Whitepaper

6 steps to a successful Web Content Management migration

xillio

liberate. innovate. elevate

Are you considering switching to a new Web Content Management (WCMS or WCM) system to manage your website content? Or have you already implemented a new WCM? Then you will know that a website migration is complex. However, they never happen without a good reason; there are always strong, overarching reasons to embark on such a migration process.

Complex? Yes, a migration of an outdated WCM system to a new system is extremely complex. Whether you are migrating from TYPO3 to Drupal, from Adobe Experience Manager to Eloqua, or from Liferay to Oracle WebCenter Sites, the pitfalls are always the same. There are many aspects you have to take into account before you start a website migration project, such as determining the scope of your project and how to transform content (mapping). That's why a step-by-step plan for a WCM migration will be helpful to you.

Using a detailed step-by-step plan, this whitepaper gives you guidance on how to successfully complete a migration of a web content management system.

Just to clarify, this whitepaper should be helpful if you are:

- Replacing an old WCM by a new one
- In charge of a web content migration
- Looking for a clear WCM migration strategy
- Looking for ways to keep the migration process predictable and controllable.

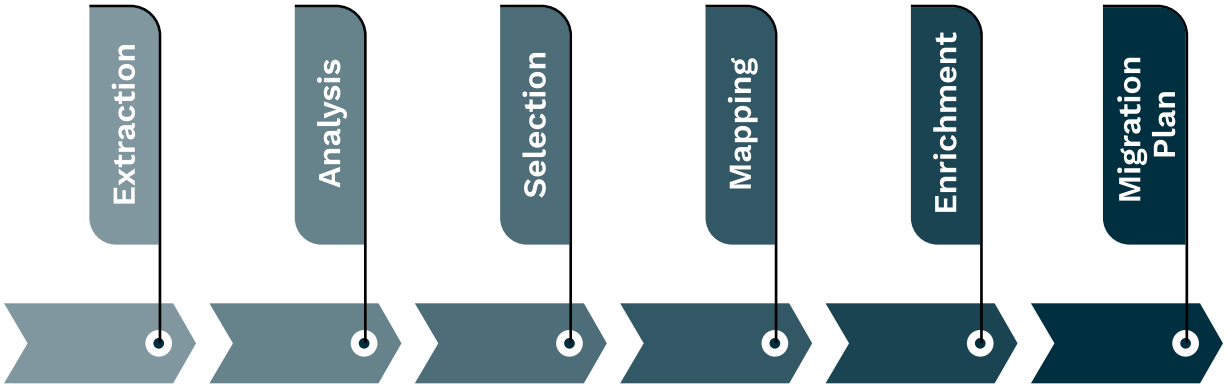
Step-by-step plan

The step-by-step in this whitepaper is a general guideline. Of course, organizations and projects differ and it is essential to tailor the plan to your situation. In addition to different types of organization, there are numerous content management systems and multiple methods of migration.

As we said earlier, our experience tells us that implementing a new website and the associated migration to a new WCM system is rarely simple. It almost never happens that the existing and the new WCM correspond exactly in terms of technique and structure. The migration of an individual page is only the tip of the iceberg. There are many deeper challenges such as importing and connecting documents, creating relations, metadata enrichment, restructuring, cleaning HTML and resolution of the internal links and dividing, merging and connecting of components.

To overcome these challenges, it is necessary that the migration process is performed in a governable and controlled way. Furthermore, the process must be repeatable, since in many cases there will be multiple environments, such as development, test, acceptance and production (DTAP). Following the step-by-step plan in this whitepaper will results in a substantiated and controllable migration strategy that can make the content migration to your new WCM system a success

Steps of a Web Migration:



Step 1: Extraction

To prepare for a content migration to a new WCM, the first step is to get insight into all existing content, or at least all the content that needs to be migrated. This gives basic information about the scale of the project, costs and time-frame. This logical first step is not as easy as you might think. The level of difficulty depends on where within an organization the WCM is maintained, but chances are that there is no easy way to generate a complete overview of the existing content and content structure.

Unfortunately, simply going through the current system and counting files is not an option, as you may have multiple versions of the same document, or files that are no longer live on the website may still exist in the source system. Also, some systems build pages dynamically from multiple reusable components and this makes getting a complete view of the situation even more complicated.

Extraction of content types

Fortunately, there is tooling available that can extract all website pages and content (e.g. content types, content, relations, metadata, location) from both front and back-end of the WCM system. Occasionally, the website contains important information that cannot directly be derived from the WCM – for instance, when the navigation is built by the site according to rules and dependencies. On the other hand, the WCM can contain metadata or unpublished content (for example, seasonal content or content only used in case of a crisis) that is not available on the front end or visitor's view of the website.

Inventory of content types

During the extraction phase, all pages and content types are extracted, including relevant information such as:

Metadata

All metadata that is associated with the content item in the current CMS, such as the date, publication state or language

Fields and structure

What are the fields that make up a page or content type? What are the associated formats, syntax, value lists, restrictions and dependencies in place?

Relations

What content links to this item and to what content does this item refer?

Location

Where is the content item found on the site? Is it part of the navigation and does it occur on more locations?

Inventory of documents

Aside from text pages, an inventory must be made of all documents. This includes files such as Microsoft Office, XLS, PDF, images and video files. At least the metadata, relations and locations must be extracted and documented for each document.

The extracted content in this phase can be used to generate several reports. See step 2 for more information.

Important tips for step 1:

- Try to get a complete overview of all existing content;
- For a full inventory, extract both the back-end site of the WCM and the visitor's view of the site;
- Generate reports on the inventory and share with the content owners (see step 2).

Step 2: Analysis

After the extraction and documentation of all website content, the extracted data can be used to generate several standard or client-specific reports. These reports give information about the current content, structure, metadata and links on, for example:

- Content structure front + URL's
- Content types, fields, syntax and value lists
- Relations between pages, for phased implementations and go-live
- Totals per content type
- Metadata
- Reusable content
- Google Analytics insights
- Specific content types such as CTA

These analyses and reports can be used to make a number of decisions. First, to offer insight into the current situation and to make decisions about cleaning up content. This can be determined in workshops with the project stake holders and content owners. Second, it is a perfect starting point for defining the content types and templates of the new systems, as the analysis contains critical information when deciding on the new content structure. Finally, these reports are often used to determine the selection and migration rules. See step 3 for more information about selection rules.

An advantage of these reports is that they can be actualized at any time in the migration process or expanded with new insights. In addition, all properties of the content are saved, which means that in addition to the standard reports, very specific reports or new cross-sections can be generated

Important tips for step 2:

- Use the extracted data to analyze the current content landscape;
- Involve the original content owners;
- Use report to decide on cleaning, content structure and migration rules

Step 3: Selection

A migration is an excellent opportunity to clean up content. Such a clean-up will result in less content that needs to be migrated to a new environment and insights in the resources needed to perform the actual migration. To perform a clean-up, you need to define the rules for deleting your Redundant, Outdated, and Trivial (ROT) information. For smaller sets of content, rules may not be needed and evaluating/identifying each page individually from spreadsheets listing all content may work.

But for large website migration it's better to automate this process and generate an Excel file, for example, with the complete structure of the website, including metadata (author, age, versions) and filter on the different selections. The migration rules can be created at different levels. For example, at a high level: all pages from a certain section of the site do not have to be migrated or move to a new location. Or all pages of the content type 'newsitem' are deleted. On lower level: per section or per page is indicated which content (including metadata) has to be transferred.

Example rules

One rule could be that everything that an intern entered three years ago can be safely deleted or that website pages that have not been looked at in the last year can be archived. In the rules set below we see that content that has been created before 1st of January 2010 will be moved to an archive.

Property	Rule
Date	Creation Date is before january 1st 2010

Action	Results
Move to archive library	41.224 of the 165.321 documents are impacted

Since all content is stored in the extraction phase, we can apply these migration rules directly on the content. This way, even before the migration has started the impact on, for instance, numbers of pages per content type, broken relations and content structure is clear.

Important tips for step 3:

- Carefully confirm that the data upon which you rely is reasonable before you start deleting;
- Test the migration rules to know the impact on the project.

Step 4: Mapping

The mapping phase is a logical continuation of the previous phases. During the extraction, the content and the content structure of the current situation are documented; during the mapping phase the current situation is placed next to the new situation. What are the similarities, what are the differences with regard to structure, document types and fields, and how are we going to handle these differences? This mapping is used as input for which content can and cannot be migrated automatically, where editors may have to supplement or correct or what has to be migrated manually.

If, for example, it is already known that the new WCM has new mandatory fields for a specific content type, then the approach to fill out these fields can be determined. In this case, editors have the following options:

1. Enter default values
2. Complete manually in Excel lists
3. Draw up rules to automatically assign metadata
4. Automatically assign metadata based on the content of files with an auto-classification tool

A prerequisite of the mapping phase is that the new content structure must be available as documentation or in a live environment. This means that the templates for the new WCM are determined in a technical design or that the development has already started.

Mapping the content types

The mapping should be done on both content type and field level. When mapping on content type attention must be paid to the following:

- Will a content type be migrated 1-on-1?
  For instance, content type A moves to content type A
- Will multiple content types be combined?
  For instance, content type A and content type B merge into content type C.
- Will content types no longer exist?
  When it is decided that content type A will no longer exist in the new situation.
- A part of the items from a content type will be migrated.
  A mapping rule is created that determines which content will be migrated, for example content that hasn't been touched since 2012 will not be migrated.

Field mapping

After mapping the content types, mapping is done on field level. For each field in the current situation it must be decided to what field in the new WCM it will be migrated. This is an intensive step but it pays off during the migration path and the project itself as it identifies possible issues and challenges early in the process. On the field level, the following details will be reviewed:

- Will a field be migrated 1-on-1?
  For instance, field A migrates to field A. Per field, the format and the list of possible values of the current and the new situation must be compared. For example, a publication date can be saved in a different format in the current system than is the case in the new WCM.
- Will several fields be merged or alternatively be split?
  Often the current system contains a block of text that must be split in separate paragraphs in the new WCM or the other way around. During the mapping, decisions must be made which rules must be applied to be able to split or merge fields.
- Will a field disappear?
  When it is decided that a field will not return in the new situation, this must be documented in the mapping rules.
- The new WCM contains a field that is not available in the current situation.
  It is possible that the new WCM contains fields that are not available in the old WCM.
  When a field is required, then a mapping rule needs to be created on how to deal with this situation

Image: Example of field mapping

Current WCMS	Target WCMS	Remarks
Title	Title	No changes
Date	Date	Syntax dd/MM/yyy
Body	Paragraphs	Split based on migration rules
Direct links to PDF	Page for each document	Create page for each and link to document
Links below body content	Related Links	Move links to related links
Subject from site breadcrumb	DC: subject	Mandatory in target enviroment

Retaining links

One of the most important aspects of a migration is retaining the internal links. During the mapping phase, a strategy must be implemented to retain the links and relations after the actual migration. During the migration, the links and content items must be registered and in the new WCM the items must be converted to relative URLs, or connected based on the IDs of the relevant CMS.

End result

The end result of the mapping phase is a document showing a listing of the content types and fields of the current situation in one column, and a second column with the content types and fields in the new situation with a list with mapping rules (see illustration above). This will be the input for the following phase in which the content will be enriched with metadata and migrated.

The information in this document can also be used for reporting and logging. It's a useful report of the old and new location of pages; it shows what content has not been migrated, what content cannot be migrated manually, or which new features and new field lack content and have to be filed out manually. It's better to have this information now then after the migration.

Important tips for step 4:

- Make sure the content structure of the new WCM is defined in advance of the mapping
- Use the information received in this stage for reporting and logging

Step 5: Enrichment

Metadata contains important information about the properties of the content and the explicit registration of metadata has the advantage that content is easier to find. Missing or incorrect/incomplete metadata can still be added in the enrichment phase. There are four techniques available to enrich content with metadata:

- Metadata from the context
  Based on the inventory results, migration rules can be created to collect or determine metadata based on the context. For instance, using the breadcrumbs, the navigation, the related content or by getting insight in the places where content is used, metadata can be automatically assigned.
- Metadata from other sources
  It is possible to create a connection with other systems to look up metadata and combine them during a migration.
- Metadata from files
  Metadata can be retrieved from other files such as web pages, Excel, Word and PDF documents, databases, etc. In this way, it is not only possible to bring across existing data, but also to enhance the collection of metadata.
- Metadata based on patterns
  Metadata can be retrieved by looking at specific patterns or text in the contents of the document.

Step 6: Migration plan

All the steps finally result in a migration plan. This migration plan contains, among other things, information about:

- What content is being automatically migrated, what not or partially?
- How can software facilitate the editorial staff in manual work?
- Detailed planning of the work to be performed
- Suggestions to split up the migration to make it manageable.

The migration process touches many parts of the organization or the project team. For example, during the configuration of the transformation rules there is a lot of alignment with the implementation team about functionality and final releases. And during the execution of the test and acceptance migrations, there is a lot of contact with (functional) management, testers and editors. These dependencies are recorded in the detailed planning in order to increase controllability and predictability.

The actual migration

Migrations can roughly be divided in three types. Depending on the migration strategy chosen, one of the following methods will be used:

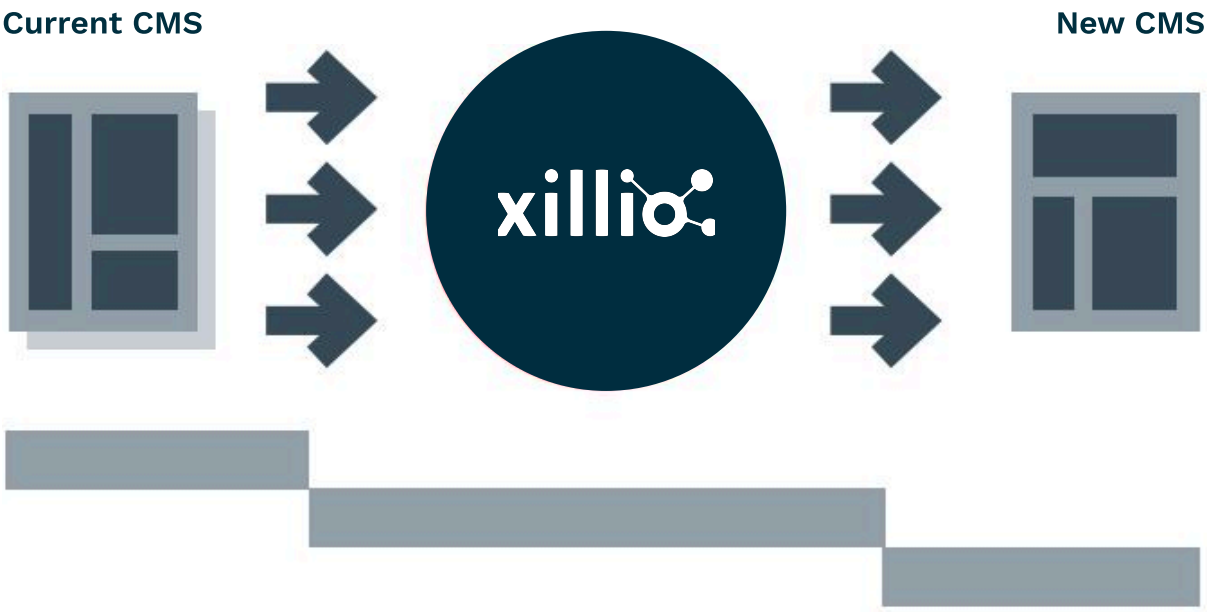
1. One-time migration

The content will be migrated in one iteration. Preceding the migration, some test and evaluation migrations will take place, but the final migration will take place in one big bang.

One - time migration

2. Step-by-step migration

The current website will be logically divided and migrated piece by piece, for example per section or per content type. This creates extra time for the performance of the migration tests and training the organization and will make the migration less dependent on the setup of the new WCM and the introduction of new functionalities.



Step-by-step migration

3. Synchronization/integration

When the current CMS will be available for a certain period parallel to the new CMS, it is possible to synchronize the content. This means that a change, made by an editor in the old CMS, will automatically be migrated to the new CMS and published on the site.

xillio

liberate. innovate. elevate.

Want to learn more or schedule a session with our experts? Contact us:

sales@xillio.com

+31 (0)35 622 95 45

www.xillio.com

www.linkedin.com/company/xillio/