Original Article

Cost-benefit analysis: Integrating an enterprise taxonomy into a SharePoint environment

Mike Doane

is Content Strategist at Ascentium Corporation. He brings over 20 years of information organization and content management experience to a project. He is currently working on engagements that require Internet/Intranet content strategy/management and taxonomy implementations. For Ascentium, he has worked on projects for NASA, Del Monte, Microsoft, Southern California Edison, University of Washington Medicine, Providence Health & Services, First Mutual Bank, Premera, Sterling Savings Bank, and over a dozen other clients in the Pacific Northwest. He has worked with clients to develop content strategies, enterprise taxonomies and search systems, metadata strategies, content optimization and content management migrations. He has worked on numerous projects implementing Microsoft Office SharePoint (MOSS) 2007, Interwoven TeamSite and Plumtree systems. Before joining Ascentium, he worked for Avenue A – Razorfish as a Content Strategist on projects for Nike, Microsoft, Philip Morris, SCE, Visa, Washington Mutual and Weyerhaeuser. He holds a Masters degree in Library and Information Science from the University of Washington and teaches graduate courses there in the Information School.

ABSTRACT For many enterprises, the need exists to thoroughly integrate a corporate-wide taxonomy or taxonomies into every aspect of the search experience on a corporate portal where documents are regularly created, maintained and retired. The reality is that many large organizations have isolated, 'one-off' taxonomy efforts within different groups or divisions that effectively function for that group or division, but would not scale out appropriately for the entire enterprise, due to discrepancies in term usage, term definitions and term relationships. (Think of how many different ways that the term 'Business Intelligence' could be used across multiple business units in an organization, for example.) Only occasionally are there attempts to scale the taxonomy solutions across the entire enterprise portal. In many cases, multiple search systems and platforms are being used, and the multiple (often redundant) taxonomies have been designed specifically for those systems/platforms. This article will examine how a cost-benefit analysis was performed to determine the costs and benefits of enabling a division-centric taxonomy that was developed specifically for the Google Search Appliance to be integrated into a SharePoint platform used by multiple other divisions, thus enabling those SharePoint users the benefit of a significantly improved way to discover tagged digital assets in SharePoint search results.

Journal of Digital Asset Management (2010) 6, 262–278. doi:10.1057/dam.2010.28

Keywords: enterprise taxonomy; SharePoint search; cross-platform tagging; Google Search Appliance; cost-benefit analysis; ROI

OVERVIEW

This article examines how a government agency (hereafter referred to as the 'Agency') with a roughly 3000 employee user base could benefit from improved document findability by integrating an 18-month-old taxonomy of terms that is currently used by a Google Search

Appliance in one part of their portal into their less-used SharePoint environment.

The Agency, like many government and private companies alike, does not have its portal completely integrated. There are documents and web pages existing in the 'official' portal, along with a separate and less used SharePoint portal.

Correspondence:
Mike Doane
Ascentium Corporation
10500 NE, 8th St. Suite 1300
Bellevue, WA 98004, USA
E-mail: mike.doane@
ascentium.com

The desire of the Agency is to eventually integrate these two separated experiences into one all-up portal, but there are significant political, security and logistical reasons why this hasn't taken place yet.

Users, however, are expressing a need to find documents across both portal areas, thus the desire to find out if the existing taxonomy, which has improved search results on the official portal, could be used to index documents in the SharePoint portal and return a hybrid or 'blended' search results page.

Currently, the Agency taxonomy is maintained by the Agency taxonomy team. It was designed and implemented using a commercially available taxonomy management tool, which provides the ability to regularly update the taxonomy related to topics, activities and people associated with work performed at or related to the Agency.

This taxonomy is used to inform the indexing process of the Google Search Appliance as it relates to the content on the Agency portal, used by many Agency employees. In combination with the Google Search Appliance, the Agency taxonomy in the taxonomy management tool provides enhanced search results for Agency employees, since it has a more contextual understanding of the content produced by Agency employees.

The Agency taxonomy is consistently being updated by the taxonomy team with new topics and term relationships that continually enhance the search results from the Google Search Appliance.

SCOPE

The scope of this project only includes evaluating the existing Agency taxonomy built and maintained within the commercially available taxonomy management tool, the existing Google Search Appliance integration with the Agency taxonomy and the SharePoint portal environment and the existing capabilities of those systems.

The following are considerations on how this study was limited in scope:

- The study was limited by the budget of the study. Although fairly representative of the broad spectrum of portal users looking for documents, the number of respondents isn't

necessarily statistically significant enough to accurately measure the entirety of portal users.

- The study assumed that the users at the Agency used both the existing portal, and the SharePoint environment. The SharePoint environment isn't colloquially referred to as the 'portal'. Instead, it is often referred to as 'the SharePoint', meaning the documents and information stored in a SharePoint site collection, for example.
- The study assumed that not all of the users have access to the Agency SharePoint sites. There are security and personnel reasons why many of the SharePoint sites at the Agency are restricted, which may have an impact on users reporting they could or couldn't find the information they were looking for. The information simply may not be accessible to them.

OUT OF SCOPE

Out of scope for this cost-benefit analysis is

- estimating the effort of modifying the Agency taxonomy to meet the needs of the SharePoint environment users;
- estimating the effort of the continued maintenance of the taxonomy to maintain relevancy for the documents in the SharePoint repository.

METHODOLOGY

For this cost-benefit analysis, we surveyed a selected range of Agency employees or contractors who were asked a series of questions about how they engage with the Agency portal and/or the Agency SharePoint environment during a typical work day/week. These users provided the data used to evaluate the following:

- How do users currently look for information in these systems?
- How much time do users spend during a typical workday looking for information?
- Do users find the information they need?
- What search system(s) do they use?

We were then able to ascertain a cost associated with the effort to find needed information, and project associated costs of having to recreate

information that might have already been created, but not found.

The budget of the project allowed only enough time to review data on 23 responses. More users were surveyed, but for various reasons (vacation, unwilling to participate and so on, did not respond. The data are therefore limited to the users sampled, although the range of users who did respond was fairly broad. A more extensive survey should be done to gain insight into a broader range of users. However, for this study, the limited number of respondents did provide enough data to be useful.

EVALUATION CRITERIA

For this study, the cost-benefit analysis is meant to evaluate the cost-effectiveness of integrating the existing Agency taxonomy into the SharePoint environment, and how much value employees would gain in discovering and using information in the SharePoint portal that they cannot find or use now.

There are approximately 3000 information workers who could access the SharePoint and existing portal environments. The two environments are now disconnected, and the information workers at the Agency have to perform two separate searches, on two separate repositories. There are no known dates on whether or not the SharePoint environment will be integrated into the existing non-SharePoint portal, nor are there any projections about usage rates (percentage of the total of approximately 3000 workers who would use the portal).

The alternatives to integrating the existing Agency taxonomy into the SharePoint environment include:

- Leaving the integration as is, and not use the Agency taxonomy to further tag the SharePoint content.
- Improving the existing SharePoint search system to integrate into the Google Search Appliance search system.

In order to gain an understanding of the costs associated with finding information for portal users – for both the Agency portal and the SharePoint portal environments – an initial survey was created to query a selected set of users. This survey was meant to find out how Agency employees search, what barriers to

finding information they encounter, and what successes they have, in order to establish costs associated with finding (or not finding) information.

Interpretations of the data follow. Details of the responses are contained in the Appendix.

Evaluation data

An e-mail was sent out to individuals identified by the taxonomy team as likely users of the portal. The instructions sent are below:

Thank you for taking a few minutes of your day to answer these questions. Your answers will help us determine whether or not we should continue to consider integrating the existing Agency taxonomy into the SharePoint environment. Your name will not be used in our report, nor will any response you provide be associated with your name.

Following are the questions sent to the respondents, their responses, and an analysis of their responses as they relate to our cost-benefit analysis.

Question 1

How much of your day is spent simply looking for information you need in order to do your job? (For example 'I spend 30 per cent of my day looking for information'.)

Overall, the individuals spend an average of 26 per cent of their day looking for information. Given an average 8-hour workday, this equates to slightly over 2 hours per day looking for information. On the basis of a study by IDC¹ in May 2009, the average information worker spends 8.8 hours a week looking for information, at an average cost of US\$14209 per year. The Agency worker spends, on average of 10.4 hours a week looking for information.

The interesting anomaly was that 5 of 23 respondents (20 per cent) said they spend 5 per cent or less of their day looking for information.

Given that the employees surveyed held jobs that would normally be considered 'information workers' (not workers in such areas as cleaning staff or general maintenance, for example, who don't use computers as part of their daily toolset), this could imply that the respondents

didn't understand the question, or simply didn't respond carefully.

Outcome

Agency workers spend more time than the average information worker during a workday looking for information, as compared to the IDC report. This could be for a number of reasons, ranging from an inability to find the information on the first or second search, to the fact that a lot of the information created at the Agency is research based and requires a lot of time spent on finding information from multiple sources.

Whatever the end reasons, the fact is that Agency workers appear to be in need of better search results in order to reduce their time spent looking for information, better search results or both.

Question 2

What search system(s) do you mainly use? (For example Google, Yahoo!, SharePoint.)

Nearly 70 per cent of the Agency workers who responded said that they use Google search, both internal searches and external searches. Only 13 per cent who responded said they use SharePoint search. Others use internal search tools or internal databases.

The overwhelming use and acceptance of Google search means that users are aware of how Google search works are amenable to Google search results and might already be seeing/using search results provided by the Google Search Appliance integrated with the Agency taxonomy. This might mean that users will be more amenable to search results that include SharePoint content.

Outcome

Google search is already widely accepted as part of the search experience for Agency employees. It is clear from the survey results that improving the search results within the Google Search Appliance environment would not pose a risk for user adoption for a search system, since most users are comfortable with the Google search experience. This could indicate that if a Google Search Appliance-based solution was considered for improving the search results, users would already be comfortable with using Google instead of

adapting to a new or currently underused search system, such as SharePoint search.

This does not mean, however, that users would abandon SharePoint search. It just means that extending the range of content that the Google Search Appliance indexes means that users would not have to 'learn' how to use a new search system.

Question 3

Do you use any other supplemental search systems? (For example 'I use Google, then SharePoint'.)

Three respondents first searched Google and then SharePoint. Two respondents used other sources first and then used Google. Six of the respondents didn't use any other search but Google. While six other respondents used other means to gather information.

For these responses, the answers paralleled the previous question in that most respondents use Google first, then usage drops off significantly for other search systems. Six respondents didn't use a secondary search system, which could mean that they abandoned search systems altogether for other sources of information.

The more interesting response was from one respondent who said 'No, I start calling people'. This is one factor that was not considered for this analysis: the amount of time spent engaging another employee in the search process.

Outcome

Most employees use one search system, and may use more labor-intensive secondary search 'systems' such as calling other employees on the phone to get information. This indicates that their primary search engine is still not producing the results they need, and the cost of finding information can increase, or even double, if more than one employee has to help out by performing their own search for information.

This is a cost that cannot be exactly calculated for this cost-benefit analysis, since we were not able to do further studies with how much time is spent with secondary search systems.

Question 4

How successful are you at finding the information you need the first time you look for it? All the time? Some of the time? Never? (For example 'I find what I need about 30 per cent of the time'.)

Of the respondents who provided an answer, the average response was 57 per cent. This can be translated to mean that on average, Agency employees find the information that they're looking for slightly more than half of the time, the first time they search. This means that, on average, 43 per cent of the time information seekers have to search again for information, or seek an alternative source for that information.

Outcome

Although we did not determine how often it takes to qualify a search as 'successful' it can be noted that on average, Google's success rate in the marketplace is around 75 per cent for search results.² This could mean that Agency employees are having a significantly lower than average success rate for finding information, especially considering that this is internal, portal-related searches, which tend to be more focused. This could be indicative of the users desiring to find information that is in the SharePoint portal, or at least not in the Agency portal that does not include information stored in SharePoint.

To refine the cost estimate of finding information, we recommend doing a cost analysis of this 'secondary' search system, which was not within the scope of this study.

Question 5

When you can't find information, do you attempt to re-create it? Or do you ask people around you? Or ask people to send it to you via e-mail?

For this question, the majority of responses involved asking others for the information. Phone and e-mail are the two most likely ways for people to ask others about the information they need. A smaller percentage are willing to recreate the information (if they know how), but asking others via e-mail or phone is the preferred alternative method of getting the information.

Outcome

If we use the previous response rate of 43 per cent as an indicator of how successful employees are on first attempts to find information, then we can assume that close to half of the time, an Agency employee will contact another employee for help finding information. Recreating information can be one of the most costly

endeavors for a company. IDC³ says that on average, 3 hours per week are wasted by the average information worker on recreating existing content.

Given that Agency workers appear to spend more time looking for and creating information, it can safely be assumed that more time is spent, than average, recreating information, too. This adds to the overall cost of not finding the right information at the right time.

Question 6

Is your job more focused on creating information, consuming information or both? What percentage of each?

For this question, the goal was to determine what the information behaviors are, and whether or not the audience tends to create more content, which would mean they might do multiple searches for source material, or whether they mostly consume information, which would mean they might not spend as much time doing a search for information.

Of the 20 responses, a majority (11) indicated they consume more than they create. Six indicated they create more than consume, and three indicated about the same. What this may indicate is that a majority of the time, Agency employees are consuming more information than creating, and that searches may not be as research oriented, but more 'consumer' oriented. Searchers may just be looking for isolated pieces of information rather than a series of searches to inform the creation of other information.

Outcome

If we use the previous response rate of 43 per cent as an indicator of how successful employees are on first attempts at finding information, then we can assume that most of the time employees are not finding what they need, and a majority of the users engaging the system are consuming information, so having relevant and timely information in search results is critical.

The ability to find information in order to create more information isn't that critical for those surveyed.

Question 7

Although we are not recording each individual's salary, we need to know a salary range for you or

your peers to determine a more accurate cost for finding information. Which of these salary ranges does someone in your peer group fall into? <\$50 000, \$50 000–\$75 000, \$75 000–\$100 000, \$100 000–\$125 000, >\$125 000 (exclusive of benefits, bonuses and so on).

This request for salary information was voluntary, and no records were kept identifying individuals with their salary. Of the 20 replies, a majority (11) were <\$50 000 per year, five were \$50 000–\$75 000, three were \$100 000–\$125 000 and one was >\$125 000. There was no response for \$75 000–\$100 000. This indicates an average salary of nearly \$50 000–\$75 000. For this analysis to move forward, we're using a \$50 000 per year average for the following reasons:

- Although a majority indicated <\$50 000, we did not include costs list bonuses and benefits, which will likely increase the overall cost of each employee.
- Our sampling may not have been indicative of a broader set of employees who would make more than \$50 000 per year.
- On the basis of an average of the US Office of Personnel Management Salaries & Wages Salary Table 2009 for the Houston area,⁴ the median wage is \$42 per hour, or roughly \$84 000 per year. Understanding that the median salary isn't always an indicator of the average salary, it does help us balance our respondents' replies against a larger audience.

Outcome

In order to calculate a 'cost' to include in the cost-benefit analysis, we needed to know this salary information, since it's the only way to start to create a baseline for how much looking for and finding information is costing the Agency. The fact that we didn't include benefits in this study is because they can vary widely, based on whether an employee is a civil servant or a contractor. Since the composition of Agency employees is varied, and we didn't determine if a respondent was a civil servant or a contractor, we did not include a range of costs.

However, recent statistics from the US Department of Labor's Bureau of Labor Statistics⁵ indicates that on average, a worker's percentage of total cost includes roughly 70 per cent

wages and 30 per cent benefits. Given this average, and the average salary of those we surveyed, the total compensation package for the Agency worker would be roughly \$65 000 (\$50 000 salary and \$15 000 benefits).

On the basis of these figures, it costs the Agency on average \$338 per worker per week (or \$16 900 per year [\$338 per week × 50 weeks]) to search for information.

ALTERNATIVE SYSTEMS

Current state

The current system for integrating the Agency taxonomy within the Agency portal environment using the Google Search Appliance is outlined below (Figure 1).

1. The Agency taxonomy is built and maintained in a commercially available taxonomy management tool.
2. The Google Search Appliance is connected to the taxonomy management tool to provide the terms and hierarchies within the Agency taxonomy to further 'tag' and refine the search results it produces.
3. The Google Search Appliance enables faceted search results, allowing the user to pivot on topics and terms relevant to Agency employees, based on the hierarchies created in the taxonomy.
4. As the taxonomy is further built out and refined, the Google Search Appliance is continually updated to index content in the Agency portal, and provide more relevant and useful search results for Agency employees.

For the SharePoint environment, there is no integration of the Agency taxonomy into SharePoint search. Users of the SharePoint portal do not benefit from a search system integrated with a relevant taxonomy, and therefore tend to see search results not fine tuned for topics relevant to their work.

In addition, the SharePoint search system has not been designed to take advantage of filtered search results, so users are not able to pivot on search topics in the search results page. With the Google Search Appliance search results, users are able to pivot on related topics, which can help greatly with findability.

Figure 1: Current system for integrating the agency taxonomy within the agency portal environment.

What this means is that users have a disconnected search experience across the Agency environments. For the Agency portal, they view and use search results from the Google Search Appliance that have been created using the Agency taxonomy, which has been fine tuned to meet user needs. For the SharePoint environment, the search results have not been fine tuned for users, and the quality of the results can suffer.

In the SharePoint environment today, search is not as aligned with user needs as it is within the Google Search Appliance environment. This 'siloing' effect can mean that users are not finding the information they need, or they have to search for information in two separate systems, using two separate ways of searching.

Potential state

In order to mitigate the siloing effect in the search systems in the current Agency portals, the proposed system for integrating the Agency taxonomy within the Agency SharePoint environment is outlined below (Figure 2).

1. The Agency taxonomy continues to be maintained in the taxonomy management tool.

2. The Google Search Appliance, using the taxonomy built in the taxonomy management tool, can be connected to both the Agency SharePoint environment and the Agency portal to provide the taxonomy-enhanced search results page using information indexed from both locations.
3. The Google Search Appliance enables faceted search results, allowing the user to pivot on topics and terms relevant to Agency employees, based on the hierarchies created in the taxonomy.
4. As the taxonomy is further built out and refined, the Google Search Appliance is continually updated to index content in the Agency portal, and provide more relevant and useful search results for Agency employees.
5. The SharePoint search system has been disabled for this scenario.

COSTS

This section will calculate all costs to develop and operate each alternative described, including both one-time and recurring costs.

Figure 2: Proposed system for integrating the agency taxonomy within the agency SharePoint environment.

Development costs

In order to integrate the agency taxonomy into the SharePoint environment, there will be costs incurred from planning to define how the taxonomy will be integrated into the SharePoint environment, what the overall design of the integration will be and the cost of building out the integration. Included in these costs will be expenditures for:

- Personnel (who is going to design the plan, who is going to do the technical implementation and testing and so on);
- Equipment (is there going to be a need for one or more new servers, will the integration affect other servers/systems and so on);
- Developer training (does a technical developer need to be trained on how to install/maintain the new software, does the developer need to train on how to integrate a new server and so on);
- Development tools (are there going to be new tools in addition to the new software integrations, are there going to be new tools needed for the servers and so on).

Define phase

In order to effectively integrate the agency taxonomy into the SharePoint environment,

a define phase of work should take place to determine how the taxonomy should be used.

This phase should include a study of how SharePoint search results are generated and displayed today, along with a usability study to define problem areas and overall user needs. This study would involve an agency team that can perform interviews with current SharePoint users, and the agency IT personnel who currently maintain the SharePoint environment.

Typical activities for this phase can take from 2–4 weeks in duration and involve 40–60 hours of effort from the taxonomy Subject Matter Expert, who would interview the users to determine specific needs for the search results page. Cost estimates for resourcing are \$75/hr with a total cost of approximately \$3000–\$4500 for this phase.

Design phase

During a design phase, plans for how the new software will be integrated into the both Google Search Appliance and the SharePoint environments should be created.

This design phase should include wireframes for the design of the search results page, with details about how to integrate the results on the

page from both the portal and SharePoint environments. It should also include a functional specification for how the taxonomy will be used to index the content in both environments. If there has already been a functional specification done for the Google Search Appliance search results integration, and/or if there are existing wireframes for the search results page, then those can be used to inform the new design document.

During the design phase, the software and all the necessary equipment should be acquired and prepared for the build phase.

A design phase can take from 2–6 weeks in duration and involve 60–120 hours of effort from the team. It typically needs the expertise from subject matter experts in taxonomy and IT. Cost estimates for resourcing are \$75/hr with a total cost of approximately \$4500–\$9000 for this phase.

Build phase

The build phase of the project would include setting up any necessary hardware, such as a new server, and installing and integrating the software to extend the taxonomy management tool into the SharePoint environment. The build phase typically includes installation, based on requirements from the design phase, in addition to testing the new design with end users. The end users can, but don't have to be, those users interviewed during the define phase.

A build phase can be quick, anywhere from 1–3 weeks, depending on the availability of the hardware and software, along with the IT personnel's availability and the availability of the systems, such as the SharePoint environment. The build phase would typically be 40–60 hours of effort, for the IT personnel to install and test. Cost estimates for resourcing are \$75/hr with a total cost of approximately \$3000–\$4500 for this phase.

Total cost for the Development Phase of a project like this could run \$10 000–\$17 000.

Operational costs

In order to integrate the agency taxonomy into the SharePoint environment, there needs to be estimation on the installation, operation and maintenance costs of the system. Included in

these costs could be personnel, equipment, operational site upgrades or changes and staff training.

Non-recurring costs

Below are the projected non-recurring costs of the solution. These costs would be a one-time investment for the agency to implement the proposed solution.

Capital investment costs

For this solution to be implemented there would need to be the following investment in equipment:

- Additional server to accommodate the extension of the taxonomy tool with the Google Search Appliance to the SharePoint environment (see Table 1 for estimated costs).
- Installation of server and other equipment to accommodate the upgrade.
- There is no anticipated need for upgrading the database system(s).

For the server install, once equipment has been received, it may take 2–4 weeks in duration and involve 20–30 hours of effort from an IT resource. Cost estimates for resourcing are \$75/hr with a total cost of approximately \$2500–\$3500 for this phase.

Other non-recurring costs

Below are the other, non-capital investment projected non-recurring costs of the solution. These costs would be a one-time investment for the agency to implement the proposed solution.

- *Software license purchase:* The taxonomy management software vendor determined that for this implementation, there would be a cost for a four CPU core configuration is \$20 000 per CPU. Total license cost is \$80 000.
- *Install and implement the software:* The taxonomy management software vendor determined that for this implementation, there would be a cost for services to implement the software upgrade of around 15 days at a total cost of \$24 000.
- *Training, travel and other personnel-related costs of development and installation (except salaries and fringe benefits):* The taxonomy management software vendor determined that for this

implementation, there would be a cost for around 15 days of training and travel at a total cost of \$24 000.

- *Internal agency direct support services:* The taxonomy management software vendor determined that for this implementation, there needs to be a resource from the agency to assist with the software implementation. Estimates for this resource are 40 hours of effort at \$75/hour, for a total cost of \$3000.

Recurring costs

Below are the recurring costs projected for the solution (summary in Table 1). These costs would be an ongoing investment for the agency to maintain the proposed solution.

- *Software fees, in-house maintenance:* The taxonomy management software vendor determined that for this implementation, there would need to be support and maintenance fees set at 20 per cent per year, for an ongoing cost of \$16 000 per year.
- *Direct support services:* There will need to be ongoing support from the agency IT group to ensure that the system is working and is still delivering relevant results. This could include the cost of working on the SharePoint

environment. This could also include updates to the Google Search Appliance system to accommodate the new content from the SharePoint environment. This will require one IT subject matter expert at 120 hours per year, for a cost of \$9000.

BENEFITS

This section will calculate the potential benefits to develop and operate each system described, including non-recurring benefits, cost reductions and non-quantifiable benefits.

Non-recurring benefits

Below are the potential one-time, non-recurring benefits to implementing the system.

Cost reduction

Implementing the proposed improvements to the system can help the agency realize non-recurring cost reductions resulting from improved system operations, such as

- reduction of resource requirements (workers can get more done in less time, reducing the need for more headcount);
- improved storage and retrieval techniques (workers can produce better work products, based on the fact that they found more relevant, vital information faster);
- improved resource utilization (workers can spend less time helping others find information and more time working on their own responsibilities);
- reduced error rates (workers will have less exposure to outdated or incorrect material).

Overall cost reduction can occur due to users taking less time to find the information they need, and less time recreating the information because they couldn't find it in the first place.

Since the cost to the Agency for each worker to search for information is roughly \$16 900 per year, and workers report only a 57 per cent success rate of finding the right information on the first search, we can assume that even an increase in search success rates (due to improved search results) of 25 per cent could have a cost benefit to the Agency of:

- On average, \$7200 per year is wasted on workers not finding the information they

Table 1: Costs summary table

Description	Cost (\$)
<i>Development</i>	
Define Phase	3000–9000
Design Phase	4500–9000
Build Phase	3000–4500
<i>Capital investment costs</i>	
Server Capital	2500–3500
Server Installation	1500–2250
<i>Non-recurring</i>	
Software	80 000
Reviews	24 000
Training	24 000
IT small and medium services	3000
<i>Recurring</i>	
License	16 000
Support	9000
Total one-time cost to implement solution	145 000–153 750
Total cost per year to maintain solution	25 000

need (cost of information searching per year $\times$ 43 per cent [unsuccessful searches]). This is the one measureable benefit that came out of this study.

- As most workers surveyed do not create, but consume information (around 30 per cent create versus 70 per cent who consume or consume/create) there should be cost savings associated with reduced time to create information.

Other

There are potential other benefits, to implementing the new search systems, such as increased adoption of portal search. Since there appears to no longer be any question about which system is better for search, workers may tend to feel more confident of the search system because they know that most of the content in the agency repositories are being searched. Most trust in the search systems and its results.

Recurring benefits

Below are potential monthly and/or quarterly recurring benefits of operating and maintaining the combined search system over the system life, including:

Personnel salaries and fringe benefits: As more content is indexed using the Google Search Appliance, more people are going to be willing to allow the Google Search Appliance to search their content. What this means is that as salaries and benefits rise, as they will inevitably do, the cost of not finding information will increase more than it does now. It's critical that the time to find information – critical information – constantly be reduced in order to control costs. As more information is created at the agency, it's going to be substantially harder to find what's needed. And what that means is that as salaries rise, it's going to cost even more to find that information.

More studies on the history of the Agency's salary and benefits package increase over time is needed to determine potential savings.

Direct support services (intra-agency services): Most IT groups do not want to be in charge of more systems than they already control. Training personnel on more technology systems can cost

an IT group a lot, both in time lost for training and the dollar cost of paying for the training. At agency, with an integrated approach to search (combining the portal and SharePoint content under one search system), the IT group will not have to maintain an ongoing search system for both SharePoint and the portal. They can spend time getting smarter on Google Search Appliance and not have to spend time and money on learning SharePoint search.

More studies on how much the Agency pays for support services is needed to determine potential savings.

Travel and training: No need to train people on keeping up with SharePoint search, especially as it relates to learning and using SharePoint 2010 search.

Overhead (Including overhead benefits that represent additional or incremental expenses attributable to not implementing the system.): There will be no need to maintain the equipment necessary for two search systems.

This figure could be calculated in a separate study.

Cost avoidance (This could describe the avoidance of future costs that would be incurred if the best alternative were chosen from a set of alternatives, compared to maintaining current operations.): As the agency SharePoint site evolves over time, eventually migrating to the SharePoint 2010 platform from its current 2007 Microsoft Office SharePoint (MOSS) platform, there will probably be tighter integration into search using the adapter from SharePoint to Google Search Appliance.

If there is a decision later to utilize the taxonomy management capabilities within SharePoint, the transition from the Google Search Appliance integration into SharePoint will be relatively low cost.

However, given that the search capabilities in the Google Search Appliance would have been already integrated into the agency portal search experience, the recommendation then might simply be to stay in the Google Search Appliance environment rather than transitioning into the SharePoint search environment.

Non-quantifiable benefits

What is described below are the benefits that cannot be quantified in terms of direct dollar values (for example, improved service, reduced risk of incorrect processing, improved information handling, enhanced organizational image) for the alternative agency search system.

Best- and worst-case scenarios: Search is often mentioned as a critical component of an online experience, whether it's a portal, an Internet site or another variation. For the agency SharePoint portal, one critical component of its eventual success could be an improved search experience.

Given that, the options for a satisfying search experience for users comes down to two basic choices: integrate with the Google Search Appliance to create a hybrid search experience that incorporates searching the main content repositories (portal and SharePoint) or spend the time and money on maintaining two separate search experiences (the Google Search Appliance in the portal and the SharePoint search experience).

The overarching theme that runs through a discussion like that is who wants to maintain two parallel search systems and who wants to have to search in two separate systems to look for information? The costs associated with maintaining two systems is not part of this analysis, but it's fairly clear that the cost of doing that will be at least 50 per cent higher, and probably much more, if two systems have to be monitored and maintained. What's more critical is really the appearance of keeping up two separate systems: users will probably trust neither one to deliver what they really want. Instead, having a unified search system that traverses all content repositories tends to build trust. If a user can't find something, at least she or he knows that most repositories were searched. A failed search experience tends to dissuade users from using search in the future, but a unified search experience, even if it has its shortcomings, will be perceived as valuable.

More studies can be done on this once a baseline of user preferences, page usage and cost per page is determined.

Trade-offs with tangible benefits: One of the key trade-offs for implementing this solution will be

the abandonment of SharePoint search. This choice has to be one made based on solid business reasons, rather than just the preference of a group of people who might favor the Google Search Appliance over SharePoint search.

As the agency moves forward with building out the SharePoint environment over the next few years, abandoning the native search system could be troublesome if the supporters of the Google Search Appliance leave the Agency or are no longer in charge of building/maintaining the Google Search Appliance. There is a risk that having to build out a proper search system in SharePoint 2007 or later versions could be costly. In addition, there is a trade-off of depending on one search system over another.

No one search system is perfect, and there are drawbacks to using the Google Search Appliance. However, the mitigating factor is the agency taxonomy. It can remain outside of the debate over which search system is better, and still integrate into either one.

RECOMMENDATIONS

Upon completion of the study, recommendations were made to the Agency as to the cost and benefit of implementing the solution.

Those recommendations are confidential and cannot be reproduced for this article. However, the recommendation did take into account the overall cost of workers looking for information (\$16 990 per year, per worker), the total number of workers at the agency would potentially be exposed to the solution (approximately 3000), the total initial cost of implementing the solution (approximately \$150 000) and the annual cost of continuing the solution (approximately \$25 000 per year) and how long it would take to recoup the costs and start deriving benefits.

The Agency does need to do more studies about secondary costs to maintaining a search system, and how to measure the ongoing costs against what tangible and intangible benefits it's receiving from the improved search capabilities on the integrated portal.

REFERENCES

- 1 IDC. (2009) Hidden Costs of Information Work: A Progress Report. Doc # 217936, May.
- 2 Seattle Post-Intelligencer Online (2010) Bing's search-success rate is rising, Hitwise says, <http://blog.seattlepi.com/microsoft/archives/191460.asp>, accessed 15 July 2010.
- 3 IDC. (2010) The hidden costs of information work, <http://www.scribd.com/doc/6138369/Whitepaper-IDC-Hidden-Costs-0405>, accessed 12 July 2010.
- 4 US Office of Personnel Management (2010) US Office of Personnel Management Salaries & Wages Salary Table 2009-HOU. http://www.opm.gov/flsa/oca/09tables/html/hou_h.asp, accessed 14 July 2010.
- 5 US Department of Labor Bureau of Labor Statistics. (2010) Employer costs for employee compensation news release, <http://www.bls.gov/news.release/ecec.nr0.htm>, accessed 18 July 2010.

APPENDIX

Question 1

How much of your day is spent simply looking for information you need in order to do your job? (For example 'I spend 30 per cent of my day looking for information'.)

<i>Respondent no.</i>	<i>% of day looking for info</i>
1	0
2	20
3	25
4	50
5	25
6	25
7	30
8	0
9	0
10	90
11	10
12	20
13	35
14	20
15	10
16	5
17	85
18	50
19	50
20	1
21	20
22	10
23	25

Question 2

What search system(s) do you mainly use? (For example Google, Yahoo! SharePoint.)

<i>Respondent no.</i>	<i>Search system used</i>
1	NA
2	Google
3	Public Google, Google, SharePoint, Yahoo and Bing
4	Federated search
5	Google, SharePoint
6	Phone Directory
7	SharePoint and Remedy
8	NA
9	NA
10	Phone Directory, database
11	Internal Google
12	Google
13	Internal Google
14	Google
15	Search; my toolbar is Google
16	Inside search engine and Google
17	Google, Science direct
18	Google, Wikipedia, Outlook
19	ShareDrive
20	E-mail, Google
21	Google
22	Google
23	Google services – web search, Google desktop

Question 3

Do you use any other supplemental search systems? (For example 'I use Google, then SharePoint'.)

<i>Respondent no.</i>	<i>Supplemental search system</i>
1	NA
2	Google, internal Google, SharePoint
3	Public Google, Google Search
4	Google, Google.com, document Index System
5	Depends: Intranet: Google; target searches in SP, I use SP
6	No
7	Remedy, SharePoint, Google as last resort
8	NA
9	NA
10	No
11	Google
12	No
13	Internal Google, Homepage, SharePoint, external Google
14	No, start calling people
15	Only Google, sometimes Yahoo looking for different results
16	See previous question
17	Google
18	NA
19	Google
20	Don't use SharePoint
21	No, I start calling people
22	No
23	No

Question 4

How successful are you at finding the information you need the first time you look for it? All the time? Some of the time? Never? (For example 'I find what I need about 30 per cent of the time'.)

<i>Respondent no.</i>	<i>Success in finding</i>
1	NA
2	60% on first search
3	90%
4	<50% in Google
5	Most of the time, 60%
6	50%
7	Successful most of the time
8	NA
9	NA
10	About 90%
11	80–90%
12	80%
13	70%
14	5%
15	10% when searching X or Y
16	Can find most of the time 80%
17	70%
18	90%
19	40–50%
20	Most of the time
21	5%
22	70%
23	Most of the time 80%

Question 5

When you can't find information, do you attempt to re-create it? Or do you ask people around you? Or ask people to send it to you via e-mail?

<i>Respondent no.</i>	<i>Steps taken if you cannot find info (recreate, ask people around, ask people to send via e-mail)</i>
1	NA
2	If I know the process and can't find a published version, but want to give the end user a written process, I re-create it
3	All of the above
4	Contact people by phone or e-mail
5	Ask people, usually get pointed to a link
6	Phone to verify information found online
7	Ask people around
8	NA
9	NA
10	Phone and e-mail are the primary means of collection info
11	Ask more educated programmers for their ideas or direction via internal IM client
12	Ask people
13	If and when above resources have been exhausted, go to others and usually will keep documented in an e-mail.
14	Ask other by e-mail/call
15	Try different searches; ask people via e-mail then call
16	Ask co-workers to send it to me
17	Ask people, co-workers, other researchers, co-interns, technical people
18	If responsible for content, recreate. If not, ask task owner for help.
19	E-mail, phone

<i>Respondent no.</i>	<i>Steps taken if you cannot find info (recreate, ask people around, ask people to send via e-mail)</i>
20	E-mail, phone
21	I normally assume the information I'm looking for will not be on Google (or any searchable engine, so I e-mail/call other people to locate the data
22	Ask people, phone or e-mail
23	Coworkers, either person or via e-mail

Question 6

Is your job more focused on creating information, consuming information or both? What percentage of each?

<i>Respondent no.</i>	<i>Job more focused on creating info, consuming info or both</i>	<i>Create</i>	<i>Consume</i>	<i>Both</i>
1	NA	—	—	—
2	Both (50/50)	—	—	x
3	70% consuming info/30% creating it	—	x	—
4	Over 60% research information	x	—	—
5	Both: 60% creating; 40% consuming	x	—	—
6	25% creating; 75% verifying info	—	x	—
7	Both; 40% creating; 60% consuming	—	x	—
8	100% consuming	—	x	—
9	NA	—	—	—
10	90% consuming; 10% creating	—	x	—
11	Consume 10–20% of the time	x	—	—
12	Creating 80%; Consume 20%;	x	—	—
13	Creating 40%; Consuming 60%	—	x	—
14	Both; 20% creating; 80% consuming	—	x	—
15	30% creating; 70% consuming	—	x	—
16	NA	—	—	—
17	100% consuming	—	x	—
18	Both 50/50	—	—	x
19	50/50	—	—	x
20	75% create 25% consume	x	—	—
21	20% Creating, 80% Consuming	—	x	—
22	Consuming 90%	—	x	—
23	Creating 80%; Consume 20%;	x	—	—

Question 7

Although we are not recording each individual's salary, we need to know a salary range for you or your peers to determine a more accurate cost for finding information. Which of these salary ranges does someone in your peer group fall into? <\$50 000, \$50 000–\$75 000, \$75 000–\$100 000, \$100 000–\$125 000, >\$125 000 (exclusive of benefits, bonuses and so on).

Respondent no.	Individual salary	<50	50–75	100–125	125+
1	<50K	x	—	—	—
2	<50K (exclusive of benefits, bonuses and so on)	x	—	—	—
3	50K–75K	—	x	—	—
4	—	—	—	—	—
5	—	—	—	—	—
6	<50K	x	—	—	—
7	50K–75K	—	x	—	—
8	—	—	—	—	—
9	—	x	—	—	—
10	<50K	x	—	—	—
11	<50K	x	—	—	—
12	< 50K	x	—	—	—
13	<50K	x	—	—	—
14	100K–125K	—	—	x	—
15	100K–125K	—	—	x	—
16	<50K	x	—	—	—
17	<50K	x	—	—	—
18	50–75K	—	x	—	—
19	50–75K	—	x	—	—
20	> 125K	—	—	—	x
21	100–125K	—	—	x	—
22	<50K	x	—	—	—
23	50K–75K	—	x	—	—