

Successful Content Migration

When Content Migration Becomes an Issue

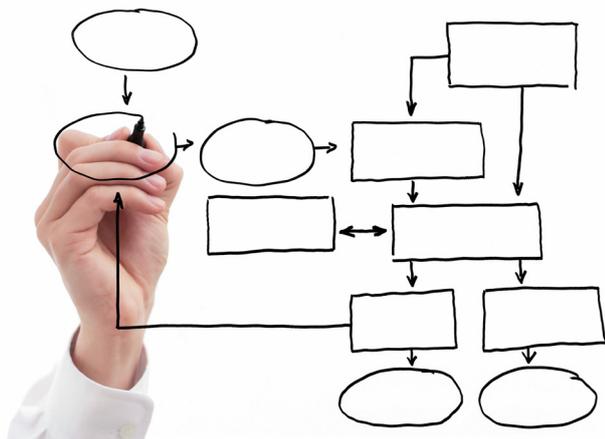
Companies today rely almost entirely on digital information for transacting business with customers and suppliers, for communicating, planning, researching and sharing ideas, for controlling workflows, for record keeping and regulatory compliance; in other words, digital content has become the essential medium for every aspect of day-to-day business operations.

The question though is, how well is this valuable enterprise content organized? How quick and easy is it to get to the right document? How supportive is the storage structure for different critical business processes? And how well protected is all this content? Who is allowed access to it and who isn't?

Enterprise Content Management (ECM) is of course the answer, provided the data is well organized, furnished with adequate rights protection and still accessible enterprise-wide for all who need to view and work with it, each document stored in the right location and tagged with meaningful attributes (metadata) for quick retrieval.

Reality, however, is all too often quite different.

Many departments still keep huge volumes of information in various file share locations with little metadata to indicate purpose and content to any but the most initiated. Certain business areas may have implemented electronic document management of some kind in the past, but without coherent ECM strategy the enterprise ends up with isolated systems, possibly even from different vendors, some of them quite outdated.



Planning your migration project thoroughly is the key to success

What's more, through successive corporate acquisitions over the years the system environment may have become more complex than ever, a costly nightmare for ECM managers to maintain. And for the user the chance of finding anything in a hurry is quite remote. Not only is enterprise-wide search in such an environment impossible,

the structure of the data itself may no longer be adequate for the growing needs of an evolving business reality.

At this point a proper reorganization of the enterprise information base is highly indicated. And to do that, consolidating repositories and mass transfers of documents will be required – a content migration project is born.

What is Really Involved?

When we say content migration – we of course don't simply mean moving masses of data from one medium to another. That might be like poring clusters of unidentified objects from one bucket to another, not much improving the overall situation.

Ideally we need all enterprise data stored and classified in a way that supports operational business needs, easy to be searched, keeping documents under full security control and observing all applicable regulations and compliance rules. In order to achieve this some preparatory work is required:

- A consistent repository structure and attribute model needs to be designed, including a clear taxonomy, naming conventions for every business area, common security and access policies.
- Local and departmental needs have to be observed, legal requirements, government regulations, other compliance issues or records management policies.
- Document types must be defined, folder structures, business units, locations, user groups, retention cycles, maybe even workflows, etc.

Assuming the new central repository has been set up in this fashion, the issue now is moving and storing the different source data from around the company in a way consistent with this newly designed content management model. Here we may be faced with a number of issues:

- The source data to be extracted will probably exist on quite different platforms or systems and in incompatible formats. Is there a single tool to extract and collect all this varied data?

- Most likely, source data will carry little, different or even no metadata. Whatever information can be gained from each document has to be somehow analyzed, transformed and mapped to create the meaningful metadata required by the new unified attribute model.
- What about directing each document to the right folder, attaching the correct access rights and other characteristics? How to handle such things as Virtual Documents, known in some systems?
- What about auditability of all those transfers and transformations?
- On very large data volumes, load speed may become a critical issue.
- And last not least, daily business operations must not be interrupted while all this work is going on. After all, life does not stop because of a content migration project.

Writing your own migration programs will not be a solution for this kind of complexity. And even the majority of the available ETL tools (Extract, Transform and Load) fails on most of the above points or shows at least very significant weaknesses.

This is especially true when it comes to metadata transformation. While some tools can handle simple substitutions, anything requiring complex data analysis and sophisticated logic algorithms to reconstruct the required metadata is usually not possible. This means, for very important document types additional and expensive programming and scripting becomes necessary. The rest of the data is then simply dumped into »legacy data pools« from which individual documents can later only be retrieved via endless full text searches.

Fortunately, quite another story with migration-center!

What's Different About migration-center?

Already the architecture and the key design criteria of migration-center make it apparent how some of the above issues are solved.

Separate source and target adapters

The nucleus of the system, the transformation engine, is completely separate and independent from interacting with source and target systems. Theoretically, therefore, with the appropriate adapters, migration-center can be used for any variety of source and target systems, since the internal number crunching required to process metadata remains the same.

Central object database

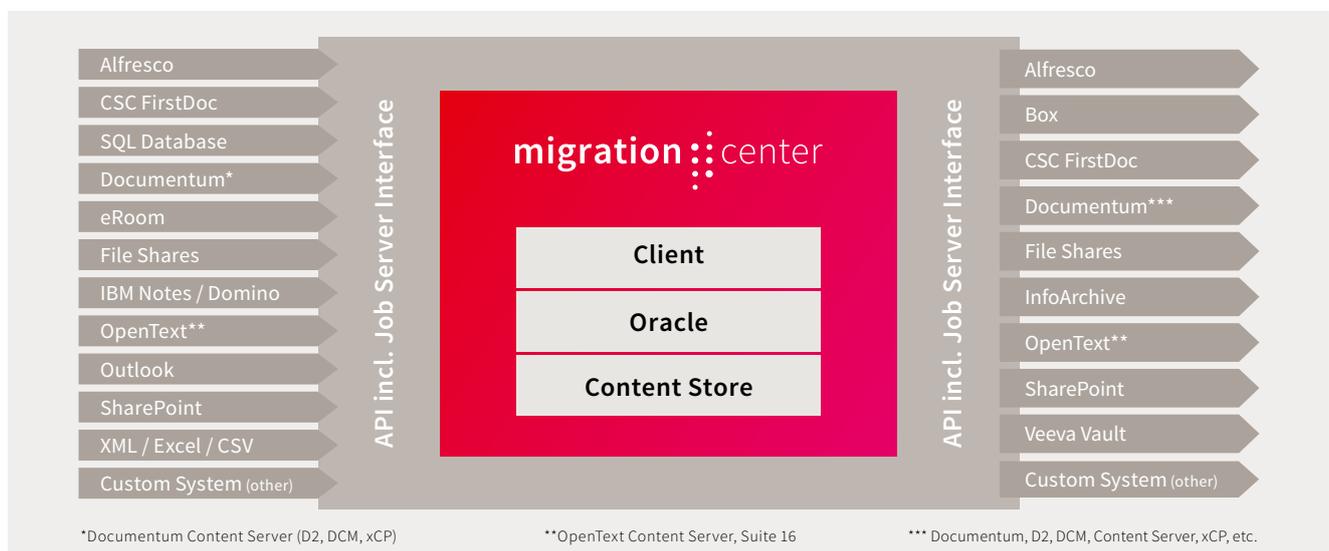
Source data is initially analyzed only and information about documents is recorded and maintained in migration-center's database, along with information about folders, object types, details about the target attribute model, source and target metadata, rules and value lists governing extraction and transformation, a complete audit trail for each object and more.

Intelligent rule based metadata transformation

Metadata transformation is carried out employing a rich collection of commands, algorithms and value lists, which allow the analyst to define different rules for different individual sets of documents, handling metadata construction, automatic assignment of folder structure, access rights and relationships.

Testing and simulating before committing

Those rules are initially run against migration-center's database only, while the actual documents continue to be available on their respective source systems for normal, ongoing business activities. Only after having fully simulated and tested the transformation rules of a given document set, the documents themselves are finally imported into the target system, which is thus being built up batch by batch as the work progresses.



Progressive delta migration

Changes occurring meanwhile in the source systems are being recognized by migration-center and carried over on a regular basis, so that in the final run only a few remaining documents are left to transfer. This delta type of migration approach makes it possible to carry out such a project over weeks without causing any disturbance to daily business operations.

Volume performance and auditability

The migration-center has been especially designed to run at high performance when processing huge data volumes. And of course every activity carried out against a document is left with an entry in the audit trail.

The MC Step by Step Approach

The migration-center is based on best practice principles developed over hundreds of migration projects. Looking at the typical migration process, it consists of a number of clearly defined steps. First the source data is being analyzed as to type, format and metadata. Then the mass of the data is being organized into manageable batches, each containing documents, which can be processed in a similar way, using the same logic. For each batch a set of transformation rules is being created, then tested and validated against the predefined target model. Corrections are being applied and finally the data is being imported into the target system.

In fact, migration-center is built exactly that way. Working with it, the migration analyst follows six clearly defined process steps, which are explained in more detail as follows:

Analyze

Analyzing the nature of the existing data is an essential first step while planning a migration. It will define the approach taken and possibly even lead to further ideas regarding the intended target model.

The migration-center analyzes object and content types, the available metadata, access rights if any, users and groups, it provides statistics, computes check sums for each document and identifies duplicates. Individual source scanners can be configured by parameter (path, query, etc.) and filters allow exclusions (for example: extension = ».log«).

For better understanding, the results should be reviewed with the »owners« of the data, the representatives of the various user departments. Only when the nature of the content is well understood should the next step be taken.

Organize

The data is split into manageable batches or sets, as they are called in migration-center. To import huge data sets in one go takes of course some time. It therefore makes sense to keep them smaller. But the main reason for doing this has to do with metadata transformation. The idea is to identify those types of documents, which can be processed by the same transformation logic. The migration-center lets you define a wide range of selection criteria for organizing your data. Maybe you wish to group all objects of a certain business unit with a particular file extension; all finance documents of a given year; all those with an identical file path; or any other criteria combination.

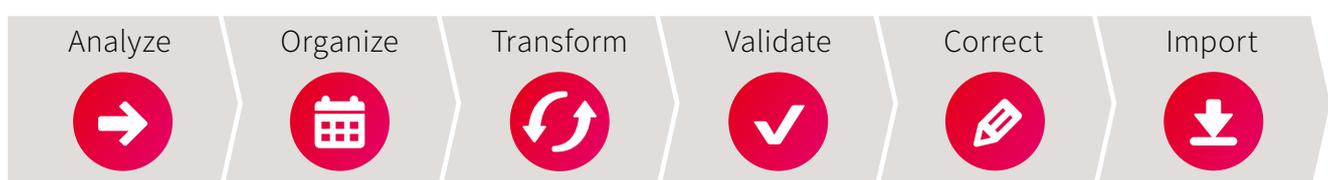
Once the individual sets have been defined and selected, the next steps are carried out for each set separately. Many sets can thus be worked on concurrently and by different analysts at the same time. Some sets may still be in the »Transform« stage, others are being validated or have even been imported already into the repository. This is useful when working in a team where every migration analyst can work on his part of the data at his own pace.

Transform

In this step the migration analyst has the opportunity to modify metadata. In practically every document migration there is a need for modifying or creating metadata. New bulk imports need to be tagged with the correct attributes, legacy objects require metadata modification or a solid cleanup is required during repository consolidation. The rules for doing this are defined in this step.

If the source of a document is a file share location, for example, information besides system attribute can be extracted from the file name, extension and even the file path of the document and then used to create the metadata needed.

Some source metadata may just need to be mapped to a different element on the target system, other may require extensive manipulation or value substitution such as: old document owners are to be replaced; the value list of branch offices has been modified; all documents from project C are now to be assigned to department X and receive also a new access rights classification; depending of the value combination of two input attributes the retention cycle and folder location may be altered. Value lists can be defined in migration-center, but lookups also work against Excel sheets or even an external database table.



The migration-center's rich command set provides functions for almost any kind of manipulation: text concatenation, substring, split string, value list lookup and substitution, procedure and logic definitions. The value result of one function can be used in another. The analyst can even create his own individual functions.

The effect of these specifications can be tested out on the spot and immediately corrected and refined. At this stage one of these manipulations are carried out on real documents or require any real import. Rule definition and testing is executed against the migration-center database only.

This is where, in comparison with other ETL tools, the true power of migration-center is revealed. The rich command set allows not only any kind of even very complex manipulation, it all can be simulated and tested dynamically, then progressively refined. No heavy programming or scripting is required, no lengthy bulk imports to be run, only to find that the script still had some bugs.

With migration-center the analyst has a highly configurable interface to work with on any given document set. He can dynamically view source values and compare them to the results produced by his rules, then correct and refine them until he is satisfied. All of this is done against the migration-center database only, leaving the documents themselves untouched. They continue to be available on their source locations for normal business operations.

Validate

Prior to actual import, a document set as modified during the »Transform« step needs to be validated against the actual target repository in terms of mandatory fields, multi-value fields, naming conventions, object types, field minimum and maximum lengths, customized regular expressions and any other constraint defined in the repository's content model.

This process represents in fact a simulated import run. Any errors, incorrectly filed documents or folders are indicated on a

report. Again, traditional ETL tools do not provide such import simulation. Discovering, after a long import run, that the script was buggy can be very frustrating. With migration-center, each individual document set can be validated within moments.

Correct

During »Correct« the document set is rolled back to »Transform« or »Validate« stage. Error lists can be analyzed and the rules adapted so as to eliminate them. The migration-center offers maximum functionality and ease to facilitate the iterative process of defining rules, testing them dynamically, validating the result against the constraints of the target system and correcting any errors until none are left.

Error lists, for example, can be exported in spread-sheet format and discussed with users in order to determine the best solution for solving metadata issues as they occur. Preferably of course through a rule, but even correcting the metadata of individual documents is possible.

The important thing to remember is that the transformation of any set can always be painlessly rolled back from any stage, even after import, corrected and then recommitted.

Import

Importing documents into the target repository is the final step when handling a migration set. Typically the data is imported set by set whenever they have been validated and proven to be error free.

Especially on large data volumes it is important not to overload the system with too many imports going on at the same time. Therefore a gradual import of documents as the project progresses is advisable in order not to hinder or interrupt any ongoing user activities.

The migration-center's »delta migration« feature makes this approach possible. Through regular re-scans of the source systems, migration-center picks up any new or changed documents to be processed and imported as the project matures, so that in the end only a small amount of documents need to be migrated before the new system goes live.

The migration-center also provides the possibility of time scheduling imports, say at night or over weekends, when few users are active. Import processing can be spread over several job servers in order to get peak performance from the system. And import process can at any point be interrupted and continued at a more convenient time without any negative impact.

Depending on the features of the target system, the migration-center handles folder creation, correct document storage inside the folder structure, metadata, versions, renditions, rights/ACL assignments, links and virtual documents.

Every activity on each document is logged for auditing purposes. And if, despite all previous testing and validation, an error has occurred, any set can be entirely rolled back and the effected documents deleted from the target repository.

ID	Status	Is Update	Processing Message	Content Location	Is In Source System	Is In Target System	Parent Version Object
1966	Validated	0		C:\Temp\1166\990436380003418.doc	990436380003418	990436380003418	
1969	Validated	0		C:\Temp\1166\990436380003418.doc	990436380003418	990436380003418	
1985	Validated	0		C:\Temp\1166\990436380003417.doc	990436380003417	990436380003417	
1999	Validated	0		C:\Temp\1166\990436380003417.doc	990436380003417	990436380003417	
2054	Validated	0		C:\Temp\1166\99043638000340c.doc	99043638000340c	99043638000340c	
2057	Validated	0		C:\Temp\1166\99043638000340c.doc	99043638000340c	99043638000340c	
2068	Validated	0		C:\Temp\1166\99043638000340c.doc	99043638000340c	99043638000340c	
1971	Validated	0		C:\Temp\1166\99043638000341f.xls	99043638000341f	99043638000341f	
1981	Validated	0		C:\Temp\1166\99043638000341f.xls	99043638000341f	99043638000341f	
2063	Validated	0		C:\Temp\1166\99043638000347n.xls	99043638000347n	99043638000347n	
1975	Validated	0		C:\Temp\1166\990436380003423.xls	990436380003423	990436380003423	
1976	Validated	0		C:\Temp\1166\990436380003423.xls	990436380003423	990436380003423	

ID	Status	Is Update	Processing Message	Content Location	Is In Source System	Is In Target System	Parent Version Object
1966	Validation error	0	There are attributes that could not be val	C:\Temp\1166\990436380003413.xls	990436380003413		
2018	Validation error	0	There are attributes that could not be val	C:\Temp\1166\990436380003463.xls	990436380003463		
1946	Validation error	0	There are attributes that could not be val	C:\Temp\1166\99043638000340c.doc	99043638000340c		
1114	Validation error	0	There are attributes that could not be val	C:\Temp\1166\99043638000324n.xls	99043638000324n		
2117	Validation error	0	There are attributes that could not be val	C:\Temp\1166\99043638000326n.xls	99043638000326n		
2074	Validation error	0	There are attributes that could not be val	C:\Temp\1166\990436380003077.mpg	990436380003077		
1976	Validation error	0	There are attributes that could not be val	C:\Temp\1166\990436380003427.doc	990436380003427		
1979	Validation error	0	There are attributes that could not be val	C:\Temp\1166\990436380003427.doc	990436380003427		
1976	Validation error	0	There are attributes that could not be val	C:\Temp\1166\99043638000326n.xls	99043638000326n		

Conclusion and Summary

With Enterprise Content Management becoming more and more the central strategy for enterprise-wide information management there is an increasing demand for a sophisticated content migration tool able to handle a whole range of requirements and use cases.

Moving existing documents into a new ECM application

– this is of course the classic migration project when a user department converts to ECM and needs to move its documents into a well-organized environment.

Consolidating repositories across the enterprise –

keeping different isolated repositories, possibly even by different vendors, is expensive to maintain, requiring too many different skill sets for users and ECM management. Time to consolidate all this into a single enterprise-wide information base.

Replacing older legacy systems – older systems are being phased out and the data integrated into the central repository; data on file share is being brought under control.

Cleaning up your repository – over time business changes occur or a new company acquisition makes it necessary to reorganize the corporate information base. The easiest way to do this is to set up a new repository with a revised content model and to then migrate documents from the old to the new.

Upgrading to a new repository version – running a software upgrade on existing data carries a certain amount of risk to the integrity of the content. It is safer to create a new repository and do a migration. And while you are at it, why not clean up your repository in the process?

Time scheduled batch migration – migration-center can be set up to run on a time schedule for the purpose of regularly picking up and storing system generated documents, say from SAP or any other application.

Test data management – migration-center is ideally suited to create a repository subset for the purposes of testing a new application.

Repository synchronization – a mirror subset of the repository may be created to share documents with business partners. As the content of these folders changes, the migration-center will automatically pick up changed content and move it to the repository subset for external access.



Try out the free version now! www.migration-center.com

Archive migration – The replacement of an out-of-date archiving system by a more modern solution may become necessary. Or certain documents are to be extracted and integrated within an already existing ECM system. The migration-center is designed for mass import. Even certain format conversions can be handled as required.

A powerful ETL-tool for content migration like the migration-center is clearly mandatory in every major

ECM installation today. Its main strength can be summarized as follows:

- Managing large data volumes with good performance
- Central transformation database
- Successive delta migrations ensuring no business interruptions
- Intelligent rule based metadata transformation
- Step-by-step migration and parallel set processing
- Real-time test and import simulation
- Open connectivity for many source and target systems
- Full audit capability for sensitive environments

The migration-center is by all standards the most advanced migration tool available on the market today and has proved its value in countless projects. Several large corporations use it as a standard for all of their migration purposes.



fme group | Germany · Romania · USA

T +49 531 238540 · info@fme.de

T +1 475 3292398 · info@fme-us.com