

Healthwise uses TopBraid Enterprise Data Governance–Vocabulary Management (EDG–VM) to create a unified knowledge graph across domains

Unifying Data Silos to Integrate Knowledge Delivery

“We knew that integrating our metadata into one authoring system would bring great benefits to our knowledge engineering team, but we didn’t expect to see such big gains in simplifying our content manufacturing applications, too. If it weren’t for TopQuadrant’s broad utilization of open semantic standards, this work would have never been possible.”

– Adam Kimball, *Director of Engineering, Healthwise Inc.*

PRODUCTS USED

TopBraid Enterprise Data Governance–Vocabulary Management (EDG–VM)

– a flexible, web-based solution that enables management of business vocabularies as semantic information models. TopBraid EDG–VM was designed to meet the needs of organizations that are doing more sophisticated vocabulary management to support enhanced content delivery, search, navigation, aligning the meaning of data across data sources, database integration, and other applications that leverage metadata to get more value out of the enterprise’s data.

TopBraid Explorer is a web-based, browse-only solution, for exploring vocabularies managed with TopBraid EDG–VM. TopBraid EDG–VM users can publish vocabularies for use in this browsing environment. TopBraid Explorer users can view published content and can create comments for the editorial staff. TopBraid Explorer is installed on a separate server.

Both of these products are also available within the Vocabulary Management package of TopQuadrant’s most comprehensive product: **TopBraid Enterprise Data Governance.**

BENEFITS

For Healthwise, the benefits of using the TopBraid EDG–VM platform include:

- Providing a consistent authoring environment across all metadata
- Allowing downstream systems to access metadata through a unified API
- Enabling a repeatable way to release and version enterprise metadata
- Tracking fine-grained changes to taxonomies across many users and roles
- Providing knowledge engineers with the tools to do their own data imports/exports, custom reporting, and UI customization

“The new solution has a simplified architecture where everything is in RDF and centrally uses TopBraid EDG-VM.”

Challenges

Healthwise users fall into two basic groups. The first is end-users, who use navigation and search tools to find and select content for themselves or a loved one. They are often working within a portal managed by a care provider or insurer. These users are exploring a knowledge base and seeking to answer questions that matter to them. The second group of users are healthcare professionals who are finding information on behalf of an end-user. This class of users typically knows what they want and can often describe it using formal taxonomies such as ICD10, or SNOMED CT. They are unlikely to use consumer terminology.

Meeting the needs of these two different user groups often requires very different metadata. And with that metadata came the software to support, manage, version, and publish it. Over time, roughly a dozen independent software systems were created to handle these needs. This not only fragmented the data, but also pushed a great deal of complexity into underlying systems that needed to query, transform, and render asset metadata for product needs. With no one single source of truth, each application had to cobble together its own version of the truth.

Solution

In response to these challenges, Healthwise built the Healthwise Consumer Vocabulary (HWCV), effectively eliminating the gap between consumer and clinical metadata.

Key points:

- Nearly 4,000 concepts (derived from existing metadata structures)
- Directly leverages the National Library of Medicine’s UMLS Metathesaurus project
- Relates SNOMED CT, CPT, ICD9, ICD10, RxNorm, and Loinc to Healthwise concepts
- Provides clinical and consumer labels
- Formalizes relationships between concepts
- Relates content assets to concepts via relevancy predicates

Surprisingly little new information was required to create the HWCV. Existing sources offered implicit relationships to other datasets and were harvested as needed. Healthwise was able to write custom data cleansing/migration scripts because TopBraid EDG-VM utilizes standards that cross platforms and technologies. In the end, the most significant work was done in Java and heavily utilized the Apache Jena framework.

Results

The first and most obvious benefit of this work is the creation of the HWCV itself. Where a number of separate taxonomies existed before, now there is one knowledge graph that crosses many different domains and mental models.

This consolidation provides benefits to everything from analytics to core domain knowledge management. Furthermore, Healthwise offers the entire taxonomy to clients via a lightweight JSON Taxonomy API. Users can walk the knowledge graph, finding assets via related concepts. It is a core component of the Healthwise product offering.

A less obvious benefit of TopBraid EDG-VM was that it simplified the integration tier from a software perspective. EDG-VM’s functionality replaced over a half-dozen internal tools that required maintenance and updates. Written in a variety of languages, these tools had complex integration requirements and utilized a metadata model that had become obsolete as metadata requirements continued to move forward. Legacy XML web services were migrated to consume JSON from EDG-VM. In some cases, older services were replaced by modern microservices written in Node.js.

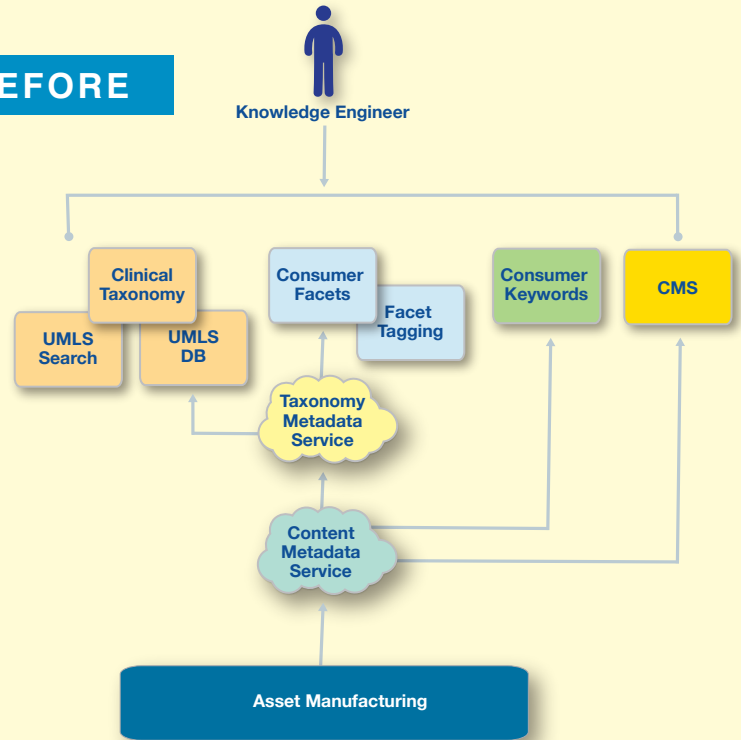
Healthwise's Knowledge Metadata Delivery Architecture – Prior to Deploying TopBraid EDG-VM

Prior to transitioning to TopBraid EDG-VM, Healthwise had different taxonomies* for different users, designed and developed at different times with different levels of functionality.

Each taxonomy had its own associated internal software system, persistence layer, and user interfaces. That meant a specialist tagging assets had to touch multiple systems to describe what a content asset was about. They would often use as many as four different apps (via the web) in the course of doing this. The data was then deeply siloed in each application since the knowledge engineering team lacked the skills to work in relational databases.

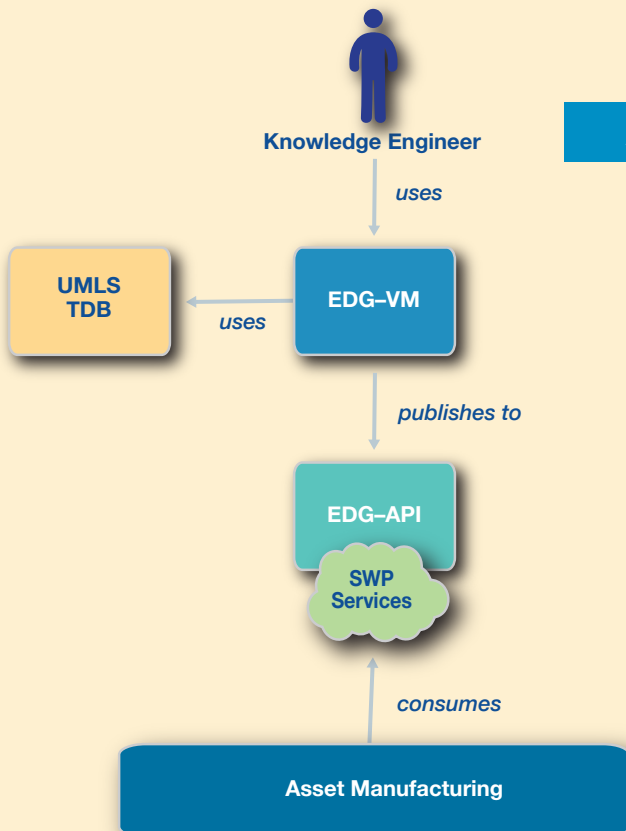
**Healthwise provides comprehensive metadata for many of its content assets. Over time, Healthwise found itself managing many separate taxonomies that had considerable overlap. Clinical metadata lived in one system, while consumer metadata lived in another. Lexicon codes gleaned from the National Library of Medicine's UMLS Metathesaurus project were siloed in systems managed by software development teams. Though Healthwise's Knowledge Engineering team could see how these taxonomies could be integrated into an ontology, the walls between the systems made the work nearly impossible.*

BEFORE



UNIFYING DATA SILOS TO INTEGRATE KNOWLEDGE DELIVERY

AFTER



Healthwise's Streamlined Knowledge Assets Management Architecture – using TopBraid EDG-VM

The new solution has a simplified architecture where everything is in RDF and centrally uses TopBraid EDG-VM. The diagram reflects the in-progress state of the architecture and will evolve with the needs of the organization, but overall it suggests the broad changes that EDG-VM helped them make.

Improvements include:

- The UMLS database is converted to a RDF data repository standard format (TDB). It is then imported and managed in a read-only mode by TopBraid EDG-VM. Knowledge engineers utilize a variety of custom properties to link lexicon codes to Healthwise concepts.
- The clinical and consumer taxonomies have been integrated into an ontology with custom properties providing relationships between concepts
- All content assets are represented and include a mixture of read-only reference data and managed metadata. Assets are related to concepts via relevancy and semantic properties.

UNIFYING DATA SILOS TO INTEGRATE KNOWLEDGE DELIVERY

Healthwise achieved a greatly simplified metadata delivery architecture after data consolidation.

Highlights Include:

- Provides lightweight JSON web services to a variety of consuming systems
- Utilized TopBraid technologies, SPARQL Web Pages (SWP) and SPARQL Inferencing Notation (SPIN) to deliver metadata to downstream systems in schemas they expect, allowing Healthwise to retrofit existing technology rather than rewrite it
- Separate production environments for human users and service consumers
- Push-based updates from TopBraid EDG-VM to dependent systems eliminates usual latency issues with job-based pull designs

Background on Healthwise

Healthwise has been delivering on its mission to “help people make better health decisions” for over 40 years.

In that time, they have touched nearly 2 billion people and delivered unbiased, evidence-based information through a variety of media. Healthwise content is available in over 20 languages and deeply integrated into the systems at the heart of the medical workflow.

Today, Healthwise is rebuilding its content library to support the next-generation of content delivery applications.

Healthwise believes that getting the right information to the right person at the right time is best solved by offering well-structured and consistent content backed by a comprehensive, consumer-based health ontology.



TopQuadrant helps organizations succeed in data governance. Its flagship product, TopBraid EDG, delivers easy and meaningful access for all data stakeholders to enterprise metadata, business terms, reference data, data and application catalogs, data lineage, requirements, policies, and processes. TopBraid Enterprise Data Governance–Vocabulary Management (EDG-VM) supports business stakeholders who need to collaborate on defining and linking enterprise taxonomies, ontologies and thesauri. They use these to improve search, enhance content navigation and align the meaning of data across data sources.

Visit www.topquadrant.com, or contact us at edg-info@topquadrant.com or by phone at +1 919 300 7945.

Ask us about scheduling a demo to explore how TopBraid EDG-VM meets your specific requirements.

TopBraid Products



**TopBraid
Enterprise
Data Governance™**

TopBraid Enterprise Data Governance™ is an agile data governance solution for today's dynamic enterprises. A modular, extendable, standards-based solution, it provides integrated data governance across the ever-growing types of assets and governance needs.



**Vocabulary
Management**

TopBraid EDG-Vocabulary Management™ supports collaborative management and use of enterprise taxonomies and ontologies to provide enhanced search and content intelligence.