

Key questions rarely asked about website content migration.

Benjamin Franklin said, “By failing to prepare, you are preparing to fail.” Obviously Ben wasn’t thinking about migrating web content from one CMS to another. But anyone who has spent time in content management implementation knows how his words apply here.

The reality may surprise many: Content migration is not about tools; it’s about preparation. I can assure you that some important questions are never considered. The first mega-question should be obvious: How do you know how to plan for something that you have never experienced? If it isn’t part of your experience, you simply don’t know what you don’t know.

#1: What are the metrics for success and value drivers for the project?

This is distinct from the CMS implementation and the redesign. Simple: What will success look like? Are the metrics about defect rate? Performance against schedule? Other metrics? This step is especially important when working with an agency or migration tool vendor. More on that on some later post.

#2: What is the QA threshold for the migrated content?

Do you plan to spot check your content using a representative sampling of migrated content, or will you need to check every single page for formatting and content issues? That’s an important decision. QA is a time intensive process, so this can have significant impact on the time line.

#3: How long can the site be frozen?

Typically the legacy site must be frozen to any updates during content migration to ensure that the latest content is successfully migrated. What is the maximum freeze time you can accept?

#4: What are the major milestones and time lines for the CMS implementation and redesign?

A migration is dependent in several ways on the CMS implementation and redesign. Some examples:

- Wire frames completed
- New detailed sitemap completed
- CMS implemented
- New templates coded
- CMS configured (with templating, workflow, meta tags etc.)
- Content migrated
- Site launch

#5: What is the total file count?

You’ll need a complete breakdown of all of the assets in the system, either in database or spreadsheet/csv format. This breakdown should list all files by category, such as HTML, JSP, SWF, JPG. Don’t miss any categories.

#6: What are the shared files?

As above, a complete list of shared files is needed. This includes files such as JS and CSS as well as server side includes (SSIs) and content items in the CMS that are shared (typically shared navigation, spotlights and ads, right hand channel boxes, etc.).

#7: What meta data is required versus optional in the new CMS?

This complete list of meta data needed in

Endstate Solutions

808 Bidwell St.

Suite E.

Folsom, CA 95630

888.569.2982

info@endstate.com

the new CMS system should identify new meta data fields (those not present in the old CMS). A spreadsheet should identify the number of files that are missing their required meta data, listed by each meta data field. The meta data needs to be identified as “human supplied,” “automated,” or “assigned by rule.” Each meta data field should also be identified as “free form,” “controlled vocabulary,” or “system generated.”

#8: What are the legacy content schemas?

In the simplest case this means the semantically transparent XML files that sit behind every piece of content in the legacy CMS. The exercise then is to identify all of the allowed elements for each schema. In the more typical and complex case, the content in the legacy system is not stored as semantically transparent XML, but rather it is stored as HTML fragments inside of the XML elements that make up the various content templates in the legacy CMS. The goal in this case is to identify all of the different permutations of HTML code that make up the legacy content.

#9: What are the target content schemas in the new system?

Again, in the simplest of cases you will migrate into semantically transparent XML schemas. The exercise is to identify the elements that make up the different schemas. In this more typical case, however, you will need to identify the new templates based on the redesign, especially noting where the coding style differs from the legacy content, such as table based layouts versus CSS based layouts. You will also need to identify the allowed elements for these new schemas, too.

#10: Will any new content need to be created?

Often a web site redesign, including an IA change, requires new content creation. This

can also include images converted to text. If new content needs to be created, will this be done before the migration project?

#11: What are the languages of the content?

Does the estimated content count include local language versions? If double byte characters are used, what are the different encodings of the legacy content? What is the localization process in the legacy CMS?

#12: Do you need to migrate the historical versions of the content?

Typically a CMS will include version control. Must past content versions be migrated into the new CMS?

#13: What features of the legacy and new CMS are being used?

Each CMS provides its own set of features and functionality. Which ones are used in the legacy CMS and which will be used in the new CMS?

#14: Can all content needing migration be accessed externally?

Will an agency need to get behind a firewall or VPN connection to access any of the content needing migration, including access to both the legacy and new CMS?

#15: Do you have a complete site map of the legacy sites?

You will need to generate a complete site map, including all of the unique URLs on the site. This should include a note on each URL to identify if it is a static URL or database driven.

#16: Will redirects need to be created?

If the IA is changing, it’s likely that the URLs will change and redirects will need to be put in place. To what extent do you need to preserve your SEO, user bookmarks, and in-bound links from sites and advertisements?

The key here is to understand if you will use domain, category, or file level redirects.

#17: Will content be referenced in the system through unique document IDs?

Will the new CMS assign unique IDs to all of the content assets and require their use when linking to content within other content (for example, links to PDF files or embedded images)?

#18: How does your legacy content link back and forth (both SRC and HREF attributes)?

Does current content use Relative, Root Relative, or Absolute linking? Does the content use URL aliases or CMS-generated unique identifiers? This question directly influences your link resolution process when you move the old content into the new CMS.

Eighteen big questions to ask. Miss one, and you may not hit your project deadline, your budget, or your management's dream image of the new website.