Policy Content at SSA: Using XML and Semantic Metadata

Terry Hynes, PolicyNet Program Manager, Social Security Administration

Duane Degler, Design Consultant, Lockheed Martin/IPGems

Abstract

A new content management and delivery system has been growing over the past years at SSA. It has arisen out of the needs of the SSA to respond to the knowledge management demands of the staff retirement wave, and the significant changes in front-line technology for SSA staff and directly for the public. From a content perspective, users have been asking for "simple answers, with all supporting information, relevant to my situation." From an organizational perspective, the role of content is increasingly seen as integrated with transactional systems in order to sustain quality service delivery in an increasingly complex business environment. From a technology perspective, the use of emerging tools that are based on XML and semantic technologies provides opportunities for simpler systems that control content maintenance more effectively, improve integration, provide easier content access, and allow migration as systems evolve over time. This case study shows the application and discusses design considerations.

1 Introduction

The Social Security Administration (SSA) in the United States is responsible for the retirement, survivors and disability benefit programs for covered workers. They also are responsible for the Supplemental Security Income (SSI) program and for administering portions of the Medicare program. SSA operates over 1,300 field offices, 140 offices for claims appeals, and supports an additional 100 state Disability Determination Service (DDS) offices. SSA provides retirement and disability benefits to about 50 million Americans. Nearly 80,000 staff and state employees use the Agency's policies in order to process benefit claims accurately. The range of Agency operations creates a need for widely diverse policy and instructional information.

What are the challenges that the Agency faces, for which there is a need to support staff with clear, relevant policy instructions and improve the technology used for content management?

- Anticipation of a dramatic rise in the number of people SSA will serve, due to the aging "baby boomer" population (more possible disabled individuals, followed by more retirees), as well as an increase in the public's expectations of the service to be provided
- Management of complex, specialized or infrequent subjects, such as Black Lung, Supplemental Security Income, and Medicare interviewing (which is done in SSA field offices)
- Upcoming changes in the staff profiles within the Agency, as a result of increasing retirements and other changes in the US economy and workforce
- Changes resulting from legislation
- Commitment throughout the US government for standardization and sharing of content and public services, generally known as the e-Gov initiative

1.1 The Content Management Challenge

There are a number of policy instructions that are geared toward specific operational areas, such as field offices, the state DDS offices, hearing offices, telephone service centers, and payment centers. In the past, one set of instructions (for field office and state DDS staff) required 38 binders. Printed documents required an annual printing and distribution budget of about $1.5 million. The move to CD-ROM in the mid-1990's made the process a little more efficient, but still required a production delay of up to 6 weeks, making it more challenging for the Agency to respond rapidly to policy changes. The publications are now all on the Intranet (and many are available on the Internet, as well).

As more of this content became available on the Intranet, people believed that search technologies would be sufficient to help users get to what they needed. However, search has not proven to be the panacea everyone thought it would be. The magical ideal has not become real – where the user thinks about the subject they are interested in, and one or more of the document binders floats off the bookshelf and opens directly to the page with the relevant information. Hundreds of "hits" are indicative of the closely inter-related nature of the content and the problem users face identifying the most successful search terms to use.

The challenge for SSA is partly one of scale, and partly one of usability. How can such a wide range of content be made more relevant and targeted to user needs? In the age of "Web content" (with all that implies for users about ease of reading, rapid access, close cross-linking, and interaction), how can these large legacy publications be made more accessible and usable for staff? How can the integration between publications (and other information repositories, such as training, system help, and local Intranet content) be facilitated in the most efficient way? And particularly, how can this large body of content be made flexible in the face of organizational change?

The recent introduction of the electronic disability claims handling process has prompted the need for new procedural information aligned with the new computer systems. The information needs to be responsive to particular locations – the new claims systems are being rolled out geographically over time, so each type of office and each state may be at a different point in implementation and use. The information also needs to recognize that there are a wide range of user needs – from very experienced people who only need short, precise instructions on the actions to take in the new systems, to less experienced people who would like to have the system policies and instructions cross-linked with other existing operational information and procedures.

1.2 The Need for New Content Management Approaches

Users have been asking for "simple answers, with all supporting information, relevant to my situation." From an organizational perspective, the role of content is increasingly seen as integrated with transactional systems in order to sustain quality service delivery in an increasingly complex business policy/process environment. From a technology perspective, the use of emerging tools that are based on XML and semantic technologies provides opportunities for simpler systems that control content maintenance more effectively, improve integration, provide easier content access, and allow migration as systems evolve over time.

The long-term goals are wide-ranging:

- Create a usable and accessible content resource for up to 80,000 internal users
- Use organizational ontology for navigation to bring the content closer to the organizational/user context and experience
- Move from a static document-based publishing model with over one hundred thousand items to a dynamic resource of simple, granular, reusable content
- Simplify content creation processes
- Gather organizational ontology from diverse resources, systems, and formats
- Establish technical formats that can migrate in the future
- Build a framework that encourages integration of policy, procedure, system help, training, and other useful resources
- Allow direct access and integration with transactional systems

This case study describes the main elements of the application and discusses the design considerations.

2 Metadata-Driven Content Management

The issues above drove the approach to eDG Online. We knew we had to move beyond the basic use of search, tables of contents, training and memorization. In 2004, we started laying the foundations for future content architecture and interfaces, with the first version now preparing for pilot. The focus is a highly usable experience for both end users and content providers, based on a detailed user-centered design approach that integrates accessibility throughout the application design process.

The new application relies on richer metadata to help identify user needs and provide more relevant information. At the heart of the system is a metadata engine based on semantic technologies, born out of more than two years of work using Topic Maps. The content itself is to become more granular (smaller, more inter-related pieces of content) in order to be shared and reused, reducing content duplication and the possibility that multiple versions of the same instruction may get "out of sync" as things change. The foundation is an XML content management database. The focus has been on end user delivery systems, supported by a suite of content administration tools that manage content elements, rich metadata, ontology development, and link management.

Basing the content management and metadata approach on standards reduces the risk of becoming trapped by a technology or format as things change technically within the Agency over the years. The use of metadata standards to model and manage ontology and content mapping will also help authors and policymakers within SSA to maintain a large – and growing – content library.

2.1 What is "Semantic Metadata"?

Semantic metadata uses a structured vocabulary – an ontology – to describe the meaning of the content in terms of the business context that the content relates to. So when a user is in a complex situation, such as helping a widowed mother process a set of disability applications for two disabled children in a particular state, content that is relevant to that particular situation can more easily be retrieved, even when all the pieces are stored in different locations.

It allows us to more effectively localize content to particular offices, states, and types of claims, so staff see content related to their specific state or local issues and policies. It also allows the computer application to "understand" a little more about the content and the business than would otherwise be the case. A search engine will not easily allow us to find "transfer claim" when searching for "jurisdiction" because the phrases are so different, and their meaning may be slightly different in different contexts. Our semantic metadata helps us relate those terms to each other in context, and to relevant content.

3 End User Access to Information

eDG Online has two main entry points from the home page (see **Figure 1**). "How Do I?" provides a selection of information geared toward quickly helping the user get "unstuck" when faced with task-related or system-related questions. The comprehensive "Help Me with My Situation" uses a faceted taxonomy to help the users describe the situation that prompts their question, in order to locate relevant information.

Figure 1: End User Home Page

During design, we asked ourselves (and our users) the question, "what are people likely to know when seeking information?" The answer is that they know about their situation, so it is important that they can phrase their information request in those terms. The faceted taxonomy is drawn from the ontology, and presents common descriptions of organizational situations under recognizable categories: your location, your tasks, computer systems used, claim issues being worked on, and types of information you want to see. Users often find it easier to recognize the common business phrases in the list, rather than having to remember or think about what search terms would appear in the content (which implies familiarity with the content and the terminology used by authors).

Users can select as many or few items that help them precisely describe their complex situation. For example, in a situation where a field office worker is transferring a claim to another state when the claimant moves, tasks include notifying various parties, changing the address in a system or systems, and possibly taking and filing new evidence for the claim (such as letters of referral from new treating physicians). There may be special additional tasks based on specific claim-related issues (such as supplemental income

requirements or critical medical needs). There are also considerations for the state DDS employee, such as whether the claim decision has been made yet or not, what processing stage has been reached, system-to-system transfer protocols needed, and whether this is the only claim the claimant has made. With eDG Online, users see lists of query criteria related to their particular circumstances, and select combinations relevant to their context.

The selections are processed to identify associated content, and the results list is presented to the user in further categories, in order to help them manage large results sets (rather than be confronted with a list with an unspecified ranking formula).

Figure 2: End User Results Page

The interaction to view categories is similar to that used by other transactional systems users find familiar. When results lists or content pages are presented, they always include appropriate warning and reminder messages on the right side of the page. For important information, it's important to present that to the user proactively rather than relying on the user selecting a link.

The reminders are associated with the content by matching the semantic metadata, so current reminders will be automatically presented with the relevant content rather than requiring manual associations on the part of content authors (which would significantly increase content maintenance).

Figure 3: End User Content Presentation Page

Along with the reminders, content makes use of a range of cross-referencing links, to be sure that additional content resources are easily available to users, reducing the need to return to the query results lists or having to search again and again.

External references and sites are presented in new windows, but include a clear signpost for what the site/information is and how to return to eDG Online.

Figure 4: End User Display of External Linked Information

It was important to plan an approach that, where possible, created a more seamless user experience across internal and external content – recognizing that the definition of external is likely to change over time, and may initially include other internal SSA content managed by different groups, regional content, state content, as well as links to content outside of SSA. Managing metadata for links to external information as if they are internal content allows a wider diversity of information to appear more consistently to the user, improving their experience. It also means we are holding metadata for a range of content, which reduces authoring rework if that content is incorporated directly into the application in the future.

4 Content Management and Authoring

The same emphasis on creating a usable application has been applied to the design of the authoring application, which manages the content and metadata.

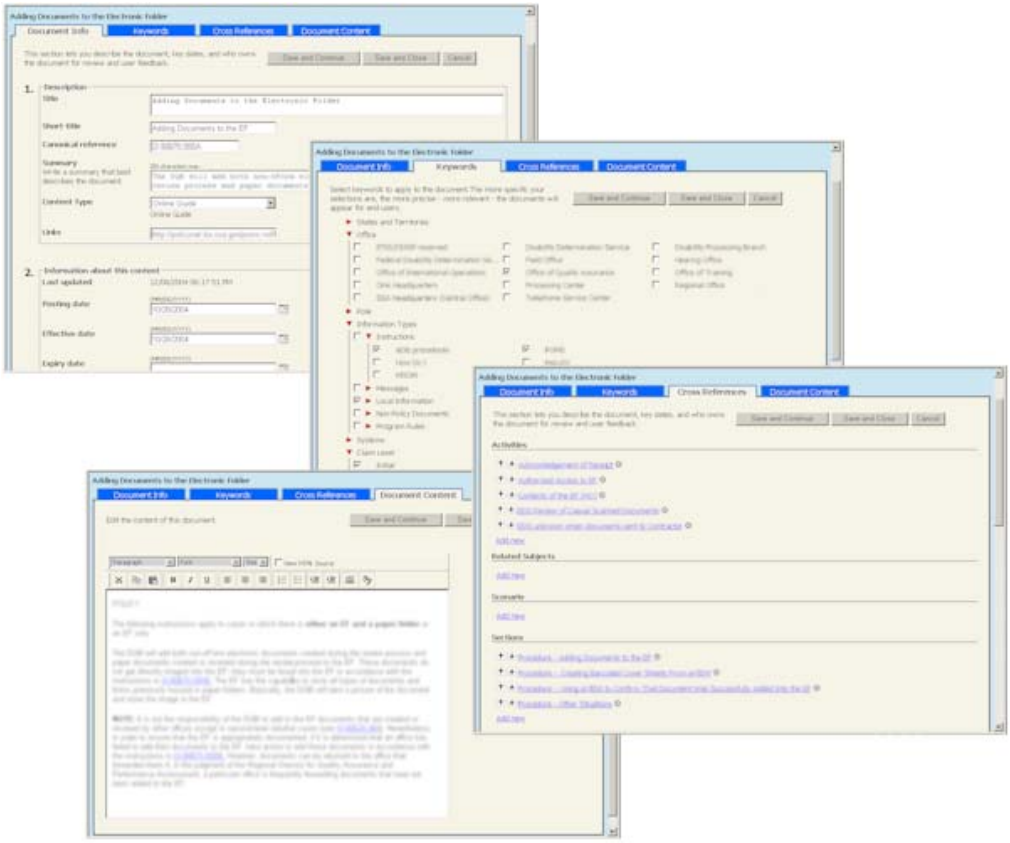

Figure 5: Content Authoring/Metadata Management Application

We've made a distinction between the data that describes the content itself (the "document info" properties) and the data that maps the content to the organizational situations it serves (the topics, presented to users as "keywords").

For the authors, answering the question "what is this content *about?*" can be done by simple selections from a keyword list (derived from the ontology), rather than typing in long, often conflicting, words and phrases. The interaction is very similar to the end user interface because authors are also users of the system, so it is important to increase familiarity and reduce the number of applications they need to learn.

It is also simple to identify cross-references between content items. Setting cross-references, assigning them to categories, and ordering lists of references are simple point-and-click actions.

Finally, draft content can be dropped in from Word into a simple browser-based WYSIWYG editor that stores HTML, allowing authors to preview content on the site and, in future, to share drafts with collaborators for comments and review.

5 Ontology Management

Specialized interfaces have been developed for common ontology maintenance tasks. These include:

- Viewing the overall set of items as a single full list (presented as the faceted taxonomy)
- Adding and changing items
- Creating, changing and describing relationships between items
- Sorting and arranging sets of items within the facets that are presented to end users
- Managing "templates" – sets of items that relate to each other in common ways

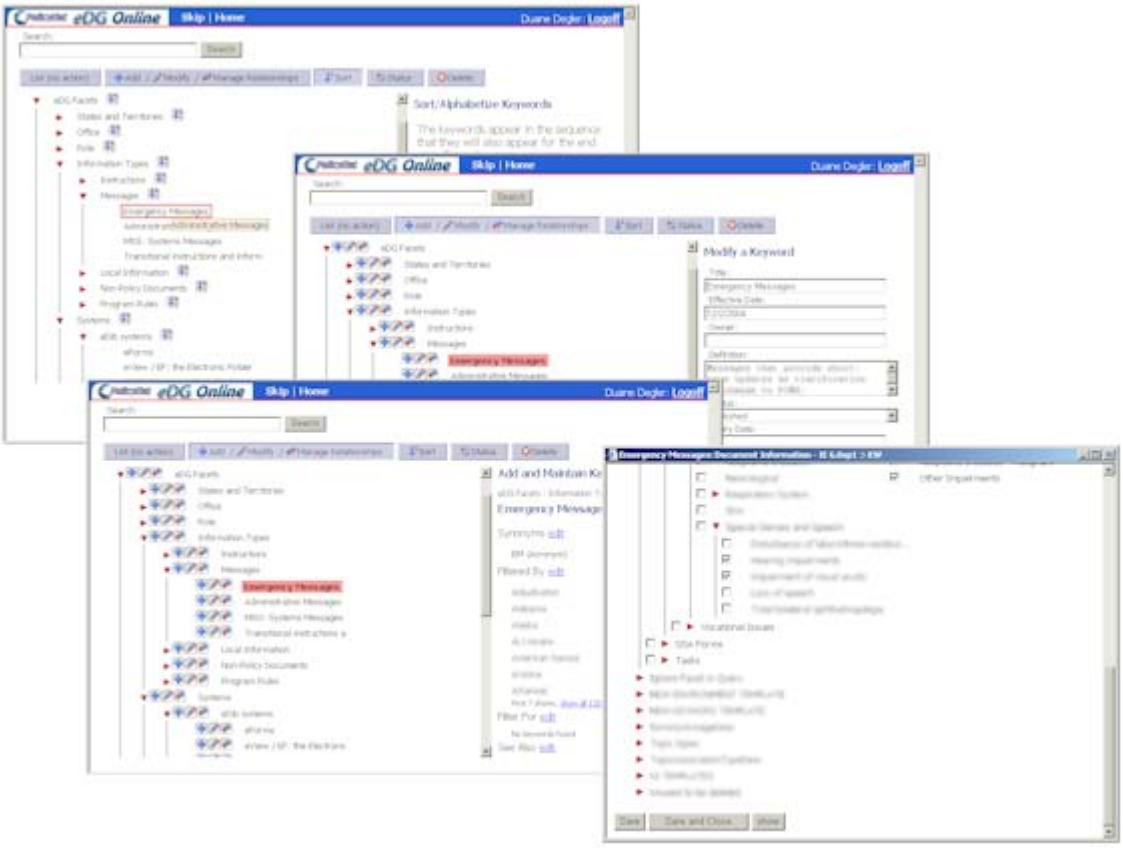

Figure 6: Ontology Management Application

The design approach allows for the easy introduction of changes to the way the Topic Map is managed, with minimal disruption to existing interactions. It also allows different items to be imported from other applications over time, so they can be managed in the same way.

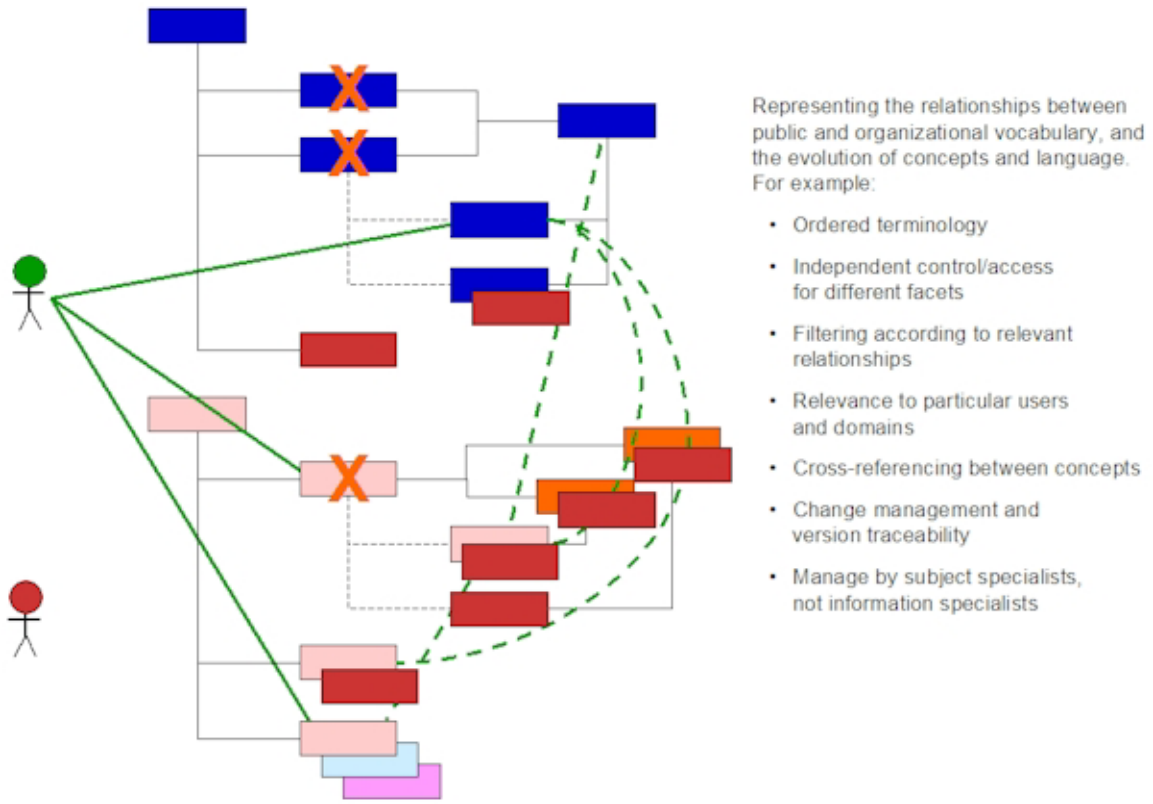

Figure 7: Requirements for Ontology Maintenance

Why go beyond simple taxonomy lists for categorizing policy information? Using the Topic Maps Reference Model during design and also maintaining alignment with standards for ontology representation allows us to build more flexible controls for managing changes over time. Examples of the situations that we expect when creating and managing the ontology include:

- Presenting ordered terminology to the end user.
- Allowing the user (and also software applications) independent control/access for different facets in the overall map (useful for end user selection of terms, but not in itself sufficient for effective querying of the content).
- Filtering between terms and facets, to be able to identify and present only relevant relationships to users, which focuses content queries more effectively. For example, certain SSA forms are only used for information capture relating to disabled children. Some instructions are directed by a Circuit Court so they only apply to particular states covered by the circuit. Filtering means that the user sees only relevant subjects and content.
- Extending filters and templates to user and role level, in order to keep the relevance to particular users, domains, and applications through programmatic management of preferences for both end users and authors. For example, the SSA's implementation of the electronic claim folder is being done over a series of months across different states. For a period of time, different states, offices and users will have different systems and be at different levels of experience. These

state and user "profiles" can be facilitated through the ontology and be used programmatically by the content management application.

- Cross-referencing between concepts. For example, the difference between "public" language and "Agency" language. Managing associations between terms and their acronyms. Also creating a synonym mapping between references, titles, and terms (such as "SSA-3820" which is titled "Disability Report – Child" and is referred to by many as the "child disability application").
- Mapping changes as they arise, to support ontology management and version traceability. For example, changes in medical criteria surrounding disability often result in changes to medical classifications. If operational departments reorganize, the responsibilities for content will change, split and merge, with new mappings and new department names that must be managed. For users, they want to be able to identify terms from the ontology and locate content using both old and new terms.
- Providing an interaction approach that allows the ontology to be managed by subject domain specialists, not information/classification specialists.

6 Functionality and Technology

There are a large number of "moving parts" that make up the content management functions of eDG Online.

6.1 Functional Overview

It has been important to create an architecture designed to manage smaller, more interconnected items of content and the linking between them.

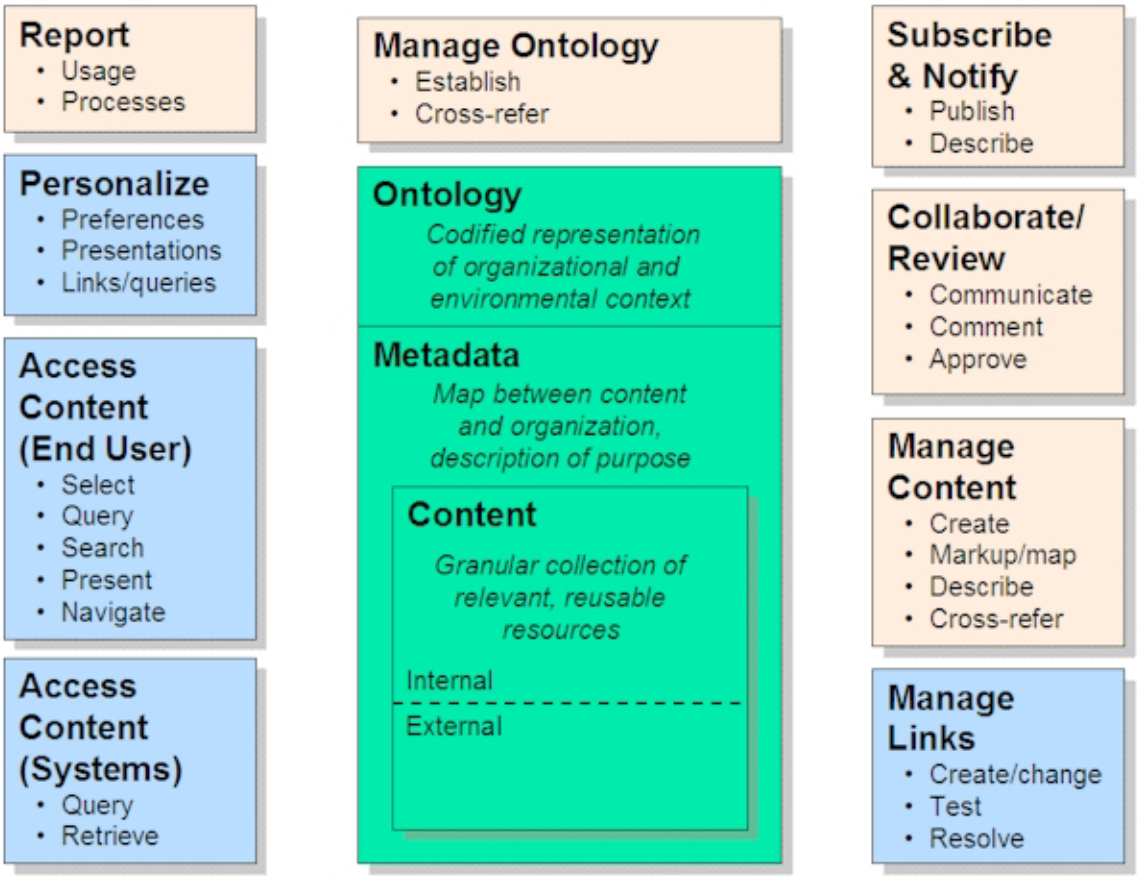

Figure 8: eDG Online Functional Overview

The use of metadata "wrappers" for both internal content and external links has helped achieve a more integrated experience, which was one goal of the design. The metadata provides the map between the user's situation and the appropriate content for their task. The ontology provides a common, controlled vocabulary.

Functions presented to the end user include the ability to access and view content, as well as to personalize the site and the types of content retrieved. There are also functions being designed for direct content requests from other SSA systems, particularly for transactional applications. Link management is a key functional piece to create a seamless, closely inter-related experience for the user.

Along with functions for the end user, other tools support content authoring and ontology management. SSA has about 600 authors, most of whom only do that a small percentage of their time, or they are brought in from the field for a short period of time to focus on writing something particular to their expertise. These people are supported by a larger number of people who collaborate, review and advise on particular policy issues. So if authors and collaborators are only working on something two or three times a year, they can't be burdened by complex technology or processes.

The same is true for people who manage the ontology. It is vital to the effectiveness of the end user content resource, but it doesn't change very often so has to be easy to pick up and use. The "keyword library" ontology editor is being designed to support the ongoing evolution and maintenance of the language of the organization and its staff.

There are additional facilities being mapped out to help the overall content management system be more proactive with the authoring community and SSA management. These facilities include notification services to alert on changes to content or the ontology, richer and easier collaboration, and reporting on content use and authoring processes.

6.2 Technology Overview

An application of this size and complexity has been developed over time using a number of different technologies and standards. This illustration shows some of the technologies and standards for the data (green), services (tan), and user applications (blue). The choices, as is common for any project, are based on best use of existing software tools, alignment with SSA's overall architecture, and greatest suitability for the SSA's purpose and environment.

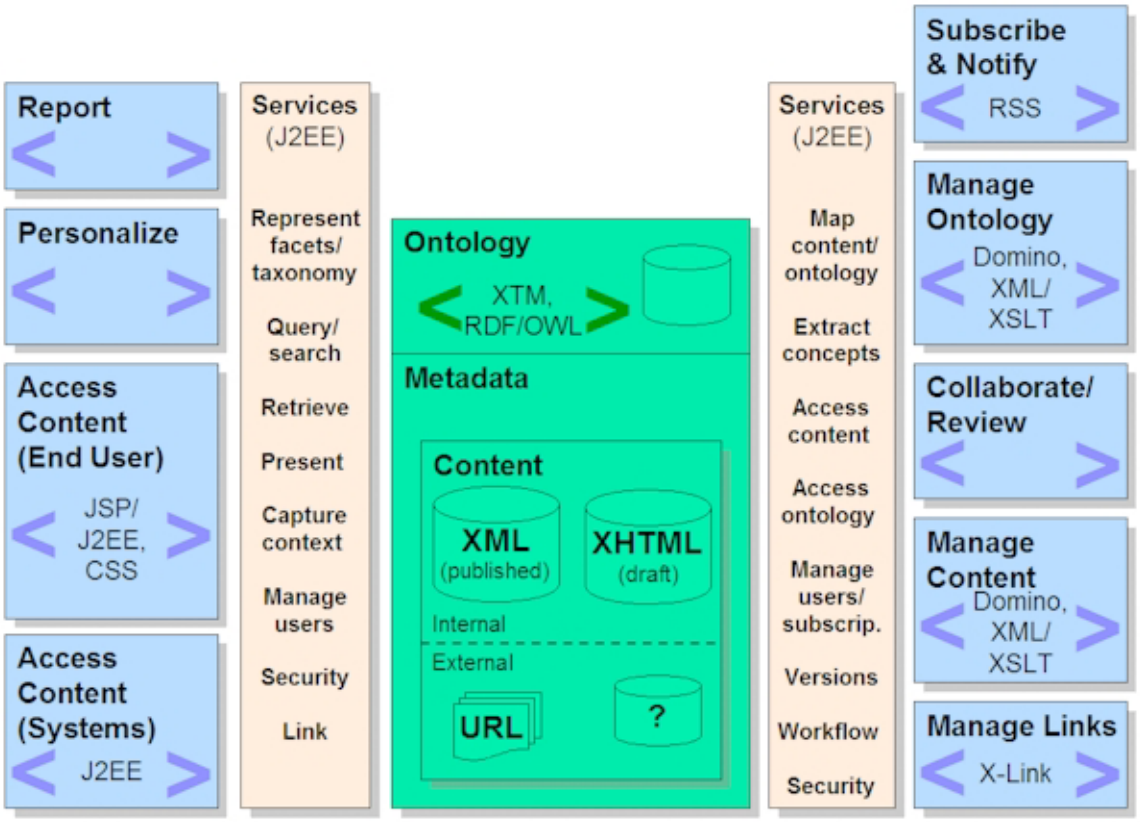

Figure 9: eDG Online Technical Overview

For the content data itself, currently we operate a mixed environment of XML and HTML content, although the plan is to use XML for the publishing repository and XHTML for draft content that can be shared through a review/collaboration viewer. For performance and compatibility in the current environment, a relational database is used for housing content metadata and for querying the end user faceted taxonomy – the underlying data has been modeled/designed using the Topic Maps Reference Model, and will be able to be shared with other systems using semantic ontology standards such as XTM (an ISO standard) or the W3C's RDF/OWL (the knowledge representation/ ontology elements of the Semantic Web).

We have been making effective use of J2EE and CSS at the end user level and XML/XSLT in content administration to modularize the interface – in this way we hope that not only the granular content will be reusable, but some "granular interface components" may also be reusable by other applications, taking advantage of the same services for content retrieval currently used within our applications. CSS also allows extensive use of dynamic positioning and formatting, which is valuable to support Section 508 accessibility compliance and some increased user control over look and feel.

There are functional areas in the application where further research is still being done to see what technologies best suit SSA's needs. There are also a number of areas where it is prudent to wait until emerging technologies mature further before making longer-term decisions on the direction – a situation that is familiar to many people working in this technology arena!

7 Initial Indicators of Success

While all this design and technology can be interesting, it only makes sense if people are able and willing to use it. At the time of this late-breaking paper, there has only been limited use of the application. However, early results from users are very encouraging. In a survey of 36 users from six different offices, the survey ratings positively showed that people felt the site would be useful and could increase the accuracy/relevance of the information they retrieved on the job.

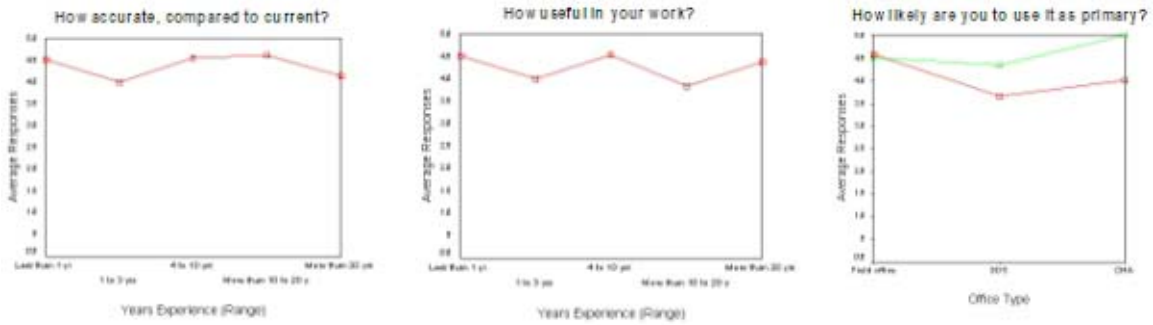

Figure 10: Initial Results from Usability Surveys

Two interesting points about the survey results:

- The comments were very consistent across all user experience levels (which to some extent correlates to ages, as well, so the site is not just effective for younger, more "computer savvy" people)
- When people got a chance to spend more time using the site, their enthusiasm for the site grew (rather than diminished after a "flashy demo")

Perhaps, to close, it's best to let the users speak for themselves:

> "*Finding information based on what you are trying to do is much faster than going through a bunch of non-specific, general information to find what relates to you.*"
>
> "*No comparison -- I have used PolicyNet for quite some time… eDG's choices… greatly sped up the process.*"