The Warrior Gateway and the Power of Social Publishing

Enriching Content for Web-wide Distribution

August 2010

by Dale Waldt and Geoffrey Bock

GILBANE GROUP
A DIVISION OF OUTSELL, INC.

Gilbane Group
A Division of Outsell, Inc.

763 Massachusetts Avenue
Cambridge, MA 02139 USA

Tel: 617.497.9443
Fax: 617.497.5256

info@gilbane.com
http://gilbane.com

The Warrior Gateway and the Power of Social Publishing

Supporting the Military Community

Across the country, many government agencies, military programs, and local organizations offer medical, mental health, employment counseling, and educational services to the military community – soldiers, veterans, and their families. Yet for returning warriors seeking to reintegrate into local communities, making sense of all these services and finding the right ones over the web is often hard. There are few digital resources that not only aggregate content about disparate services, but also make the information useful to community members.

The Warrior Gateway (http://warriorgateway.org) is designed to fill this void. Seeking to support the military community members reentering civilian life after deployments in Iraq and Afghanistan, the Warrior Gateway collects, organizes, enriches, and redistributes content about a wide range of health, welfare, and veteran-related services. It also serves as a resource for employers wanting to hire veterans and for organizations seeking to engage with the community. It appeals to a digitally savvy audience, comfortable with the web and social media tools.

More than simply an online catalog of service providers' listings, organized by topics and locations, the Warrior Gateway restructures the content that government, military, and local organizations produce, and enriches it by adding veteran-related categories (e.g., Veteran's Administration offices and services, mental health and physical therapists, etc.). Furthermore, the Warrior Gateway adds a social dimension by encouraging contributions from veterans and family members. These "voices of veterans" include short comments about the quality of listed services, ratings, online reviews, and moderated forums, together with capabilities for organizing and tagging content.

The Warrior Gateway thus augments government, military, and local organizational efforts, by adding insights and opinions from the community participants to published information. For veterans and their families, the appeal of the Warrior Gateway is curated content from multiple sources, organized into a consistent set of online resources, and supplemented with advice, commentary, and categories from communities of contributors.

Aggregating and Socializing Content

A Web Framework

From a technical perspective, the Warrior Gateway is a content aggregation and integration environment, based on a model-view-controller framework and hosted in the cloud, utilizing Amazon Web Services (EC2). Initially developed within two weeks by a four-person team, the Warrior Gateway is continuously enhanced through an agile

development process that quickly incorporates audience feedback to add new features and functions.

The Warrior Gateway currently collects content from over 30,000 service providers. It is expanding its aggregation efforts and, by the end of 2010, expects to encompass over 50,000 providers. Government, military, and local organizations publish their content in many different ways, with varying degrees of accuracy and currency. The Warrior Gateway seeks to capture and normalize the essential elements of these listings, and provide links to the remote resources.

Essentially, the Warrior Gateway is a social publishing platform that combines the curated content from authoritative sources with the user-generated content and categories provided by soldiers, veterans, and family members. Initially, the Warrior Gateway deployed network-crawlers to crawl remote service providers' web sites, extract relevant content and links, and aggregate the results within its own repository. Depending on the granularity of the content published on remote sites, the Warrior Gateway captures varying levels of unstructured and semi-structured information. With access to granular content and self-describing XML tags, the crawlers can automatically retrieve detailed information that providers publish to describe their services.

Once stored within the Warrior Gateway, the content is reorganized and reclassified to provide the veterans' perspective about areas of interest and importance. Soldiers, veterans, service providers, and others who want to contribute to the Warrior Gateway can add comments and ratings to refine the information on services listed in the directory. Volunteers working with Warrior Gateway can scrub the data and add new classifications when necessary. Service providers can claim their profile and improve their own data details.

With contributions from multiple stakeholders, this form of social publishing allows data to be enriched over time without requiring a large internal staff to add the extra information. It also recognizes the natural interest of veterans and service providers to make the information as accurate and complete as needed to meet their needs.

The Smart Content Infrastructure

The Warrior Gateway relies on the MarkLogic Server to deliver the technical capabilities for its content infrastructure. Specifically, the MarkLogic Server provides a flexible database that can easily manage large volumes of disparate content sources, including structured content (e.g., HTML, XML, etc.) as well as common unstructured formats (e.g., PDF, word processing documents, spread sheets, etc.). The MarkLogic Server powers the agile development processes. It provides search and analytics functionality to easily navigate through large content collections that are distributed over the web. Systematic indexing of these disparate sources together with XQuery-based tools provide end users with easy access to the content, organized and accessed in consistent and useful ways.

The MarkLogic Server also supports geospatial index information, enabling content to be classified and delivered according to location. Geospatial information drives

innovative delivery options, such as displaying services geographically on maps, to help veterans locate services and enable agencies to plan coverage.

Finally, the MarkLogic Server is used to select and transform the content stored within the repository into specific views and formats, which are then syndicated to external web-based resources. The Warrior Gateway, with its sophisticated content management, indexing, and transformation capabilities, can serve as a master resource for information about veteran-related services, and can easily redistribute content to third-party environments.

As a result, service providers only need to update their information once within the Warrior Gateway, and can then syndicate it to additional veteran-related sites with no extra effort. Various organizations can rely on the Warrior Gateway to store and provide access to their own content, enriched with categories, comments, and ratings that the community members generate. By redistributing directory ratings and other information from the Warrior Gateway to their own networks of stakeholders, service providers can thus leverage the enriched content without duplicating development efforts.

Improving Consistency and Accuracy

By design, the content is managed in the Warrior Gateway's central repository, rather than simply linked from remote sites, as shown in Figure 1. This improves consistency and accuracy, but also maintains references and links to the originating sites. This approach also presents challenging logistics for capturing and enhancing the collected information.

Warrior Gateway initially used automated load processes and content crawlers to automatically aggregate and enrich content. Now the information can be uploaded by service providers or entered directly using dialogs or software provided by the service. Content can also be syndicated from other sites and search queries that feed information to the Warrior Gateway.

The Warrior Gateway provides a social publishing platform that captures contributions from throughout the military community. Much of the work of providing, organizing, and enriching the information is done at the grassroots level by veterans and service providers, rather than by the internal staff. The user dialogs allows *ad hoc* updating and classification of specific directory entries over time as the information becomes available.
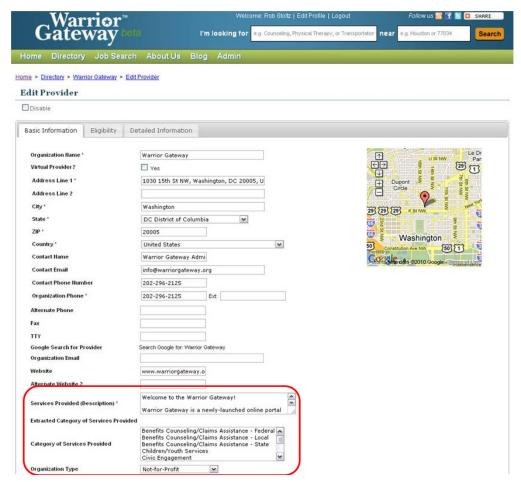
Figure 1: Social publishing web dialog allows service providers to update and enrich their listings, to improve searching and reuse of content.

The Consequences of Crowd Sourcing

Aggregating and organizing content in a consistent way and adding a social dimension to the information makes it easy for users to find what they want and need. Augmenting the content with metadata using easy-to-use tools enables targeted searches by topic or location, as well as innovative mashups with mapping and other third-party resources. Adding user comments with social media tools helps retiring warriors evaluate the services described and make informed decisions on the services they seek.

The end results are more useful and valuable information resources than would otherwise be available. There are several additional benefits enabled through the use of crowd sourcing.

First, there is the power of network connections. Only a small team is needed to support the core platform and provide the framework that facilitates content aggregation and social publishing. Service providers join the effort and make their information available to veterans and their families. By providing the tools and services to enrich content contributed by others, the Warrior Gateway manages web-wide resources for spreading the word and amplifying the contributions of local organizations.

Second, crowd sourcing removes many of the obstacles to enriching content. For instance, government agencies may produce the raw information, while third parties add assessments and evaluations of the listings. The end results are more informative online contributions than the information produced by any single source.

Third, crowd sourcing enables the content to be continually enhanced. Increasingly, web users have come to expect content to be refreshed frequently, to be integrated in a variety of ways, and to be interactive with commenting and other features. A large audience of contributors enriching the content helps to keep it timely and relevant.

Finally, once content is collected and enhanced by multiple contributors, new uses for the information and new services for the user are possible. Warrior Gateway can help identify current gaps in services based on the needs of the military community located in a particular geographic area. The site can help reveal areas that are underserviced or in need of a particular service or specialty, as well as those areas that may be over served based upon the community needs and population size.

For example, using metadata that identifies all the services from a single region, the Warrior Gateway can produce a map showing where all the services are located in a veteran's area. Or, the site can list certain types of services being sought, such as physical therapists, psychologists, or even day care and lawn mowing. The result is a rich, flexible, current, and accurate resource for soldiers, veterans, their families, and other stakeholders within the military community.

Smart Content Insights

The Warrior Gateway represents an exciting and innovative approach to collecting large amounts of disparate information, in a scalable and feasible fashion and within a limited budget. The Warrior Gateway also illustrates how social publishing enables content contributors and various service providers participate in enriching and classifying content to improve its relevance for soldiers, veterans, and their families. Lastly, the enriched content can be easily repurposed (or multi-purposed) and/or queried on other sites and even combined with additional content and applications in mashups, therefore making Warrior Gateway content even more valuable and accessible to the military community it serves.

Content Delivery: Service providers can have their information delivered in a highly targeted manner, by geospatial coordinates, service types, or through other types of enrichment techniques. Content that includes rich metadata can be sorted, organized, and searched more intelligently, repurposed and combined in ways that unstructured or more generic content cannot. At the Warrior Gateway, searching becomes much more accurate and likely to return relevant results. Powerful query tools and rich smart content enable flexible delivery and syndication features.

Content Enrichment: The Warrior Gateway contains not only listings of services, but also descriptive metadata that is used to provide powerful and accurate delivery to the returning warriors. Content can be enriched using a combination of automated and user generated methods including social media tools without requiring an internal team to add to and classify content.

Content Creation: Tools such as web forms and dialogs ease the collection and enhancement of content to create a robust data source. Automated collection and syndication tools ease timely and high-volume data gathering. A robust repository that can manage large volumes of heterogeneous information provides the platform for aggregating and integrating smart content.