KM COLUMN

Beyond the standard content audit

Content is at the heart of all intranets. Pages, documents, and files populate the architecture with varying degrees of elegance and are managed under different authoring and publishing regimes.

Over time, an organic process somewhere between glacial and viral overtakes the well designed structures (assuming they were there to begin with) and starts to clog the pathways and interfere with the effectiveness of the intranet. You may have watched this process over time or simply inherited it because of a new role, perspective or sense of responsibility.

Regardless of how you arrived here, you are faced with a significant quantity of content, but with little visibility of its structure and make up. Diving into a redesign without this visibility is both daunting and ill-advised.

But where to start?

The standard approach is to list all of the content, carve it up, distribute among assumed owners, remove the unwanted, and after several iterations of this, recombine the list. Valuable though this is in identifying ownership and reducing the volume of content, it many not have added any insight around existing or proposed structures.

When managing larger audits with many hundreds, thousands or even tens of thousands of items, it becomes increasingly valuable to overlay the standard process with a process to manage the audit at different degrees of granularity. Not only does this provide visibility in a dynamic way, it also simplifies the steps when conducting the design steps to follow.

Stephen Byrne is an organisational psychologist and senior consultant at Step Two Designs, an intranet and content management consultancy based in Sydney, Australia. Stephen specialises in intranet strategy and design, knowledge practices and organisational change.

By imposing structure on your content, you gain insight

Going beyond the standard audit

By definition, the unruly content on the intranet site has become unstructured. The inventory and audit is the first step in helping to restore structure. Using a process that allows you to maintain visibility of the different content clusters throughout assists in many ways by:

- providing clarity around quantity, distribution and existing structure of the different content types
- providing an initial cull of duplicate or otherwise redundant information
- helping determine the scope of work related to the planned content
- providing input into the information architecture (IA) development
- assisting in mapping the migration following the IA development

This paper provides a series of steps and techniques that allow you to take this higher level perspective. Even if a redesign is not planned, the principles and techniques presented can be used to gain a greater understanding of the existing intranet and guide governance and upkeep practices.

This paper will take you through the steps of listing your content, supplementing the list with additional information, then summarising the list to deliver the desired structures and insights.

Specifically the steps covered are:

1. Structuring your content list
2. Populating and preparing the list
3. Categorising the content
4. Creating dynamic summaries

1. Structuring your list

A content inventory is little more than a list. It may be a long list, with many columns, but a list nonetheless. To ensure you can get the most from this list you need to use the most appropriate tool, a spreadsheet.

Spreadsheets have powerful tools that allow you to easily manage vast lists of content. The following are some of the actions that can be used to help you work with your content listing:

- sort your content based on alphanumeric, date or even by colour
- filter your list to allow you to examine and manipulate selected portions
- perform automated processes such as removal of duplicates, and extraction of relevant information
- create instant summary tables across multiple categories of interest
- rapid copy, paste and fill functions allowing operations across large blocks content

For those unfamiliar with spreadsheets, adhering to a few simple principles will maximise what you get from your list when it comes to manipulating and analysing the information.

To get the most from your content list, the following principles should be followed:

- avoid the temptation to divide up the lists across multiple sheets, since the information can not be analysed as one list if the information is not together
- ensure every item of content has one and only one row associated with it, and in particular do not merge cells
- add a new column for each additional topic or category of interest and fill this for all content items
- be consistent in how information is represented – spelling, date format, surname/first name format etc.
- avoid deleting items unless you are sure they will never be relevant again (a better approach is to include a column with an 'exclude' marker so they are not included in the analysis, but can be referred to later if required)
- label your headings in a meaningful way by asking yourself 'if I come back to this in 12 months will I understand what this means?'

Fig 1: Lists are most useful when organised in a consistent manner. Here the individual components of the raw file name with path (1) is divided into individual columns (2). A formula can be used to extract the 'content type' from the file extension (3). Sorting and filtering is used on 'source path 3' to clean the list and generate a more consistent use of terms (4) with a new sub grouping. Additional data is added too, in this case ROT (5).

2. Prepare the content list

Using the principles outlined above, list the content you need to analyse or audit. Luckily, many intranet platforms can generate a list of content automatically. The key is not so much getting the list perfect, but getting a list with additional data that allows you to analyse in an insightful way. Additional information may be available automatically, but even simple content lists contain more than just file names.

Even beginning with a simple file list as shown 1 in figure 1, provides a good deal of valuable information:

- file name

  This is a key indicator of what the file contains on an individual basis but may also contain specific terms such as 'archive', 'sales' or 'safety' that can be can be automatically isolated, examined, and categorised to make them easier to evaluate in bulk form.

- file type

  It can be useful to distinguish between documents, graphic files, system files and data files. If this information is not explicit in the listing, ensure the file type identifier (file extension) is included with the file name.

- file location (path)

  In addition to helping interpret the meaning of the file by providing context, this information provides source information which helps identify ownership and is important in migration planning.

  The utility of this information can be enhanced if divided into individual path elements across columns. This allows summaries to be created in a more fine-grained way, and will be particularly valuable when mapping current state to future state relationships.

Other information that is often available directly from the system includes:

- date of creation / last modified

  Though not foolproof, the age of a file or recency of changes can indicate how valuable a file might be today.

- file author/owner

  This can be a department and/or a person, but ideally both. Note that the publisher of a file may not be the person who is responsible for its content.

A repetitive spreadsheet task usually has a built-in short-cut

Your initial listing will probably require some degree of cleaning and culling to optimise the list for analysis. First clean, and then cull:

- clean the list

  The more consistent the list information is, the more you can make use of automatic spreadsheet tools, avoiding the need to manually work through thousands of content items.

  For example, are all the dates in the same format? If some are in day-month-year and others are in month-day-year automatic analysis becomes difficult. It is preferable to 'clean' this category so all dates are formatted the same rather than manually analysing the dates. Other typical cleaning includes ensuring consistent word usage and spelling, removing leading and trailing spaces. Much of this cleaning can be done using built-in spreadsheet data cleaning tools.

- cull the list

  There will often be extraneous files that need to be separated from the items that require more in-depth analysis. Using the familiar ROT (redundant, outdated or trivial) or similar identifier, rather than removing the actual file name, just tag each content item using a 'status' column.

  Immediately you can probably cull system files, and all but one of a duplicate set. Another large list of files to cull are graphic files. Generally a graphic file will sit within a content page which is the item of interest. Data files need to be carefully examined as they may be few in number but can indicate a very rich vein of information that needs to be fully considered.

3. Categorise the content

Categorisation has already begun, but as new and more valuable groupings become apparent they are used to chunk the list into manageable groups.

Using sorting and filtering techniques mentioned earlier, as well as the table summaries described next, it is possible to begin to get a broad sense of what the content contains. These techniques help draw together related content, and enabling you to categorise them in a more meaningful manner.

For example, a category that lists document type (form, policy, procedure) can be useful, but much more valuable is the 'theme' of these documents such as employment, expenses or travel. These more insightful groupings will emerge based on your understanding of the organisation and the planned needs of the intranet. These groupings may also relate more to audience groups and user needs than to inherent properties of the content itself.

Larger categories can be divided further, and this is best done by adding columns and populating with second and third level groupings. When these multi-level groupings are summarised they can resemble an information architecture but should not be relied upon as such until a formal IA design step has been conducted.

However, when a formal IA has been created, this can also be included in the content list, where accurate 'before and after' maps can be created.

Following this broader level data preparation, what remains should be a smaller list of files that can now analysed more surgically.

4. Create dynamic summaries

Great pains have been taken to ensure that the content and related information is contained in a well structured list. In addition to making use of the filter and sorting tools, one of the key benefits is the ability to now create simple yet powerful summary tables. These tables can take millions of content items and summarise them across the different categories created through the analysis process.

For example, analysing even the small sample data set shown above can be tedious, but when 'pivoted' into a summary table (figure 2), it becomes much more manageable. This table immediately makes the following information visible:

- the new grouping is presented in a stacked form, which makes the new multi-level categories immediately visible
- the relationship between the old and new grouping structure is plain to see
- a count of the number of files is created automatically

This technique can be scaled instantly. The table in figure 3 shows this technique ex-

Fig 2: The simple pivot table shown creates columns and rows based on the list content, and in this case counts the number of items in the respective categories. For example, there are 10 documents from the policies and procedures source that are in the new grouping Employment: Pay and leave.

	policies and procedures	communications	IT	finance	safety	legal and governance	forms	useful information	HR	Grand Total
About										
Who we are		7				3				10
What we do	1	12		1	2	4		6		26
Board and governance	23			9	1	13		12		58
Media and public announcements	3	234				5		23		265
Employment and recruitment									12	12
Employment										
Entitlements and benefits	120			3					16	139
Training and development									16	16
Code of conduct					5				45	50
Pay and leave				6					24	30
Corporate support										
Building services	16		5			1	17	24		63
Procurement				24						24
Technology			65							65
Guest relations	1	4						16		21
Event marketing		16						14		30
Travel and expenses	8	3		12		2		22	1	48
Safety and emergencies	36	3			85					124
Duplicates	48	3								51
Eliminated	368	6	3		26	62	67	1	30	563
Unidentified		48				38				86
Grand Total	624	336	73	55	119	128	84	118	144	1681

Fig 3: The larger the content list, the more value and insight pivot tables deliver. Even this relatively small content listing of 1681 documents would be impossible to gain oversight of manually. Shown here are the original file locations listed across the top and the proposed structure in the left providing instant perspective.

tended to include over 1600 content items, resulting in a relatively simple table. Millions of content items can be managed just as easily.

From these summaries, patterns and points of interest emerge.

For example it would be worth exploring the *Media and Public* announcements content to better understand the value of the 234 documents. It could be further divided:

- by file type to see if the same types of news are published in different formats
- by date to see what the publication rate is and if there are patterns in it
- by owner to see if there is one (central) or multiple owners (divisions, teams etc.)

Also, looking at source folder (top categories) and themes (left hand categories) might be insightful. For example, the items in the IT folder all relate to technology so planning later project steps for this might be straightforward, however the 'useful information' category might be more problematic to deal with.

Note that the example shown is only one of numerous summaries that can be created. Any of the content categories contained in the content list can be used to summarise.

Even more insight can be gained to help inform governance practices or to support decisions made during design or migration.

Early effort simplifies the task overall

Conducting a content audit is a tedious but necessary task. Given the significant effort that must be spent in gathering the data, engaging stakeholders and managing the process, significant additional value can be extracted by overlaying the standard process with good data practices and using common spreadsheet tools.

It is not necessary to be a spreadsheet 'guru' to make best use of your spreadsheet. Taking the time to become familiar with the tools outlined in the panel 'making the most of your spreadsheet' below will help remove much of the tedium and make time and head-space available for more insightful and rewarding aspects of intranet redesign.

Making the most of MS Excel

MS Excel is the world's most popular spreadsheet and yet does not have the usage of applications like Word or PowerPoint. If you are planning to use Excel to manage your content inventory then the following might be helpful for those less familiar with the application.

Must use

Excel is designed to work with lists of information and has a number of simple, menu-driven features. Become familiar with these to start to take the tedium out of manually working lists:

- *sort* (arrange the list in the order you want)
- *filter* (select sections of the list)
- *series* (automatically fill cells with 1,2,3 etc.)
- *fill* (fill cells with the same content)
- *find* and *replace* (individually or en bloc)
- *find duplicates* (identify identical cells)

Added value

Formulas add another level of capability, and while many shy away from them for being too 'technical' they can be as simple as 'Clean' which removes unwanted text from cells. Formulae to begin with:

- *&* (also known as 'concatenate' – combine cell content together)
- *trim* (remove spaces from the ends of text)
- *clean* (remove unwanted characters)
- arithmetic: +, -, / , x, *sum* (does your sums)

A little more advanced

The advanced feature that is most worth the effort of learning is *pivot tables*. Demonstrated above, these summarise vast amounts of list information instantly.

Some formulas beyond the basics you will find particularly valuable include:

- *find*, *left*, *right*, *length* (for manipulating text and extracting pieces of interest)
- *vlookup* (match cells from different lists and use to combine lists together)
- *if* ('if this, then do that' – a formula to make all the rest even more valuable)

More articles!

Find out when new papers are published:
www.steptwo.com.au/subscribe

Did you like this article?

Send your thoughts and feedback to:
stephen@steptwo.com.au

Step Two DESIGNS

www.steptwo.com.au • contact@steptwo.com.au
intranets • usability & IA • content management • information management