

# eRoom Introduction

The eRoom connector will crawl content from an eRoom server instance (site) with an enabled XML Query option (Allow XML queries and commands from external applications).

- [Features](#)
- [Content Retrieved](#)
  - [Include](#)
  - [Exclude](#)
- [Limitations](#)
- [Operation Mode](#)

## Features

---

Some of the features of the eRoom connector include:

- Performs incremental crawling (so that only new/updated documents are indexed).
- Fetches LDAP (including Active Directory) access control lists (ACLs) for document-level security (including users and groups).
- Metadata extraction.
- Is search engine independent
- Runs from any machine with access to the given eRoom site.
- Designed to support early binding mechanisms and group expansion of nested permissions.
- Filter the crawled documents by file names using regex patterns.
- Supports Windows/Linux/macOS file shares.

## Content Retrieved

---

The eRoom connector retrieves several types of documents, listed below are the inclusions and exclusions of these documents.

### Include

- Folder
- Discussion
- Note
- Poll
- Calendar and Events
- Project Plan and Tasks
- Database and Rows
- Inbox (basic information)
- Other Files
- Link
- Dashboard

### Exclude

- Comments on items
- Emails on Inbox item
- Content of Link addressed pages

## Limitations

---

Due to API limitations, the eRoom connector has the following limitations:

- Multi-Threads Technical limitation

With Aspire 3.3 we introduce a multi-threading platform to perform crawls, but due architecture / API limitations the eRoom Connector basically work as single thread connector. Seems to be that the server does not allow multiples connections performing queries at the same time.

At some point of the crawl, the connector get this errors from the server "**No target objects were found evaluating the command's select attribute**" but the query that we are trying to execute works well (we tested using a SOAP/XML Test program provided for EMC).

Also if you look at the lasted comments in this post from the EMC forums:

[eRoom Forum](#)

Seems to be that eRoom is not multi-thread safe so when multiple processes are spawned you will get unpredictable behavior. So because of that, the eRoom Connector is single thread and that affect the performance due Aspire 3.3 is multi-thread.

## Operation Mode

The connector will use SOAP / XML over HTTP or HTTPS to acquire information of eRoom content. The connector acquires content by doing the following:

- Go recursively through all items and documents of an eRoom site, creates sub-jobs for each object discovered. Each sub-job contains all metadata available, including ACLs.
- Saves the item states into a MongoDB instance in order to compare and perform the incremental crawls with added, updated and deleted items.