

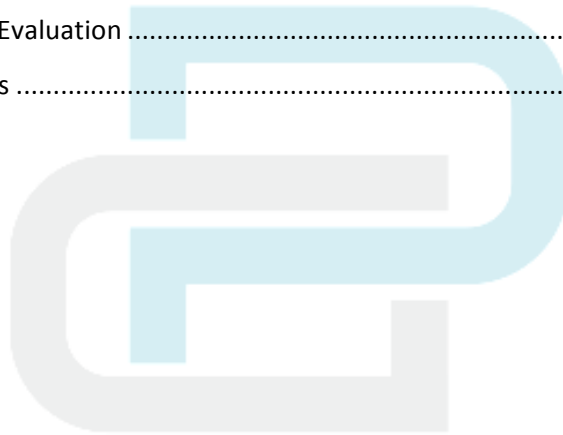


PG public services, LLC

**Desktop Enterprise Search Tools:
Comparison and Recommendations for
Federal Agencies**

Table of Contents

1. Introduction	3
2. Assumptions and Definitions	3
2.1. Assumptions	3
2.2. Definitions	3
3. Business and Organizational Considerations	5
3.1. Business Functionality	5
3.2. Organizational Considerations	6
4. Technical and Security Challenges	6
5. Preliminary Product Evaluation	7
6. Suggested Next Steps	9



PG public services, LLC

1. Introduction

PGPS has learned from its Federal clients that there is a significant business need for desktop search capabilities. As one client's has recently stated

“Current search capabilities are time intensive and inadequate to meet organizations’ needs. Using the existing options available, personnel must conduct multiple searches of multiple sources of data (including multiple e-mail archives per user, shared drives, and desktop hard drives) and then sort through large amounts of irrelevant data to find the information required. Agencies require search software that will enable users to conduct a single search e-mail, e-mail attachments, shared drives, network drives, and hard drives, and to provide ranked search results to the user (similar to an internet search engine).”

Tools such as Google Desktop, Enterprise Edition for could be used to improve efficiencies in daily operations and Freedom of Information Act (FOIA) compliance activities. However, for this tool to be implemented on a small or large scale, there are a number of technical and security considerations that need to be addressed at a minimum by business owners, divisions of IT and the Office of Privacy.

Granting greater access to more of the right information in a more efficient manner for a majority of users necessitates the effort required to identify and quantify risks along with probable mitigation and monitoring strategies. A balance of business needs and acceptable risks will determine the extent to which search engine tools can be used.

The requirement for desktop search capability should be addressed within the context of longer-term planning for IT tools to better access the information in an agency's files, common files/applications utilized within their offices, and multiple agency-wide applications and databases that can be shared when appropriate.

2. Assumptions and Definitions

2.1. Assumptions

The tools being discussed in this paper are only for internal desktop and intranet search. Internet search engines which can be accessed through a federal intranet to the internet and have tools that address the need for searches by the public on the outside facing federal web sites are outside the scope of this discussion.

2.2. Definitions

Definitions and the business, technical and security considerations in the next sections are taken from “What’s the Right Search Engine” and “Quick Start – Enterprise Search” from the Burton Group.

To offer context for this discussion of enterprise search, the four major types of search are characterized by the location of the content: web search, enterprise search, application search, and desktop search.

Application Searches

Application search tools include Business Intelligence (BI), which provides historical, current or predictive views of operations. Security is managed by the operational application and the underlying file formats are controlled and understood by the primary users (after training). The major challenges are the limited scope of the search (it surfaces information only within the application) and an application-centric UI that can rarely be altered or a BI tool that needs to be applied as a bolt-on.

Enterprise Search

Enterprise search is based on an agency's official digital documents and is organized in a way that meets the agency's needs. Security is set up so that employees who do not have permission to see certain documents cannot read them or even discover their existence via search.

Desktop Enterprise Search

Desktop Search brings together the entire digital workspace under one user interface. When desktop enterprise search is implemented, there needs to be a monitored balance between enterprise search tools for multiple users and the use of BI by application for specific users. Keeping the BI environment separate from the operational environment prevents the operational systems from becoming bogged down in queries. However, BI typically retrieves structured data, not unstructured data and BI architectures can limit the ways to look at data in new and novel ways. BI typically uses database technology, which optimizes transaction updating rather than retrieval speed.

Web Searches

While Web search engines are generally outside the scope of this discussion, they provide users a very fast response rate, ease of use due to lack of security, and ranking of appropriate results. Web searches are enabled through a common digital format (HTML). Web search engines must provide ranking functionality and be scalable.

In general, desktop search needs to allow for search of all formats encountered by the user and have the interface tailored to that individual and business unit's needs.

3. Business and Organizational Considerations

3.1. Business Functionality

The basic functionality of enterprise search engines creates challenges for an agency which has requirements to control access to data at many levels of sensitivity. The Burton Group categorizes overall search engine parameters into User Requirements, Performance Needs and Content Considerations

User requirements include but are not limited to:

- Relevance – the delivery of pertinent results consistently.
- Ease of use – the user interface cannot be too simplistic or too complicated.
- Identity – the linking of a user to a role or group via the user's behavior or to control access based on role.
- Fit to task – the search engine allows the user to quickly perform the task.

Performance needs include:

- Speed – the speed at which result are returned. Web search users expect subsecond response time and may be unaware of the behind the scenes indexing task needed to control enterprise search.
- Scale - the number of documents and the expected number of concurrent users.
- Required Time Relevance – the timeliness of information required. Does the user need the most recently available information loaded and indexed?

Content and content access considerations are the most challenging for any organization to balance and include:

- Scope of Accessible information – which content must be indexed (and which can be ignored).
- Scope of Search – the ability to let the user restrict the search to a specific repository, set of documents, or attribute.
- Recall – the ability to retrieve as many relevant documents as possible,
- Precision - retrieving only relevant documents and nothing more.
- Security - enforce the necessary security, by ensuring that documents are found and are read-only by those with the appropriate permissions.

3.2. Organizational Considerations

The Federal implementation of this type of industry and consumer accepted software requires an “owner” to focus on documenting and achieving specific business needs while coordinating with other relevant offices such as divisions of the IT, Office of Privacy and acquisition.

All areas of IT will need to identify risks, mitigation plans and pathways for success. The Office of Acquisition will need to be engaged to determine the applicability of Federal Acquisition Regulations for software that may be either “free” or “included” in other licensing agreements. Total life cycle costs will need to be projected and possibly allocated to offices depending on policy and use.

Training in use of the desktop search tool and surrounding security considerations will be a priority for any agency. Users should be taught about the security differences (and impact on performance) between web search engines and a desktop search tool which access information from various applications across secure networks.

4. Technical and Security Challenges

The overall significant technical and security challenges associated with the implementation of an enterprise search tool reside at the compatibility at the desktop image level. The challenge is to balance the need to provide effective projected and on-going system workload analysis with the need to address overall security control parameters and monitoring requirements. The system workload and information assurance considerations are normally documented and addressed during the Systems Lifecycle Management process for a given project or program. For a desktop and or an enterprise search engine these would include but not be limited to:

- Individual data leakage – information intentionally or unintentionally passed from individual’s computer to entities outside the agency. The “re-storage” of relatively controlled data or indexes of available data creates a potential for security breaches that are much harder to control than data that is only secured within a system with specified controls.
- Application Programming Interfaces (APIs) may not be available on every application to allow for search.
- Operations and maintenance of the search engine and its components need to be planned and budgeted – “free” software will have some possibly significant costs such as hosting, administration, and security review.
- Tools will be needed to track search and overall system performance along with security related tools to monitor use and compliance

Compatibility relative to the desktop image must be considered. Anything considered for addition to all or a set of user workstations would need to be implemented via an agency’s

patching cycle. Therefore, consideration should be given to the approximate time-frames for use and roll-out of versions of an agency’s desktop image.

5. Preliminary Product Evaluation

For the purposes of this sample preliminary evaluation, we focused on the requested Google Desktop, Enterprise Edition, Microsoft Windows Desktop Search (Called Windows Search in Windows 7), and Yahoo Search (which is the X1 sets of tools) based on Gartner’s evaluation that these are the top three vendors with free or included desktop search offerings. The use of a stable desktop search tool that also allows for controlled searches across allowable and appropriate network shares should be explored.

The research for this preliminary evaluation was done through open sources without utilizing procurement methods and without vendor talks or demonstrations. The initial cost for the desktop and enterprise editions were assumed to be “free” or included with the understanding that there would be costs for hosting, administration, and security review.

The following table was adapted from a special study conducted by the State of North Dakota that considered the same three tools.

Requirement	Google Desktop Search	Microsoft Windows Desktop Search	Yahoo/X1 Desktop Search
Windows 2000/XP Compliant	Yes	Yes	Yes
Ability to configure and administer centrally (i.e. Group Policy)	Yes	Yes	Unk
Ability to apply Group Policy at a very detailed level (allows locking down search locations - users can’t add additional local or remote searches)	Yes	No	Unk
Ability to apply Group Policy in restricting remote searches but allows users to add local searches	No	Yes	Unk
Document Security remains enforced	Yes	Yes	Yes
Use a minimum amount of client resources	Yes	Yes	Yes
Ability to search local drives	Yes	Yes	Yes

Requirement	Google Desktop Search	Microsoft Windows Desktop Search	Yahoo/X1 Desktop Search
Ability to search network shares	Yes	Yes	Yes
Ability to search Outlook email (on Exchange Server)	Yes	Yes	No
Ability to search .pst files (locally stored Outlook archives)	Yes	Yes	Yes
Ability to search within attachments on e-mails	No	Yes	Yes
Ability to search Calendar and Contact information	Yes	Yes	Yes
Ability to search within attachments on Calendar entries	No	Yes	Yes
Ability to restricting to certain file types	Yes	Yes	Yes
Contain and embedded content viewer	Yes	Yes	Yes
“Free” to install for the enterprise	Yes	Yes	No

The above table is a summary of the vendor data available, shows significant similarities and highlights some important differences. Essentially it focuses on the security and control features offered at the enterprise level:

- Each have a variety of search features that the others do not – one size does not fit all
- Yahoo Desktop is really X1 and is not “free” at the enterprise level, it was eliminated, and, therefore, the remainder of the information for the important security features were not explored at this point
- Security is initiated through group policy – there are a variety of benefits and limitations to the way the Google and Windows products allow control

Both the Google and Windows products should be considered for the short and longer term needs of an agency’s users.

Note: Previously the Google product allowed a Search Across Computers feature. As of January 20, 2010 this was no longer available although some of the open source documentation still includes descriptions of the features.

6. Suggested Next Steps

- 1) An agency should consider exploring implementation of a pilot in the use of an enterprise search tool such as Google Desktop Enterprise Edition. SLM should be applied and tightly controlled. PGPS recommends a “no assumptions/no stone unturned – positive or negative” approach and a willingness to terminate if risks deemed too high and/or cannot be effectively tracked and mitigated.
- 2) An agency should fully document and understand the individual, office and organizational risks associated with this desktop enterprise tool. Lessons learned, especially those related to security should also be documented.
- 3) Upon preliminary and continued success of pilot use of and desktop enterprise search tool, an agency should consider agency wide use of the tool within the context of its larger enterprise architecture.
- 4) As technology continues to evolve, an agency should continue to research longer term cost/benefits/risks of available market desktop and desktop enterprise search applications including future versions of existing products.

PG public services, LLC

This white paper includes data and information that shall not be disclosed outside the government and shall not be duplicated, used, or disclosed—in whole or in part—for any purpose other than consideration of this white paper. In no event shall any part of this white paper be used in connection with the development of specifications or work statements with respect to any solicitation subject to full and open competition requirements. This restriction does not limit the government's right to use information contained in these data if they are obtained from another source without restriction.

PG Public Services, LLC

www.pgpublicservices.com

USA

1069 West Broad Street
Suite 724
Falls Church, VA 2204
Telephone: 877-350-8886