Using Scenario Analysis to Predict the Future of the Semantic Web

August Jackson
Ernst & Young

Think about the complex relationships information professionals have with search engines. They give our stakeholders an ability to self serve many of their information requirements. Regardless of what we may think of their value, many of our stakeholders think search engines are "good enough" for many of their information needs. And they're right.

"Good enough" for your end user is getting a lot better. We'll be expected to deliver higher value add to justify our salaries.

Today we'll use some of the very methods that can help us deliver higher value to explore the transformation the semantic web will entail.

I work for Ernst & Young

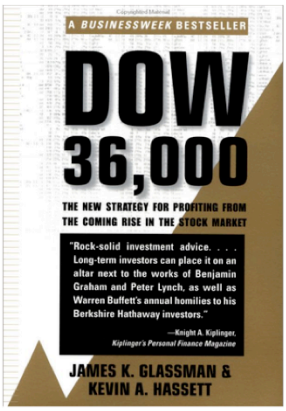
One of the largest audit, tax and advisory firms in the world

This material does not constitute advice related to any of the services the firm provides

Mentions of firm clients do not constitute an endorsement or recommendation to invest

<u>All opinions expressed here are solely my own</u>

"It's tough to make predictions, especially about the future."



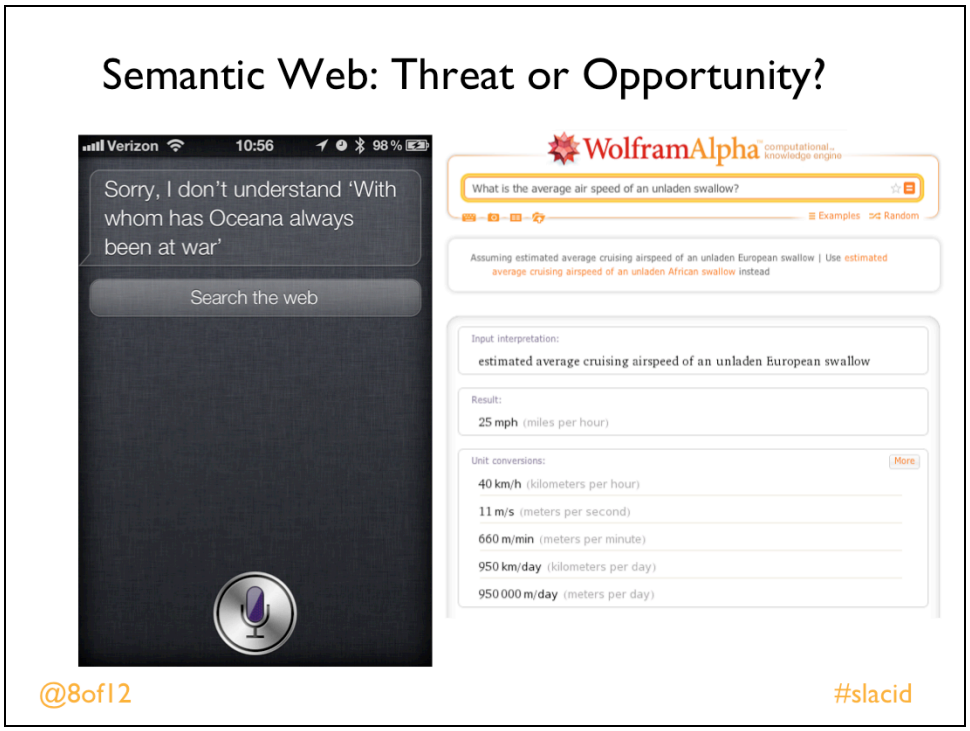
@8of12 #slacid

Specific cognitive biases make humans bad predictors:

- We actually tend to over-predict the pace of change
- Pan Am flights to the moon in "2001: A Space Odyssey:
- At the same time we under-predict the impacts of change, especially indirect implications
- Arab Spring
- We are over-reliant on our past experiences as templates for predicting the future
- Collapse of the Soviet Union
- We struggle to grasp complex causal systems that drive change and instead seek single-factor explanations
- 2008 Financial crisis

Quote source:
Yogi Berra as quoted at
http://www.famous-quotes-and-quotations.com/yogi-berra-quotes.html (16 June 2012)

Images
http://bigpicture.typepad.com/comments/images/dow_36000_1.png
http://binblog.files.wordpress.com/2009/08/thelongboom.jpg



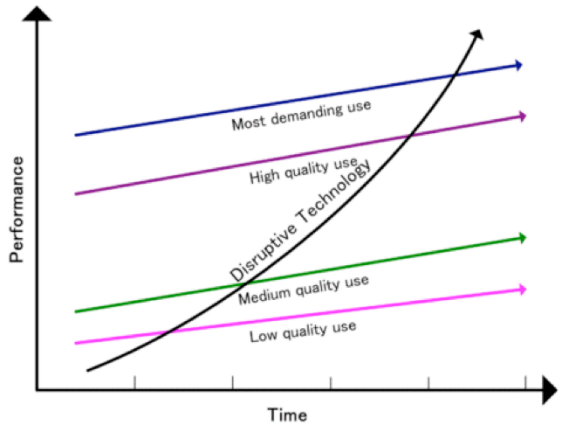
"Semantic web" is an imprecise term for technologies that are unifying the meaning of digitally-published content with descriptive meta-data so that web information can be processed by computer much the way numerical data are processed.

- A computer can draw meaningful conclusionsMeaningful connections among dataExplicit representation of meaning

Development of the semantic web follows a curve similar to Moore's Law, so sorry state of today's technology should not give anyone solace. Functionality will improve on a non-linear scale.

Images:
August Jackson iPhone screenshot of Siri
August Jackson computer screen shot of Wolfram Alpha query

Disruptive technologies follow non-linear improvement curves



@8of12 #slacid

The model for disruptive innovation gives us a framework for how semantic technologies are likely to develop and impact existing solutions to key problems. The key issue to remember about disruptive technologies is they tend to improve on a logarithmic scale. This tends to surprise incumbents that expect new technologies to progress at the same linear pace existing solutions tend to follow.

- Organizing information and data
- Increasingly automated creation of taxonomies and organizational structures
- Automated application of organizational structures onto existing knowledge sets
- Retrieving information
- Support for more complex queries
- Tolerance for vague queries
- Simplification of query syntax approaching plan spoken language
- Information presentation
- Automated summarization of single and multiple information sources
- Reconciliation of information across multiple data sources
- Reconciliation of distinct schema and context
- Visualization of large volumes of information for faster clarity
- Ability to drill down for detail, scale up for “30,000 foot view”

Questions?

Scenario analysis enabled us consider possible futures and create meaningful early warning systems

High Uncertainty

Disagreement

Significant Events

@8of12 #slacid

Scenario analysis allows us to build a shared baseline for strategic thinking and develop meaningful strategic early warning.

Scenario analysis is a useful methodology for understanding and anticipating the future:

•In situations with high uncertainty
•The timeframe under consideration is beyond the ability to forecast using traditional means
•When trying to anticipate the outcome of unlikely and significant events
When there is a great deal of disagreement among decision makers and stakeholders, especially when multiple perspectives have similar merit)There is a lack of a common language to describe uncertainty-- no shared strategic vocabulary
According to the consulting firm McKinsey scenario analysis is becoming one of the most important tools for corporate strategic planning:
•Most famous and one of the earliest examples is Royal Dutch Shell
•Harley Davidson of Canada

Sources:
http://www.shell.com/home/content/future_energy/scenarios/
Harley Davidson case is based on a presentation by Kenneth Sawka and Dom Bovalino "Linking Intelligence to Strategy with an Early Warning Indicators Framework" presented at the Strategic and Competitive Intelligence Professionals (SCIP) annual

A clear problem statement is critical to a quality scenario analysis

Subject Domain	Customer Segment
Timeframe	Geography

@8of12 #slacid

All intelligence projects begin with a requirements analysis, and scenario analysis is the same. It's important to have a clear problem statement that aligns with Key Intelligence Topics (KITs) that are based on the decisions stakeholders are expected to make. This is important to help focus the scope of which questions your scenario analysis will be expected to illuminate.

- Subject domain: The product or service market, industry or topic that will be examined.
- Customer segment: It is often helpful to clarify whether or not the customers in questions are individuals or businesses as well as clarify specific attributes within either of those categories (e.g. income level or industry sector)
- Timeframe: How far into the future do we intend to look? Scenario analysis is generally considered useful for looking 3 – 10 years into the future.
- Geography: It is very useful to understand how local or global the scope of our analysis is intended to be.

STEEP framework aligns trends and uncertainties

Social	Technological	Economic	Environmental	Political
• Demography • Family life • Public health • Religion • Culture • Beliefs	• IT • Biotechnology • Materials science • Manufacturing	• Drivers of growth • Inhibitors of growth • Monetary environment • Business cycles • Wealth distribution	• Air quality • Water quality • Arable land • Climate trends • Resource availability	• Laws • Regulation • Elections • Political power distribution • Conflict • Litigation • International Relations

@8of12 #slacid

STEEP is a flexible framework for organizing the trends, events, drivers and inhibitors in any given macro-market. Without a framework like this it can be difficult to conceive of the diverse trends impacting a nation, industry or market. Alternatively STEEP provides sufficient structure and guidance to focus those considerations.

An initial STEEP analysis can be done from secondary sources. Good STEEP analysis gets beyond the mainstream press to focus on specialized sources. Emphasize solid quantitative data and qualitative information. Employ a very healthy sense of skepticism and look for independent perspectives to confirm trends.

Avoid trying to draw conclusions about the data at this stage. Considerations of implications will come later.

STEEP is purely external to the institution, organization or firm. Think macro.

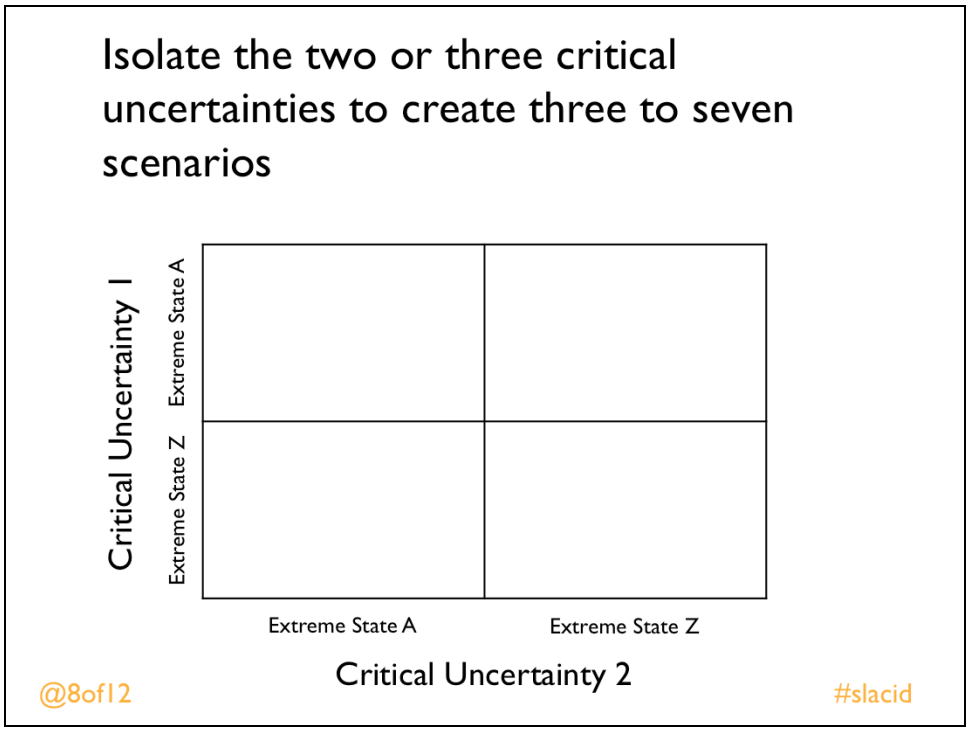
Source:
Garland, Eric. Future, Inc.: How Businesses Can Anticipate and Profit from What's Next. Washington, AMACOM: 2006.

Moving to the next phase of your scenario analysis is to engage with experts to augment, qualify and rank the trends identified in your initial STEEP analysis.

- Are there any new trends you need to add to your STEEP analysis?
- What details can your experts add to your identified trends and uncertainties?
- Separate known and understood trends from uncertainties
- Rank uncertainties in terms of their level of importance

Many scenario analysis exercises engage in a Delphi technique by sending out questionnaires to engage experts in multiple rounds of review and refinement. Key stakeholders should be engaged in this phase of the process. Their involvement throughout the active stages of the scenario exercise makes it more likely that they will accept some or all of the ultimate analysis.

External experts (and perhaps even non-experts) should have their opinions sought as well. It may be appropriate to develop sufficiently generic versions of your problem statement. External expert participation is actually better if these external experts are thinking in more general terms that allow them to apply their own experience and



One of the outcomes of your expert interviews will be a prioritized set of critical uncertainties. You'll choose 2 or 3 mutually-independent uncertainties to use as the basis for formulating your scenarios.
For 2 uncertainties create a 2 x 2 matrix. For 3 uncertainties you can create two separate 2 x 2 matrices.
Each scenario will consist of an extreme state for each of your uncertainties such that each uncertainty is considered as a binary possibility (e.g. this technology will be adopted by everyone or no one will be using this technology)
In some instances some may choose to create a 3 x 2 scenario matrix to describe a high/medium/low state for one of the identified critical uncertainties.
No more than 3 uncertainties should be considered because then scenarios cease to be distinct.
It's important to avoid the notion of scenarios as best case/worst case/most likely case because this approach tends to limit the strategic discussion that is the most important output of a scenario exercise.
People usually have difficulty keeping more than 7 scenarios in one of these exercises. In the event you choose 3 uncertainties you will want to eliminate some scenarios that are considered the least plausible.

Questions?

Here's the problem scope we'll address in a very high-level analysis

Subject Domain: Application of semantic technologies to models for research and analysis	Customer Segment: Knowledge Professionals and their Stakeholders
Timeframe: 3-5 years	Geography: Global

@8of12 #slacid

As with all intelligence projects, we need to begin ours by defining a clear scope.

- Subject domain: We are interested in understanding how models for research and analysis will change based on how widely or narrowly semantic technologies come to be adopted and applied to specific sets of information.
- Customer segment: For the purposes of this study, let's consider ourselves and our stakeholders as the customer "segment."
- Timeframe: Just for the sake of our purposes let's say 3 to 5 years. Be aware that timeframe is particularly important if you're going to incorporate quantitative elements into your trend analysis. For the simplicity of this exercise we will be much more general in our analysis.
- Geography: Semantic technology has a global scope, so for our purposes we'll consider a global scope. In reality defining a geographic scope will help you narrow the number of relevant trends, be very specific about the social, economic and political factors and also get to a greater level of specificity with the quantitative details related to your trends and uncertainties.

STEEP framework aligns trends and uncertainties

Social	Technological	Economic	Environmental	Political
• Increased use of social media • Corporate adoption of industry standards and general ontologies • Growing frustration of time spent seeking information • Consumerization of IT	• Inclusion of semantic standards in applications • Adoption of non-relational databases • Adoption of natural language processing • Availability of easy-to-use tools to create ontologies • Ubiquitous high-speed network connectivity • Growing volumes of data • Moore's Law	• High valuation of "big data" companies • Cloud-based platforms move IT spend from capex to opex • Availability of semantic expertise • New revenue models and services based on data	Drive to conserve energy with smart grids and smart appliances	• Public budget constraints drive push for cost-savings • Regulations related to personal data • Desire to increase transparency of some government operations

@8of12 #slacid

The STEEP framework helps us identify and organize the relevant trends and uncertainties we should consider related to our problem scope. In a full scenario analysis these trends and uncertainties would be much more specific in their nature, including quantitative forecasts, references to specific laws and regulations and deeper detail on specific technological standards and frameworks. For example, a full scenario analysis of the semantic web might incorporate a perspective as to whether or not RDFa or OWL are the more appropriate standards to move the state of the semantic art forward.

A good scenario analysis will capture the opinion of a diverse set of external and internal experts, customers, users and the like about what trends should be considered and the forecasts for those trends. Full scenario analysis projects have many more factors than you see here.

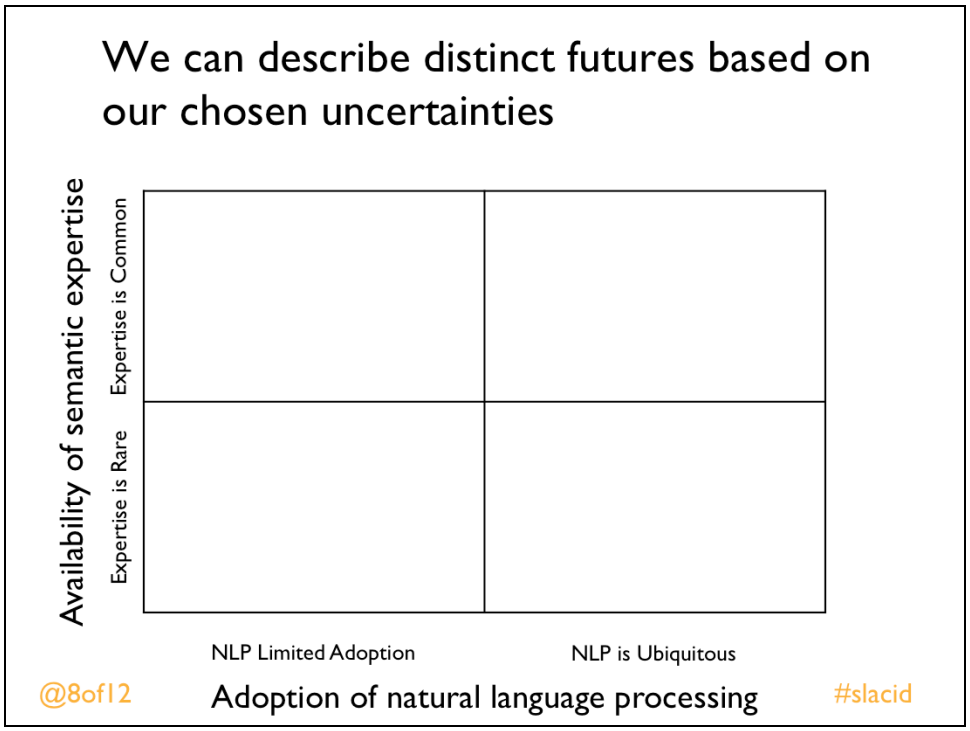
1. Availability of semantic expertise
2. Adoption of natural language processing
3. Corporate adoption of industry standard ontologies
4. Inclusion of semantic standards in applications
5. Adoption of non-relational databases
6. Availability of easy-to-use ontology editing software
7. Valuation of "big data" companies
8. New revenue models and services based on data
9. Regulations related to data privacy and protection

@8of12 #slacid

One of the goals for your Delphi rounds is going to be to rank the trends and uncertainties in terms of their order of importance or impact on your subject domain. Using this ranking you will choose the top 2 or 3 uncertainties to create your matrices that will enable you to produce between 3 and 7 scenarios.

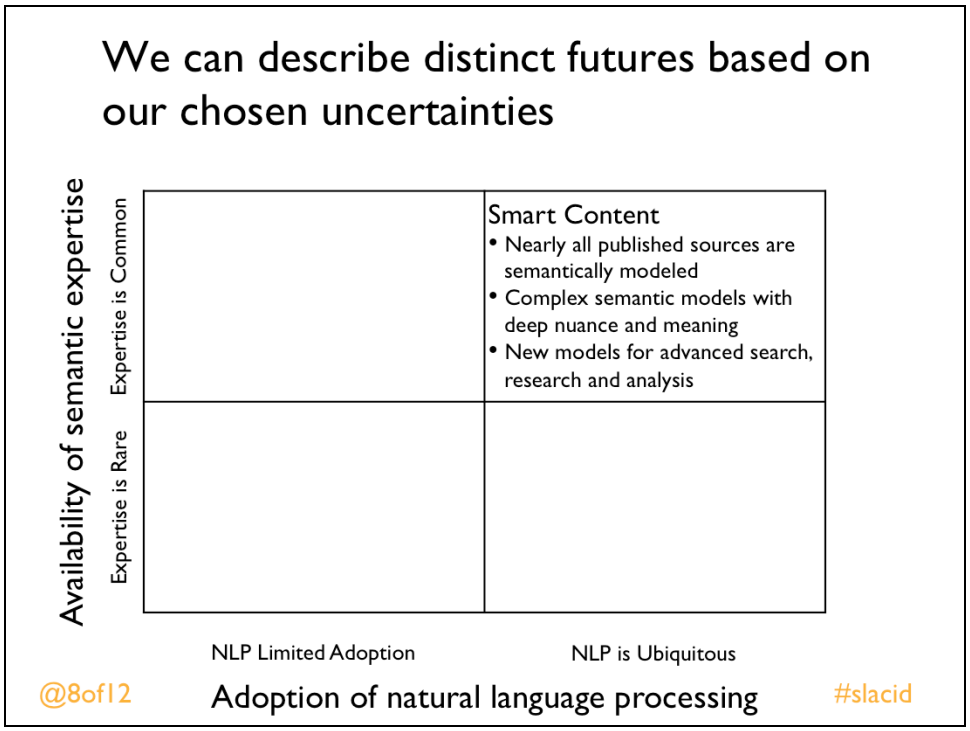
For the purposes of our very simple exercise, I've identified two uncertainties as our most critical:
- Availability of semantic expertise. How widely held is the ability to create the rich taxonomies and ontologies that are the basis of semantic models. How prevalent is the programming expertise to build necessary semantic software platforms.
- How widely will natural language processes software be adopted? These are software tools that are able to convert written text (and even transcribed audio and video) into semantic models.

You may agree or disagree with these choices. You may even agree or disagree as to whether or not one of both of these selections are actually uncertainties. This disagreement is valid and illustrates the importance of the diverse panel and multiple Delphi rounds to develop a consensus around these choices. Creating buy-in with the customers



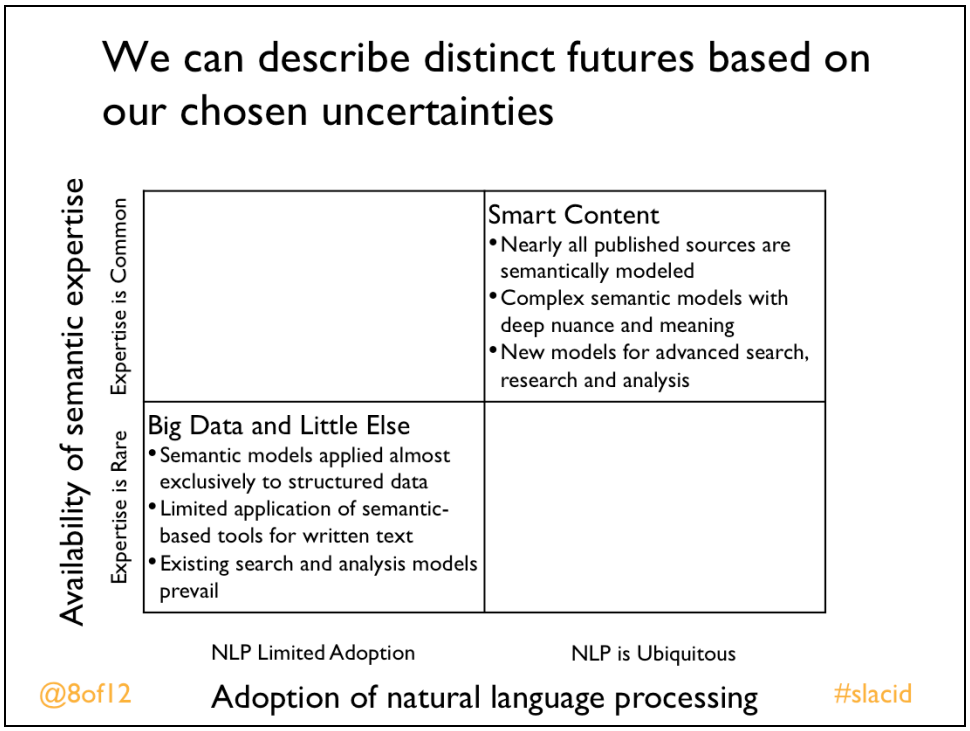
Now that we've chosen our two uncertainties we can create our matrix. With this matrix we will be able to draft four scenarios based on the extreme states of our two uncertainties.

A real scenario analysis would provide a much clearer quantitative or qualitative distinction for the state of each uncertainty. It may be necessary to develop quantitative measurements in order to get buy-in on your analysis from stakeholders that seek quantitative rigor.



So now we describe a world in which natural language capabilities are ubiquitous and the expertise to develop semantic models and platforms is common. What does that world look like? What are the features of that world that are most pertinent to our initial problem statement?

- Certainly we are concerns with which data sources will have semantic modeling and tools applied to them. In this world we should expect semantically-rich data structures to be available for data as well as domains that are more commonly associated with unstructured data (e.g. text).
- How rich are those models? Because there will presumably be a large number of people with the requisite skills we can expect there to be rich semantic models designed by people for many subject domains in business, academia and science. In this world semantic models are not limited to specific industries or topics.
- So what this means for knowledge professionals and our roles is that new models for search and analysis will be available to our stakeholders. Much of what we do for them today in very labor-intensive and time-consuming projects they will be able to more easily do themselves with comparatively easy-to-use software tools.



Next we'll look at the alternative state where there are few people with the skills to build semantic models and tools and there is very limited adoption of natural language processing.

Semantic models will be applied almost exclusively to structured data. Because the expertise to build these models will be hard to find and expensive, semantic models are likely to prevail only in those applications where there is a very high ROI, e.g. financial applications

This is the world that is probably most similar to our current state of affairs. This likely does not present many expectations for change on the part of our stakeholders. Innovation in research and analysis may happen around the edges of the average user's experience or at the very high end of the market, but the current token-based model for search would likely prevail in this world. Collation of information from diverse sources and analysis will remain time- and labor-intensive.

We can describe distinct futures based on our chosen uncertainties

Availability of semantic expertise

	NLP Limited Adoption	NLP is Ubiquitous
Expertise is Common	**Big Data Everywhere** • Semantic expertise applied primarily to structured data • Meaningful ontological models for structured data become the norm • New search and analysis models for data, less so for text	**Smart Content** • Nearly all published sources are semantically modeled • Complex semantic models with deep nuance and meaning • New models for advanced search, research and analysis
Expertise is Rare	**Big Data and Little Else** • Semantic models applied almost exclusively to structured data • Limited application of semantic-based tools for written text • Existing search and analysis models prevail	

Adoption of natural language processing

@8of12 #slacid

Next let's consider a world where semantic expertise prevails but natural language tools are still rare or lacking. This, to me, seems like the most unlikely of the four scenarios.

In this scenario we will apply semantic models to structured data sets-- probably a good majority of those data sets. This will mean much greater flexibility for consolidation across diverse data sets. However, with limited application to unstructured information sources (e.g. text) the challenges to which semantics might be applies will still be somewhat limited. They will become common-place in practices and fields where structured data are the norm (e.g. finance, science) but less so where unstructured data prevails (e.g. humanities, strategy).

This scenario will require some level of data expertise out of all knowledge professionals dealing with problems in those data-intensive domains.

We can describe distinct futures based on our chosen uncertainties

Availability of semantic expertise	NLP Limited Adoption	NLP is Ubiquitous
Expertise is Common	**Big Data Everywhere** • Semantic expertise applied primarily to structured data • Meaningful ontological models for structured data become the norm • New search and analysis models for data, less so for text	**Smart Content** • Nearly all published sources are semantically modeled • Complex semantic models with deep nuance and meaning • New models for advanced search, research and analysis
Expertise is Rare	**Big Data and Little Else** • Semantic models applied almost exclusively to structured data • Limited application of semantic-based tools for written text • Existing search and analysis models prevail	**Statistics = Meaning** • Building ontologies is expensive and requires clear ROI • Reliance on statistical methods for concept mapping • New search and analysis methods based on word co-occurrence

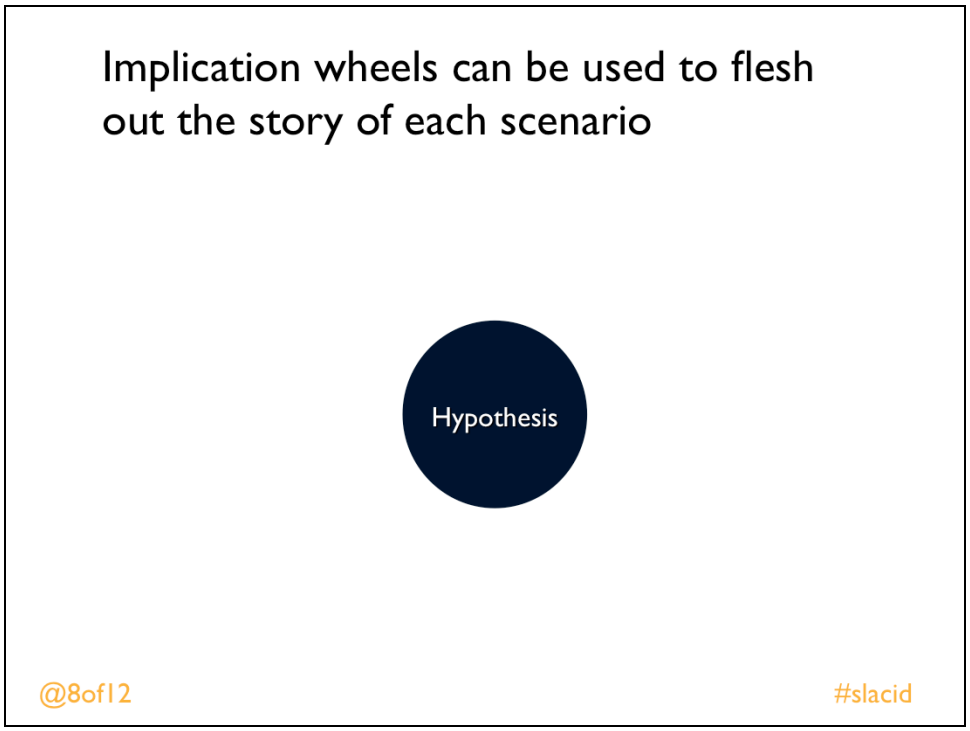
Adoption of natural language processing

@8of12 #slacid

Finally we come to the scenario where semantic expertise remains rare but natural language processing capabilities are common.

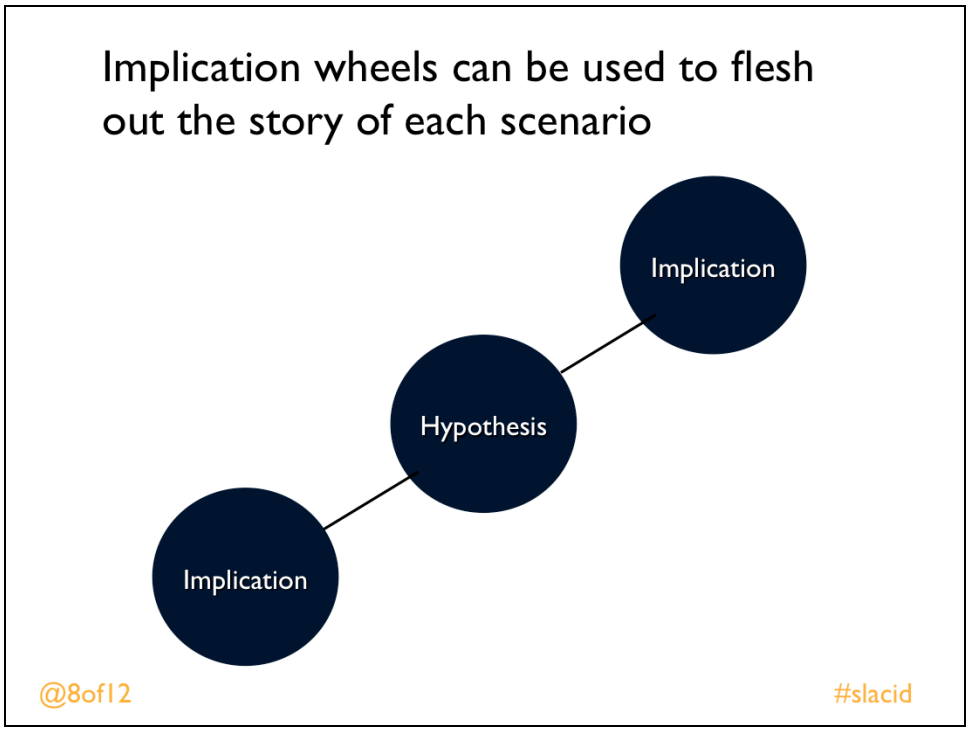
In this scenario we can expect NLP tools to take the place of expensive semantic expertise. The models for applying semantic tagging to unstructured information (e.g. text) will be based primarily on what NLP software is able to develop. Co-occurrence of words within a corpus of text will be interpreted as relationships among those concepts by NLP software. The context of those relationships is more difficult to capture without engaging humans with expertise in creating semantic models.

This scenario will see new models for research and analysis. The role of the information professional would require a greater understanding of statistics in order to evaluate the relevance and importance of linkages among words identified by NLP software.

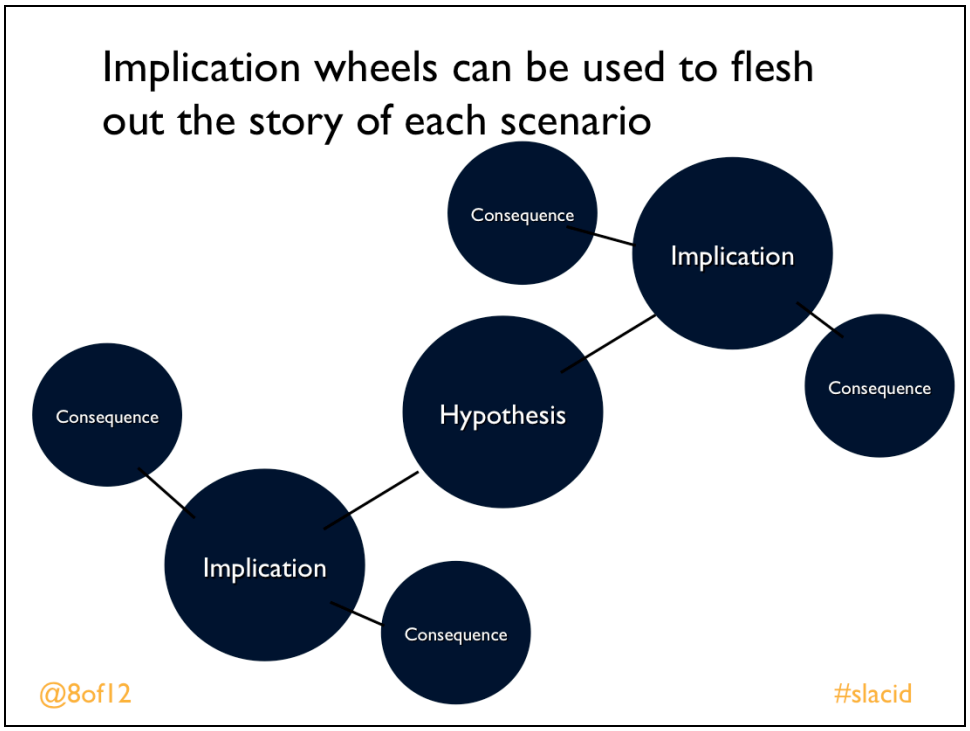


So how did I develop the story associated with each of those scenarios.

One of my favorite methods to build out the story for each scenario is an implications wheel. You start out with a small set of hypothesis about your scenario, for example “There will be few people with expertise in building semantic models or developing semantic functionality for software.”

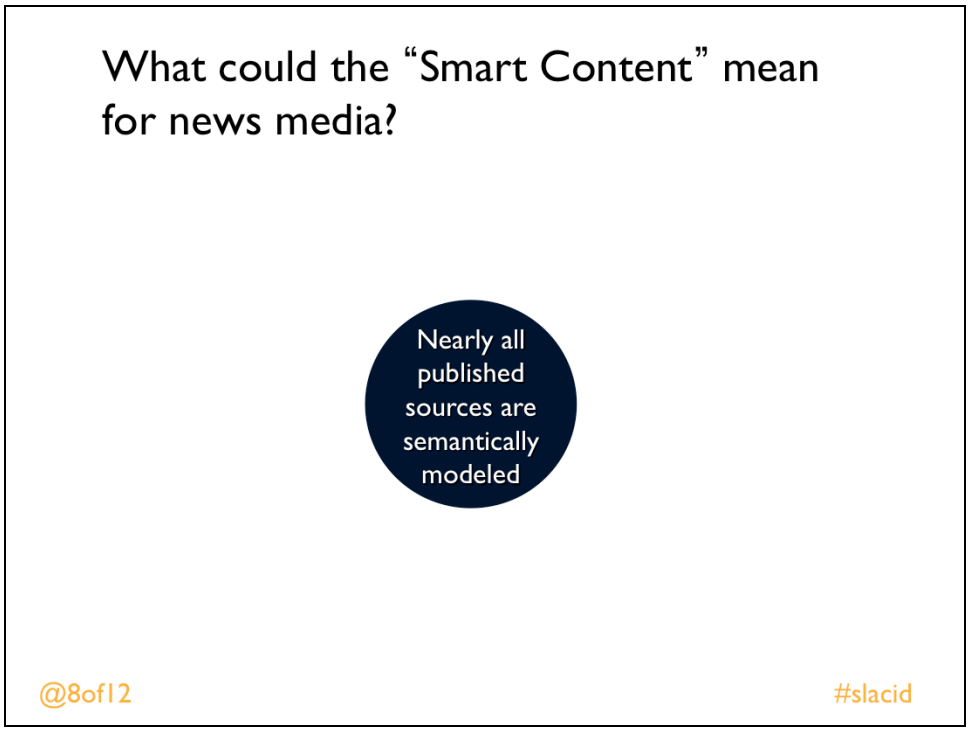


Then you ask yourself “What are the implications of that reality?” It’ s useful to recall your original problem statement when you think about those implications.
In the case of limited semantic expertise, it will be expensive to complete semantic projects



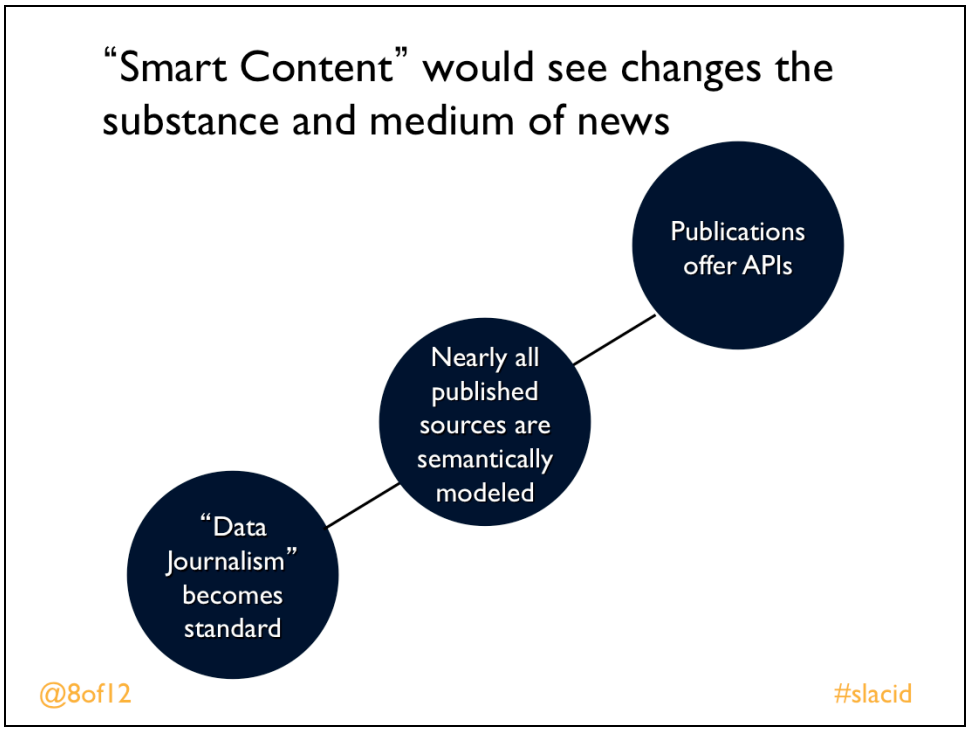
Then you ask yourself “What are the consequences of that implication?” You can also think of consequences as a next second round of implications-- implications of implications.

If it is expensive to hire for semantic expertise, then only projects with a very clear and immediate ROI are likely to get any attention. The focus will probably be on data-heavy industries with lots of money to invest (e.g. healthcare) or those for which data models can deliver outsized competitive advantage (e.g. finance).



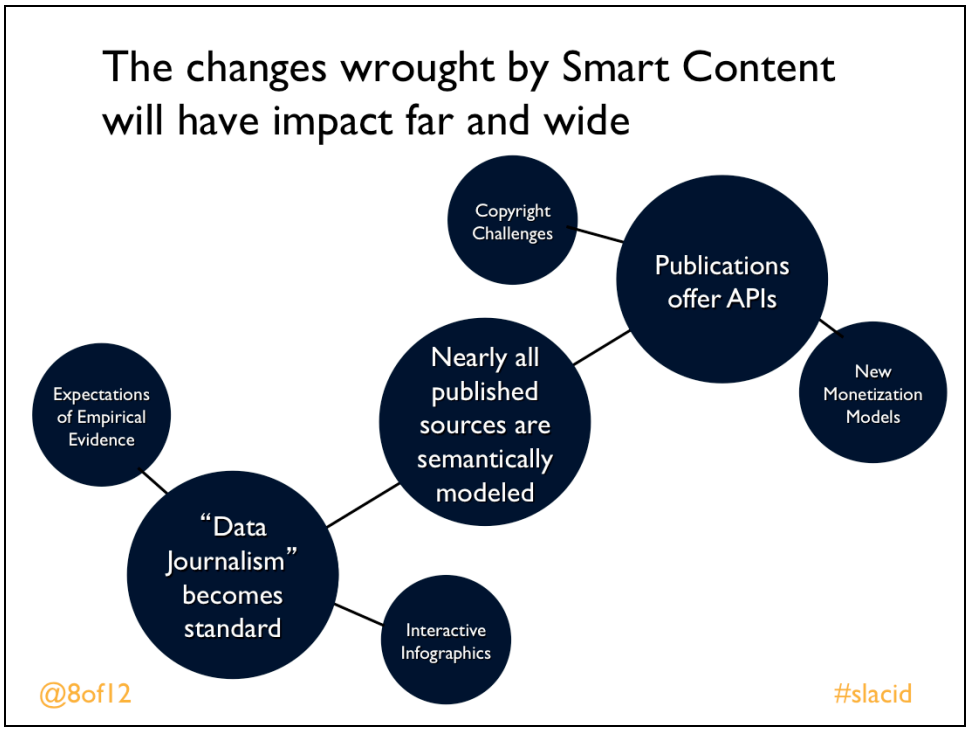
In the description for this session I mentioned that scenario analysis could be used to predict the future of specific industries, including news media and telecommunications. To really examine the impact to the industry using scenario analysis we would incorporate the industry focus into our problem scope.

In the interest of time we're going to take a different approach and use our implications wheel to explore how one of our scenarios will impact the news industry. So let's think about what the consequences and implications would be for this industry if semantic modeling becomes the norm.



One implication of ubiquitous semantic modeling of published sources will be new models for accessing that published content. In addition to traditional reading, content will be able to be processed by software, or Application Programming Interface (API). One model that already illustrates this is the New York Times APIs.

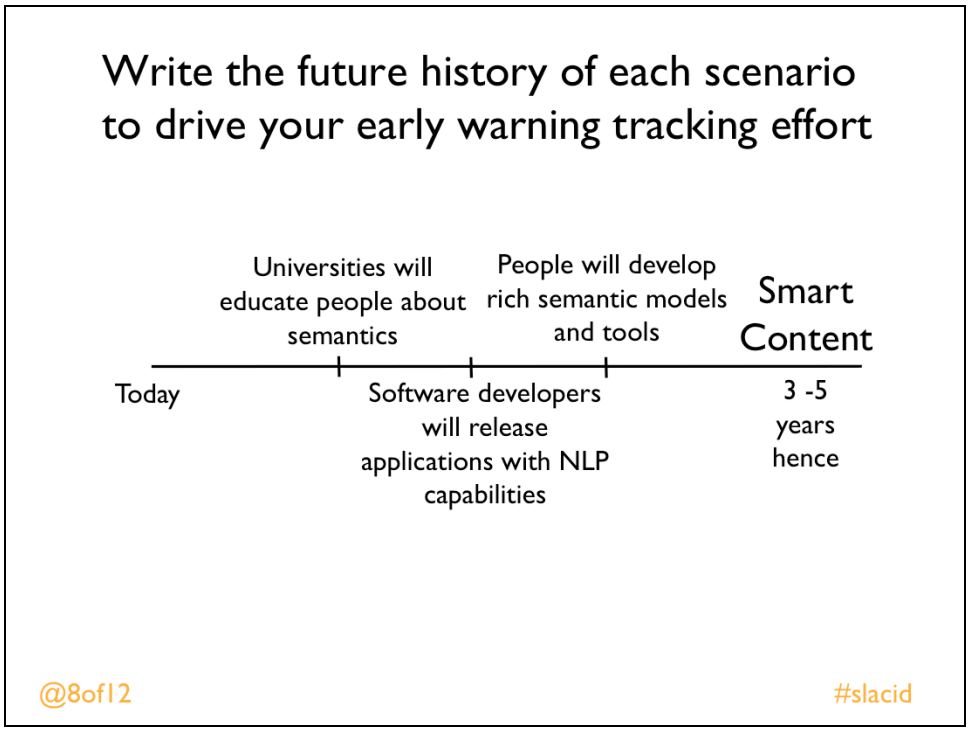
Another implication is what that will mean for the change of the substance and value of news. A transition that we're already seeing is the growth in "data journalism" practices for publicly available data sets. With semantic modeling of structured data will make new methods of analysis possible. Extending these data analysis practices to unstructured sources. Data journalists would be able to use software tools to analyze and cross-reference huge volumes of published sources quickly.



Two immediate consequences of publication APIs become apparent are around the business models and ownership of content. Existing concerns that arise from digitization of content will become more problematic as individual data points and facts begin to have a life and value beyond the article in which they were published. Some of the value of those data will only be realized when it is cross-referenced with other data and facts owned by other people. What does this mean for copyright, fair use and concepts we take for granted today?

The flip side of that very issue: some users will see tremendous value for access to news sources not by the article or the drink but to entire archives and for commercial purposes. This could make new monetization models possible.

The ability to analyze these volumes of data with fidelity will greatly increase expectations for evidence-based reporting and empirical evidence. The product of journalism will incorporate interactive infographics and robust data visualizations.



For each of your scenarios, you'll want to work backwards from your imagined future and plot out the important milestones and events you would expect to see to enable that future to come into being.

In a full scenario analysis this future history will be very detailed. For example:

- How many students will need to graduate with semantic expertise?
- What are the timelines for existing professionals to develop the appropriate competencies.
- How many universities or training programs would need to be in place and by when?
- Based on standard technology adoption models, how long would it take for natural language software functionality to become mainstream?

These future histories will be the basis for your early warning system. Once you have a detailed set of milestones for each scenario, you will be able to develop a research and monitoring plan. Who are the thought leaders in each domain? Where are they speaking and publishing? Who are the actors who will influence the outcome of each milestone?

Questions?

Thank You!

August Jackson
august (at) augustjackson (dot) net
http://augustjackson.net
@8of12