SMART CONNECTOR TECHNOLOGY FOR FEDERATED SEARCH

VERSION 1.4 • 26 MARCH 2018 • EDULIB, S.R.L.
MUSEKNOWLEDGE™ FEDERATED SEARCH PLATFORM

Delivering content integration technology since 1998

- Building and delivering the widest range of federated content through the industry’s most flexible content integration platform
- Seamlessly integrate limitless content sources into applications and services
- Experienced and proven technology powering scaled applications
- 15 years of continuous development and integrations

We rapidly deliver comprehensive applications without substantial in-house development

- Muse Source Factory™ of 5,000 + content source connections enables partners to scale platforms and services

Muse is the only product of comprehensive, plug & play content integration technology

- Federation, Harvesting, Transformation, Enhancement, Security, Source Maintenance, Multiple Delivery Mechanisms, Analyzed extracted data
MUSEKNOWLEDGE™ AND THE ENTERPRISE SEARCH PLATFORM

Muse’s Value to Enterprise Search

• Expanding source connections, especially for information outside corporate repositories
• Providing enhanced (enriched) records from multiple sources
• Performing an array of advanced post-processing on results sets
• Multiple integration architectures and information delivery modes

MuseKnowledge™ Federated Search is Complementary to Enterprise Search

• MuseKnowledge™ Federated Search is a content harvester, normalizer and feeder
• MuseKnowledge™ Federated Search is designed to be a pass-through technology, with no repository functionality
• Feeds directly to Enterprise Search Engine and, or repositories

MuseKnowledge™ can Federate the Enterprise Search Engine and Other Sources in Real Time for End Users
MUSEKNOWLEDGETM FEDERATED SEARCH

Multiple Domains of Applicability
• Library, University, Hospitals, Legal, Police
• Business (B2B), Government (B2G), Consumer (B2C)

Various Types of Sources
• Search, Writer, Enrichment, Content Mining, Inter Library Loan, Hold, Patron, DeDupe, Ranking, Shopping Cart

Various Protocols Supported
• Atom, HTTP/HTML, HTTP/XML, JSON, NCIP, OAI-PMH, RSS1.0, RSS2.0, SIP2, SQL, SRU, SRW, Telnet, Z39.50

Maintenance and Management at Runtime
• Application grouping and execution; parallel execution, thread control, post processing, DeDupe, Ranking, Content Mining, hot deployment, sandboxing

Scalability
• Single Muse instance vs. Shared Load Balanced Environment; NFS and rsync for synchronization
MUSEKNOWLEDGE™ FEDERATED SEARCH

• User searches via browser User Interface or client system
• Access to free and authenticated Sources
• Search translated for each Source
• Multiple Content Sources, Same Search Query, Single Integrated List
• Results enhanced from multiple Sources
• Results reformatted and normalized
• Result Set sort, rank, export, deduplication, processing

• Refine results functionality
• Search history and saved searches
• Personalization of functions and Sources
• Enterprise customization of UI and Sources
• User Interface for Mobile Devices
• Linking to the native detailed record
• Application server, sandboxing
• Statistics
• Vendor neutral
Layered Architecture
Muse Can Integrate in Multiple Ways Because of:

- Protocol handling Bridges
- Message based interaction (loose coupling)
- Symmetric message handling (listens and talks)
- Session management (convert a transaction into a dialog)

Multi-Processing & Multi-Threading Kernel Allows:

- Handling multiple input message streams (user sessions)
- Efficient resource usage through asynchronous processing
- Robust operation with different speed external systems
THE FEDERATED SEARCH PROCESS

<table>
<thead>
<tr>
<th>QUERY</th>
<th>SEARCH</th>
<th>POST PROCESS</th>
</tr>
</thead>
<tbody>
<tr>
<td>End users initiate searches through customer’s or Muse Knowledge’s system.</td>
<td>Dedicated “Bridges” integrate each front-end system with Muse Knowledge.</td>
<td>The federal query is authenticated &amp; semantically mapped to selected sources via specific connectors.</td>
</tr>
</tbody>
</table>

**CUSTOMER SYSTEM**

**MUSE KNOWLEDGE SYSTEM**

**CONTENT PROVIDER SYSTEM**

**SEARCH RESULTS**

**AUTHORITY CONTROL**

**POST PROCESS**

**SEARCH RESULTS**

**LINK**

**DOCUMENTS/TRANSACTIONS**
Sophisticated search and content integration solution, with advanced post-processing and a highly productive user environment, supported by a full complement of powerful management tools

**Searching**
- Source Selection
- Source Capabilities
- Source Limiters
- Pre mapping
- Stats & usage tracking

**Admin & Management**
- Authentication & DRM
- User Interface customization
- Management consoles
- Source maintenance

**User Environment**
- Post-search processing
- Personalized workspaces
- Alerts
- Advanced exploration
- Content Mining

**Integration**
- Content integration
- Application integration
- Message Passing Environment
- Session Management
THE MOST EXPENSIVE METADATA SUPPORT AVAILABLE

Sampling of key Metadata fields supported by Muse

- Control numbers (ISBN, ISMN, ISSN, etc.)
- Author, Title & Journal Title
- Citation Data (title, volume, issue, page, year, ISSN)
- Assigned Keywords & Derived Keywords
- Abstract
- Subject Headings
- Classification
- Category
- Location (geographical, within building, shelving, filing)
- Printing/publishing details (year, publisher/printer, location, etc.)
- Material Descriptors & Thematic Descriptors
- Target Audience
- Size (pages, bites, seconds, etc.)
- Linking (material, subject, location, thematic, etc.)
- Use Limiting Data (access rights, price, embargo, etc.)
- Availability Data (inventory, comparison shopping, shipping, etc.)
- Author Affiliation
- References (to other material)
- Object Linking (full text, image, video, audio)

Thousands more...
ENRICHMENT

• Create an enhanced record with content from more than one input record
• Secondary searches may use search parameters derived from main record
• Uses field level merging and whole record merging
• ExParsers (Extended Parsers) process record components for normalized data
• Dynamic selection (content based) of ExParsers
• OpenURL resolvers
PLUGGABLE LOGIN MODULES

Muse comes with a collection of login modules to define the desired authentication scenario. The Login Modules can be combined to form an authentication stack.

Login Modules Parameters & their Semantics

- **Required** The Login Module is required to succeed. If it succeeds or fails, authentication still continues to proceed down the Login Module list;

- **Requisite** The Login Module is required to succeed. If it succeeds, authentication continues down the Login Module list. If it fails, control immediately returns to the application (authentication does not proceed down the Login Module list).

- **Sufficient** The Login Module is not required to succeed. If it does succeed, control immediately returns to the application (authentication does not proceed down the Login Module list). If it fails, authentication continues down the Login Module list.

- **Optional** The Login Module is not required to succeed. If it succeeds or fails, authentication still continues to proceed down the Login Module list.

Existing Login Modules

- ICELoginModuleXML - username/password authentication
- ICELoginModuleIP - IP authentication
- ICELoginModuleLDAP - authentication against LDAP
- ICELoginModulePPMS - personal user authentication
- ICELoginModuleHTTPReferer – referrer authentication
- ICELoginModuleSQL - authentication against SQL database
- ICELoginModuleHMAC – HMAC signature authentication
- ICELoginModulePropertiesExtractor – SAML authentication
PLUGGABLE LOGIN MODULES

Authentication scenarios for MuseKnowledge™ Applications:

• **Standard Username/Password authentication.** The end-user enters a username and password at the Muse Applications login form;

• **Standard IP authentication.** If the IP of the end-user is authenticated he/she is allowed in the Muse Application;

• **Standard LDAP authentication.** The end-user enters his LDAP username/password details;

• **IP authentication** for on campus users and Username/Password for off-campus;

• **IP authentication** for on campus users and LDAP for off campus;

• **IP authentication** for on campus users and personal user details for off campus;
AN EXPLORATION WORKFLOW

MuseKnowledge™ User Interface – Home Page and Search Form

Linking to Native Full Text Record

Personal Workspace

Search Results

Refine & Post Processing
CUSTOMER FEDERATED SEARCH

This screenshot shows the initial home page customized according to Partner’s requests.

The Search Form allows building complex queries.

Widgets to display recommended list of books and articles, RSS feeds and the available list of resources.

Customizations can be done to fit the look and feel of the rest of the Partner’s website and other customer facing systems.
CUSTOMER FEDERATED SEARCH

The “Search Sources” drop menu has been opened and some Sources selected in addition to the defaults. A simple search is typed in ("book") The user now clicks “Go”...
CUSTOMER FEDERATED SEARCH

A Navigate and manipulate the whole result set. Selected records can also be saved, exported, etc.

B Related Queries widget

C Search results are listed in the selected display format.

D All Sources are listed with a record count. Clicking any one displays records from that Source only.

E Special panel shows terms extracted from the results set which can be used to refine the results on categories like Authors, Years, Keywords, Companies, Locations, Products, Languages. Refine by Year 2016.

F Term highlighting

G Side search widgets with specialized sources.
A Breadcrumb functionality allows navigation between post-search action results and the original search results.

B ... and then end up with 3 records. The desired record is a click away.
CUSTOMER FEDERATED SEARCH

- **Muse Navigation Manager re-writes** the page for restoring the conditions of the initial search (cookies and session related data).
- **Original full text, from free** and subscription Sources.
- **Publisher’s** platform complete set of features are now available.
MUSEKNOWLEDGE™ APPLICATION SEARCH SOURCES

- Select individual, multiple sources for searching
- Select entire groups of sources
- Quickly locate the source(s) by the quick find functionality
- Alphabetical listing
- Restore to default sources selections
- Show only selected sources
- Access to sources descriptions and native platforms
### Search Options

- **Remove Duplicates By:**
  - None  
  - Title  
  - DOI

- **Display Duplicates:**
  - Show Duplicates  
  - Hide Duplicates

- **Results Per Source:**
  - 5  
  - 10  
  - 25  
  - 50  
  - 100

- **Results Per Page:**
  - 5  
  - 10  
  - 25  
  - 50  
  - 100

- **Results Display Level:**
  - One Line Record  
  - Brief Record Detail  
  - Full Record Detail  
  - Text  
  - Raw Data  
  - XML  
  - Atom  
  - RIS  
  - EndNote

- **Show Search Progress:**
  - Yes  
  - No

- **Show Progress Details:**
  - Yes  
  - No

- **Sorting By:**
  - None  
  - Relevance  
  - Title (A-Z)  
  - Title (Z-A)  
  - Author (A-Z)  
  - Author (Z-A)  
  - Date (Oldest)  
  - Date (Newest)

### Available Options

- **Handle Duplicates**
- **Set how many results to retrieve per Source**
- **Set how many results to display**
- **Set Display Level:** One Line, Brief, Full, Text, Raw, XML, Atom, RIS, EndNote;
- **Control Search Progress and Details**
- **Sorting Options**
• **Set General Limits:** Language, Material, Date, Full Text, Peer Review

• **Source Specific Limits**
Manage your Results

- Display Level
- Deduplication
- Sort & Filter
- Handle the Records: Keep, Delete, Save to WorkRoom, Email, Distill, Enrich, Place Holds
- Save Records to Disk as: HTML, Text, RAW, XML, Atom, RIS, EndNote, PDF
- Place Orders
- Export to RefWorks, delicious, QuickBib, Dropbox, Mendeley, EndNote

Search Details

- See extraction status for all searched sources
- Stop All searches or individual searches
• Store your Personal Search Options
• Set your individual access details for Sources
• Overwrite the default selected, hidden sources with your individual ones
• Set own search limits
• Administer the Saved Searches and saved WorkRoom
• Administer the Alerts

My Account - WorkRoom

My Account - Saved Searches

This is your personal workspace where you can view and modify the saved lists of records retrieved when searching.

Find your saved list:

• Result Set Name: [your search]
• Time: 12/3/17 1:14 PM, Query: [your query]

Records: [number] Description: [description]

• Alert: [your alert]
• Time: 12/17 1:43 AM, Query: [your query]

Records: [number] Description: [description]

My Account - Saved Resultsets

This section contains all the searches saved by the user during various sessions. It presents all the details of the searches being saved. Users can re-run queries saved with the possibility to be re-run automatically.

Find your query:

• Query: [your query]
• Time: 12/17 1:43 AM, Description: [description]

• Alert: [your alert]
• Time: 12/17 1:43 AM, Description: [description]
Configure every general aspect of the application through the MuseKnowledge™ Administration Consoles

- Application Name and Description
- Email and contact information for the emailing features
- Default Navigation Management settings used by the Application
- Define the settings used for Proxy Configuration
- Manage the HTTPS certificates used by the application’s sources
- Define the OpenURL resolver settings
- **Configure user interface options such as:** Default skins, Languages, Banner, Logo, Search Options, Search Sources, Search Limits, Application Functionality, Logoff Behavior
- Configure the Application’s Login Modules
- **Configure the Application Modules:** Search, DeDupe, Jitterbug, Ranking, Circulation, ILL, Shopping Cart, Writer, Content Mining
- **Other Application settings such as:** Name, Expiry Date, User Concurrent Sessions, Maximum User Emails, Default Locale, Properties, Components Paths, etc
Native code applications, both iOS and Android, supporting a large variety of devices, phones and tablets.

Features

- Select which source to search
- Search by many fields
- Sort by relevance, title, author, source or date
- Specify how many records to retrieve from each searched source
- Set the desired display level of the records
- Link to the detailed record on the publisher’s platform
- Search History and many other features
MUSEKNOWLEDGE™ MOBILE APPLICATION

Available on Google play

Available on the App Store
MUSEKNOWLEDGE™ SOURCE CONNECTIONS

Unparalleled ability to extract value & relevance from disparate content sources

- **Cover full range of** content and source types
  - CMS, search engines, repositories, database systems
  - Magazine, news, journal, library archives, books, articles, images, web formats, videos, blogs, real objects
  - Traditional and online publishers (subscription, premium content)
  - Online content aggregators
  - Standard Web search & deep web information
  - Native database content (numeric and text-based)
- **Support multi-level** metadata density and complexity
  - From popular, consumer through to academic, research
- **Supported by** Automated update mechanism
- **Administered through** central Source Factory
- **Managed via** browser based Consoles
- **Highly Automated** change reporting and fixing
- **Retrives at any** of the three levels
  - User
  - Application
  - Data
- **Access through** API or User Interface
- **Can present** metadata, abstracts and “full text”
  - Link to native sources in real time
WHAT IS A SOURCE PACKAGE

• **Something that enables** content from external Sources to be used by a technology platform, application as though it were native.

• **A facility that creates ‘clouds’** of content to be available to users and systems based on need rather than format.

• **A means of supporting** the integration of information at a business layer instead of a technological layer.

• **“Plug-and-play” bundles of code** that address in a very consistent manner authentication, translation and linking from a Muse system to the target data service that the Source Package was built for.
SOURCE PACKAGE BUILDING BLOCKS

Authenticator
- Allows for different methods of credentialing for users
- Interchangeable between compatible sources
- Uses data in Source Package Profile to make connection

Connector
- Code for communication between Muse and a Source
- Defines protocol for messaging back and forth
- Uses data in ISR and IDR to interpret searches and responses

ISR & IDR Maps
- Establishes translation for searches to native language of source
- Defines data-element tagging and normalization for specific data elements
SOURCE PACKAGE SUM OF PARTS

AUTHENTICATOR

IDR & MAPS ISR

CONNECTOR
SOURCE PACKAGES TYPES

- Business
- News
- Art & Music
- Health Care
- Religion & Philosophy
- Environment
- Biomedical
- Technology & Standards
- Law
- Science
- Humanities
- Politics & Sociology
- Computing
- Education

54% COMMERCIAL DATA SOURCES
14% ENTERPRISE APPLICATIONS
32% FREE WEB SOURCES

COMMERCIAL DATA SOURCES
ENTERPRISE APPLICATIONS
FREE WEB SOURCES
SOURCE PACKAGES SUBJECTS (SPECTRUM)

- Health Care: 13%
- Business: 13%
- Science: 10%
- News: 9%
- Humanities: 8%
- Technology & Standards: 8%
- Politics & Sociology: 7%
- Science: 6%
- Biomedical: 3%
- Law: 3%
- Religion & Philosophy: 2%
- Environment: 2%
- Computing: 3%
- Education: 13%
- Art & Music: 33%
SOURCE PACKAGES CONFIGURATION

Manage the MuseKnowledge™ Source Packages through the MuseKnowledge™ Administration Consoles: MuseKnowledge™ Console for Applications Administration and MuseKnowledge™ Console for Customer Support.

- Available Source Package Actions
- Installed Source Packages
- The build date of the Source Package
- The date when the Source Package was installed
- The test status of the Source Package: Working, Not Working, Unknown, etc.
- The version of the installed Source Package
- Filters for quickly identifying the desired Source Package(s)
SOURCE PACKAGES CONFIGURATION

Configure every aspect of a MuseKnowledge™ Source Package:

- **Identification and Description**
- **Search and Home URLs**
- **Connection Parameters**: User Agent, Time Slice, Connect Time Out, Read Time Out, Time to run, Encoding, Database Name
- **Mapping Files**: ISR, IDR
- **Authentication Settings**: Authenticator, User Name, User Password, User Pin
- **Extended Parser Settings**: Use Extended Parser, Extended Parser Class, Extended Parser Encoding, Extended Parser Configuration File
- **Proxy Settings**: Use Proxy, Proxy Host, Proxy Port, Proxy PAC, Proxy Authorization Scheme, Proxy User Name, Proxy User Password
- **Server Settings**: HTTP User Name, HTTP User Password, HTTP Authorization Scheme, SSL Certificates
- **Navigation Manager Settings**: Link URLs

Import configuration values from Profile
“Configure more Sources like this” feature
Backup, Restore a Muse Source Package
SOURCE PACKAGES CONFIGURATION

Control entirely the behavior of a Source Package through configurations.

[Image of a configuration interface with details such as Source ID, Name, Description, Search and Home Addresses, and Connection Parameters.]
Over the last decade, Muse has amassed a library of more than 6,000 Source Packages. This global library, called the Muse Source Factory, serves as a central repository of Source Packages that are licensed to Muse implementations.

Built-in consoles in Muse let system managers choose from the vast array of sources in the Source Factory, and they can be downloaded seamlessly into local Muse installations.

The seamless, bidirectional integration of the Source Factory into the administrative consoles of Muse mean that whenever a Source Package is corrected and published, the implementations of Muse out in the world get notification of the available update.

Because Sources can change as their providers enhance them, Source Checking can identify those that need attention from the Muse development team, and they are flagged for testing and update.
## Muse Sources

<table>
<thead>
<tr>
<th>Source ID</th>
<th>Source Name</th>
<th>Status</th>
<th>Date Created (UTC)</th>
<th>Build Date (UTC)</th>
<th>Data Service</th>
<th>Type</th>
<th>Host</th>
<th>Protocol</th>
<th>Access</th>
<th>Package Version</th>
</tr>
</thead>
<tbody>
<tr>
<td>FourTo40</td>
<td>FourTo40</td>
<td>Test</td>
<td>2006-11-15 06:15:01</td>
<td>2014-01-11 08:13:14</td>
<td>4to40</td>
<td>Web portal</td>
<td>Four to 40</td>
<td>HTTP/HTML</td>
<td>Free</td>
<td>1.94</td>
</tr>
<tr>
<td>AAASSO</td>
<td>AAASSO</td>
<td>Test</td>
<td>2010-12-20 03:18:05</td>
<td>2014-01-11 08:09:33</td>
<td>Science Online</td>
<td>AAAGS Datapages: Petroleum Abstracts</td>
<td>HTTP/HTML</td>
<td>Subscription</td>
<td>1.65</td>
<td></td>
</tr>
</tbody>
</table>
SMART CONNECTOR ECOSYSTEM

Connectors need a world to live and work in:

- **Deployment**
  - The Muse Source Factory contains details of all Connectors
  - Automated Source Update handles endpoint deployment

- **Monitoring**
  - The Source Checker operates constantly
  - Results of user operations are used for early warning

- **Repair**
  - Automatic and user notification
  - Tracking, testing and building system
TOOLS FOR SOURCE PACKAGE MAINTENANCE

Source Packages Upgrade
MuseKnowledge™ Control Center
Task File

• No need to manually update the MuseKnowledge™ Source Packages, now it can be handled automatically;
• Very useful when administering a big number of applications;
• Complex algorithm for upgrading the Source Packages;
• Email notifications for task completion/error/failure with logs attached.
Source Checker MuseKnowledge™ Control Center Task File

- Perform regular checks of the installed Muse Source Packages
- Store the test status for having accurate and up to date test status values
- Get detailed report with the test results
- Email notifications for task completion/error/failure with logs attached

Detailed HTML report of the Source Packages check.
TOOLS FOR SOURCE PACKAGE MAINTENANCE

Report Broken Source Packages for Fixes, Updates

- Easily done through the Muse Administrator Consoles
- The Source Problem Report is sent upon submission to Muse Technical Support department

If the Source Package fails in retrieving results or if the parsed information is not correct simply click on the “Problem Report” button.

Just fill in the requested information and submit the problem report.
Report a MuseKnowledge™ System Problem

- Easily done through the Muse Administrator Consoles
- The Problem Report is sent upon submission to Muse Technical Support department

Click on the “Problem Report” menu item to send a Muse Problem Report.

Just fill in the requested information, allow the report to collect log files, attach any relevant information and submit the problem report.
MONITORING THE MUSEKNOWLEDGE™ SYSTEM

Real Time Monitoring

- Check servers status and system lifespan through the Muse Knowledge™ Console for Applications Administration
- Advanced monitoring of Muse servers through JMX
- Historical JMX graphs with RRD Grapher
MuseKnowledge™ Statistics Monitor

- **Desktop tool** for manually generating statistical information from dedicated log files
- **Can be connected** with the Muse Knowledge™ Control Center for generating regular automated usage statistics
- **Allows filters** to be specified: date filters, regular expression filters, etc.
- **Multiple analysis modules** are available, depending on the requested statistics: Memory Usage, System Lifespan, Session Lifetime, Connector Activity, Network Analysis, etc.
- **Various output formats**: CSV, XML, Graphical Tables/Charts/PDF files (only when running with the desktop GUI version)

Statistics are generated for 4 main areas of Muse activity:

- **User sessions** for gathering overall usage statistics such as number of sessions logged on, length of sessions, IP addresses of sessions, failed login attempts, etc.
- **Muse Instructions** for gathering information about the activities within Muse - searches - including queries, databases searched, parameters used
- **Muse Modules** more detailed statistics from individual search source or transaction modules including numbers of hits, time taken for query, download and processing time, etc.
- **System information** available and used memory
MuseKnowledge™ Statistics Monitor
SMART CONNECTOR TECHNOLOGY FOR FEDERATED SEARCH