Strategic Plan:
Providing High Precision Search to NASA Employees Using the NASA Engineering Network

By Jayne Dutra and Lisa Smith

Jet Propulsion Laboratory
California Institute of Technology

January 23, 2006

1.0 Introduction

Workers today expect to access material in a self service information environment where the information is delivered to them seamlessly. Because the trend towards content publication on the Web is accelerating, it is apparent that more efficient ways to manage content are critical to any enterprise wanting to be successful in a dynamic information environment. The goal of this plan is to briefly describe new technologies available to us in the arenas of information discovery and discuss the strategic value they have for the NASA enterprise with some considerations and suggestions for near term implementations using the NASA Engineering Network (NEN) as a delivery venue.

The amount of content produced and published to the Web continues to grow at a brisk pace. According to a recent study by the University of Berkeley, there are now 250 megabytes of content for every person on the earth and we are continuing to create new content at the rate of one to two billion gigabytes a year.[1] The sheer amount of content makes finding any one piece of relevant data a daunting and time consuming task. Additionally, big IT investments in the late nineties encouraged the use of proprietary technologies. These technologies meant that creating linkages between systems required APIs needing hard coded interfaces. The results have been fragmented information spaces in legacy applications that do not fit with current NASA business goals. Also, security concerns have led to the implementation of firewalls that silo information around Center-specific environments. Thus the ability for NASA engineers to share and re-use the accumulated knowledge of the agency over the Web is greatly hampered.

Up until the present, NASA engineers have had to guess where a particular piece of information might reside in order to query the system and retrieve it. Not only do users need to know where exact types of data are located, but they also need to know the correct keywords to enter into the search box. Because there isn't yet a consistent set of controlled vocabularies in use across the Agency, keywords associated with documents can be generated from anywhere by the author and they might be expressed in highly technical language that is unknown to the worker trying to access content further down the information food chain.

Clearly this is an environment where the chances of finding exact information at the time it is needed are very low. In the NASA 2003 Strategic Plan, one of the Agency's primary goals is "to develop new capabilities and revolutionary technologies that will change the definition of what is possible". It goes on to say that "we will assemble new tools and architectures to provide an intuitive, highly networked engineering design environment that will unleash the creative power of engineers and technologists across the Nation." In order to achieve such lofty goals, we need to rethink how we are providing information to NASA employees.

2.0 Vision and Applicable Technologies

Instead of workers having to search repositories individually, we should work towards an environment where information is delivered to them via the NASA Engineering Network portal based on rules derived from task completion, work processes and roles. Although we will probably never completely lose our need for individual search engines, in the future we should come to rely on them less and less. As an alternative, information should be clearly marked with the attributes necessary to move it where it is needed at the time it is needed. In situations where a formal search is required to find relevant data, the results should be clustered and presented in

[1] M. Strohein, S, Stearns, *Content Management That Fuels the Real Time Enterprise*, Outsell and Inmagic, 2003.

ways that allow guided navigation through large information sets depending on the user's point of view and context.

The NASA Engineering Network (NEN) task includes several components designed to improve information accessibility for NASA employees including an Experts Locator, a knowledge base called the Lessons Learned System and a suite of tools designed to facilitate virtual collaboration among teams. These applications are housed in a knowledge portal serving various NASA communities of practice. In order for these products to be most successful, a robust information infrastructure should be in place. The infrastructure should be made up of both conceptual elements and technical application architecture that is meant to facilitate the interchange of data across the internal information landscape.

NASA has recently seen some internal developments key to the implementation of a cohesive information architecture. The creation and delivery of a stable NASA Taxonomy in spring of FY 04 marks the first time the Agency has adopted a consistent data reference model for its content. The release of the Agency Taxonomy provides a common semantic framework that developers can build to while being sure that their components integrate into a larger architecture. The NASA Taxonomy is meant to act as an upper level brokering schema allowing the many repositories across the Agency to be mapped into a common reference model from local domains for re-use and consumption by various customers.

Taxonomies contain descriptors that are used to mark content chunks. They are composed of discrete branches also known as *facets.* Facets are made up of consistent sets of attributes for labeling content components and can be repeatable. Facets allow taxonomies to be multi-dimensional, which accommodates a wider range of content types. Taxonomies that are designed to be modular and extensible will be the most robust.

The NASA Taxonomy delivery also includes a set of recommended Dublin Core metadata tags and some extensions which together make up the first NASA Core Metadata Specification. The specification has been rendered and published as XML schema. The Core Metadata Specification provides a framework for consistency in the handling of electronic information. The Core Metadata Specification is intended to define the fields for NASA information and the Taxonomy is intended to express the values that should populate the metadata fields. Both the NASA Taxonomy and the NASA Core Metadata Specification[2] are freely available to all NASA Centers and Enterprises for use in the building of applications and content repositories using new technologies that enable the Semantic Web.

The Semantic Web

In 1998, Tim Berners-Lee authored the Semantic Web Road Map. In this document he describes an Internet space that has been enabled for use by machines and automated systems. One of his primary assertions is that information available to us through the Web up until now has been "designed for human consumption, and even if it was derived from a database with well defined meanings (in at least some terms) for its columns, that the structure of the data is not evident to a robot browsing the web." [3] Through the use of mechanisms such as Resource Description Frameworks (RDF), appropriate metadata and schema frameworks, he proposes the evolution of a Semantic Web that is composed of machine-understandable information.

---

[2] NASA taxonomy Web Site http://nasataxonomy.jpl.nasa.gov
[3] Tim Berners-Lee, *Semantic Web Road Map*, 1988, http://www.w3.org/DesignIssues/Semantic.html.

For information to be found and acted upon by multiple systems, it is necessary to pre-define its scope and meaning. These data definitions are expressed through schema and reside within Uniform Resource Identifiers (URIs) which are easily referenced and found by machines on the Web.

Enterprise Framework Ontologies

Once terms are defined through the use of taxonomies and RDF statements, their relationships to other terms can be specified through the use of ontologies. Ontologies for the Semantic Web are most commonly composed of a taxonomy tailored for the data and a set of inference rules. Taxonomies allow us to define classes of objects and the relations among them. Implemented together, classes, subclasses and relations can be used to express a wide range of information through the use of properties. By allowing subclasses to inherit the properties of their more general parent classes, systems can deduce the proper meaning of derivative terms even if the system does not have a direct link to the original database[4].

Taxonomies can overlap information spaces and allow them to interrelate. This higher meta level of taxonomy formation is expressed through an ontology. An ontology is defined as “an explicit specification of a shared conceptualization.”[5] A conceptualization consists of relevant concepts of a domain, the relationships between these concepts and agreed upon knowledge about these concepts by a group. A formal ontology enforces well-defined semantics on a conceptualization, which can be described through XML elements. Content produced in such a way is referred to as being "semantically enabled".

Ontologies can be used as interchange formats, enabling mediation between systems in a Web Services model. When implemented with controlled vocabularies and taxonomic underpinnings, ontologies enhance reusability, search results, reliability, and knowledge acquisition. Ontologies and topic maps can allow us to catalogue what we know and what we don’t know, helping to shape our future research efforts as an Agency.

Using Metadata About People to Increase Knowledge Sharing

As we develop useful metadata attributes about content, we are also developing an understanding about the metadata set we need to know about people. Role architecture is in its infancy at NASA, but it is key to creating an environment where the right content can be matched to the right person at just the right point in their work when they need it. Associations and relevancy based on role, work breakdown structures, discipline and access rights will all provide valuable attributes to push appropriate content to various flight teams or communities of practice.

The POPS Ontology was developed at NASA Headquarters to leverage information regarding its employees and their core competencies in FY 05. It is designed to act as an additional layer of data on top of the NASA Directory enriching our understanding of which skill sets are needed for successful mission completion. The POPS Ontology was developed from the FOAF (Friend of a Friend) ontology and utilizes content from the NASA Competency Management System based at Johnson Space Center It is a stable foundation for source information regarding the disciplines and areas of expertise utilized by NASA employees.

Organizing data to inform mission judgments by providing associated content helps us examine problems in order to mitigate risk and become a proactive, learning enterprise. This allows each

---

[4] Tim Berners-Lee, *The Semantic Web*, Scientific American, 2001.
[5] Tom Gruber, *Stanford University.*

mission to build on the missions that have come before. This is the foundation architecture for a true NASA knowledge base.

Content Processing Architecture

The Unstructured Information Management Architecture recently released by IBM provides a framework for describing and invoking the operations involved in processing content for knowledge applications, including the production of metadata.

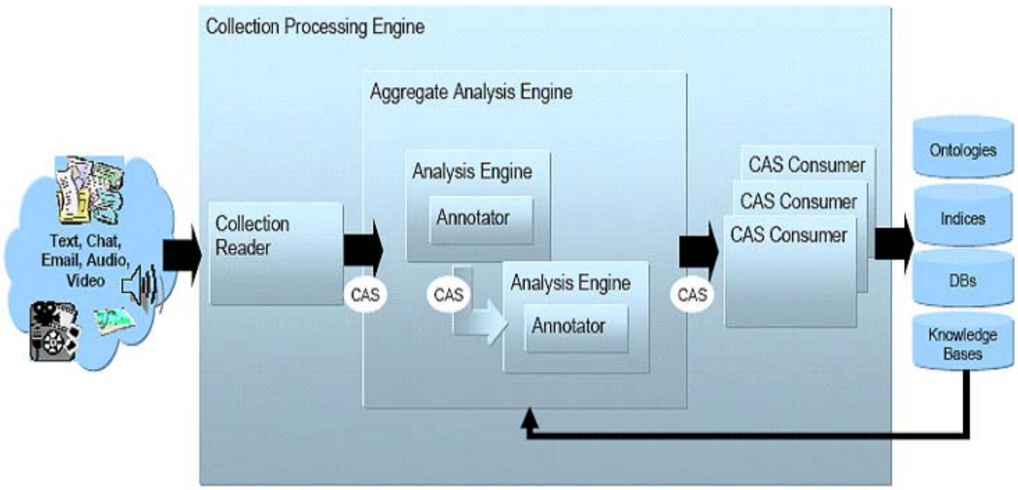

Figure 1.
High Level UIMA Component Architecture from Source to Sink[6]

The above diagram illustrates how a variety of source content types can be processed. The Collection Processing Engine contains three operations for processing content—reading the content, analyzing the content, and consuming the content.

The first process involves reading the content. This process is performed by the Collection Reader and may include some form of conversion to text if the content is in a proprietary or encoded format such as PDF, as well as extraction of previously applied metadata. After converting and reading the content, the Collection Reader creates a record of the content called a Common Analysis Structure (CAS).

The CAS is passed on to the second process which involves analysis of the CAS record. Analysis can be performed by one or more Analysis Engines, each of which may contain one or more Annotators. When more than one Analysis Engine is used, an Aggregate Analysis Engine is defined. An Aggregate Analysis Engine specifies whether each Analysis Engine will run serially and in what order or specifies whether each should run independently. The CAS record is updated with metadata according to the Annotators used in each Analysis Engine.

Finally, the CAS is passed to one or more Consumers. Consumers may apply collection-level analytics or otherwise aggregate the CAS records in some way which generates another level of metadata. Common Consumers may be search or visualization applications. Note that an Analysis Engine may also be a Consumer depending on its placement in the Content Processing Engine.

[6] IBM, *UIMA SDK User's Guide and Reference*, 2005

A crucial aspect of the UIMA framework involves the use of various knowledge resources such as Ontologies and Knowledge Bases. Processed CAS records can be used to inform the development or maintenance of these resources. Meanwhile, the knowledge resources themselves are used by the Analysis Engines to perform content analysis and annotation. This kind of feedback loop is what makes it possible to dynamically and accurately process a variety of content sources.

Metadata Auto-Population Techniques

The auto-population of metadata is fast becoming significant to knowledge management practices at enterprise and government organizations worldwide. This is because there is so much legacy content that lacks metadata altogether and also because there is so much new content created that it will typically only receive cursory metadata as an afterthought. Content with limited or no metadata is difficult, if not impossible, to discover and re-use. The solution is to add metadata via automated means.

There are now a number of metadata auto-population applications available on the market. Through a combination of text analytics, statistical modeling, and algorithmic processing, these products can auto-populate metadata for a variety of content assets.

Text analytic methods include tokenization, stemming, term frequency, term co-occurrence, and part-of-speech tagging, among others. Statistical processors and algorithms use this information to determine keywords, key phrases, key sentences, conceptual categorizations and content clusters.

How do auto-population applications fit into UIMA? Auto-population applications typically come with Content Readers to convert content to ascii text and extract previously existing metadata. In addition, they contain a variety of Analysis Engines (with Annotators) to perform analysis and indexing of the content. Some also include Consumers to cluster and aggregate content.

The most common processors (Analysis Engines) supported by auto-population applications are listed below.

- Recognition – Matches terms or alternate forms of terms from a controlled vocabulary.
- Classification – Classifies to terms in a controlled vocabulary by referring to exemplar content or by applying collection-level clustering algorithms (or a combination of these) to conceptually categorize the content.
- Rules Engine – Applies user-specified rules based on logical statements such as "If…, then…".
- Entity Extraction – Matches to user-specified patterns such as dates, money, or significant proper nouns.
- Metadata Extraction – Extracts previously applied metadata.
- Keywords – Specifies keywords according to statistical algorithms.
- Summarization - Specifies and aggregates key sentences according to statistical algorithms.

Though many enterprise-class auto-population applications provide all of these functions, not all applications do and it is possible to buy several different smaller applications with specialties in one area or another to meet the auto-population needs of the organization.

Controlled Vocabularies and RDF

Metadata, either automatically or manually applied, tends to be most useful when provided from a controlled vocabulary simply due to consistency. Because of this, organizations need to spend significant time and resources in developing and maintaining the vocabulary.

Because business is ever-evolving, controlled vocabularies in many organizations tend to be fairly complex. They are usually expressed as taxonomies or ontologies depending upon their scope and purpose. They contain a multitude of terms and term inter-relationships. As such, the popularity of expressing these in RDF is increasing. By making statements about the terms and their inter-relationships in RDF, inferences can begin to be made by machines when they encounter the terms as metadata associated with content.

Inference rules provide the foundation for machines to manipulate terms in ways that are much more meaningful to the human user. Once the work flow models are defined and documented, much can be done with today's technology to embed content along the way, making it appear at the right moment in the worker's business routines.

Service Oriented Architectures

In order for the infrastructure to move information between systems and individuals, an appropriate technical architecture must be in place. In the past, enterprise applications needing to interact had to be tightly bound with proprietary APIs. These interfaces were built one at a time for specific task enablement and could be easily broken by a change in configuration at either end of the information transaction chain. The new service oriented architectures (SOAs) depicted in Figure 2 are based on Web services. There are three fundamental components to Web services:

- SOAP (simple object access protocol) - the transport layer for XML; it is the means of moving content from one application to another.
- UDDI (universal description, discovery and integration) - this is a kind of central "yellow pages" where a Web application can seek and discover other Web services it may need in order to complete an electronic transaction.
- Web Services Description Language (WSDL) - allows a service to describe how it functions and how another application can invoke its services

Unlike previous interfaces which are usually bolted together with proprietary APIs, the new service oriented architectures are typically loosely coupled. SOAs are self-describing and bind together dynamically at the moment that the components are needed. This provides a more flexible and granular application interface.

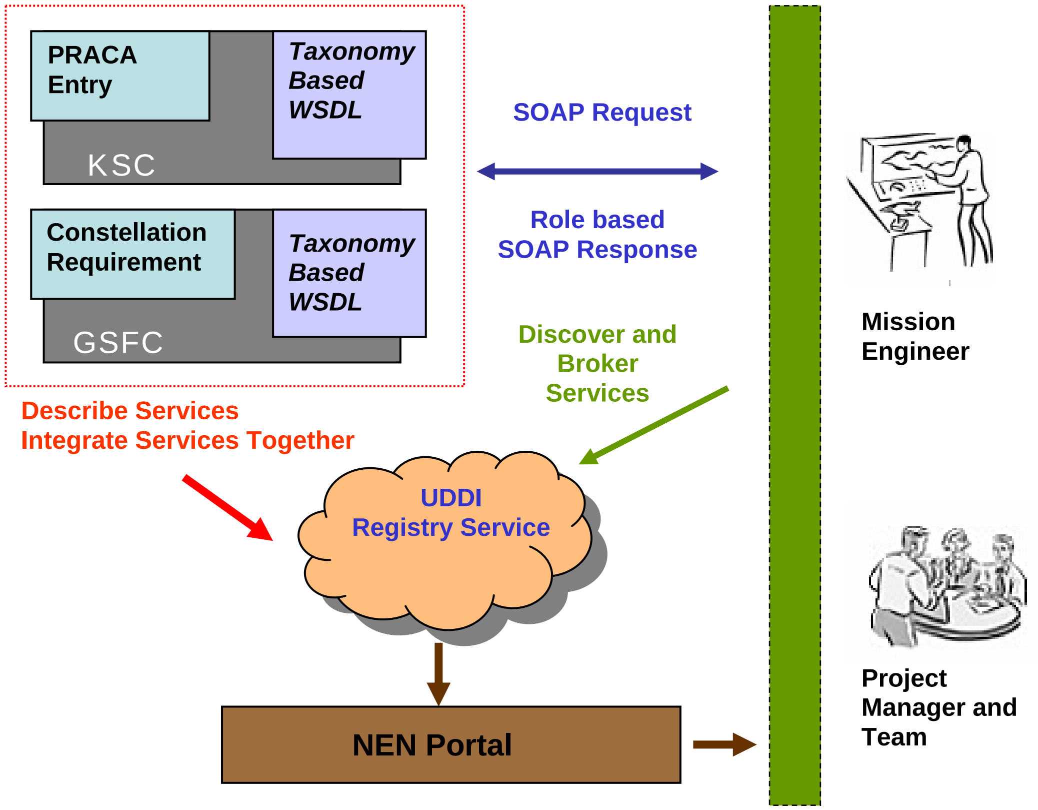

Figure 2.
Utilizing Semantic Web Components To Enable Information Interchange

3.0 Goals and Objectives

One of the fundamental goals behind taxonomy formation and adoption is to develop a consistent methodology for handling NASA's electronic content. Information objects need to be described with a standard classification scheme that follows a predefined hierarchy. This will enable users to see correlations between subject areas. It also allows search engines to retrieve information with more precision and relevancy. Each time an engineer or scientist finds and reuses a piece of content, the return on investment (ROI) of the work to originally produce the material increases 100 percent. This cycle of reuse directly impacts the Agency's bottom line. It also pushes the pace of development forward at a greater rate as teams build on previous work instead of reinventing the wheel.

For the NEN portal to provide a seamless delivery venue for NASA content, the architectural design work should focus on the analysis of processes used in the core business of NASA: the development of flight missions meant to further human knowledge through scientific experiments and observations. As flight projects refine the processes involved in their development activities, semantic models can be built that describe the business entities and their relationships to each other, including the logistics of needed resources. The evaluation of processes enables us to specify work flow models that result in robust business products (i.e., a propulsion system appropriate for the mission's science goals, or an instrument designed to capture critical data).

In addition, NEN developers need to understand information models at use at each NASA Center. It is recommended that Memoranda of Understanding be agreed to for each system holding significant NASA material and that local owners be encouraged to modify their systems to enable sharing of information. This can be done through a phased process described below. It might also be advisable to develop Service Level Agreements with system owners that detail the service profile of each system, its API for data interchange and the cycles of export activity.

4.0 Proposed Strategy

Moving Forward

In order to achieve the Vision for Space Exploration as articulated by President Bush, NASA is moving ahead with the design and development of the Constellation Program. The program calls for a model based engineering approach using a schema called NexIOM. The schema is meant to be expressed in XML and its attributes can be shared across the Web in RDF or OWL files.

Stakeholder Goals and Objectives

The goal is to implement the notion of an infrastructure environment for machines to find and retrieve information relevant to the individual independent of a situation-specific human query, but rather designed to be initiated upon the worker arriving at a certain point in his daily work process. This type of content delivery builds on the ideas of the Semantic Web. If funding can be gained, specific deliverables include architectural components and navigation interfaces designed to perform with high-level usability.

Phased Approach

Phase 1 consists of the effort currently under way to set up a pilot NEN site with sample data from a couple of repositories. This “quick-and-dirty” effort shows the potential value of the network and is useful in gaining or solidifying stakeholder support. The pilot is already acting as a basis for identifying issues and determining appropriate solutions for moving forward and has been significant in informing the direction of this strategic paper. Because the NASA Lessons Learned Information System has been re-implemented by the NEN team using a content management system, it is feasible to attach appropriate metadata to the content relatively easily, allowing it to be directed to the targeted community. Recommended metadata fields from the NASA Core Metadata Specification that could be applied to the content would include Author, Date, Description, Competency, Subject, Organization, Content Types, Mission/Project and Instrument.

Matching up metadata about the content with metadata about people using the NASA POPS 2 schema will enable the ability to pass the information to the appropriate NEN community and channels.

A modest next step (Phase 2) would be to identify a core group of source repositories and establish interface agreements with repository owners specifying the data to be exported, the schedule, support contacts, and other foundational logistics. One suggestion which is detailed later in this paper, is to use the output from the NASA PRACA Data Mining and Trending activity. Because this task is funded by the NASA Office of the Chief Engineer and because the content generated is particularly relevant to NASA engineers, it would be a good way to start identifying and repurposing materials for re-use.

Probably one of the most significant aspects of this project—and one that will need to be ongoing—requires developing the means to effectively export repository metadata into a central NASA Engineering Metadata Catalogue. In this phase, it is not the technical architecture for how to get data into the catalogue that we are concerned with, but rather the question of semantics that comprises the foundation for the information architecture. With the development of the NASA Taxonomy and Metadata Core Specification, NASA has begun to build a foundation for data interchange; however, there will be a need to first map the local systems to a local model or taxonomy for the data reconciliation process to work smoothly.

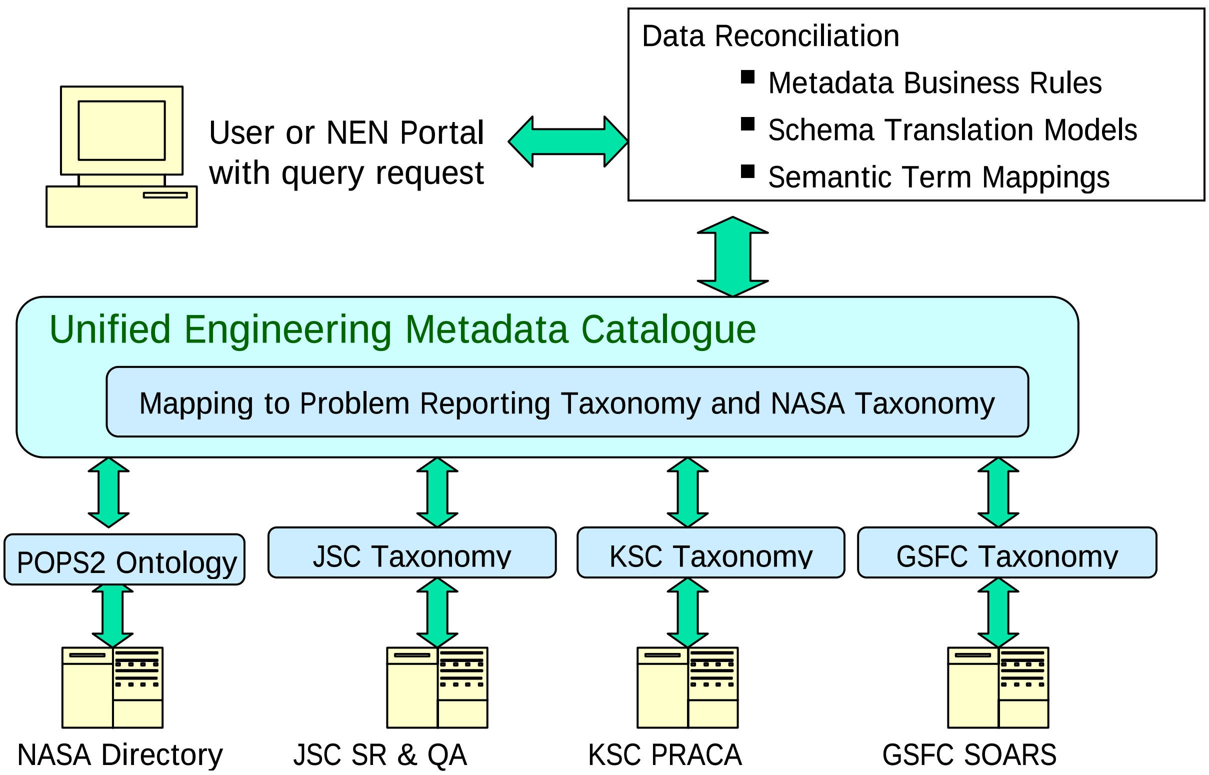

Figure 3.
Notional Layered Semantic Architecture for NEN Information Portal or NASA Search Across NASA PRACA Systems

As the diagram above details, using the NASA PRACA systems as an example, the owners of each repository would need to take steps to map the values in their repositories to their local taxonomies. These local taxonomies should map up to the NASA and Problem Reporting Taxonomies, thus making data interchange possible at the Agency level. Failing this approach in the case where local taxonomies have not yet been developed, repository owners will need to map repository values directly to the NASA and Problem Reporting Taxonomies.

As this kind of data reconciliation can be difficult, tedious, and time consuming, especially when done manually, it is recommended that some kind of content processing mechanism be implemented that corresponds to the UIMA framework. The following refers to a Verity K2 solution for example purposes, but that is by no means a required application. There are many other COTS products which would provide similar—and potentially better—solutions.

Using UIMA, the Verity K2 solution will be described first. The UIMA Content Processing Engine in this solution includes the three content processing components: Collection Readers, Analysis Engines, and Consumers. Verity K2 provides a DDA (Dynamic Data Access) Layer which provides Collection Reader functionality including repository Gateways, document Filters, and language Locales. Additionally, a Tokenizer used in creating a word index is employed. The outcome is a Verity "Collection" which supports a CAS record for each content item. The CAS record is then passed to the Analysis Engine(s). One form of Verity K2 Analysis Engines will use the NASA and Problem reporting Taxonomies as Verity Knowledge Trees to automatically categorize content. Others may use Verity's Feature Extraction and Summarization techniques to add metadata to the CAS record. Finally, the CAS records may be aggregated and clustered using the Verity Clustering technique. Through the use of these Analysis Engine(s), content metadata appropriate for use in a centralized metadata catalogue will be automatically populated.

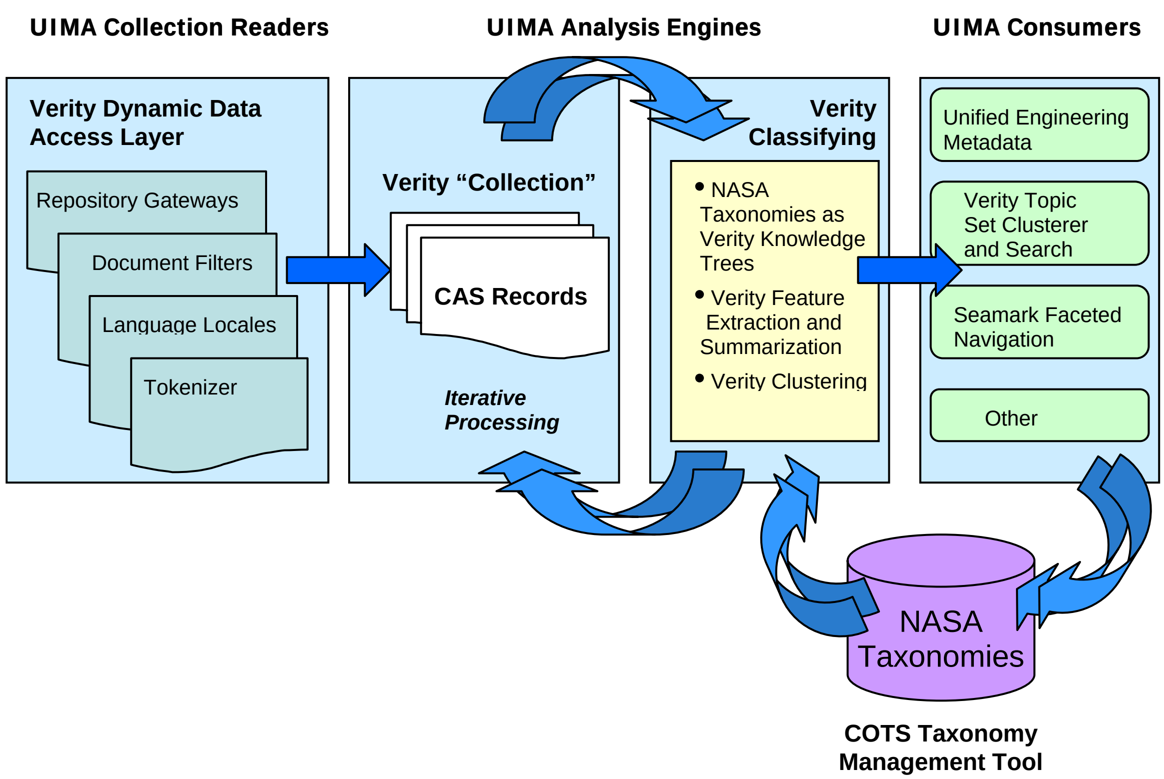

Figure 4.
UIMA Framework Applied to Verity K2 Capability

The primary CAS Consumer will be what outputs metadata to the Unified Engineering Metadata Catalogue as detailed in the previous diagram. Other CAS Consumers may include the Verity Topic Set Clusterer as an option, but should result in the hierarchical display of the NASA and Problem Reporting Taxonomies with content items appropriately categorized under these browseable models. Additional Verity Consumers include the various Text and Parametric search capabilities built into the K2 product.

As specified by UIMA, the results of content processing, particularly the auto-population of metadata from the Analysis Engines, should inform the continued development and maintenance of the NASA and Problem Reporting Taxonomies. Work on these taxonomies will also be ongoing and should be thought of as a parallel phase of content processing. This work includes the addition of new terms and the deletion of expiring terms as well as making hierarchical, equivalent, and associative relationships among terms to effectively support semantic processing such as auto-categorization and inferencing.

To be most effective, a taxonomy management application of some sort will need to be procured. As Verity K2 does not provide a taxonomy management component that meets the rigors of the Z39.19 ANSI/NISO standard[7], other COTS products should be considered. One example is Teragram[8] which provides robust taxonomy management software. In addition, Teragram provides content processing software components which could be used in place of the Verity K2 solution described above.

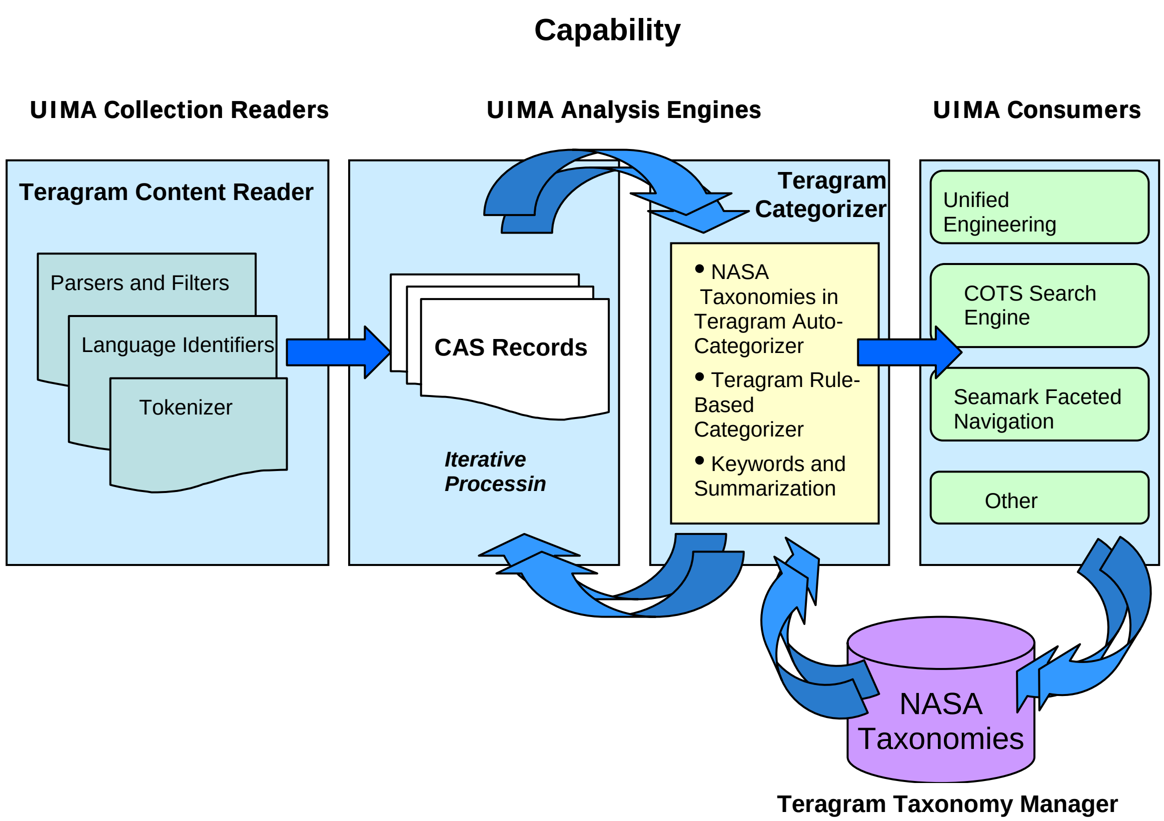

Figure 5.
UIMA Framework Applied to Teragram

[7] American National Standards Institute/National Information Standards Organization, *Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies*, Bethesda, MD, August 2005.
[8] Teragram Corporation, www.teragram.com.

To support a semantic web-enabled solution, both the taxonomies and the content metadata need to be expressed in RDF. In the long term, using RDF will allow for inferences to be drawn by machines when accessing data. This will mean more relevant content being pushed to the right people. In the near term, it will allow for other means of accessing information beyond the traditional search capabilities offered by Verity. An example would be the use of the RDF in Siderean Seamark's guided navigation application[9], which provides multiple avenues of entry to content items depending on user perspective.

The next phase will include the standardization of the architectures and processes described above to allow for additional repositories to be added at any time. With the architectures in place and the processes established, it should be a relatively simple act to add new repositories.

The step-by-step process is summarized in the table below:

Step	Task
1	Identify target content at NASA Center
2	Sign MOA with content owner specifying data interchange mechanisms
3	Document content data structure and/or architecture
4	Map data reference models of targeted content to NASA Taxonomy
5	Extract content and associated metadata from Center repository into central NASA Metadata catalogue
6	Perform UIMA techniques for iterative metadata auto-population
7	Once content is properly tagged, deliver into various venues
8	Continue looping this process for more richness of vocabularies
9	Develop SOAs for continuous content streams over time as resources allow

5.0 Next Steps

As these steps are repeated over time, it is recommended that the NASA team procure a taxonomy management tool in order to store all of the terms in use at NASA and track the mappings from individual Center repositories to the brokering layer of the NASA Taxonomy. In addition, as recommended earlier, an auto-population tool should be implemented in order to perform the processes necessary to flesh out any missing metadata. There are many products available today and a survey of current market offerings is appended to this Plan.

Once thesaurus relationships between terms has been well established, the semantic mappings will grow richer over time and be able to handle a broad diversity of content. If the technical scope of the content is very narrow, it may be desirable to create a specialized taxonomy for the NASA Engineering domain. This could in turn be normalized at the higher level of the Meta NASA taxonomy depending on the customer base needing content delivery. This will become evident as more content is mined and extracted.

[9] Siderean Software, www.siderean.com.

As the content is tagged, it becomes available for a multitude of distribution venues including the NEN community portals. As mentioned above, one suggestion for early implementations might be to work with the NESC PRACA Data Mining and Trending Working group to identify anomaly reports and other content gleaned from the various NASA problem reporting systems. These reports would then be tagged in such a way that they could be sent to a NEN portal channel with the appropriate engineering audience. Thus, a recurrence of problem failure reports regarding the power generation system on one of the shuttles might be delivered to a NEN portal for the thermal engineers. In addition, the taxonomy could be used to tag filtered RSS feeds from other NASA sources, such as the GIDEP alerts and other Agency and industry publishing services. By tagging content appropriately and reconciling it with metadata about NASA engineers, many feeds can be directed to the proper consumption point for a tailored self service information environment. This suggestion would demonstrate good synergy between NESC teams and leverage the efforts of both tasks.

Summary

Understanding complex technical information requires the presence of relevant supporting data which makes up the context for accurate data interpretation. Missing factors can be a crucial source of errors in the decision making processes of knowledge workers. Without a complete picture of past experience, we risk repeating mistakes and reinventing the wheel. Therefore, IT work flows that facilitate relevant content aggregation and robust search are key for an effective work force to operate efficiently.

However, in order for the appropriate content to be delivered to the right consumers, we must first build out a technical infrastructure able to handle the demands of an information architecture tailored to NASA's core business tasks. These tasks are most often associated with the engineering work surrounding successful space mission development.

This plan has offered a high level framework for the development and implementation of an Agency-wide search and content delivery architecture that would benefit the NEN project and NASA engineers. The plan calls for a technical infrastructure deploying Web Services and Service Oriented Architectures (SOAs) for a continuous stream of content that can be regularly harvested, tagged and made available for downstream distribution through portals, email or other desk top applications in order to guarantee the best knowledge sharing of NASA's intellectual assets.

It further calls for the consistent documentation of metadata regarding people and their attributes. By combining these types of information together, NASA can increase its ability to utilize the lessons learned from its experience and apply it proactively going forward. In cases where metadata is missing or incomplete, an auto-categorization tool can be employed to fill in the gaps. Appendix 1 provides an analysis of requirements for such a tool and a survey of current market products that might meet those requirements.

The NEN project offers an opportunity to begin building the foundational and infrastructure elements necessary for a robust information environment capable of delivering data where and when it is needed by NASA workers to successfully do their jobs. The vision of knowledge management is to capture, organize and re-distribute information in a way that increases the quality of an organization's output, products and final results. NASA should deploy the components described above to move towards making that vision a reality.

Acknowledgements

This work was performed at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration.