

BulkMetadataManager 2.2.1 Release Notes

2024/07/26

1.	Introduction.....	4
2.	About BulkMetadataManager.....	4
3.	Supported Environments	5
3.1	Supported Operating Systems.....	5
3.2	Supported SharePoint Versions.....	5
3.3	Supported Authentication methods.....	5
4.	Releases.....	6
4.1	Feature and bug release (2.2.1)	6
4.2	Feature and bug release (2.2.0)	6
4.3	Feature and bug release (2.1.4)	7
4.4	Bug release (2.1.3).....	7
4.5	Feature and bug release (2.1.2)	7
4.6	Feature release (2.1.1)	8
4.7	Feature release (2.1.0)	8
4.8	Bug release (2.0.7).....	9
4.9	Bug release (2.0.6).....	9
4.10	Feature release (2.0.5)	9
4.11	Bug release (2.0.4).....	9
4.12	Feature and bug fix release (2.0.3).....	9
4.13	Bug fix release (2.0.2)	9
4.14	Bug fix release (2.0.1)	9
4.15	Feature release (2.0.0)	10
4.16	Bug fix release (1.0.13).....	10
4.17	Bug fix release (1.0.11).....	10
4.18	Bug fix release (1.0.10).....	10
4.19	Bug fix release (1.0.9).....	10
4.20	Bug fix release (1.0.8).....	10
4.21	Bug fix release (1.0.7).....	11
4.22	Feature release (1.0.6)	11
4.23	Feature release (1.0.5)	11

4.24 Feature release (1.0.4)	11
4.25 Initial release (1.0.0).....	11
5. Known Issues	12

1. Introduction

This document describes the release notes for BulkMetadataManager 2.2.1 and provides the latest information on the SharePoint application.

2. About BulkMetadataManager

BulkMetadataManager is a Node.js application that uses REST API to interact with SharePoint. The application can extract properties from documents stored in SharePoint and capture the values into SharePoint columns. This allows for optimal use of the metadata for searching, filtering, sorting etc... and this results in increased user productivity and adoption. It also allows for properly managing the document life cycle (records management).

The documents that need to be processed are selected using a CAML query. The application is optimized for performance by using parallel requests, caching, and batch requests. It also caters for SharePoint throttling (429 responses) and the application can handle large list exceeding the list view threshold. The CAML query needs to include a condition to avoid processing the same document multiple times.

The WhatIf option allows for running the application without the final step to add the metadata into SharePoint columns. The extracted metadata is saved into a csv for further review and analysis.

The target audience for this document comprises SharePoint architects, SharePoint administrators and site collection administrators.

3. Supported Environments

This section provides information on the supported environments. The application is a node.js application and runs on a separate system, it does not require installation of software on SharePoint. The installation instructions are described in “BulkMetadataManager 1.0. Administrator Guide”.

3.1 Supported Operating Systems

BulkMetadataManager is supported on the following operating systems.

Operating system	Supported	Notes
Windows 10	✓	
Windows 11	✓	
Windows Server 2016	✓	
Windows Server 2019	✓	
Windows Server 2022	✓	

Operating systems not listed are not supported.

3.2 Supported SharePoint Versions

The following SharePoint versions are supported.

Versions	Supported	Notes
SharePoint Online (Office 365)	✓	

SharePoint versions not listed have not been tested and are not supported.

3.3 Supported Authentication methods

The following authentication methods are supported.

Method	Supported	Notes
SharePoint App-Only* authentication	✓	

*Using Client Id and Client Secret at site or tenant level

Other authentication methods not listed are not supported.

4. Releases

This section provides an overview of the new features and bug fixes per release.

4.1 Feature and bug release (2.2.1)

1. Support for capturing the version of BulkMetadataManager into a SharePoint column.
To use this feature add the following line to the BulkMetadataManager
`BulkMetadataManagerVersion=<ContentType name>:<Internal SharePoint column name>`
For example,
`BulkMetadataManagerVersion=Email:Email_x003A_UpdatedByBMMVersion`
The column needs to be created a single line of text.
2. Fix bug in BulkMetadataManager 2.2.0 that fails to update the content type.

4.2 Feature and bug release (2.2.0)

1. Support automatic selection of all sites associated with a hub site.
In the BulkMetadataManager.ini file set the GUID of the hub site as value for the “Sites” property. BulkMetadataManager will process all sites that are associated with the hub site (including nested hub sites).
Note #1. Make sure the SiteClientID or tenantClientID have ‘write’ access to each of the associated sites. This is not necessary when using the tenantClientID.
Note #2. Separate credentials are required when using the hub sites’ GUID because the associated sites are retrieved via search. This is described in section 3.1.4 of the Administrator Guide
2. Support for using an ini file from a different location (e.g., C:\Apps\BulkMetadataManager\ini)
This allows for placing the ini file in a separate directory with controlled access.
Note. To use this feature provide us with the full path to the ini file.
3. Support for specifying the locations for the log files and the temp folder used for processing Office files. *This is controlled by the following properties in the ini file:*
`LogFilePath=C:/Apps/mylogs`
`TempFilePath=C:/Apps/mytemp`
The default behavior is to use the subfolders logs and temp in the current location of the BulkMetadataManager exe file.
4. Support for using multiple content types.
Note #1. The content types need to be specified in the BulkMetadataManagerProperties.txt file in the property DocumentContentTypeName. The values are separated by commas.
For example,
`DocumentContentTypeName=Email,EmailExt`
Note #2. Each of the mapped metadata fields need to be prepended by the content type name plus a : character as separator:
`Subject=Email:Email_x003A_Subject`
`Subject=EmailExt:EmailExt_x003A_Subject`
Each available metadata property can be mapped to multiple content types.
Note #3. Renaming documents using the property DocumentNameSyntaxBulkMetadataManager

is independent from the email content types used. The replacement tags refer to the *BulkMetadataManager* properties like `<Subject>` and not the internal name of the SharePoint column.

5. Minor changes:
 - support for using the replacement tags `<CurrentDate>` and `<CurrentTime>` in the property `DocumentNameSyntaxBulkMetadataManager`

4.3 Feature and bug release (2.1.4)

1. Cater for http 503 responses when retrieving documents or updating the metadata.
2. Handle processing of corrupt zip attachments without crashing the application.
3. Minor changes:
 - include the SharePoint document ID in the log and Whatif log files
 - allow for configurable separator characters for the log file and the Whatif log file. For example, add the following lines to the ini file to use this feature:
`SeparatorCharacterLog=#`
`SeparatorCharacterWhatIF=~`
The default is to use a tab.

4.4 Bug release (2.1.3)

1. Add missing display labels such as “truncated to” and “characters”.

4.5 Feature and bug release (2.1.2)

1. Fix bug with saving incorrect date values into SharePoint columns when email metadata extraction is used and the SharePoint site uses a time zone other than UTC. In previous versions the extracted date values (e.g. Sent Date, Received date, ..) were uploaded in UTC without taking into account the time zone of the site.
2. Fix bug affecting the extraction of emails with multiple pdf or Office attachments resulting in `ERR_UNHANDLED_REJECTION` an error.
3. Include the pdf properties (e.g., keywords) into the list with extracted words.
4. Minor changes:
 - extend error handling for the metadata update response
 - automatically create folders for logs and temp if they do not exist
 - use subfolder (temp) for temporary saving Office attachments
 - remove multiple dots (e.g.,) from the extracted tokens of attachments
 - insert the text “pdf attachment without tokens” into the extracted “Attachment Content” column for pdf attachments without unique words (e.g., scans)
 - show “Average time per document” in ms
 - when a batch request to update metadata fails log the details of the affected emails
 - limit the maximum number of characters extracted from attachments to 100k characters

4.6 Feature release (2.1.1)

1. Support for extracting properties defined as list columns.
BulkMetadataManager 2.1.0 and earlier versions only extracted properties into columns defined at site level. This limitation has been removed. Add the following line to the BulkMetadataManager.ini file to use columns defined at list level:
ContentTypeSource=list
The default value is site.
2. Support for combining the user name plus smtp email address. For example, this will show Paul de Jong <paul@slimapplications.com>
Combining fields is supported for the From, To, CC and BCC fields. To use the feature add one or more of the following lines to the BulkMetadataManager.ini file (or update the tenant properties):
AppendAddressFrom=FromAddress
AppendAddressTo=ToAddress
AppendAddressCC=CCAddress
AppendAddressBCC=BCCAddress
Note: copy the lines as is.
3. Use of a separate support site (<https://<tenant>.sharepoint.com/sites/bulkmetadatamanager>) for storing the metadata mapping file. The file "BulkMetadataManagerProperties.txt" needs to be uploaded to the support site's siteassets library.

4.7 Feature release (2.1.0)

1. Support for indexing attachments in msg emails (optional).
SharePoint Search does not crawl all attachments in msg emails. This results in blind spots in the search index that affects users and security/compliance officers using search or advanced tools like Purview to find content.
BulkMetadataManager extracts unique words from pdf, docx, xlsx, pptx, vsdx, csv and txt email attachments and captures the values into a SharePoint column. This feature also caters for the above file types in (nested) zip attachments.
Searching for a word in an email attachment will then include the msg email in the search result list. This feature allows users and security/compliance officers to search across all existing email content in SharePoint.
2. Support for extracting the number of attachments (optional).
This number includes the files present in zip attachments and msg attachments. It excludes inline images like image001.png that are typically used for logo's and small icons.
3. Support for extracting the attachment names (optional).
The inline image names like image001.png are excluded.
4. Simplified deployment by packaging the BulkMetadataManager application, node.js and required modules as a single exe file.

4.8 Bug release (2.0.7)

1. Fix bug affecting the extraction of Create and Modify dates from email, Office and PDF files.

4.9 Bug release (2.0.6)

1. Fix bug affecting the generation of the log file.

4.10 Feature release (2.0.5)

1. Support for extracting Create and Modify dates from email, Office and PDF files and using the values as SharePoint creation and modification dates.

4.11 Bug release (2.0.4)

1. Fix bug affecting the metadata extraction from multiple document properties into a single common SharePoint column where one or more of the document properties are empty. For example, when both the email's From and FromAddress properties are mapped to a single SharePoint column named Sender and the From property has a value and the FromAddress is empty the resulting value for the Sender was empty.

4.12 Feature and bug fix release (2.0.3)

1. Allow trimming of X500 addresses by using the optional parameter `TruncateX500Address=true`. For example, instead of using the X500 value `/O=OEXCH082/OU=EXCHANGE ADMINISTRATIVE GROUP (FYDIBOHF23SPDLT)/CN=RECIPIENTS/CN=6C83EA5270B446858D39D8713DE2D5A1-someuser` the value `someuser` is extracted.
2. Fix bug preventing the use of sp-request module 3.0.0

4.13 Bug fix release (2.0.2)

1. Same logic as 2.0.1 with the exception that the Office and PDF metadata extraction is not supported. The 2.0.2 version can only be used for extracting properties from msg and eml emails. This allows for using existing configurations working with BulkMetadataManager 1.x. The following modules are therefore not required:
 - npm install xml2js
 - npm install adm-zip
 - npm install dom-parser

4.14 Bug fix release (2.0.1)

1. Use separate credentials for root and tenant credentials. See section 3.1.4 in the Administrator Guide.
2. Extract sent date and received date from emails accepting a meeting.
3. Fix bug affecting the use of `Sites=*` that depends on search to list all sites present in the tenant.

4.15 Feature release (2.0.0)

1. Support for using a single set of Client Id and Client Secret at tenant level.
2. Support for granular control of sites and libraries via wild cards, using specific site/library names and exclusions.
3. Support for extracting metadata from additional document formats:
 - Office files (docx, xlsx, pptx and vsdx)
 - PDF
4. Support for specifying parameters via an .ini file.

4.16 Bug fix release (1.0.13)

1. Change in parameter naming convention to circumvent conflicts with Email Manager. The optional field EmailNameSyntax containing the email naming syntax needs to be changed to EmailNameSyntaxBulkMetadataManager.

4.17 Bug fix release (1.0.11)

1. Support for using a default user account as “Modified By” when the item’s existing “Modified By” value is “SharePoint App”. This feature also circumvents errors when updating the “Modified” and “Modified By” columns using the “Modified By” value app@sharepoint. This username cannot be resolved by SharePoint and will result in an error for both columns. The default user account needs to be specified via the tenant properties or the static properties.txt file and the key is named “ModifiedByDefault”.
2. Fix bug preventing the extraction of email metadata in emails with attached msg files.

4.18 Bug fix release (1.0.10)

1. Circumvent bug in SharePoint’s \$batch requests where URL’s are case sensitive. This avoids http 401 responses with “invalid request”.

4.19 Bug fix release (1.0.9)

1. Fix bug affecting msg files where metadata is present as non-Unicode characters.

4.20 Bug fix release (1.0.8)

1. Use the tenant properties (via StorageEntity) as primary source for configuration settings. If the tenant properties are not available, the application will then try to open the static BulkMetadataManagerProperties.txt file in the SiteAssets library in the root of the tenant.
2. Improved error handling when the specified Content Type is not available.

3. Omit the Client Id's and Client Secrets from the summary log file.

4.21 Bug fix release (1.0.7)

1. Fix bug with processing large number of emails (5000+) in large lists.

4.22 Feature release (1.0.6)

1. Increase the maximum file size for msg file from 50 MB to 75 MB.

4.23 Feature release (1.0.5)

1. Increase the maximum file size for msg file from 25 MB to 50 MB.

4.24 Feature release (1.0.4)

1. Support for keeping the original "Modified" and original "Modified By" details.
Note: this requires an extra request per item and negatively impacts end-to-end throughput. It should only be used when it adds value.
Include "Modified", "Modified By" and size details of the original items in the log file.

4.25 Initial release (1.0.0)

1. Support for SharePoint Online with App Only authentication (using clientID and clientSecret).
2. Support for extracting email properties (From, Subject, Sent, ...) from emails (msg and eml) stored in SharePoint Online and capturing the values into SharePoint columns.
3. WhatIf functionality to test the application without updating the email metadata in SharePoint. The extracted email metadata is written to a csv file for further review and analysis.
4. Select documents for processing using CAML.
5. Optimized performance by using parallel processes, caching, and batch requests.
6. Control performance by setting the number of download and metadata threads.
7. Support for large lists (exceeding list view threshold).
8. Support for continuing after being throttled by SharePoint.
The requests will automatically start after the specified retry period has expired.
9. Logging.
10. Support for Windows Server 2016 and Windows Server 2019.

5. Known Issues

This section lists known issues.

1. The optimum number of parallel threads for downloads and setting metadata depends on several factors:
 - sizes of the documents involved
 - latency between the system running node.js and SharePoint
 - available bandwidth to the SharePoint system
 - time of the run.

The throttling limits seems to depend on the load on SharePoint. During non-Office hours the limits seem to be less strict.

It is recommended to start the application with 1 download and 1 metadata threads and adjust these values based on the findings.

2. Extracting metadata from PDF files using encryption is not supported.
3. The application works with a single content type (configured via the properties file or tenant properties). The application needs to run multiple times if multiple content types need to be used.
4. The log file log_YYYYMMDD_hhmmss.csv contains duplicate entries when a value (e.g., length 500 characters) exceeds the length of a single line of text (255 characters) and is automatically truncated.

SLIM Applications B.V.
Rotterdamseweg 402Y
2629 HH Delft, The Netherlands
www.slimapplications.com
info@slimapplications.com