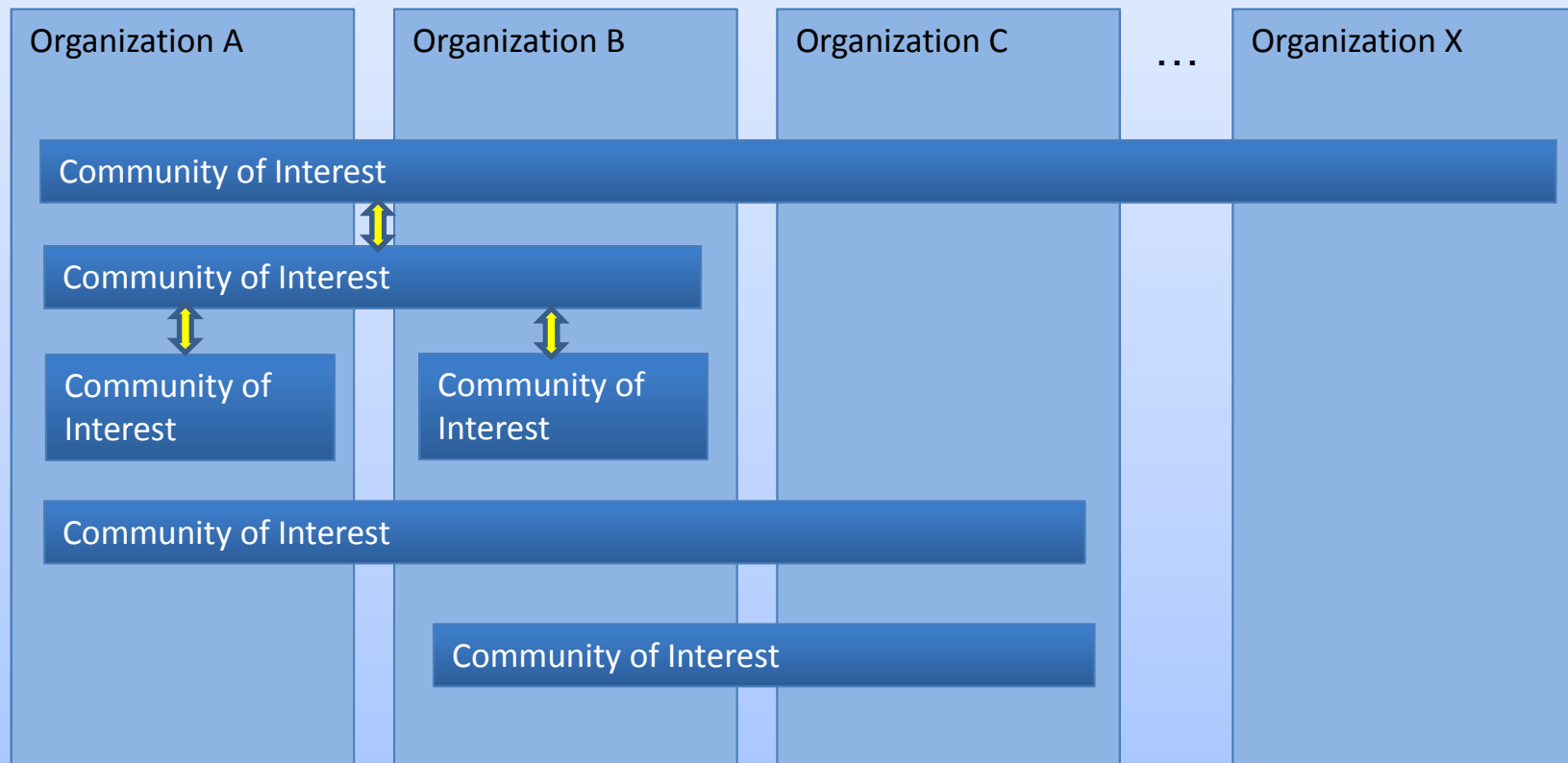
Examples:

- Avian Flu
- Environmental Data
 - Carbon Data
 - Estuaries

Why are Communities of Interest at the Focal Point?



A Extended Enterprise



The Basic Concepts from the DRM

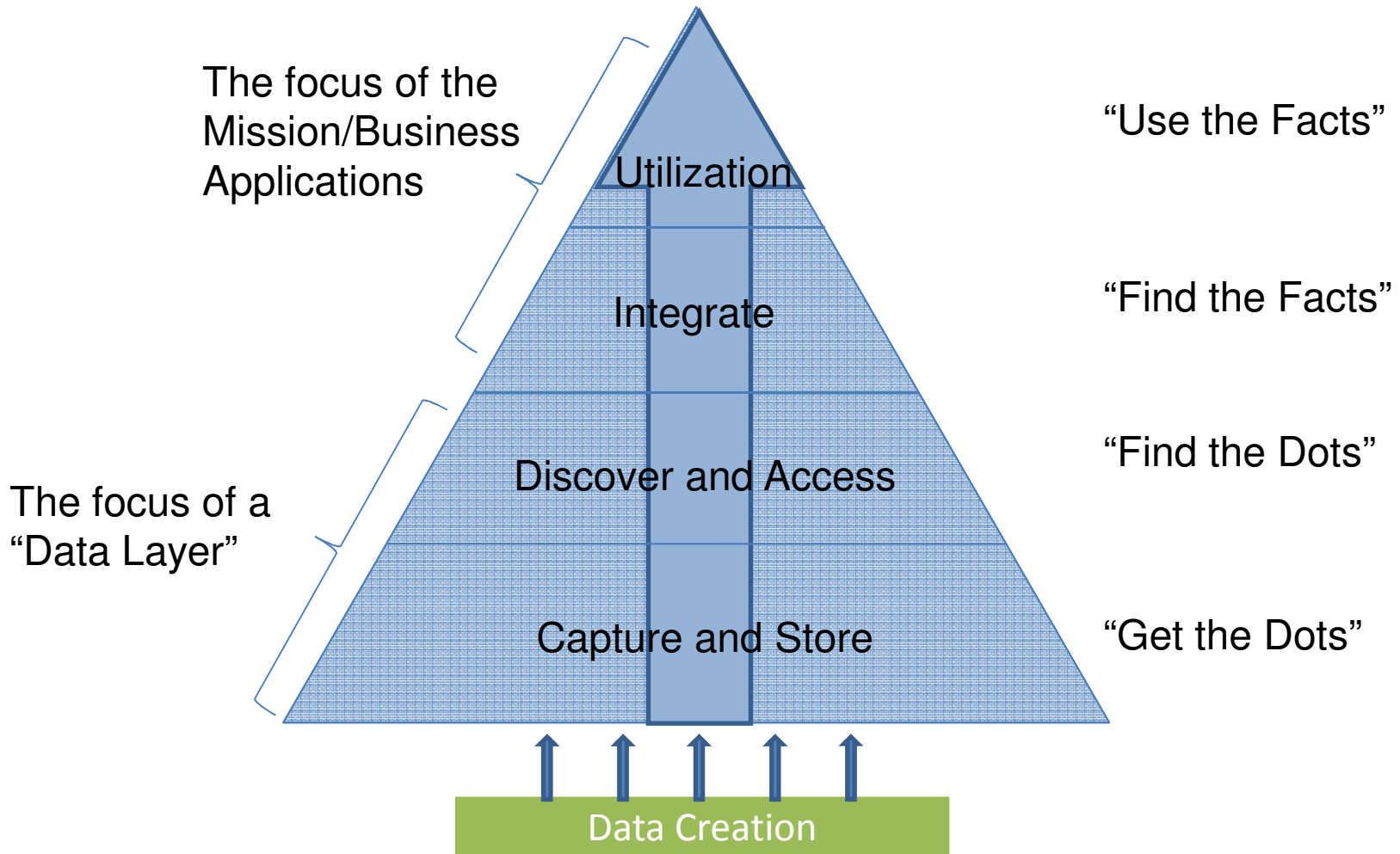
The basic Implementation Framework...

		DRM Chapters		
		Context	Description	Sharing
DRM Sections	Introduction	<ul style="list-style-type: none"> - What are the data needed to support the business/mission needs of a COI? - What core information does the COI need to make the data discoverable and establish governance? 	<ul style="list-style-type: none"> - How will the meaning and structure of the data be conveyed? 	<ul style="list-style-type: none"> - What is the data sharing architecture? (i.e., How will the data be made sharable)
	Guidance	<ul style="list-style-type: none"> - Define subject areas and entities of interest - Identify data sources and stewardship - Establish governance 	<ul style="list-style-type: none"> - Establish semantic and syntactic standards 	<ul style="list-style-type: none"> - Establish the Data Sharing services required to support the data sharing needs of the COI
	Abstract Model	<ul style="list-style-type: none"> - Document in accordance with the DRM abstract model 	<ul style="list-style-type: none"> - Document in accordance with the DRM abstract model 	<ul style="list-style-type: none"> - Describe services specifications in accordance with the DRM abstract model

DRM Implementation Framework

Within the COIs...

An Hypothesis: Information Architecture as a Hierarchy of Needs





The DRM Implementation Framework has a Number of Implications...

COIs need to be able to...

- Discover information assets that are relevant to their mission/business;
 - COI missions, and the information they require to fulfill those missions, cannot necessarily be defined a priori
 - Discovery has to take place across broad federated “extended enterprises”.
- Reconcile and harmonize their semantics;
- Define common syntactic standards; and,
- Deploying data sharing services that permit the discovery and access of information.



In Simpler Terms, Within Any Given Extended Enterprise...

COIs need to:

- Know what we know
- Know what the data mean
- Know how they are structured
- Know how to get to them

*This is basically the
DRM Concept of
Data Context*



A Short Discussion of Search Engines

- One primary focus within the Federal Community has been making data discoverable by and accessible to the citizens of the nation.
- Search engines have played a prominent role.
- Some have argued for a seemingly logical step: Make everything discoverable through search engines.
- However, search engines have limitations as a complete solution.



A Short Discussion of Search Engines

- Simply and obviously, search engines can only make the information that they can access available for discovery, and...
- There are any number of types of information that we would not want to be accessible to everyone:
 - Private information that would enable identity theft.
 - Medical records
 - Proprietary Company information
 - Information sold as a value added service
- *So, how do we strike the right balance?*



Now, back to the question of COIs...

- The intent of the Data Context described within the DRM was to address...
 - Knowing what we know
 - Knowing what the data mean
 - Knowing how they are structured
 - Knowing how to get to them
- The intent of Context Awareness Services in the DRM was to make that information discoverable by COIs within an extended enterprise.



DRM Description of Data Context

The DRM uses an accurate, albeit esoteric definition for data context. Specifically:

“The Data Context standardization area facilitates the discovery of data through an approach to the categorization of data according to taxonomies, and provide linkages to the other FEA reference models.”

DRM Description of Data Context



The FEA DRM provides the following guidance on what context data to make available for discovery:

“Typical examples of Data Context for a given Data Asset may include a Topic identifying a subject area, a data stewardship assignment, sources of record, etc. At a minimum, the Data Context for a given Data Asset should answer the following questions:

- What are the data (subject Areas/Topics and entities of interest) contained within the Data Asset?*
- What organization is responsible for maintaining the Data Asset?*
- What is the linkage to the Federal Enterprise Architecture Business Reference Model?”*
- What services are available to access the Data Asset? (See Data Sharing)”*

DRM Description of Context Awareness Services



The FEA DRM goes on to define “Context Awareness Services” as follows:

“A context awareness service allows the users of a collection (a set of data assets) to rapidly identify the context (as defined above) of the data assets managed by the Community of Interest. Context information may be captured in a formalized data architecture, a metadata registry or a separate database.”



The Way Ahead

(or a least a glimmer of one)

- The current approaches for data context:
 - Organizational Metadata Registries to capture some context information (e.g., DDMS)
 - Organizational Service Registries to capture service descriptions and endpoints.
 - Formal/semi-formal registration requirements and supporting process within organizations.



The Way Ahead

(or a least a glimmer of one)

- It's not clear (to Bryan, anyway) how such approaches will migrate to Internet scale.

- Why?

- Our information sharing requirements are increasingly:

- Intra-organizational verses inter-organizational, or worse;
- Intra-enterprise verses within an enterprise.

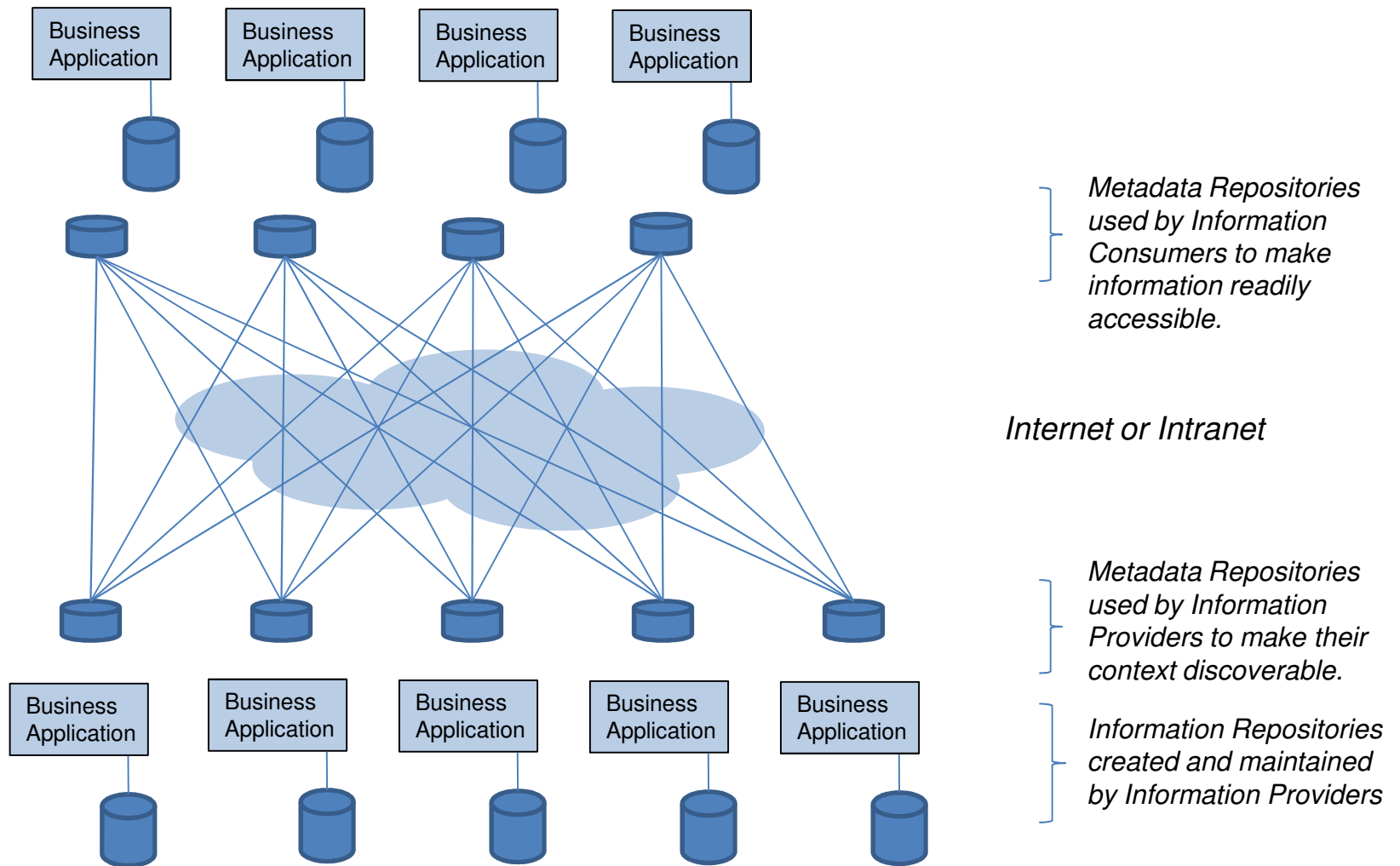
Any such registries will not be aligned to serve the wide variety of “extended enterprises” that any given organization must support. Examples:

- Health and Human Services
- Environmental Information

- Large scale enterprise-level registries will meet a broad “least common denominator” approach to data asset and service inventories, and therefore they will add value.

However, COIs will need and will create independent registries to capture the data asset and service descriptions that they need to do their jobs.

In Fact, the Data World is Starting to Look More Like This:



Think about approaches used for geospatial data.



The Way Ahead:

Some Suggested Parameters For a Target State Solution

- Data Context should be discoverable by enterprise-level search engines within an extended enterprise.
 - Provides a rapid means for COIs to discover that data assets relevant to their mission/business.
 - Provides a rapid means for COIs to deploy metadata & service registries to align to their interests.
 - These registries become platforms for business/mission applications.



The Way Ahead:

What would be required to implement?

- Common Standards for Context Metadata for a Data Asset:
 - **Topics within the Asset:** “Knowing what we know”
 - **Definitions/Semantic Descriptions:** “Knowing what the data mean”
 - **Data Models/Syntax Descriptions:** “Knowing how they are structured”
 - **Service Descriptions, Access Policy, Stewardship Assignments:** “Knowing how to get to them”
- Common Methods/Practices for Making Context Metadata Discoverable and Accessible:
 - Generic XML Schema: Data Asset Providers publish an instance
 - Style Sheets
 - RSS Feeds
- A Certification Process managed by an independent organization, preferably an automated one.



The Way Ahead:

Advantages of This Approach

- Supports the concept of COIs within one or more extended enterprises.
- Comparatively low barrier to entry:
 - Doesn't require (multiple) registration
 - Shifts the onus to the information producer to make information assets discoverable
 - If done right, this becomes a standard that can be incorporated into commercial products. Packages can exposed their internal data assets for discovery and reuse. Like an API, put discoverable and useable on Internet scales.
- Allows COIs to focus on the information they care about.



The Way Ahead:

Proposed Next Steps

- Set up a “Community of Interest” for Context Metadata Standards
- Engage Stakeholders
- Define the standards through an appropriate standards body



Acknowledgements:

I would like to thank the organizations for their intellectual contributions and sponsorship:

- Association for Enterprise Information: (*See Federated Governance of Information Sharing Within the Extended Enterprise* on the AFEI Web Site)
- The Mark Logic Corporation
- The Original DRM 2.0 Authors



Questions?