

Whitepaper

NewVantage Point: Unleashing Unstructured Data's Value

Mobilizing Us, Not Just Our Machines

NewVantage Point! is published by NewVantage Partners, a boutique management consulting firm which serves as trusted advisors and senior consultants to Fortune 1000 clients. NewVantage provides independent, expert, perspectives through executives and industry thought-leaders who have “sat in the chair”. Our goal is to help organizations mitigate the risk of large-scale business and technology transformation initiatives by providing a critical link between business and technology strategy and capabilities. NewVantage Point! is published on a periodic basis to share our perspectives with a broader executive audience.

Questions and comments can be directed to Randy Bean, Managing Partner, at rbean@newvantage.com.

NewVantage Point Summary

Unstructured data are everywhere. During 2010 the internet saw 107 trillion emails, 152 million blogs, and 25 billion tweets¹. The lion's share of our verbal communications and broadcast media has been moving to our websites, intranets, social media and emails. As a result, the productive economy is moving faster than our analytical models can keep up. Being proactive with our messaging and getting the right information into the right decision-makers and buyers isn't just a matter of buying tools. Success hinges on our ability to mobilize our attention, processes, policies, and incentives into a metadata program – a robust information architecture that is the basis for making data into meaning.

Too much volume and not enough meaning

In our daily web experiences – on the internet and intranet – there is too much volume, and not enough meaning. We feel it as individuals, communicators, marketers, and customer service planners. Organizations and users are expecting to communicate in a targeted way, find knowledge easily, and be informed just in time. However, sociologists claim, many are so overwhelmed with that volume that they are less curious and less able to use any unstructured data effectively for decision-making. For organizations to deliver quality service and to know what's happening with their brand, they need a way to access and manage unstructured data.

¹ RoyalPingdom.com <http://royal.pingdom.com/2011/01/12/internet-2010-in-numbers/>

The promise of unstructured data

One way of defining unstructured data is “the elusive meaning inside the structure.” Many technologists will tell you that, in the final analysis, everything is structured because it all lives within digital systems – in the tweets, the emails, the attachments, the file folders, the podcast, even the humble notes you took on your iPad at this morning's meeting.

But the reason we talk about unstructured data is because there is a grey area between the tidy columns and rows where structured data reside (e.g., in a database of web pages or documents), and that sea of bytes. Words, sounds, colors, images, styles and fonts – all these could tell us important information about our work, our customers, our competitors, and environment. That is, if only we could put that information to work.

What business challenges could we solve if we better managed unstructured data?

- We could learn and anticipate, if we could use analytics to see patterns in the unstructured data.
- We could entice people to buy, if we could recognize their preferences.
- We could inform our employees and customers, if we could “hear” their questions.
- We could influence decisions if we could see the behavior patterns of our customers, competitors, and governments.

What's the state of meaning-making today?

Knowledge Managers and intranet champions have been grappling with these thorny problems for fifteen years. Communications executives, brand owners and customer service managers have also been grappling with these problems recently as digital channels have become more relevant and secure. The prognosis for both groups has improved: More and more content is in html and readily indexable, minable, and presentable. And, tools to quickly tag content for easy retrieval, or to index content against an agreed-upon thesaurus (more on this later) are maturing. Services that produce statistics on the behavior or sentiment of our target communities are available for anyone with the budget.

The challenge is more about personalities and processes than it is about technology. Introducing a management capability for unstructured data requires consensus, discipline, and, often, a shift in mindset. This capability – sometimes called “knowledge management,” “content management,” or “digital asset management” – is frequently defined more based on tradition and politics than on the needs of decision-makers and employees. In many organizations the function straddles IT and the business, and suffers from lack of awareness, resources, and leadership.

Where can we start?

Following are some best known methods for unstructured data management that we've drawn from experiences with intranet, internet, social media and ecommerce:

1. Identify business priorities and plan to invest

incrementally: Agree to what you want to accomplish. Focus on specific employees, customers, partners, or target content, and paint a vivid picture of what you want them to do, learn, decide, or buy:

- Find content by searching, navigating, or subscribing
- Reuse knowledge (even discussion threads and blogs) to inform and streamline work
- Alert decision-makers to patterns
- Anticipate or respond intelligently to employee or customer interests
- Capture and store content for easy retrieval

A “metadata strategy” involves a formal process of defining where metadata can provide this value, and choosing the processes, tools, and resources necessary to get there sustainably. A metadata program is a collaboration between business and IT. It matches the program investment with your business planning, your content, your talent (including your customers' and partners' talent), and your company's desired position in the marketplace.

For example, customer complaints mounted as a large credit card company repeatedly corrected credit contract terms. The over 50 applications that were used to set up and manage their credit and merchandising products had inconsistent unstructured product data. Errors and revisions to these data

were costly, as was the damage control for negative customer experiences. By framing and addressing the problem as metadata “silos,” and establishing standards, they rapidly eliminated inconsistencies and accelerated time to market for new products.

2. Design for change but practice standards. Here's the paradox: Metadata strategy calls for standardizing, while providing enough flexibility to adapt to the inevitable changes in business. Unstructured data is about meaning, after all, and how we express meaning changes every day.

In the call-out boxes are the main metadata “standardization” ingredients.

Putting in these standards begins with gap identification. Many de-facto taxonomies exist throughout your internet or intranet, and mind-bending differences may exist between your unstructured and structured data. Even web page names and folksonomies (free-text tags) contribute to inconsistencies, and can reduce the effectiveness of employees' and customers' navigation and search, text analytics, and knowledge-reuse.

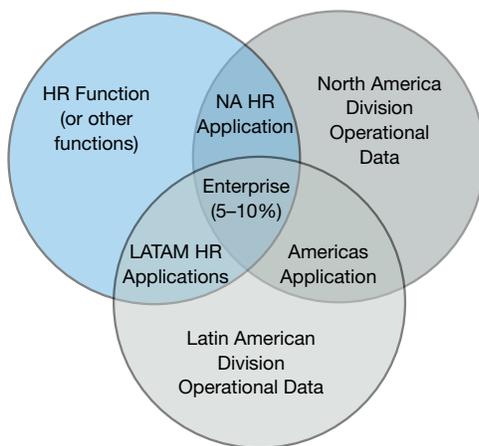
The “Basics”:

- A taxonomy is a set of hierarchical or conditionally related metadata. For example, in a financial taxonomy an analyst could drill down into Cost Centers for each Division. It imposes some rigor on what's permissible. Taxonomy is most valuable when it includes not just terms and parent-child relationships, but narrative about where the vocabulary comes from and which systems or groups are impacted.
- A facet or dimension is a field within that taxonomy. That is, something that can have a commonly recognized value. Division and Cost Center are facets.
- A controlled vocabulary is a finite set of terms which a facet or dimension could take. A good example is the drop-down list we see when we fill out a form.
- A tag is the actual value a facet or dimension holds for a particular information asset.
- Metadata are “data about data.” The term is so generic that it's better to use these more precise terms.

3. Have a federated operating model. Shared-metadata thinking is very often a shift in mindset for content creators, brand managers, knowledge workers, and even business intelligencia.

Setting up and maintaining these require time and attention from some of your most busy subject matter experts. Getting them to agree how to describe a fast-moving technical discipline is difficult even without busy schedules.

A federated governance model is preferable to a centralized model because it takes advantage of unique vocabularies and subject-matter expertise around the organization, even while centralizing such activities as publishing thesauri, indexing schedules, maintaining tools, and reporting analytics. Enterprise-level shared facets are often few in number, such as divisions, cost centers, country codes, and employee IDs. Many divisional or functional facets may be locally managed as seen in the Venn diagram below.



As little as 5–10% of metadata are common across operational and application data

The metadata program needs great facilitation and communication, and work should be transparent and graphical. With a sufficiently large metadata model you may need an auto-indexing and taxonomy management tool from Concept Searching, Data Harmony, SmartLogic, Synaptica, or Teragram, or a master data or thesaurus management tool from Oracle or SAP.

4. Keep it simple: Strive for simplicity for all people inside and outside your organization producing or working with unstructured data:

- Use automated indexing wherever possible.
- Use text analytics to down-scope the metadata program efforts. For example, watch for themes that internal and external users are raising.
- Use type-aheads from your thesauri for folksonomies on blogs or other social media.
- Let your users select their metadata for alerts.
- Keep the thesauri current and borrow from “industry” thesauri where it is more cost effective (and perhaps less politically-charged) than creating your own.

5. Keep the metadata program alive and vital. No metadata strategy can stand still. It must evolve with changes in your customers, your products, and your employees’ interests. Staying relevant requires standard, visible processes and ongoing discovery and analytics. Even more important are the regular, meaningful conversations about where your business climate is changing, and about whether your current thesaurus does or doesn’t effectively index new content. Learn from your users about how meaning is evolving, through their content, preferences, and click-paths. Let them inform your ongoing investments in tagging, maintaining metadata, indexing and measurement.

The ones to watch

Unstructured data management, or metadata management, is as much business operation as taxonomy wizardry. It’s about getting subject matter experts to contribute their unique knowledge, while continuously advocating for shared language, and shared improvements in revenues and productivity. Follow the lead of successful web and intranet organizations like Amazon (internet commerce and social selling), Pfizer (Intranet), and Deloitte (Social Media) who have channeled unstructured data into competitive products, loyal customers, and operational efficiency. Through better meaning-making from unstructured data, they’ve increased employee engagement, customer retention, and profit.

More Advanced Concepts:

- An ontology is the set of logical relationships among terms and among hierarchies. Those relationships can enable you (or the system) to make inferences.
- Indexing involves automatically crawling the words in a document or web page (nouns, phrases, surnames, dates, context, etc.) and creating a “profile.” Indexing engines may also show term-frequency statistics within or across documents or web pages.
- A thesaurus is a point-in-time list of terms and their importance. Thesauri hold the controlled vocabularies for each facet or dimension. Thesauri include synonyms for terms, and weights for term-values so that when we index a document, the resulting profile can tell us which terms appear, which are most relevant, and what it’s “about.”

About NewVantage Partners

NewVantage Partners is a boutique management consulting practice established in 2001 and comprised of former C-Level business and technology executives, and senior subject experts.

Our work comprises up-front planning – current state assessment, future state design, business case, execution roadmap, as well as the development of business and technical requirements, business capabilities, and business architecture. We are frequently engaged to provide a critical link between the business and technology organizations of our clients.

NewVantage fosters a commitment to executive thought-leadership through a series of small group executive dinners, and through our executive advisory board comprised of current and former Fortune 1000 business and technology executives and well-known industry thought leaders.

Contact NewVantage Partners LLC: Info@newvantage.com

Boston | San Francisco | New York
Tel: 857-991-1404 | info@newvantage.com | www.newvantage.com

NVP
NewVantage Partners