



# Developing an enterprise search strategy

Intranet Focus Ltd Research Note 02/13

February 2013

Intranet Focus Ltd  
12 Allcard Close  
Horsham RH12 5AJ, UK  
+44 1403 267030  
[www.intranetfocus.com](http://www.intranetfocus.com)

# Developing an enterprise search strategy

## Summary

A number of surveys have indicated that few organisations have an enterprise search strategy. The high profile of Big Data and the availability of a high-performance search application within SharePoint 2013 will both require consideration to be given to an integrated approach to enterprise search. In this Research Note a framework is presented for an enterprise search strategy that includes business objectives, user requirements, content (based around in information life-cycle management model), technology, governance and management.

## Contents

1. Introduction	3
2. Business objectives	3
3. User requirements	5
4. Stakeholders	5
5. Governance	5
6. Content	6
7. Big Data	7
8. Metadata and taxonomies	7
9. Search application audit	7
10. Infrastructure	8
11. Federated search	8
12. Disaster recovery	9
13. Security	9
14. Help Desk	9
15. Usability and accessibility	9
16. Training and support	10
17. Search support team	10
18. Performance	10
19. Risks	11
20. Communications strategy	11
21. In conclusion.	11

## Research Notes

*This is one in a series of Research Notes published by Intranet Focus Ltd. since January 2012. For further information see <http://www.intranetfocus.com/resources/downloads>.*

# Developing an enterprise search strategy

## 1. Introduction

Given the potential benefits and challenges of enterprise search it is surprising that the 2012 Findwise Enterprise Search and Findability survey indicated that only 14% of respondents had a search strategy, though 30% were planning to develop a strategy in 2012/2013. This result is consistent with the Digital Workplace Trends report from NetStrategyJMC and tends to support the view that search is not seen as a business-critical element.

Search does need to be planned. It is technically challenging, users have high expectations and a high dependency on the success of search and there is going to need to be a substantial investment in personnel for the search support team. It is one of very few enterprise applications that probably everyone in the organisation will use weekly. Enterprise search also bumps into many business operations. The search engine will need to interface with other applications and there are some legal and compliance issues. In the future the boundaries between search, business intelligence and content analytics are going to become increasingly blurred and delivering access to enterprise search through mobile devices is going to be essential within a year or so.

This Research Note sets out a framework for a search strategy but of course the details will be specific to each organisation. It is based on 'Enterprise Search – enhancing business performance' by Martin White, Managing Director of Intranet Focus Ltd. It is published by O'Reilly Media.

## Business issues

### 2. Business objectives

It is important to set out the objectives of the organization as these could have a major influence on the way that search develops. The acquisition strategy is especially important as this could require the search application to index substantial new repositories at short notice and also result in the negotiation of new license deals with a number of different vendors. There could be challenging divergences in metadata values and consistency.

It is not possible to write a search strategy in terms of the return on investment that it will bring. Rarely is this done for any other application and in the case of search the problems of doing so are insuperable to the following reasons

- Search is not a workflow application. There is no underlying process (such as creating a new supplier record) where there may be some quantifiable benefits from reducing the number of process elements or the time taken for some of these elements.
- Time saved is not a metric of success. Outside of some professional services organisations no one maintains time sheets at a level that will show that good search will save a measurable number of hours.
- Information is not the output. Just finding "x results" does not help if they do not help the employee make a more informed decision. Frustration with search quality is often because of a lack of confidence in the quality of the content, for which the search implementation gets blamed
- The business could change in a matter of weeks. A reorganisation, takeover or acquisition could radically change the requirements for information.

There are only two reliable ways of building a business case for internal enterprise search. The first is to look analytically at the business risks and see how an effective search application could reduce these

## Developing an enterprise search strategy

risks. This can result in some form of quantitative measure because the risk scores will be reduced.

The second is an historic lost opportunity analysis. A group of senior customer-facing managers will soon come up with a set of lost opportunities or poorly executed projects/products which arose because the organisation could not find the information it needed at the time. Only later did someone say “But we already knew that our product would not work well under those conditions. Didn’t you find the engineer’s reports?” Summing the value of the missed opportunities and taking just 10% of the figure will still be substantially more than the potential investment in enhancing the search application.

Nothing carries more weight than a story from a respected manager about how they failed to find information that could have made a positive impact on the organization. It’s a trick used by many management authors, who use call-out stories to make an impact in an otherwise mundane book on some aspect of business operations. Beginning the business case document with a really strong search success or search failure (or both!) is a guarantee that readers of the business plan will already be pre-disposed to agree to the investment.

In 2011 Hofmann Le Roche, one of the world’s leading pharmaceutical companies, identified five simple but challenging questions that employees were probably asking themselves and colleagues on a regular basis.

- Can I handle this?
- What is the implication?
- Can we find out sooner?
- Will it work?
- Have I chosen wisely?

If enterprise search can provide answers to those questions then the business case can be made on just a few pieces of paper.

The 2012 Findwise survey found that the top ten justifications for enterprise search implementation were the following, based on a summation of respondents ranking these reasons as ‘imperative’ or ‘significant’.

The list is in decreasing order of importance.

- Accelerate retrieval of information from known information sources
- Improving the re-use of information and knowledge
- Increasing the extent of collaboration through finding people with relevant expertise
- Eliminating information
- Accelerating the speed of finding people both by name and expertise
- Raising the awareness of what was already known
- Eliminate duplication of work because relevant information could not be found
- Improving the consistency and quality of response to queries from customers and partners
- Creating a more personalized intranet solution
- Providing support for compliance management

Unfortunately none of those are easy to quantify.

However it is important not to restrict the scope of ‘enterprise search’ to internal information resources. Web site search also needs to be taken into consideration as the expectations of site visitors will be shaped by Google and Bing. A search on a major consumer products company for information about

## Developing an enterprise search strategy

one of their major brands has two documents duplicated in the first ten results, and three have the document heading 'Microsoft Word. Document 1.'

### 3. User requirements

Determining user requirements is very important with search applications. It is a subject in its own right so in this document only the importance of building search personas and use cases for search is highlighted. The four most common use cases are

- I need to find a specific document
- I need to find someone with specific expertise
- I need to become an expert on this subject very quickly
- I need to find out everything we know on this subject

Each presents a challenging set of issues for search managers but that is not an excuse to setting out the core use cases in the search strategy. Within an organisation there is considerable scope to undertake user surveys but for web sites site logs and search logs will set out user actions, which may not be the same as user requirements.

### 4. Stakeholders

In 2012 we published a Research Note on Intranet Stakeholder Management which set out a methodology for identifying and managing stakeholder interests. The same principles apply to search stakeholders and so are not repeated here.

### 5. Governance

If building a search team is difficult finding someone who will take business responsibility can be even harder. This is probably because there are no business and compliance-critical workflow processes that are supported by enterprise search. Look around at the main enterprise systems and they are owned by the manager responsible for the workflow; Sales Director, Operations Director, HR Director and so on. In many organizations search is owned by Corporate Communications, almost certainly when the same search application is being used for the web site and for internal enterprise search.

All too often it seems that IT departments work on the basis that search is not interesting, especially if it is a commercial rather than an open-source application. They cannot write code for it, they cannot customise it, and providing the expertise on the Help Desk to support employees who cannot find documents they know must exist are all good reasons for opting out of managing the application. On the other hand it does use servers and networks for which IT is responsible, and so perhaps IT should manage search.

It will be interesting to watch Big Data governance structures evolve. Technically these are more interesting to IT but just as in the case of search (in the narrow sense of searching unstructured information) the benefits are totally with the business.

In an ideal world search should report to the senior manager whose performance bonus is based on meeting customer requirements, either through product development and delivery or service development and delivery. This could be a General Manager, or Director of Manufacturing. All that the search owner really needs to do is fight for a sensible capital and operating budget.

Search certainly needs a Steering Group that can represent the interests of users, especially in areas

where search is likely to be business-critical.

### Content Management

#### 6. Content

A search engine needs to be instructed about the content that needs to be indexed. The place to start is a content audit based around the repositories of information that the organization holds. The content audit needs to cover the following elements for each repository and application.

The first, and often most difficult, step is to find out who owns the repository. It may have been set up some time ago and the initial owner might even have left the company. If there is a current owner the chances are that the original intention of the repository has long since been overridden. This is often the case with a departmental repository where the department has been merged or fragmented over time.

A brief description of the content should be prepared, along with a description of the user categories that will contribute and use the information in the repository. The total file size and total number of documents are important to know when it comes to sizing the search application. For the same reason the rate of addition of documents by time will give an indication of how frequently the repository needs to be re-indexed, or whether the documents are such that they need to be indexed as soon as they are added to the repository.

All the file extensions should be identified and listed out and at a minimum the maximum file size of the collection should be ascertained.

Making the assumption that all the content is in a single language and that the language is English is only a safe assumption in a very few countries of the world. It could be that the French version of a document has been stored alongside the English version in what otherwise looks like a totally English-language collection.

Overall search quality is highly dependent on content quality, as the example quoted in Section 2 above illustrates.

Ideally content management and search both need to be set within the context of an information management lifecycle approach. An information life cycle provides a framework for ensuring that information is managed as an asset within The Company

**a. Create** - This is the process of creating documents in a way that enables the document to progress through the stages of the information lifecycle.

**b. Store** - Employees know where to store documents so that related documents are stored together to support browsing through collections and to support security and disaster recovery management

**c. Discover** - Information can be found by searching through repositories, browsing through folder structures and intranet navigation and through alerting services such as wikis and blogs.

**d. Use** - Documents can be used only subject to rules on confidentiality, security and personal privacy.

**e. Share** - Documents will be shared internally, with third parties and with the public. Users of these document have to be confident that the information they contain can be trusted to be reliable, and that if needed the documents are available in more than one language.

## Developing an enterprise search strategy

**f. Review** -Processes need to be agreed for undertaking the review process and if needed creating a new version of the document to maintain its value and validity

**g. Record** -Details of retention periods need to be agreed which take into account legal and regulatory requirements, and product and service lifetimes.

**h. Dispose** - Disposal is the final stage of the information lifecycle and it is the point the document has no further value to The Company and can be deleted from all systems without any risk to the future integrity of the company.

### 7. Big Data

Distinguishing between the hype and reality of Big Data remains a challenge. We will be publishing a Research Note on Big Data later this year. The corporate position of Big Data has to be included in an enterprise search strategy because the value of both will be substantially enhanced by application and content integration. An AIIM report last year indicated that 59% of respondents regarded this as important for their organisations but that only 2% were able to do so.

### 8. Metadata and taxonomies

A few years ago at a search conference there was a presentation about a new search implementation. The search manager reported that one of the tests that had been run during the implementation was to find members of staff called Jane. To everyone's surprise most of the high relevance results were to male employees. It turned out that they had all written and submitted their cvs on a template owned by someone called Jane, and the search engine was placing more value on this metadata item than on the name field.

The problem with metadata is that the content contributor has to add it, and does so either with reluctance, or without due care or a combination of both conditions. This section of the search strategy needs to highlight the importance of metadata and how it will be generated, either automatically (e.g. the name of the content contributor from the system log-in information) or through manual addition. Entity extraction is a half-way house.

All the evidence points to the benefit of a taxonomy and metadata enhancing search performance, and especially in presenting highly-relevant information. However taxonomies are time-consuming to compile and to maintain. As with so many search-related issues a balance needs to be established between the value of the taxonomy and the benefit to users, taking into account that the users of the information may not be the people who have to add the taxonomy metadata in the course of saving the document.

## Technology

### 9. Search application audit

An important component of the technology section is to provide a list of current applications that already have search functionality, possibly as an embedded application. Examples might include document management and records management systems and enterprise resource planning systems. There are often more of these embedded search applications than most managers appreciate. These are often optimized for a specific application, and probably an enterprise application could not provide the same level of search performance and satisfaction. Having a list of these applications enables decisions to be made about whether there would be a benefit in providing a federated search environment.

## Developing an enterprise search strategy

There are a number of IT issues that need to be surfaced in a search business plan, and these include

- The use of open-source software
- Mobile access to enterprise information assets
- The adoption of cloud/software-as-a-service applications
- In-house versus external development and maintenance

In most organizations there will already be a number of search applications in use. It is easy to suggest that having just one powerful engine will solve all current search problems at a stroke. It could easily add to them. An important section of any business plan should set out how the process of migration from a number of different search applications to one single enterprise search application is going to be accomplished. This is not just a technology issue, but has to be approached both from an IT and a user perspective. Without doubt there will be some considerable change management, training and support issues that will have to be addressed and solutions put in place long before the technical migration occurs and for some time afterwards.

As with any software application a search vendor will release versions of the software to either address bugs or to provide additional functionality. This section of the business plan should set out the basis for considering whether to implement a new version of the software, bearing in mind that there could be risks with connectors to other applications.

During 2013 and 2014 perhaps the major technology challenge for many organisations is whether, and how, to implement SharePoint 2013 search. The improved search performance could be one of the reasons for accelerating the migration from SharePoint 2010 to SharePoint 2013, or for migrating from SharePoint 2007. The benefits and challenges are beyond the scope of this Research Note but need very careful consideration, especially in organisations with no search support team and using either in-house or contracted development resources who have not had previous experience of high-end SharePoint search applications, either FAST Search Server for SharePoint 2010 or a third-party application.

### 10. Infrastructure

Enterprise search can have some challenging infrastructure requirements as far as storage in particular is concerned. The topology of a large-scale enterprise search application with good disaster recovery and the minimum latency on queries will need careful planning. The issues will not just be about the size of the index relative to the size of the repository but also the write speeds of the disk arrays. Many capacity planning specialists may be in unknown territory when it comes to planning search capacity.

Network bandwidth to distant but still important offices can also present issues that need careful review. Substantial files could be downloaded very rapidly for perhaps 10-15 minutes as a user works their way through the top 50 results looking for a specific piece of information. This can be a particular issue with PowerPoint files, and a number of vendors offer a feature to render the document, or PowerPoint file, as a small HTML thumbnail image.

### 11. Federated search

Many organisations regard the achievement of a fully federated search environment with a single search box for all repositories and applications as a core objective. These organisations fail to recognise the technical and user interface challenges of delivering a fully-federated search. A search strategy should set out a set of scenarios for federated search, as there are likely to be many of them. These scenarios should be based on a good understanding of user requirements for search that reflect tasks and not information. One reason for this is the emergence of many excellent approaches to search-based

## Developing an enterprise search strategy

applications that support task completion.

### 12. Disaster recovery

It is tempting to put search down the bottom of the disaster recovery priority list but arguably it should be right at the top. It may enable the organization to keep going while other applications are brought back to life. After all the search index will contain a copy of most, if not all, of the information that the organization possesses, and if the application can provide users with an HTML thumbnail of a document that could be more than enough for business-as-usual to continue.

A disaster recovery plan usually sets out a Recovery Time Objective (RTO) defining the maximum application downtime and a Recovery Point Objective (RPO) noting an acceptable restore point. For an enterprise search application these need to be considered from basics rather than the blind adoption of objectives from other enterprise applications. It is not just a case of getting the application back and running from a user perspective but understanding and accounting for content that may not have been completely crawled or an index that has not been correctly updated. The index itself may have been distributed around the world, with a concomitant the need to re-synchronize the indexing process.

### 13. Security

Organizations are rightly very concerned about the risks from employees finding information that they are not entitled to see. Even if there is a corporate security policy the questions that should be asked is whether it is granular enough and implemented rigorously enough to ensure that ACLs can be created and maintained. The potential impact on staff and the reputation of the organization from a failure of the security policy could be very dramatic.

This is especially the case if the document being created is being indexed as soon as it is saved to a repository. At that moment in time there is in effect a duplicate of the content accessible to anyone with the correct security permissions. The index will almost certainly be backed up on a second server for disaster recovery purposes. Removing the document from the repository will almost certainly not remove the content from the index. If the document was a list of senior executive salaries then a search for this information might well disclose the amounts even if the document has been deleted.

## Management

### 14. Help Desk

A search application needs its own help desk, even if it is a virtual one, and there needs to be a Service Level Agreement both ways between the IT and Search Help Desks because it may take quite a bit of effort to work out what is causing the problem and what actions should be taken to remedy the problem.

### 15. Usability and accessibility

Despite the high profile efforts of usability experts such as Jakob Nielsen few organizations seem to take usability seriously. Search usability testing is especially important because of the complexity of many search user interfaces with a profusion of filters, facets, annotations to results and perhaps even graphical representations of clusters of search results.

Most web sites (and probably fewer intranets) are now compliant with the Web Accessibility Initiative accessibility guidelines. Search implementations seem to be overlooked and accessibility evaluations of

## Developing an enterprise search strategy

the search user interface should be included in the test regime.

### 16. Training and support

The view is sometimes taken that search should be so intuitive that there should be no need to providing training and support. This view is often based on the 'simplicity' of the Google search box, a view that ignores that books have been written on the very wide range of hacks that are available to users of Google search.

The same is true of an enterprise search application. Certainly there should be as few barriers as possible to carrying out a basic search, but in an enterprise context there are probably very few basic searches as finding most, and ideally all, of the relevant information is very important.

### 17. Search support team

Technology can be bought but a search support team needs to be the sort of people who are in very short supply inside most companies. Most of Chapter 10 is about the skills needed in a search team. It would not be too much of an overstatement to say that if you cannot find the people who will form the search support team there is really no point in making any investment in an enterprise search application.

To summarize our 2012 Research Note there are five search team roles

- Search Manager taking management responsibility for search delivery
- Search Technology Manager, looking after the IT elements
- Search Analytics Manager, running and analyzing search logs
- Search Information Specialist, with responsibility for search quality
- Search Support Manager, providing training and user support

In the initial stages of an enterprise search project these roles could be undertaken alongside other work but once the implementation begins these roles need to be filled on a full time basis. There simply is no option. If senior managers say that this is a ridiculous number of people to be supporting a single application ask how many staff support the ERP, business intelligence and document management applications. None of these are used by almost every member of staff every day.

### 18. Performance

There are many ways of measuring search performance. User-centric measurements are set out in Chapter 10 but and certainly need to be summarized in the business plan. In addition there are also some system performance measures that need to be set out. These typically include

- The optimum freshness for all categories of content
- Crawl times and crawl frequency
- The rate at which content is going to have to be ingested in order to achieve the desired freshness
- The latency between ingestion and the index being updated
- The amount of temporary disk space used in the process of creating and updating the index
- Elapsed indexing time for a new content set
- Indexing processor time, which excludes speed gains from parallel processing
- The expected number of queries-per-second
- Desired response times

## Developing an enterprise search strategy

In cases where the enterprise search application is also going to be used for site search on the corporate web sites the internal performance and external (public) performance metrics may be quite different.

When desired response times are being established a number of special cases need to be taken into consideration. Federated search will almost certainly be slower than searching on a single repository or application and the way in which security management has been implemented will also have an impact on search response times. The time taken to display a results set on the desktop is only a small element of the search process, and users will quickly become dissatisfied with any response time to open up a document in excess of perhaps 10 seconds. This latency time is not just absolute but relative. If one search provides a user with response times of only a few seconds and the next search takes perhaps 30 seconds to call up a document the user reaction will be one of considerable disappointment even if in absolute terms the search application is working to the maximum of its technical performance.

### 19. Risks

It is always advisable to have a risk management strategy for enterprise search, and this probably needs to cover off the following risks

- Lack of resources in the search team leading to poor search performance
- Search manager leaves and there is no internal candidate
- Search vendor is acquired or goes out of business
- Search vendor unable to provide an adequate level of support
- No clear roadmap for development
- Changes in senior management at the vendor result in a repositioning of the search engine
- Inadequate security management leads to a breach of access permissions
- Key development skills in the open-source contractor are not available
- Disaster recovery procedures prove to be inadequate
- Enterprise networks are giving rise to significant performance problems
- Poor performance of connectors and APIs
- Best bets are no longer best bets

### 20. Communications strategy

It is not uncommon for the search team to be invisible, often because there is no search team. Enterprise search applications are business-critical and almost always used when other approaches to information discovery have failed. Yet feedback channels are usually non-existent. Ensuring that users are aware of changes to search applications, search scopes and user interface developments is an essential element of enterprise search management.

### 21. In conclusion.

This set of topics is by no means fully comprehensive. For example we have not had space to touch on compliance and legal issues. Our objective is to show the range of issues that organisations need to consider in managing enterprise search, and Big Data, applications and in doing so emphasise the need to take a strategic view of search aligned to the needs of the organisation, its customers and stakeholders, and the information discovery requirements of all employees.

For further information on enterprise search strategies:

Enterprise Search – enhancing business objectives. Martin White, 2012. O'Reilly Media