Content Migration

Seven Steps to Success

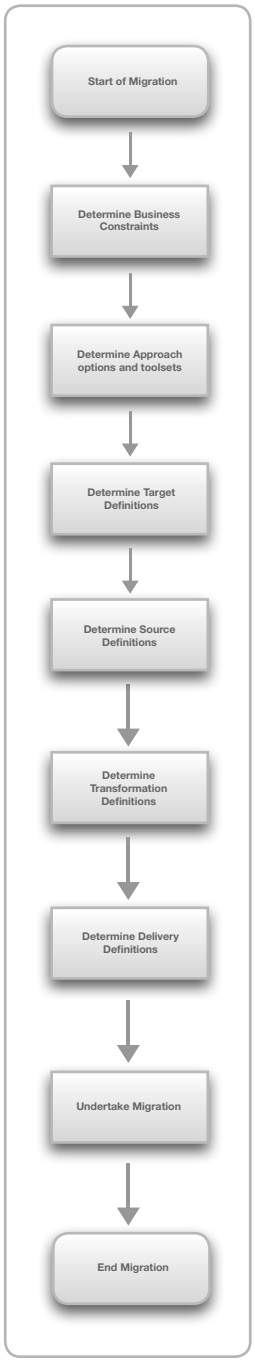
Contents

Content Migration: Seven Steps to Success

The migration of web pages or documents into a Content Management System (CMS) is often undertaken as a last minute exercise. This white paper outlines the business benefits of taking a planned approach to content migrations.

Migration Roadmap

Assurances from vendors suggest there are no issues associated with migrating content when implementing a successful CMS strategy.



Like many other aspects of IT implementations it is the correctness and validity of data entered into a system that determines its success or failure. As CMS technologies mature a successful and proven migration methodology is key in determining the suitability of one vendors product over another.

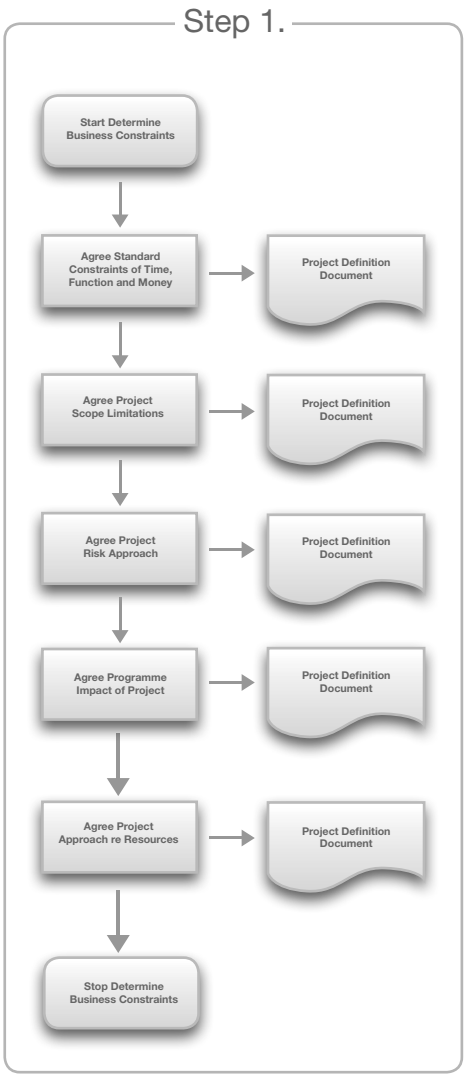
The concept of migrating legacy data is easy to understand - as is visualising the end result. What proves increasingly difficult is understanding the processes involved in moving data from a legacy environment into the target CMS.

A roadmap is needed with simple rules showing how data migration is achieved. This white paper presents that roadmap and defines the implementation steps required in order to achieve a successful migration project.

The roadmap has a number of distinct steps and following them can ensure a successful migration. Each step, or project phase, has an associated flowchart outlining the specific processes of each.

Step 1: Determine Business Constraints

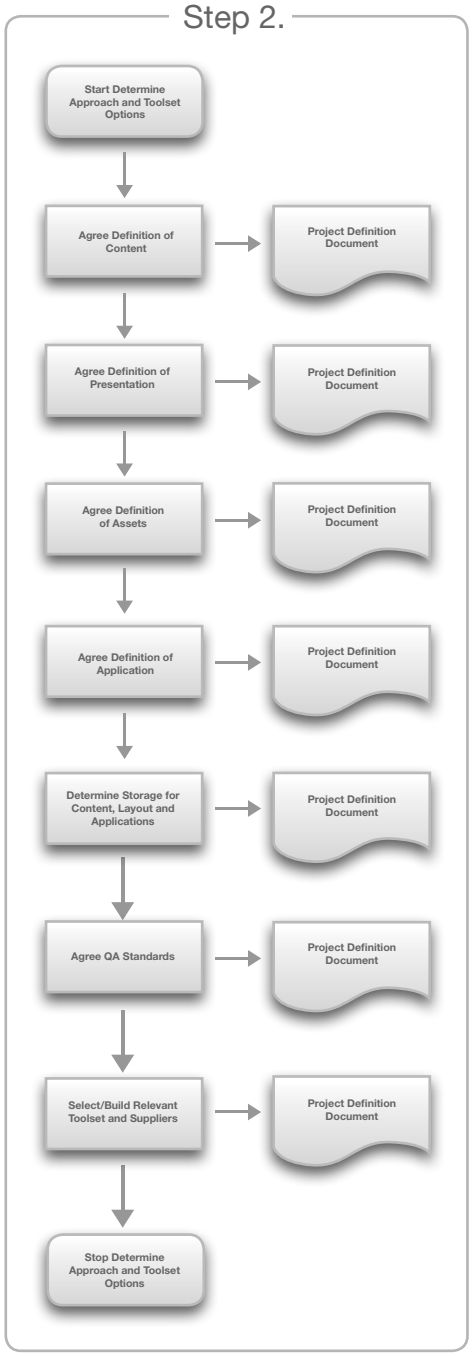
As with all projects it is necessary to outline the migration business case and the 'Time, Function, Money' constraints it must operate within. These constraints require embedding within the philosophy of the migration and are reflected throughout its operation.



The start of the planning process is the time to clearly define the measurements for success. Remember that the migration does not impinge into the area of business change – instead the overall business change project actively sets the boundaries of the migration.

Step 2: Determine Approach & Toolset

The business constraints always determine the project approach and which migration toolset to use. If the volume of migrating data is high then an automated approach is needed. Smaller volumes, typically less than 5000 pages, can migrate manually. While both manual and automated have a number of options, automated migrations are the most complex to manage successfully.



The options for automated migrations break down into four distinct types:

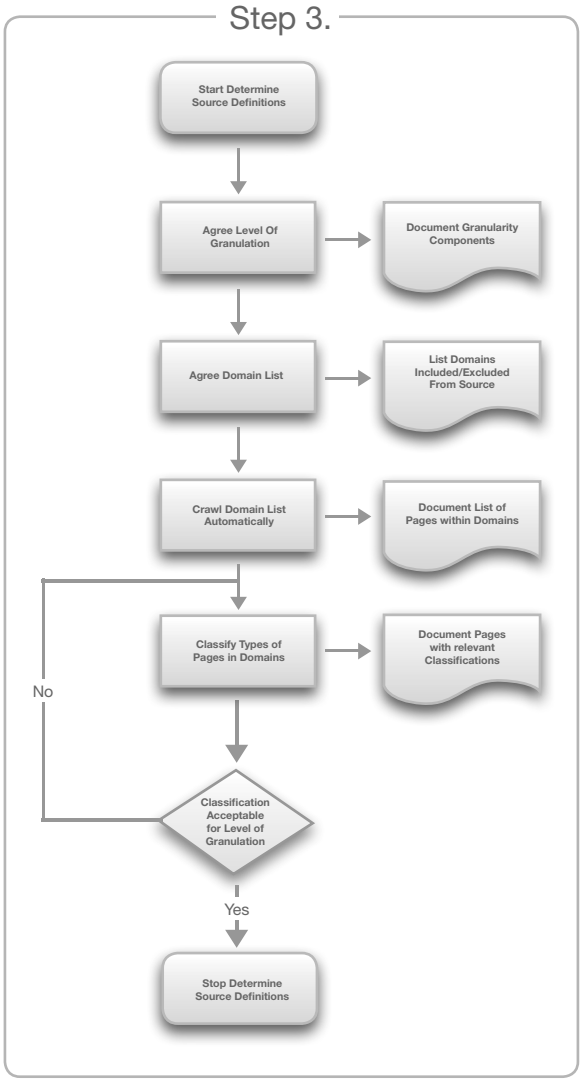
- **As-is Migration** - source pages and assets migrate with no change to their nature or additional information added to their definition.
- **Enhanced Migration** - source pages have additional metadata properties created, often as a requirement of the target CMS system.
- **Standardization Migration** - source pages break down into smaller levels of granularity which migrate and reassemble in the target CMS. This is often referred to as 'applying standard templates'.
- **Restructure Migration** - as pages migrate the navigation elements are updated to reflect a new Information Architecture (IA)

It is important to ensure the original project constraints are built into these decisions. As migrations progress they tend to attract a growing requirements list. This scope creep occurs if the original constraints and the business requirements they are founded upon are ignored. It is important to tie back the choices for which type of migration is undertaken based on the original constraints and the success criteria designed for them.

Step 3: Determine Source Definitions

The next step in the Migration Roadmap determines the following key source definitions:

1. In Scope Domains
2. Information Architecture (IA)
3. In Scope Pages and Assets
4. Number of Page Classifications
5. Available Granularity by Classification



The overall complexity of the migration is in part determined by these source definitions. Identifying the domains in scope for migration and defining the structure of the source IA is the first of five key areas.

The third area relates to specific pages and assets. All images and documents that link from web pages are classed as assets and are handled differently within the migration process. It is essential to clearly identify all in scope pages and assets within each of the domains.

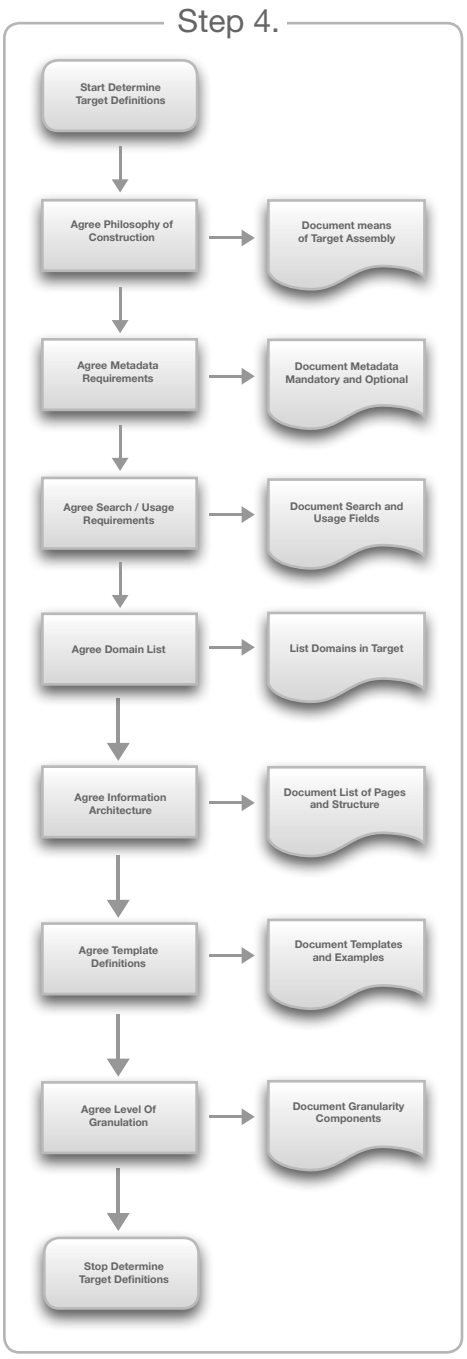
The importance of page classification and associated levels of granularity in a migration is often misunderstood. Put simply, page classification breaks down structured or unstructured html large volumes of pages into a number of common areas or templates.

Classification ignores any page furniture or navigation areas and identifies the most common presentation layer across the migrating content. The volume of pages that fall outside the common presentation layer are identified this allows the migration team to plan how to process these non-standard areas most effectively.

Step 4: Determine Target Definitions

Having defined the source definitions it is necessary to match these with the target environment. It is safe to assume that the target system requirements are greater than those enabled in the source system. The target definitions are:

1. Domain Hierarchy
2. IA Structure
3. Template Definitions
4. Granularity Specifications
5. Metadata Requirements



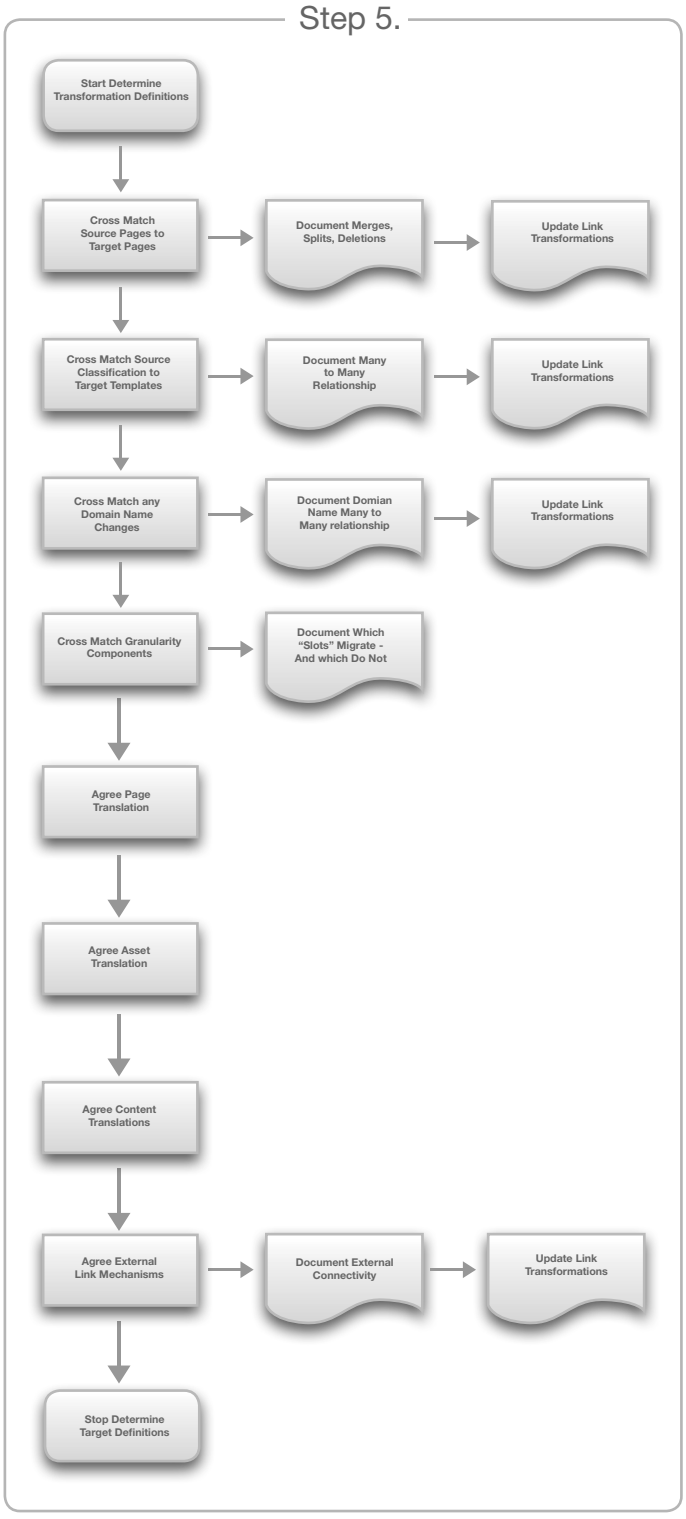
Here the specific requirements of the selected CMS are defined in detail. These may include naming conventions, ownership, security, expiry dates, classifications or taxonomies.

Definition of the target system determines the extent and nature of the transformation requirements of the source content. Unfortunately, it is easy to design a complex solution that is impossible to implement successfully.

Step 5: Define Transformation Requirements

Having defined both the source and target systems it is now possible to design the content transformation rules. The rules must address all aspects of the target system as being the primary recipient of the legacy data.

The target system requirements always drive the definition and design of the rules — working backwards from the target to the source system. The approach is to ensure the source data fits the target data and not to change the target requirements to suit the source.



The source system cannot always supply relevant data to populate the target system correctly. In these instances the rule definitions must include effective auto-generation logic. An exception handling formula that addresses these instances of inappropriate source data or incomplete auto-generation is also required.

In a simple example the target system may restrict the 'Title' field length to less than 50 characters. How does the migration transform source titles of more than 50 characters? Options might include truncating the 'Title' field length to 50 characters or flagging the data element as 'not to migrate' until the correct rule handling is defined.

Transformation rules break down into the following key areas:

1. Domain Name
2. Information Architecture
3. Classification Match to Template
4. Granularity Cross Matching
5. Page Changes
6. Asset Requirements
7. Content Enhancement

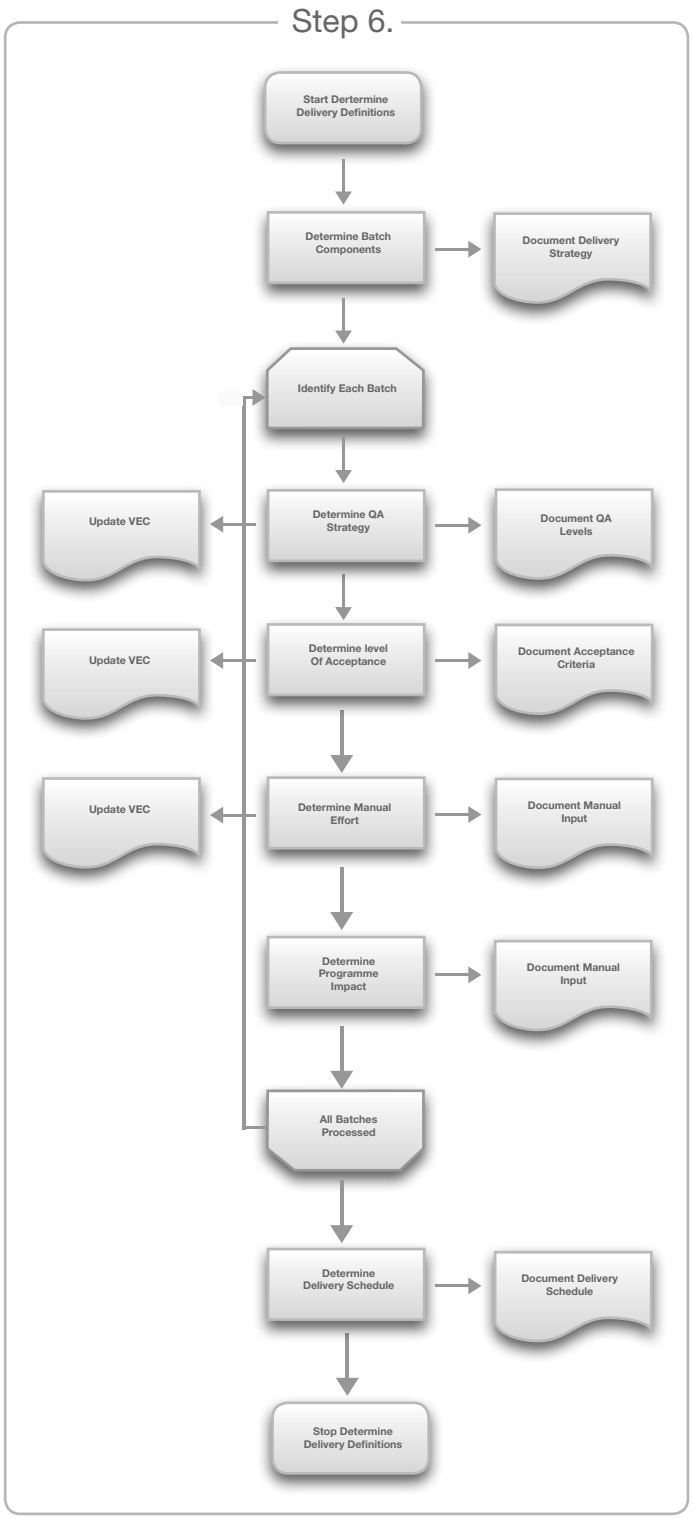
It is important to fully define and test all transformation rules prior to coding. Deskchecking actual pages identified during the classification process proves invaluable— ensuring the rules give the expected results when applied.

Each rule has an associated Verification Exception and Correction (VEC) document. The VEC determines the test verification criteria, the identified exception and the required corrective action.

In the previous example the verification checks that the 'Title' field conforms to requirements, the exception is a length greater than 50 characters and the correction truncates the title to the first 50 characters and deletes the rest. This is incorporated into the transformation rule action.

Step 6: Determine Execution Options

When the volume of objects is small executing a migration in one batch is relatively straightforward. The QA effort required to determine the success of the project is manageable.



As volumes increase the need for executing in a number of smaller batches arises. Whilst undertaking a migration of one million pages is possible in one large batch operation, implementing the QA function may prove very difficult.

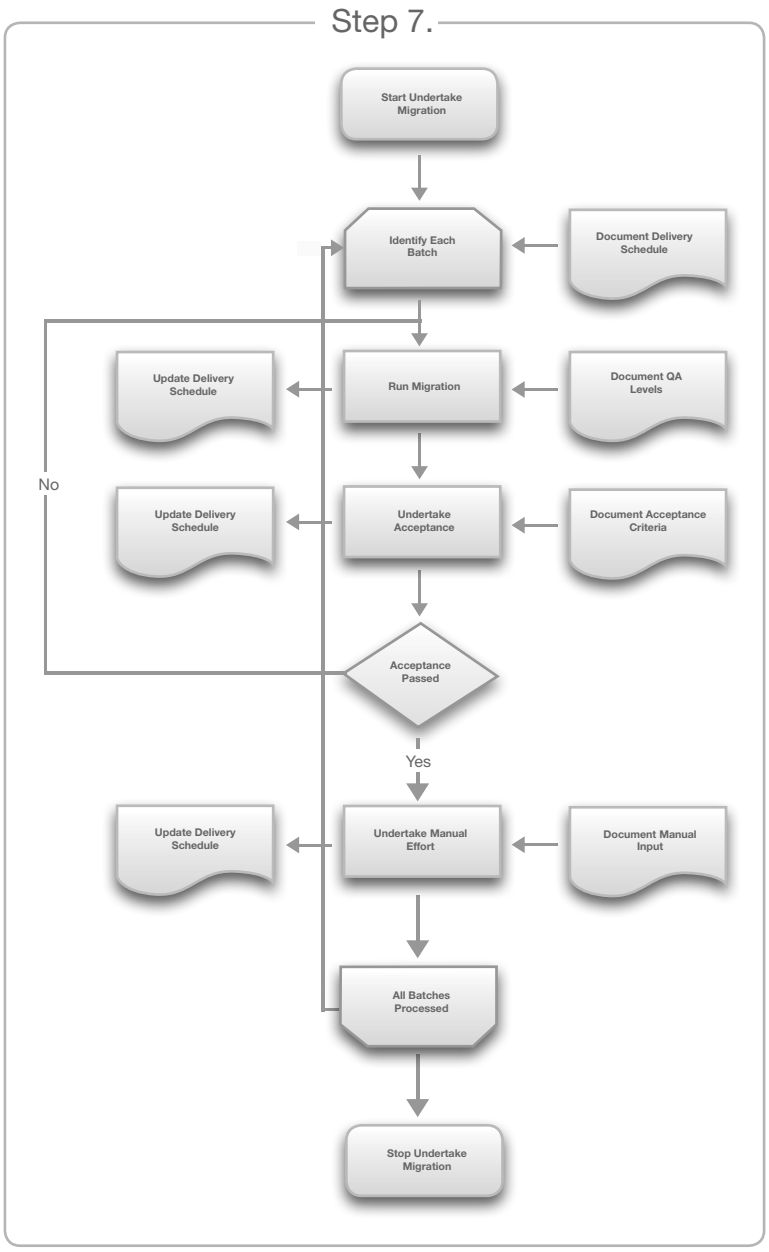
Part of 'Step 2 — Determine Approach and Toolset' identifies the exact QA requirements for the project. At this stage it is necessary to design an approach that enables the QA implementation.

This typically involves breaking the migration down into smaller manageable sub-projects that migrate in either a parallel or sequential manner.

If so, it is crucial to ensure that the co-relation of link items is maintained across all the sub-projects. The toolset used to automate the migration must support this dynamic link handling in order to maintain the target IA structure.

Step 7: Undertake Migration

The actual undertaking of the migration is now possible. All the previous definitions give an operational framework reflecting both the source information and the target system requirements.



Having defined both the 'as-is' and 'to-be' requirements, transformation rules and exception handling, combined with the QA process all the technical elements are in place to proceed.

If additional facts come to light during the migration execution, e.g. unforeseen data supplies or values, this is documented within updated source definitions and captured in modified transformation requirements.

Matching of the final migration output to the success criteria is extremely important. If a migration does fail it is because the previously defined success criteria, based upon the transformation rules, are not met.

In summary, the Migration Roadmap is vital to ensure project success. With objectives clearly stated and rules clearly defined there is little room for error.

About Vamosa

Vamosa is a sector-defining software and solutions company specializing in the emerging area of Enterprise Content Governance (ECoG). Recognized as a global leader in ECoG, Vamosa is the only company that offers an integrated suite of products and services to manage the end-to-end process of discovering, transforming, deploying and maintaining content quality. Our solutions allow the world's largest enterprises and governments to analyze, enhance, standardize, monitor and maintain all forms of business content within a policy-based infrastructure of best practice methods.

Organizations with Vamosa's ECoG solutions benefit from improved regulatory compliance, enhanced knowledge management and a lower Total Cost of Ownership (TCO) for content management. We have provided solutions and consultancy for content discovery, cleansing, migration and maintenance to some of the world's largest enterprises and government departments – including Cable & Wireless, Vodafone, IBM, Oracle and Volkswagen North America.

Vamosa has quickly established itself as the sector-defining ECoG provider. Large multinational companies and government organizations in Europe and North America are relying on Vamosa to solve content quality challenges.

Our unique solutions for both documents and web content include:

- **Vamosa Content Analyzer** - enabling organizations to gain a clear understanding of their current content inventories and content management activities
- **Vamosa Content Cleanser** - for enhancing content by automating classification and tagging while removing duplicates and managing version control
- **Vamosa Content Migrator** - for simple, automated, web content and document migration
- **Vamosa Content Quality Builder** - a high performance rules based content transformation engine to address complex content quality process challenges
- **Vamosa Check and Fix** - the world's only web monitoring tool that also fixes problems automatically, for guaranteeing governance, accessibility and brand monitoring
- **MetaPoint** - which tags document content in an MS Office/ Sharepoint environment at the point of publication ensuring that it is readily found, shared, and stored according to corporate information management guidelines
- **Vamosa Expert Services** - to ensure industry best practices are tailored to the unique needs of every business

Find out more

To find out how your organization can ensure good content quality and reap the benefits of effective Enterprise Content Governance, call **+44 (0)141 574 2500 (EMEA)** or **(617) 273 8398 (US)**. Alternatively, **visit www.vamosa.com**.

Europe

Glasgow
1st Floor, Ailsa Court
121 West Regent Street
Glasgow G2 2SD

Tel. +44 (0)141 574 2500
Fax. +44 (0)141 574 2501

London
Regents Place
338 Euston Road
London NW1 3BT

Tel. +44 (0)207 544 8661
Fax. +44 (0)207 544 8401

North America

Boston
4th Floor
470 Atlantic Avenue
Boston MA, 02210

Tel. (617) 273 8398
Fax. (617) 249 0461

New York
19th Floor
5 Penn Plaza
New York, NY 10001

Tel. (646) 378 2269