# 4. INFORMATION ARCHITECTURE

Internet users have historically employed two approaches to finding information on the World Wide Web: 1) browsing across topical subject categories, and 2) searching for specific instances of informational keywords. Both these behaviors need to be addressed in order to give users support across the full range of information navigation services.

## A. INFORMATION ARCHITECTURE: WHAT'S BEHIND THE DOOR?

A top goal of any portal is to act as a single point of access for users to find desired information generated by the enterprise on topics of particular interest to them. The portal is literally a door to an Agency information catalog rich in knowledge about the topics in which NASA specializes. NASA discoveries and research over the decades span a wealth of information in technology, planetary science, engineering, and many more subject areas. Information architecture seeks to create a topical framework that embodies and enables these areas of interest.

For the purposes of this paper, we will use the definition of information architecture first coined by Richard Saul Wurman in the 1980's

*Information architecture is the art and science of structuring and organizing information systems to help people achieve their goals.*

For this report, we will focus on the aspect of information architecture (IA) that concerns itself with designing organizational systems for content, creating consistent labeling schemes, and devising navigational pathways through sets of associated data. (This is only a portion of the larger IA that NASA needs to address.) The goal of information architecture is to facilitate knowledge access by building taxonomies, categorizing information, and creating site maps to enable user exposure to relevant material. In order to accomplish this, we must first understand our content and users.

As we examine the contents of NASA web space, it is useful to know the audience types that will be visiting the site and what their needs are. The NASA portal will be developed for two primary audiences: internal and external. Clearly, the needs of these two audiences and their many communities are different and, therefore, portal navigational mechanisms should reflect their distinctive requirements.

External user groups break down into communities. A report was recently prepared for Code P by SAIC, analyzing the external audience demographics and likely visitors to the **www.nasa.gov** site and any future public-facing portal.[5]

- Internet User Type A
    - Families with children (age six and up)
    - Students
    - Large institutions
    - Some large corporations
- Internet User Type B
    - Business professionals
    - Academics
    - Engineers
    - High school and college students
    - Government administrations
- Internet User Type C
    - Space enthusiasts
    - Scientists
    - Media
    - Researchers
    - Computer enthusiasts

Internal audiences may look very different. They may be organized by job family or by their role on a mission. Research needs to be done in order to properly define and characterize the base customer groupings. Many organizations are engaged in studies that profile their core user groups. They are detailing typical tasks that users regularly perform using the Web and identifying information repositories that must be accessed in order to complete those tasks. This type of information leads to a better understanding of the needs and requirements of the users.

Once there is a common understanding of our customer base, we can begin to address content classification and information architecture issues. A strategy to formulate a useful

[5] Pino, Chris and Brian Dunbar, *HTML 4.01 Tools, Preliminary NASA.gov web Audience Profile*, SAIC for Code P, NASA HQ, National Aeronautics and Space Administration, Washington DC, May 2001.

underlying architecture for NASA web space can be developed from the consideration of three main components (Figure 7):

- Content
- Business context
- User feedback loop for continuous improvement

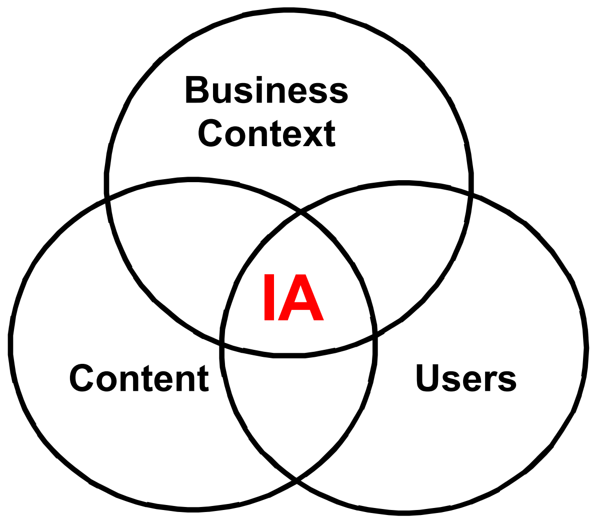


*Figure 7. An Ecological Approach to Information Architecture*

## B. GENERAL STRATEGY: TOP DOWN AND BOTTOM UP

NASA web space has many pre-existing sites on a plethora of subjects. With a *top-down approach* to topical organization, a main hierarchy can be determined that offers navigational pathways for users to take through a NASA Web space directory. This directional approach emphasizes a broad view and includes large topic groupings through which the user can drill down to desired content areas.

With *bottom-up information architecture*, individual content chunks are considered and bridged to the site through navigation from the lower to the higher levels (Figure 8). There are different levels of granularity of information architecture for any site. An information space as large as NASA's will take some time to analyze and categorize.

Both methods work together to create a cohesive web environment. Most sites are a combination of the two, though some sites are more focused on one than the other. Good information architecture is invisible if done well.

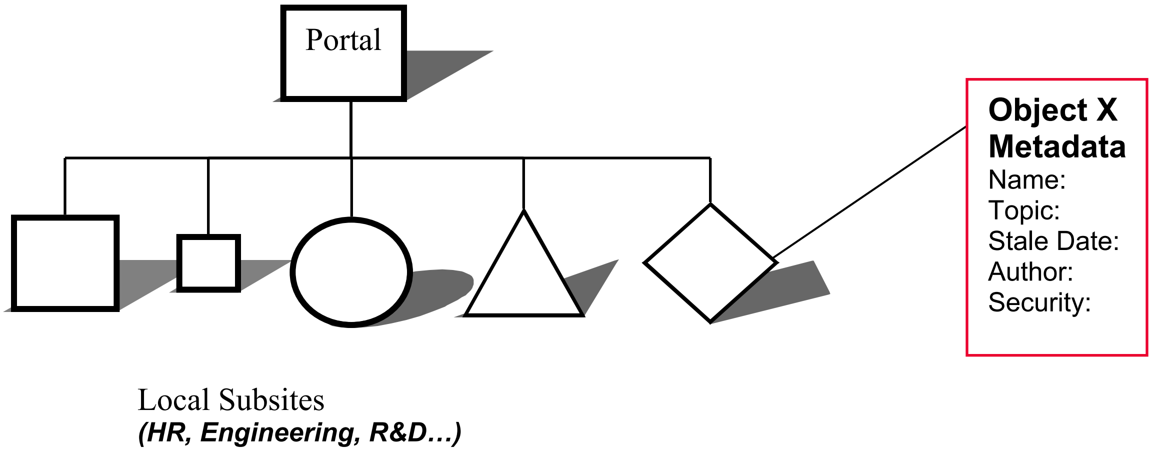


*Figure 8. Information architecture from top to bottom*

## C. ORGANIZING CONTENT AND CLASSIFICATION SCHEMA (TAXONOMIES)

After analyzing user needs, an initial step in designing information architecture is to carefully consider the content of a site. A content inventory and analysis should yield a clear understanding of content requirements. A content map is developed to facilitate the organization of content into intuitive groupings for user browsing.

Once the content is well understood, study can begin on a classification schema that fully describes electronic assets. The general trend in data architecture systems design is in a deconstructionist direction, breaking content down into information objects. Taxonomies expose relationships between data objects and provide a blueprint for an integrated view into the information space. In other words, once the building bocks are broken apart, they are ready to be glued back together in ways that reflect user understanding of the information environment.

A key part of information architecture is the design of taxonomies that introduce users to related ideas. Users may all see the same information; however, they traverse through it in individual ways that reflect their experience and need to know. The associations one makes provide a creative experience and enlarge the value of the NASA web and its possible knowledge discovery connections. Taxonomies are information access tools that encourage brainstorming, collaboration, and improved communication.

Furthermore, taxonomies are necessary due to the complexity and subtlety inherent in language and information retrieval. Most people insert keywords into a search box and click the "Go" button. Keyword searching assumes that individual terms line up with concepts. But, language is inherently ambiguous, so keyword searching often ends in user frustration. Thus, *taxonomies create a contextual framework for information retrieval while mitigating the complexities of language*.

But how do we go about creating taxonomies? Once the content map is created as described above, the next steps are an examination of the content looking for patterns and relationships within the material. The architect is seeking content attributes relevant to users. Content groupings tend to gravitate towards a natural state of affinity. Some classification schemes are known as exact schemes, such as categories that are grouped in the ways below.

- Chronological
- Geographical
- Alphabetical

Other schemes are known as ambiguous schemes. Some examples are:

- Subject
- Audience
- Task

Ambiguous schemes usually provide for more than one entry into the resulting directories, which gives the user a better chance of finding the desired term. Providing multiple pathways to information is usually a good idea since it increases the chance that users will find relevant content and improves usability.

In general, taxonomies progress from the "genus to species" classification model, meaning that they go from broad categories to more narrowly defined groupings. It is important to remember that there is not always one way to express a "best" taxonomy. Taxonomies can be cross-faceted to express attributes of significance to a varied audience. This allows for a flexibility and robustness to site directories and allows them to address the needs of a mixed audience that may have differing goals when visiting a site.

## D. BUSINESS CONTEXT AND CLASSIFICATION SCHEMA

Once a first draft of taxonomies is formulated, it is time to ask content providers, site designers, and other stakeholders if the taxonomy adequately describes their goals for searching and finding online information. In this way, we can check for the validity of the taxonomy in relation to the Agency's desired goals for the portal.

Other goals may include better information sharing among groups such as communities of practice or project teams. The IA developers should study the design of targeted taxonomies tailored to specific team needs. The NASA thesaurus may be of use to identify commonly referenced terms and popular topics.

It is important at this point to consider the scalability of the architecture. Will it suffice in one year, three years, or further? Since the underlying information architecture will be used as a framework to build or map associated metadata, we want to be sure the taxonomy is stabilized and well tested before we begin the next step in developing content management systems. However, taxonomies represent our understanding of the world around us and, as our knowledge evolves, our navigational signposts will have to change as well. So we need to consider long-term maintenance and care by information stewards for the classification schema.

In addition, we will want to consider any upcoming initiatives, branding strategies and NASA's placement in a larger Federal information architecture, such as FirstGov or other e-government projects.

## E. WEB DIRECTORIES AND USERS

Once in place, testing of the taxonomy should occur to see if it is adequately serving site users. Refined taxonomies are generally known as directories. The descriptive words used in directories are extremely important to the ease of information access. Usability testing can confirm which areas are performing at designed levels and which areas need to be re-evaluated.

Jared Spool and Erik Ojakaar discuss typical information foraging behavior on the part of users. They advocate the development of practical taxonomies and "trigger words" that users will typically employ to find certain site material. They recommend the study of user click streams to better understand how users think about site content. They also suggest that the most effective directory structures are designed to expose subcategories, thereby giving the user more clues as to what they can expect to find at the other end of the link.

In addition, they recommend a careful examination of site search logs. Many times, user queries gleaned from these logs can expose what users are really looking for when they come to a site and what words they use to describe the information they seek. Search log analysis can often provide information architects with valuable clues to user needs and typical behaviors.

Spool and Ojakaar also cite the notion of a "level of agreement" from users about what certain terms mean. The more universally accepted the term, the more successful users will be at following the logic of the directory entry to where content resides. Therefore, the more descriptive the terms used in a categorical grouping of site content, the more "intuitive" the site is said to be. This underscores the importance of utilizing users' preferred cultural language and vocabulary terms when it comes to directory design.

Taxonomies are most often developed through an iterative process, meaning that they may take several generations to evolve into their most effective version. In addition to describing the existing corpus of material, new content becomes available and users' needs may grow in unexpected ways. Therefore, it is the recommendation of this team that work on the taxonomies precedes the design and implementation of complex content management systems.

## F. THE ROLE OF METADATA

Once a stable taxonomy has been developed that meets the needs of users, we can begin to build metadata tables that express content attributes. Metadata definitions will generally develop into XML-derived solutions for content management and reuse further down the road, so they are a key piece of the information architecture to get right. SOAP, RSS, RDF, ebXML; these are all specifications that are increasingly coming into common use. In order to fully leverage the power of data interchange, we will want to be sure we have a solid foundation to build on. Metadata represents the building blocks of that foundation.

Metadata can also be optimized in a number of different ways. Metadata attributes can address the varying needs of audience and allow for flexibility within a schema. Metadata attributes that are related but include characteristics of importance to different audiences are said to be cross-faceted—they can serve several audiences at once.

Metadata can be used for helping to implement publication-related business rules. For example, many content management systems use an "expiration date" tag to remove content from the Web that is no longer considered relevant. In the same way, metadata could be used to mark a document with an "Archived" stamp, indicating to users that the information they are viewing is no longer current and should be treated as such when making engineering, design, or business decisions.

Metadata can be used to help enforce authorization business rules. Documents tagged with different levels of classified status can automatically allow a certain class of users to view them and turn away others. This would allow certain external partners (such as academic researchers) access to all public information and only that internal information to which they have been granted specific access.

It is possible that different communities will want to select tags they feel best describes their content. In that case, it would be appropriate to consider the creation of a metadata registry at the Agency level that includes an underlying thesaurus for the reconciliation of differing metadata structures. This would allow communities to control their information space at the local level, while integrating their specific constructs at the Agency level. This is an example of the top-down and bottom-up strategy described earlier.

Although we want to allow for the heterogeneous expression of data in ways that communities find intuitive, we also want to encourage the adoption of standard tags developed from an existing set, such as the Dublin Core metadata specification. A small set of commonly used core tags should be recommended to all Agency Webmasters for use on web materials.

In order to get the most out of the effort to tag web content consistently, first-level controlled vocabularies could be developed that include some simple synonym lists most commonly used by NASA personnel. Hence, mission names or Center names that are commonly abbreviated should be included in keyword descriptions.

In addition to supporting the foundation for content management systems, metadata facilitates keyword searching. Because site users utilize search extensively, we want to be sure that standards are known and implemented across the Agency by NASA Webmasters. This will take some time and resources dedicated to education and training. The NASA Webmasters Group can be instrumental in providing support in this area. Standards can also be reinforced by using the web management model being developed by the Web Management Services Team.

## G. INFORMATION ARCHITECTURE AND DATA ARCHITECTURE

A discussion about metadata would not be complete without a side bar on data architecture and how the two relate to each other. The underlying data infrastructure system is critical to how information is passed over the Web. Data must have mechanisms to move through cyberspace and to specified destinations. Part of the infrastructure is expressed by coded middleware wrappers using XML to tag the data with identification characteristics and affiliated uses.

In addition, new tools such as Web Services Description Language (WSDL) and Universal Description, Discovery, and Integration (UDDI) are becoming available to extend the power of data architecture. It will be necessary at some point in the future to undertake a study of NASA's existing data architecture model and how it can be extended to handle increased traffic flow.

Data architecture also addresses problems of interoperability between systems that have been developed separately. It provides ways for information to move across the Web between providers and consumers of needed data. The technology components of a mature data architecture compliment the information architecture and allow it to function at its peak.

## H. INFORMATION ARCHITECTURE AND THE USER INTERFACE

The art of labeling is an often overlooked component of information architecture and yet, since it defines the interface between the user and the content, it is one of the most important aspects of site organization. Labeling systems are expressed in site navigation mechanisms such as tabs and buttons, so care should be taken to see that they are thoughtfully developed, easily understood and consistently applied.

Once top-level Agency constructs are in place to give the web environment some definition, content publishers from different parts of the Agency can begin to see themselves as a part of a larger community. They can then better understand the benefit of employing language that is universally understood in their content and that its use maximizes their interface to NASA users through NASA portal data channels.

The user interface often impacts how visitors interact with a site. Information architects create wireframe models of web pages that map out functionality and navigational pathways. These wireframes are employed in testing to see if the material is presented in a framework that enables site comprehension by the user and promotes usability. This not to be confused with graphic presentation, but rather expresses a visualization of how content is organized and presented to the user. It is meant to diagram the site's underlying information flow and express the various ways that users traverse the information space.

## I. INFORMATION ARCHITECTURE AND KNOWLEDGE MANAGEMENT

Knowledge management echoes the concerns of information architecture in its desire to aggregate and order the intellectual assets of an organization. Both disciplines seek to promote greater efficiency and productivity through better management of content crucial to enterprise success. Knowledge Management seeks to encourage community interactions and knowledge discovery through the creation of common platforms that house enterprise information sharing systems. As users browse through structures of knowledge, they refine and extend what they want to know.

By defining a NASA institutional information architecture, knowledge management's goal is to determine the scope and landscape of the NASA web domain. In making the web environment an easily accessible resource, the team is enabling innovation and knowledge reuse as well as speeding the process of placing the user as close as possible to desired information.

## J. SUMMARY RECOMMENDATIONS

The steps to develop an effective information architecture for NASA web space can be summarized as the following:

1. Identification of audience types and needs – internal vs. external
2. Identification of top NASA goals for NASA web space
3. Inventory existing content
4. Content requirements analysis
5. Mapping of content to user needs and creation of information architecture blue prints
6. Development of topical taxonomies from blue prints
7. Validation of taxonomy alignment with Agency goals
8. Metadata development and iterative user testing
9. Descriptive labeling systems consistently applied across NASA web space
10. Processes established for ongoing identification of documents with metadata attributes
11. Testing for continuous improvement

The long-term methodology for developing a robust information architecture for NASA is mapped out above. It would involve a team to do research on audience types and perform a content inventory. This might be facilitated by support from the eNASA Web Services Team as well as the NASA Webmasters and take about two to three months. Content analysis and mapping could take longer depending on the scope of the effort and the resources available.

Taxonomy design evolves from content maps and usually requires expertise in classification methodology. It would be helpful to engage a professional information architect with library training from an association such as the American Society for Information Science and Technology (ASIS) to suggest groupings that reflect the best usability. The time frame for development of metadata tables depends on the scope of the project, but typically range from 3 to 6 months. The acceptability of associated metadata is usually more of political issue than a strictly IA issue. Getting buy in and agreement from all parties may take time. Implementation of tags will take more time as Webmasters will probably have to procure funding in order to do the necessary HTML production work. Financial support and support from upper management will determine how quickly this can be accomplished.

## K. GOING FORWARD: EVOLUTIONARY INFORMATION ARCHITECTURE

As NASA moves forward in its transition to a web-enabled organization, information architectures that provide scaleable frameworks for web assets are an indispensable aid to navigation by both the internal and external users. The Web Management Services Team is studying ways to better manage web publication processes. In the future, we will have to consider online web services and applications as well as content integration into our directories.

Internal and external directories will overlap, but the needs of the audience bases remain quite different. Internal groups may be centered on communities of practice that reflect technical disciplines, such as thermal engineering. Other groups may want to use the portal as a platform for collaborative partnering. Most NASA missions are spread across Centers and the world, and the portal can facilitate work for teams that are not co-located. In the future, it is possible that these groups may build their own directories that point to web assets that are of particular value to them.

The Web is a powerful communications medium, capable of carrying many types of information. In order to better structure NASA's information space and facilitate retrieval of relevant data, it is helpful to provide a foundation of information architecture. Information architecture acts as a framework for users and helps them develop a mental model of how online assets are organized (Figure 9). This, in turn, stimulates more successful interactions between users and the NASA web, improving the quality of work performed by NASA employees and effectively communicating the goals of NASA missions to the public.

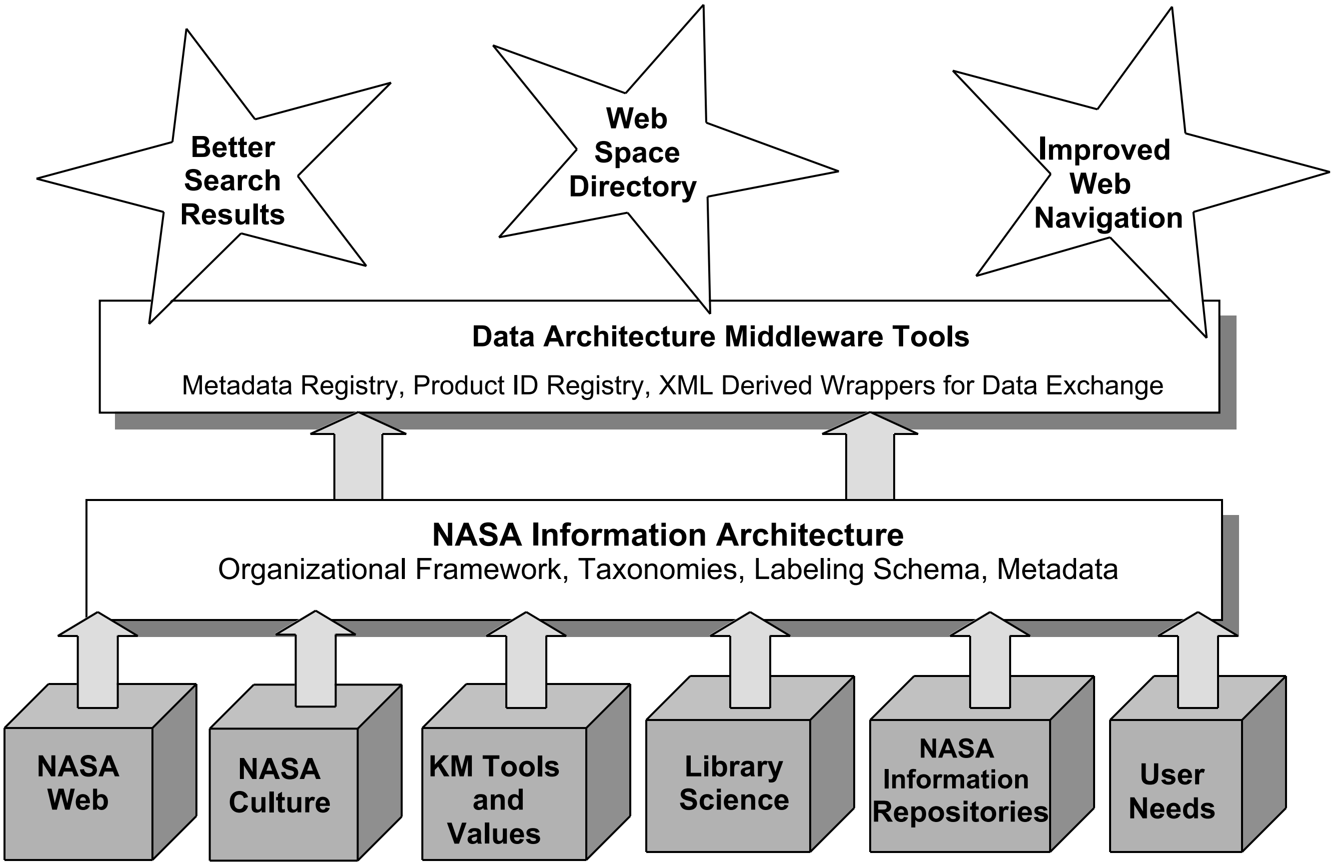


*Figure 9. Building blocks and information flow of NASA information architecture*