



In today's world, many organizations are information-intensive. These organizations have experienced challenges managing a growing collection of information, including tabular data, spatial data files, Computer-Aided Design and Drafting (CADD) files, presentations, spreadsheets, manuals, meeting notes, emails, reports, video, and images. In addition, large data streams from sensor and video feeds and Light Detection and Radar (LiDAR) point clouds have become part of the mix in recent years. As information volume and diversity grow, an organization's employees and partners find it increasingly difficult to keep up with relevant information and how to access it. Staff expect to be able to retrieve what they need from their desktop or mobile devices by typing a few keywords into a search box. Unfortunately, more often than not, conducting an information search within an agency is a frustrating and time-consuming experience.¹ As organizations transition from primarily paper-based to digital information formats, improved techniques are needed to find relevant, authoritative information efficiently. Improving the ability to locate information is not a simple exercise. The term *findability* was introduced by AIIM² and is used to characterize how easy it is to find relevant information.

Improving information findability in an organization is often difficult and is viewed as both an art and a science (AIIM 2008). It involves a variety of information organization, classification, tagging, and search techniques, as well as information governance and training. Poor findability results in wasted resources due to excessive time spent searching for information, reworking to investigate issues that have already been researched, duplicating data collection (due to lack of awareness of existing data), and decisions and actions that fail to take advantage of the organization's full knowledge base. Poor findability can also put the agency at risk, inhibiting responses to legal actions, claims, or Freedom of Information Act (FOIA) and public records requests.

Findability can be one of those "out of sight, out of mind" issues. Every day, a large number of employees (or contractors) could each be spending over an hour searching for information, but because these wasted hours are spread across the agency, the problem is invisible to agency leaders. Yet, these hours can represent large losses in agency productivity. A 2015 paper reported that a review of several surveys spanning different business sectors found that "24% of a business professional's time is spent looking for information" and that "48% of organizations felt search was unsatisfactory" (Cleverley 2015).

Fundamental impediments to findability include lack of disciplined information management practices, non-searchable information, legacy and disparate information repositories, and a lack of a consistent approach to metadata and use of terminology for information classification. Most organizations have limited resources that are devoted to information management tasks because priorities are placed on mission essential products and services. In most organizations, significant information is not searchable. Large bodies of content (books, reports, plans, maps, forms, etc.) have not been converted to electronic formats and are not searchable. Image files are stored without metadata or have not been OCR'd. Even when an organization has reasonable information management practices, information is often stored in repositories that are outside the reach of available search tools due to the lack of specific Web services or API's to these repositories and/or access restrictions. Finally, a lack of standard metadata makes it more difficult to overcome inherent limitations on the effectiveness of full-text search.

¹ Improving Findability and Relevance of Transportation Information: ISBN 978-0-309-44635-8 | DOI 10.17226/24804

² Association for Information and Image Management (AIIM)

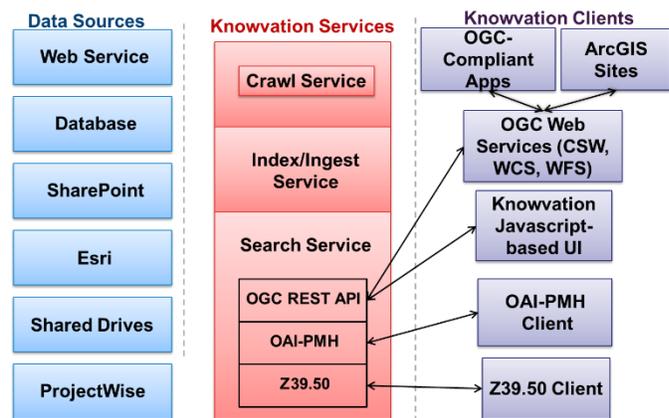


IMPROVING FINDABILITY

So how does an organization improve Findability? The most conventional method organizations have used is to implement an Enterprise Content Management System (CMS). There are, however, many reasons why organizations have not utilized an Enterprise CMS or have tried and failed. Reasons range from data migration to training, but in most cases the organization is just too decentralized or has too many political issues to invoke an Enterprise wide information management strategy that gains sufficient acceptance and becomes consistently utilized by all parties.

PTFS has recognized these issues and developed the Knowvation™ (KV) Content Services Platform (CSP) which uses a robust content management strategy along with technology that allows organizations to find and use their data regardless of its source repository and location. This means that the organization does not have to force all divisions to manage their data the same way using the same system. An information management strategy can be developed that, for example, allows certain groups to continue managing their data the way they have done for years while others may migrate their data to a centralized repository. The key is "federating" technology that allows for creation and searching of a master index of an organization's data, where the data may be stored locally in one or many places or across geographically dispersed repositories.

KV CSP provides a federated search capability to simultaneously search multiple data sources using index-in-place technology in combination with data ingestion. Using a web-based administrative GUI for configuration, KV can index data in place for existing data, file shares, repositories, and other systems such as SharePoint, and subsequently builds a federated index with full protection of the integrity and authenticity of the data. This dual function offers a "one search finds all" capability and permits data discovery and management, regardless of whether the software is hosted in the cloud or on premise in a data center.



Data Indexing and Processing Methodology

While Federated Indexing allows users to find all data, that data must be able to be indexed. KV provides tools to help make non-searchable data searchable. Tools include imbedded Optical Character Recognition that can OCR data on ingest and auto-geotagging tools that extract geospatial information from unstructured text and save this information in the metadata. KV has the ability to support multiple metadata schema allowing groups within an organization to use a data model that best suits their mission. KV also provides a browse categorization function that can be created without human resources when data is ingested. Finally, we also provide an easy to use metadata editor to create or enhance metadata. This web based tool was developed to allow crowd sourcing to permit enhancement of metadata without jeopardizing the integrity of the data.

For federated searches, KV's browser-based search interface makes it simple to find datasets, images, maps, layers, and other documents. Regardless of whether the data is ingested and managed by KV or indexed in place, one search finds it all. The KV CSP has GUI's for managing the system, along with advanced administrative features that allow workflows to crawl and identify new sources and add/define hot folders/parameters for automatic system ingestion. The KV CSP roadmap includes auto-categorization techniques that will further improve findability for unstructured information.