

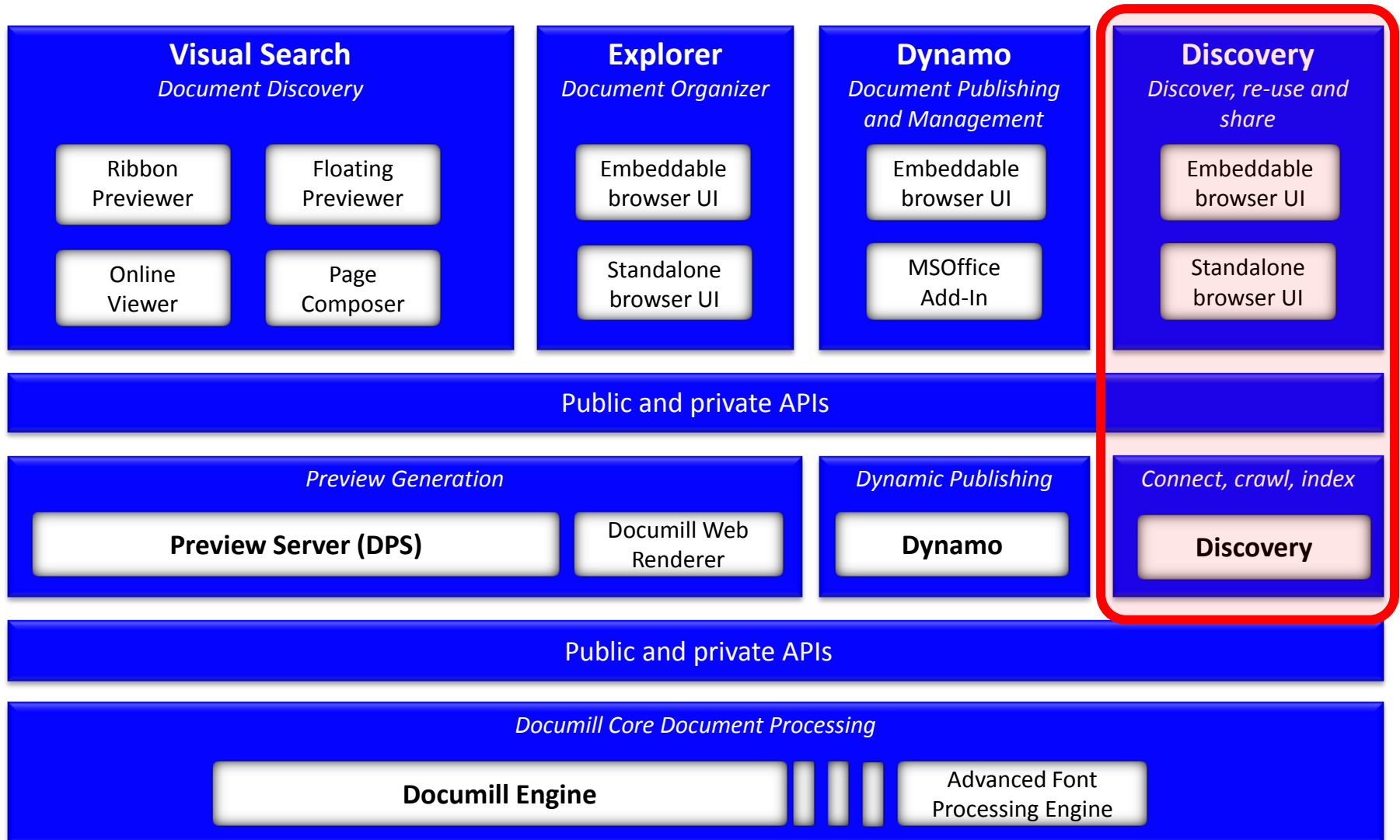
Documill Discovery

Search - Find - Take Action

Documill Discovery **Solution Summary**

2015-08-22

Documill Offering



Documill Discovery

Discovery is an enterprise search solution focused on cloud services. It includes:

- Support for most important data sources
- Data crawling
- Enforcement of access controls
- Processing content for indexing and previewing
- Fast and scalable search index
- Authentication
- Localisation
- Support and maintenance
- Upgrades

The image displays the Documill Discovery interface integrated with a Salesforce CRM. The main window shows search results for 'market documents' with 403 results. A specific document titled '2013 CRM Market Share Update_40...' is highlighted, showing its metadata and a preview of a Forbes article with charts. A sidebar on the right provides filtering options, including 'Source' with a list of cloud services and their document counts. Below the main interface, two smaller screenshots provide additional context: one shows the Documill Discovery logo and search bar, and the other shows a detailed view of the search results page with a diagram illustrating the integration of various cloud services like Box, Jira, and Confluence.

Documill Discovery

salesforce 14

Terho Laakso ▾ Help & Training

Home Chatter Files Leads Accounts Contacts Opportunities Reports Dashboards Products Forecasts Explorer Libraries ExplorerDevel + ▾

21 results Logged in as terho [\(Log out\)](#)

Search result (432) Filter results:

[Documill Apps - Explorer - Tablet.pptx](#) ↓ 1.4MB | Kimmo Liimatainen | 09/10/2013

Documill Apps - Explorer - Tablet.pptx www.documill.com 1.2.3 All rights reserved. © 2013 Documill Slide # Background Documents are stored in several different repositories reserved. © 2013 Documill Slide # Multi-repository content discovery and management service Discovery Find the documents that you need Connectivity

Show matches | 1 2 3 4 5 6 7 8

Salesforce <https://login.salesforce.com/00PD000000KbNcCKMAN>

[Documill Apps - Explorer - Tablet.pptx](#) ↓ 1.4MB | Kimmo Liimatainen | 09/10/2013

Documill Apps - Explorer - Tablet.pptx www.documill.com 1.2.3 All rights reserved. © 2013 Documill Slide # Background Documents are stored in several different repositories reserved. © 2013 Documill Slide # Multi-repository content discovery and management service Discovery Find the documents that you need Connectivity

Search Within Results

Modified

Order By

Author

Category

File Type

Language

Source

Salesforce (13)

Confluence (8)

Space

Type

Documill Discovery

What?

Integrated search solution for all the repositories where you store information, both in the cloud and behind the firewall.

To whom?

Any user who needs access to multiple data systems. Customer support, sales, product management ...

Why?

1. Better customer service with faster information discovery
2. Higher productivity with immediate access to information
3. Enablement of customer self service portals

Key capabilities

- Universal search from file shares, web pages, CMS, ECMs, databases, cloud services etc.
- Page-level insight into documents such as MSOffice, PDF and RTF files
- Enforcement of complex access right models as defined by the source systems
- Highly actionable results with accurate previews

Key value points

- Comprehensive coverage - find the needed information across all company systems
- Enables unified search experience for both structured and unstructured data
- Efficient - powerful filters to drilldown into meaningful set of results
- Mobile ready - preview content using any browser-enabled device
- Secure - source system access rights are relentlessly enforced

SOLUTION DETAILS

Discovery Search

The fastest way to the information

SIMPLE

- One integrated search covers all the places where you store information: file shares, document management systems, cloud, the intranet etc.
- Integration with existing authentication systems ensures a seamless experience and dependable security.

POWERFUL

- An industry-leading search solution covering the intranet and wiki content, as well as Microsoft Office™, RTF and PDF documents.
- Broad and continuously expanding support for different file repositories – both in the cloud and behind the firewall.

FAST

- Search results show you the exact pages that contain the relevant information. You can view documents in the browser, without having to switch between applications.
- A browser interface offers uniquely simple access to information whether you work on a laptop, tablet or smartphone.



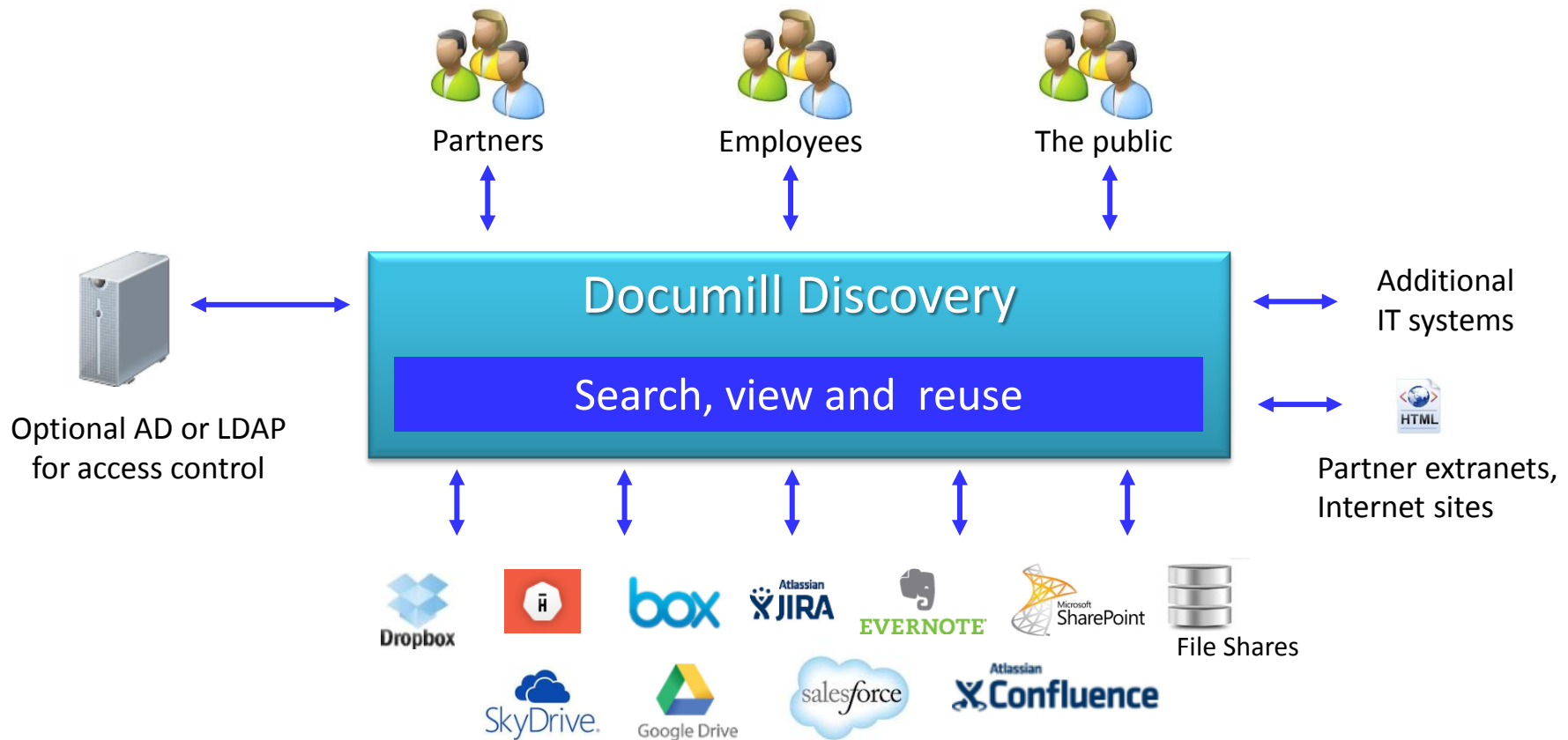
EVERNOTE



File Share



Discovery solution architecture

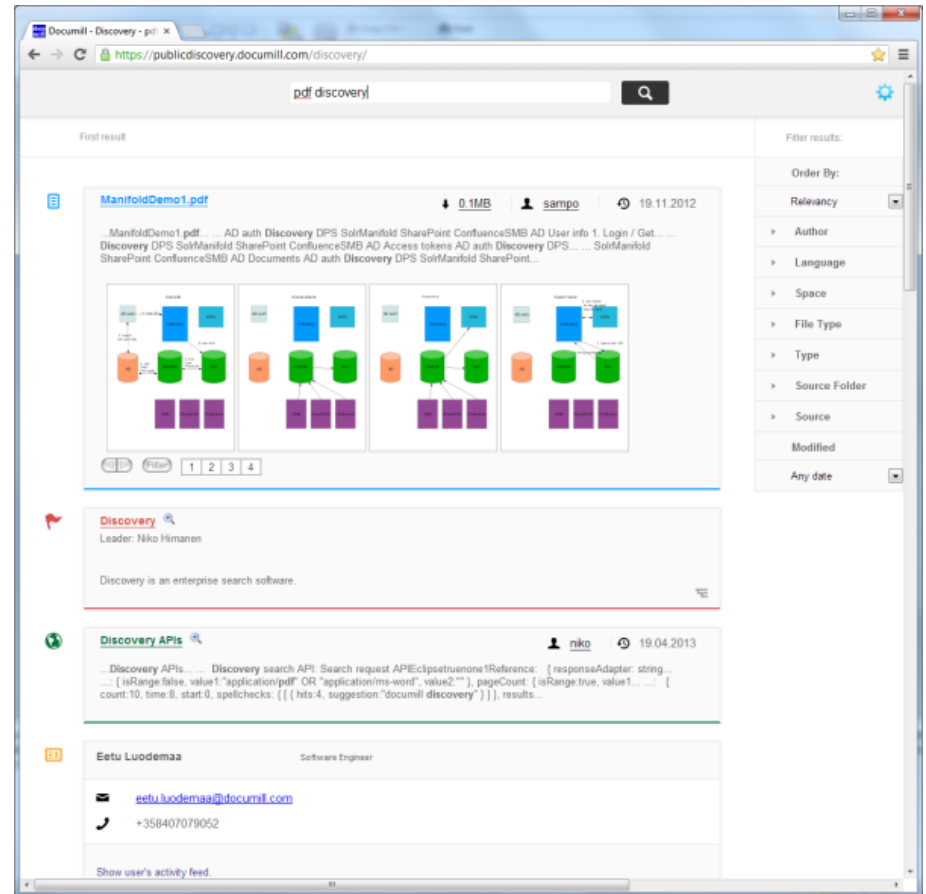


Design Principles

- Discoverability
 - Relevancy thru document and field level boosting, relevancy configuration based on connected document sources and customers' needs
- Modularity
 - Every Discovery component is designed and implemented to be a modular part of the whole architecture: HTTP REST JSON interfaces in use.
- Scalability
 - Scalability thru modularity, component modules can be distributed and while doing so, clustered for the best possible performance and responsiveness.
- Integrability
 - Functionality enhanced and modified through plug-in architecture
 - Discovery's connector framework provides tools to support third-party solutions and services as content repositories or authentication and access control authorities.
- Performance
 - Fast search performance is achieved with early binding of access rights and with aggressive caching of previous search terms and results.

Efficient search with integrated access control

- User authentication can rely on existing user identity management, such as Active Directory
- Access to previews and other data complies with the access rights of the source system
- Information about access rights related to each content item is included in the search index



Integrated document previews

- See which pages contain relevant information
- Preview documents in the browser, without having to download them or switch between applications
- Access even large documents on mobile devices
 - Previews are images or HTML, and will open with popular Internet browsers

The screenshot displays the Documill Discovery web interface. The browser address bar shows the URL <https://publicdiscovery.documill.com/discovery/>. The search bar contains the term "specification".

The results are filtered to show the "First result". Three document entries are visible:

- 2013-01-15 Documill - PM related processes.pdf** (0.8MB, user: mika, date: 31.01.2013). The preview shows a document with sections for "Terminology", "Product Planning: Deliverables", "Getting started with the software", and "UML & PDL - Getting from approved".
- end-user-documentation.pdf** (10.2MB, user: sampo, date: 27.11.2012). The preview shows a document with sections for "New! Search - End-user Documentation", "Contents", "Rationale for market definition", and "Enterprise Content management".
- Enterprise Search SaaS MRD 11.pptx** (4.8MB, user: toni, date: 15.02.2013). The preview shows a document with sections for "Contents", "Rationale for market definition", and "Enterprise Content management".

On the right side, there is a "Filter results" panel with options for "Order By", "Relevancy", and "File Type". The "File Type" filter is expanded, showing checked options for "Adobe PDF (146)", "comment (35)", "MS Word (11)", and "MS PowerPoint (2)". Other filter options include "Modified", "Any data", "Author", "Language", "Space", "Type", "Project", "Source Folder", and "Source".

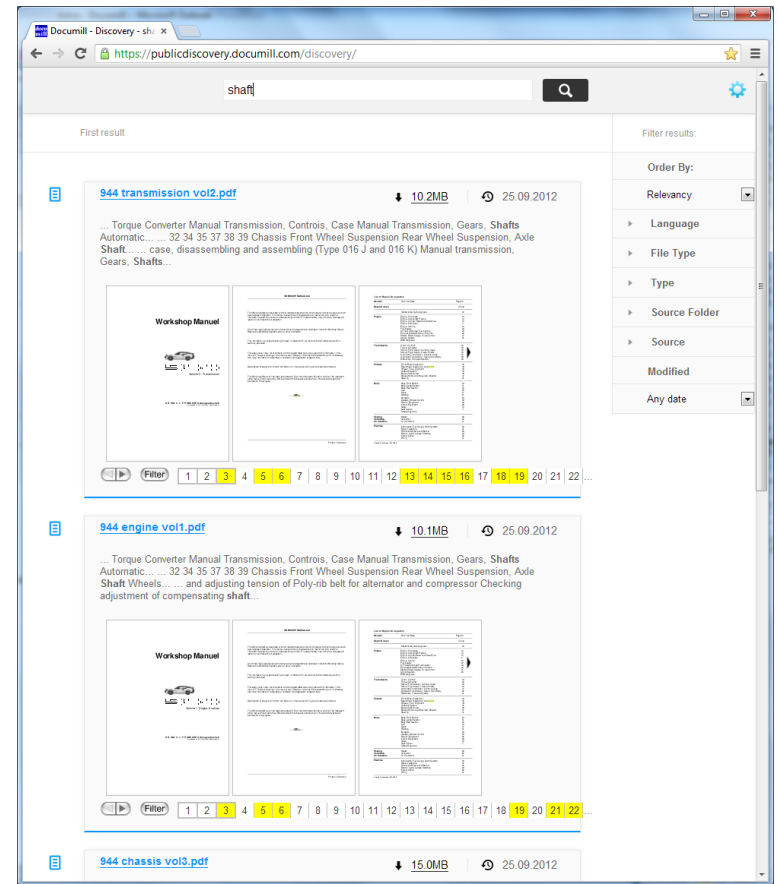
Powerful result filtering

- Filters speed up your way to the right document
 - Focus based on some known detail, such as the author, the file type, or the source repository
- The search covers both full text content and metadata
- The list of available filters automatically adjusts to the specific result set
- Use case example:
 - Content type could be 'contract' and the results could be filtered or sorted by the creation date

The screenshot displays the Salesforce search interface. At the top, the Salesforce logo and navigation menu are visible. The search bar contains the text "market documents" and shows "403 results". The search results are displayed in a list format. The first result is titled "2013 CRM Market Share Update_ 40..." and is a PDF file (0.8MB) from Documill Discovery, dated 02.10.2013. The preview shows a Forbes article snippet and a pie chart. The second result is titled "guide_box.net.pdf" and is a PDF file (0.5MB) from Documill Discovery, dated 02.10.2013. The preview shows a document snippet and a small image. On the right side, there is a "Filter results:" panel with various filters applied, including "Source" (Salesforce, Box, Web) and "Type".

Optimized search results

- The default configuration is ideal in most cases
- The results can be refined further by
 - Stemming words to their root form
 - Matching letter case
 - Splitting compound words
 - Searching for synonyms
 - Checking spelling in queries
 - Prioritizing results by language
 - Prioritizing results by data element, such as title, description, user name etc.



Supported content types

Indexing and searching

- HTML
- XML and derived formats including XHTML and ODF
- Microsoft Office document formats
 - OLE 2 Compound Document format introduced in MSFT Office 97, DOC, XLS, PPT, MFC serialization API based file formats
 - Office OpenXML Format introduced in Microsoft Office 2007 and 2008, XLSX, DOCX and PPTX
 - Outlook MSG, TNEF
 - Visio VSD
 - Publisher
- Google document, presentation and spreadsheet
- OpenDocument format, ODF
- Portable document Format, PDF
- Electronic Publication Format, EPUB
- MBOX email format
- Rich Text Format, RTF
- Text formats
- Image formats
 - Metadata from JPEG, PNG, BMP, WBMP, GIF, TIFF
- Audio formats
 - Metadata and textual information from MP3, AIFF, AU, WAV, MIDI Type 0, MIDI Type 1, and Rich Music Format (RMF)
- Flash video
- Compression and packaging formats ar, cpio, Unix dump, tar, zip, gzip, XZ, Pack200, bzip2, 7z and arj
- Java class files and JAR archives

Preview generation

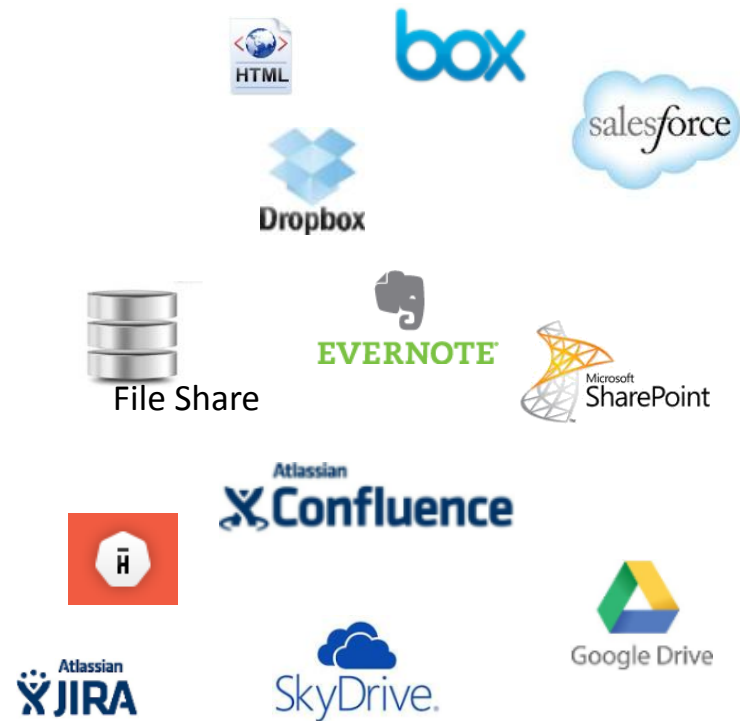
- Microsoft Office document formats
 - OLE 2 Compound Document format introduced in MSFT Office 97, DOC, XLS, PPT, MFC serialization API based file formats
 - Office OpenXML Format introduced in Microsoft Office 2007 and 2008, XLSX, DOCX and PPTX
- Portable document Format, PDF
- Rich Text Format, RTF
- Google document, presentation and spreadsheet
- Text formats
- Image formats
 - JPEG, PNG, BMP, WBMP, GIF, TIFF
- X/HTML



Wide support for different data sources

Off-the-shelf support for most relevant data repositories:

- File shares
- HTML sites, with or without authentication
- Microsoft SharePoint
- Atlassian Confluence
- Atlassian JIRA
- Salesforce
- Google Drive
- Dropbox
- Box
- SkyDrive
- SkyDrive for Business
- Evernote
- Hightail
- And more to come...



About Documill

- Documill is an ISV whose solutions enable the browser-based viewing and server-side content processing of Microsoft Office and PDF documents.
- Our competencies are in server-side large scale document processing.
- Our solutions are used by the world largest online service providers and enterprises.
- Documill HQ is in Finland, Europe.

