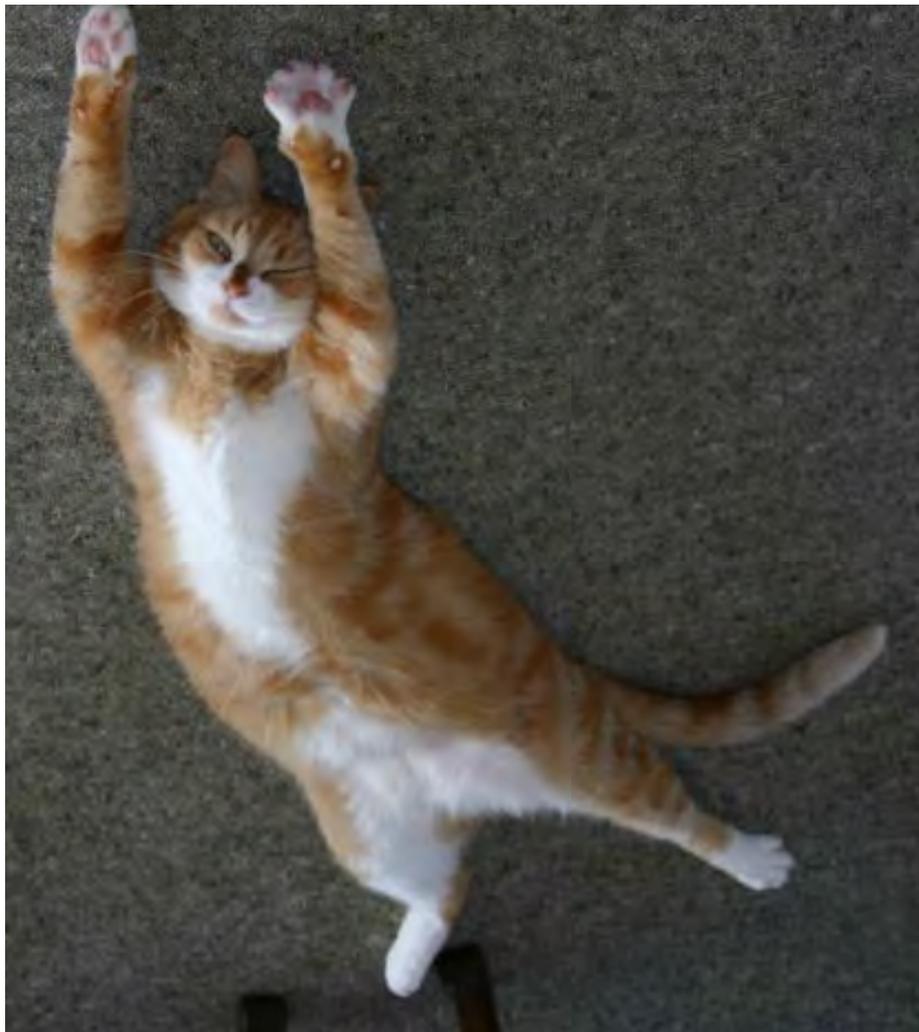# Content Analysis: Methods and Mentoring

By Chiara Fox

ASIS&T IA Summit: March 26, 2006

STRETCH!!

Chiara Fox

Senior information architect at Adaptive Path

Been an IA for 6+ years

Worked as both an "innie" and an "outie"

Spent 10 years in public and scientific libraries

And I LOVE content analysis!

And yes, I'm a little weird.

What Is "Content Analysis"?

The process of assessing the nature of the content on a given web site.

This involves determining:
- The depth and breath of the site
- Total number of documents or content items
- Content structure
- Varity of formats of items included on the site
- Identification of patterns and relationships between the content items

First, Some Definitions

What is considered content?

- Web pages (e.g., HTML, PHP, JSP)

- Text files

- MS Office documents (e.g., Word, Excel)

- PDF files

- Images

- Executables

- Audio

- Video files (e.g., QuickTime, Flash)

Pretty much anything that is available on the site should be included in the content analysis.

Why Do a Content Analysis?

Content analysis lets you understand the relationships and patterns between the pieces that make up the site.

Often done at the beginning of:

- A site migration to a new platform or content management system (CMS)

- A re-architecture project

- The task analysis and mental model process.

Content analysis involves a series of deliverables.

Content Audit
- A sampling of the site content to get a flavor and feel for the site.

Content Inventory
- An all inclusive accounting of every page and asset on the site.

Content Map
- An abstraction illustrating the types of content or genres of content available.

A content audit is a sampling of the site content.

The goal is to gather examples of every type of content that is offered. It is not necessary to examine every content item, but at least one of every type.

- Gathering the sample should not take a lot of time. About 1 minute to gather the item and 2 minutes for analysis.

- Most press releases look exactly the same. Only include a handful in your audit.

- High volume sites (such as a news site) with millions of pages make audits difficult. Focus on examples of the different templates in use.

- Be sure to capture as many "one off" examples as you can.

A formula for determining how big of a sample to take.

1. How many major sections of content are there?
   Example: 6 major sections

2. How many documents are in each section?
   Section A (1000), B (200), C (500), D (2000), E (1000), F (500) = 5200 documents

3. How much time in the budget is allocated for content sampling?
   1 person, 3 days, 6 hours per day = 18 hours = 1080 minutes

4. Amount of time to be spent on each content item.
   1080 minutes / 3 minutes per document = 360 documents

5. Number of passes through the content for analysis.
   1080 minutes / 3 minutes per document / 2 passes = 180 documents

6. Percentage of total content items.
   180 analyzable documents / 5200 total documents = 3.5%

Skew sample based upon which sections are most important.

Use the current site structure as a guide. Work your way down the hierarchy tree.

1. Visit each of the major site sections, capturing the top level pages.

2. Visit each secondary navigation page that hangs off each main section.

3. Start looking for uniqueness. Keep in mind your guidelines for how much to collect in each section.

   a) Capture anything that stands out, grabs your eye, or looks especially interesting.

   b) Gather all document and format types.

   c) Be sure to capture tools such as search, site index, shopping cart.

   d) Note things that are broken, can't access, or have errors.

4. Be sure to go to the bottom of each section until you come to a dead end.

There are two methods for capturing the sampled content.

Printed Page

- Allows you to take notes right "on" the pages
- Can move them around later during re-architecture
- Have to go back later and take screenshots for report
- Kills a lot of trees

Spreadsheet

- Indent spreadsheet rows to indicate site hierarchy
- Notes are captured all in same place
- Easy to share with a remote team
- Earth-friendly
- Have to deal with Excel
- Loses the visual impact of the printed page

Example Audit

Example of a Content Audit

id	page title	page name (if different)	url	notes	location on page	redundant
0.0	Home Page		http://www.c			
1.0	Shop	All Departments	http://www.c		Global nav	
2.0	Gift Registry	Gift and Bridal Registry	http://www.c		Global nav	
3.0	Gift Ideas		http://www.c		Global nav	
4.0	What's New		http://www.c		Global nav	
5.0	Catalogue		http://www.c		Global nav	
6.0	Stores		http://www.c		Global nav	
7.0	Gift Cards		http://www.c		Global nav	
8.0	Order Tracking		http://www.c		Utility Nav	
9.0	My Account		http://www.c		Utility Nav	
10.0	Assistance		http://www.c		Utility Nav	
11.0	Cart		http://www.c		Utility Nav	
12.0	Search		http://www.c		Utility Nav	
13.0	About Us		http://www.c		Footer Nav	
14.0	Careers		http://www.c		Footer Nav	
15.0	Our Ads		http://www.c		Footer Nav	
16.0	Corporate Sales		http://www.c	Part of the Assistance section	Footer Nav	
17.0	Customer Feedback		http://www.c	Pop-up window	Footer Nav	
18.0	Site Index		http://www.c		Footer Nav	
19.0	Privacy Policy		http://www.c	Pop-up window	Footer Nav	
20.0	Guarantees		http://www.c	Part of the Assistance section	Footer Nav	
21.0	Terms of Use		http://www.c	Pop-up window	Footer Nav	

Example of a Content Audit, II

id	page title	page name (if different)	url	notes	location on page	redundant
0.0	Home Page		http://www.ci			
1.0	Shop	All Departments	http://www.ci		Global nav	
2.0	Gift Registry	Gift and Bridal Registry	http://www.ci		Global nav	
2.1	Create a Registry		http://www.ci		2nd global nav	
2.2	Update a Registry	Edit Your Registry	http://www.ci		2rd global nav	
2.3	Give a Gift		http://www.ci		2th global nav	
3.0	Gift Ideas		http://www.ci		Global nav	
4.0	What's New		http://www.ci		Global nav	
5.0	Catalogue		http://www.ci		Global nav	
6.0	Stores		http://www.ci		Global nav	
7.0	Gift Cards		http://www.ci		Global nav	
8.0	Order Tracking		http://www.ci		Utility Nav	
9.0	My Account		http://www.ci		Utility Nav	
10.0	Assistance		http://www.ci		Utility Nav	
11.0	Cart		http://www.ci		Utility Nav	
12.0	Search		http://www.ci		Utility Nav	
13.0	About Us		http://www.ci		Footer Nav	
14.0	Careers		http://www.ci		Footer Nav	
15.0	Our Ads		http://www.ci		Footer Nav	
16.0	Corporate Sales		http://www.ci	Part of the Assistance section	Footer Nav	
17.0	Customer Feedback		http://www.ci	Pop-up window	Footer Nav	
18.0	Site Index		http://www.ci		Footer Nav	
19.0	Privacy Policy		http://www.ci	Pop-up window	Footer Nav	
20.0	Guarantees		http://www.ci	Part of the Assistance section	Footer Nav	
21.0	Terms of Use		http://www.ci	Pop-up window	Footer Nav	

A content inventory is a detailed, meticulous accounting of all items on the site.

It is very similar to a content audit except that you are recording *every item* on the site.

Used during content migration, when have to keep track of the status of moving data, identify ROT (redundant, outdated, and trivial), or need exact numbers for forecasting.

Becomes stale very quickly, because sites change all the time.

Sadly, no good way to automate, though database dumps can help.

Can be built off of content audit spreadsheets or done in a database.

Example of a Content Inventory

Link ID	Link Name	Link	Subject Type (not summary or keywords)	Document Type	Topic
2.0.0	products	http://www.peoplesoft.com	nav		
2.1.0	applications	http://www.peoplesoft.com	nav		
2.1.1.0	eBusiness	http://www.peoplesoft.com			
2.1.1.1.0	peoplesoft portal solutions	http://www.peoplesoft.com	marketing	paragraphs	intro to portal solutions
2.1.1.1.1	campus portal	http://www.peoplesoft.com	feature/benefit	paragraphs	portal solutions
2.1.1.1.2	customer portal	http://www.peoplesoft.com	feature/benefit	paragraphs	portal solutions
2.1.1.1.3	employee portal	http://www.peoplesoft.com	feature/benefit	paragraphs	portal solutions
2.1.1.1.4	enterprise portal	http://www.peoplesoft.com	feature/benefit	paragraphs	portal solutions
2.1.1.1.5	government portal	http://www.peoplesoft.com	feature/benefit	paragraphs	portal solutions
2.1.1.1.6	supplier portal	http://www.peoplesoft.com	feature/benefit	paragraphs	portal solutions
2.1.1.1.7	portal solutions	http://www.peoplesoft.com	feature/benefit	paragraphs	links to above sections
2.1.1.1.8	customer profile: anadarko launches emp	http://www.peoplesoft.com	customer	paragraphs	customer article
2.1.1.1.9	press release	cross-link			
2.1.1.1.10	events	cross-link			
2.1.1.1.11	brochure: portal overview	http://www.peoplesoft.com	product description	brochure	portal solutions
2.1.1.1.12	white paper: portal technology of PS 8	http://www.peoplesoft.com	marketing	white paper	portal solutions
2.1.1.1.13	white paper: integration technology	http://www.peoplesoft.com	marketing	white paper	portal solutions
2.1.1.1.14	white paper: security features of PS intern	http://www.peoplesoft.com	marketing	white paper	portal solutions
2.1.1.1.15	learn about the ps8 release of this product	http://www.peoplesoft.com	feature/benefit	one-sheet	portal solutions
2.1.1.1.16	take an in-depth look at the features and f	http://www.peoplesoft.com	product description	paragraphs	links to pdf with product description
2.1.1.1.17	featured stories: portals: nirvana, bubbly, c	http://www.peoplesoft.com	marketing	paragraphs	article reprint
2.1.1.1.18	featured stories: is there a portal in your fu	http://www.peoplesoft.com	marketing	paragraphs	article reprint
2.1.1.1.19	peopletalk: building a portal experience	http://www.peopletalkonlir	marketing	paragraphs	links to peopleTalk
2.1.1.1.20	demo: see how exec users can benefit form the ps employee portal		demo	flash demo	portal solutions
2.1.1.1.21	demo: see how ps campus portal gives yo		demo	flash demo	portal solutions
2.1.1.1.22	demo: see how the customer portal helps		demo	flash demo	portal solutions
2.1.1.2.0	eprocurement solutions	http://www.peoplesoft.com	marketing	paragraphs	intro to eprocurement solutions
2.1.1.2.1	eprocurement	http://www.peoplesoft.com	feature/benefit	paragraphs	eprocuremment
2.1.1.2.2	purchasing	http://www.peoplesoft.com	feature/benefit	paragraphs	eprocuremment

The content inventory is for keeping track of information about each content item.

Examples of information included in an inventory:

- Content owner

- Person responsible for migration/updating page

- Freeze status

- ROT status

- Current location in site

- New location in site

- New file names/URL changes

Example of a Site Migration Table

| New Architecture Link ID	New Architecture Link Name	URL	Current site content object title	Current Site	Content Approver	Freeze Phase
D5.0	Support & Training					
D5.1.0	Customer Service Guide		customer service guide	cc	Carrie Koch	
D5.1.1	eCenter Customer Support Description					
D5.1.2	Global Support Center Ts & Cs					
D5.1.3	MarketPlace Support Description					
D5.2.0	PeopleSoft Customer Services		customer services	ps.com	karen	
D5.2.1	Platinum Customer Support	http://www.peoplesoft.c	Platinum Customer Support	ps.com	karen	1
D5.2.2	Premium Customer Support	http://www.peoplesoft.c	Premium Customer Support	ps.com	karen	1
D5.2.3	Standard Customer Support	http://www.peoplesoft.c	Standard Customer Support	ps.com	karen	1
D5.3.0	Training		education and training	ps.com	sandi	
D5.3.0.1	Training Needs Roadmap	http://www.peoplesoft.c	Training Needs Roadmap (also in Asse	ps.com	sandi	Phase 2
D5.3.1.0	Overviews			ps.com	sandi	Phase 2
D5.3.1.1	Why PSU?	content to come	Why PSU?	ps.com	sandi	Phase 2
D5.3.1.2	Frequently Asked Questions	http://www.peoplesoft.c	Frequently Asked Questions	ps.com	sandi	Phase 2
D5.3.2.0	Solutions		Solutions	ps.com	sandi	Phase 2
D5.3.2.1	Implementation Training	http://www.peoplesoft.c	Implementation Training	ps.com	sandi	Phase 2
D5.3.2.2	Optimization Training	content to come	Optimization Training	ps.com	sandi	Phase 2
D5.3.2.3	Upgrade Training	http://www.peoplesoft.c	Upgrade Training	ps.com	sandi	Phase 2

A content map is a graphical representation or abstraction of the site content.

Hundreds of rows of a spreadsheet or stacks of printouts can be hard to "grok" or understand. The map helps to get you out of the trees and shrubs so you can see the forest.
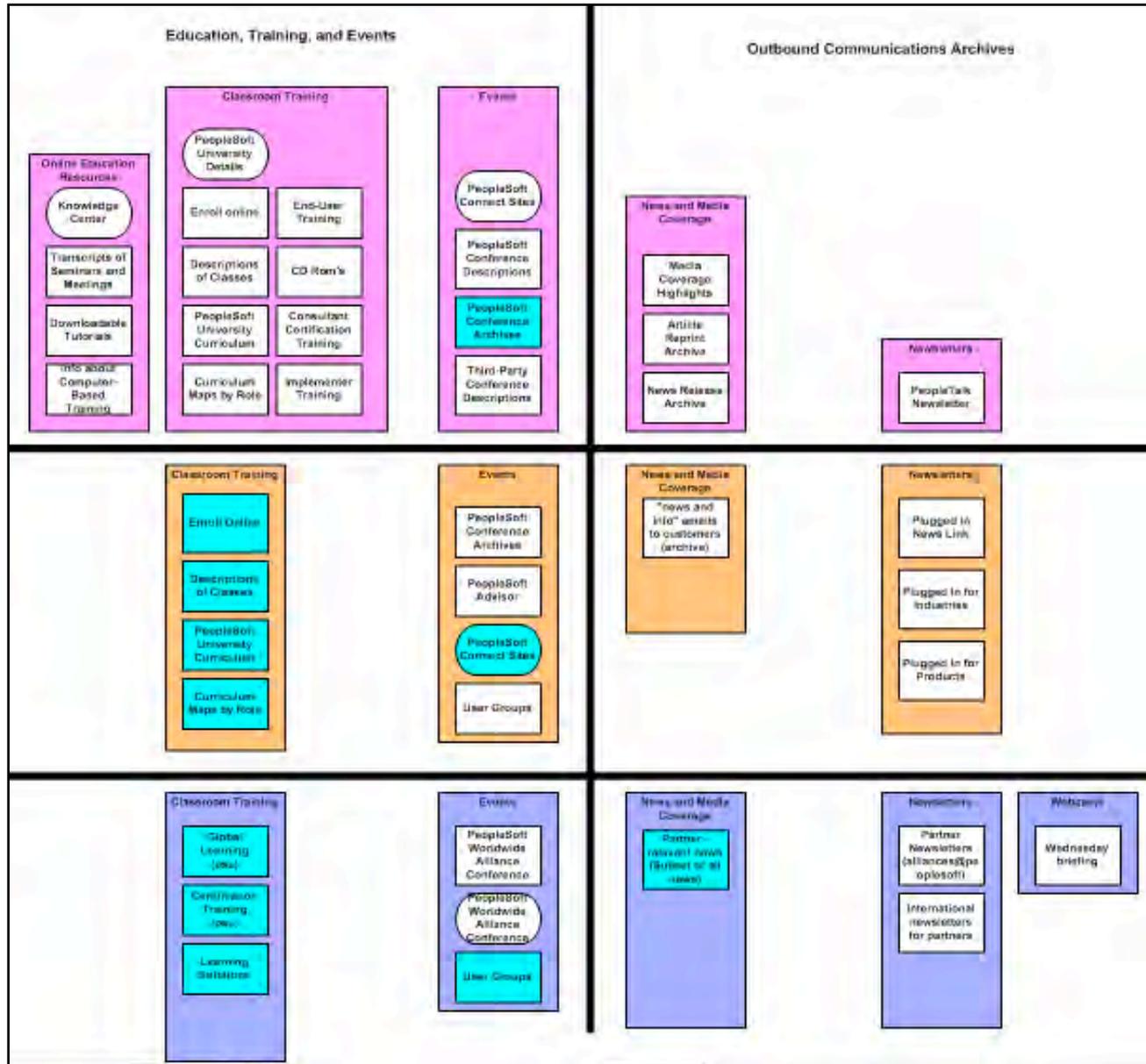
The content map may take different forms and have different uses:

- As a list of the content types/genres for placement into a mental model
- Illustration of the redundancies and uniqueness between sites
- Content model for metadata and database design
- Aide for communicating with executives and other members of the team

Content Map for a Mental Model
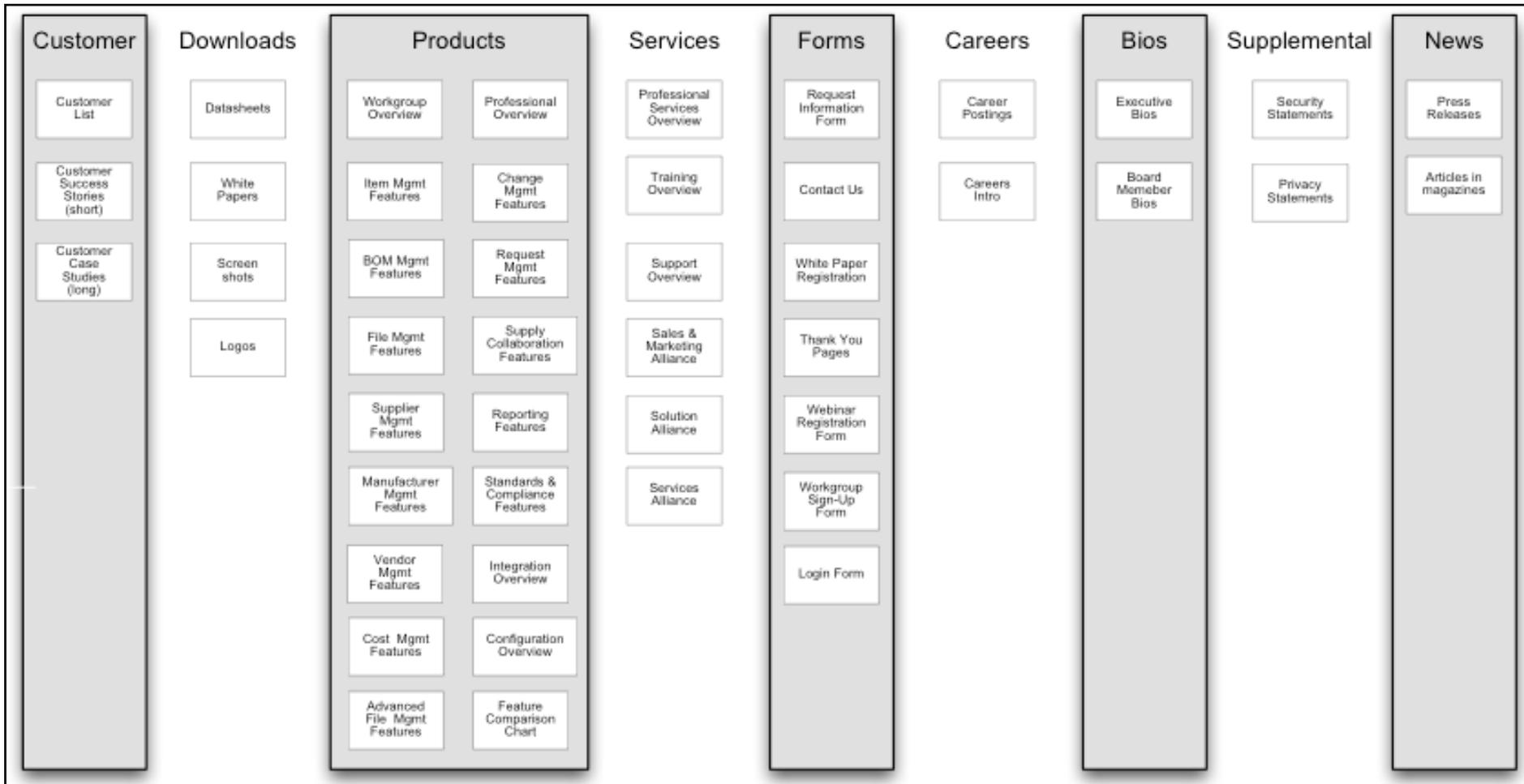
We've talked a lot about capturing data about a site, and how to record what we've found. Is that "analysis" though?

Content analysis happens:

- In your brain as you are collecting and scouring the site. You are building a deep understanding of the relationships between the content items.

- During the identification of the content types or genres that are illustrated in the content map.

- As you think about new groupings and pairings of the content.

What to Look for During Analysis

Content analysis is all about patterns and relationships.

Patterns - similarities among content
- Genres or types, security requirements, audience
- Physical formats, dynamically generated vs. static
- Just a "feeling" that two disparate pieces of content "go together"

Relationships - connections between content
- Hierarchical relationships such as parent-child
- General documents point to specific or detailed documents
- Dependencies, such as sequences, procedures

The patterns and relationships will suggest groupings for the different content types.

You will start to see how things fit together, such as all service related documents together.

If the site has been around for awhile, these broad groupings probably already exist in the current site structure. You'll be looking for a finer level of groupings and connections.

- Cross-sell and up-sell opportunities between products.
- Links between sections, such as products and the training on how to use the product.
- New ways to use supporting materials such as brochures and data sheets.
- Related content, such as linking a background piece on the history of Hamas to a news story on the Israeli-Palestinian conflict.

The analysis and observations will feed and influence your IA deliverables.

Wireframes and schematics

Site blueprints

Page flows

Metadata schema

Gap analysis

Ideally performed by the information architect.

Gives the IA a deep understanding of the content and the relationships and patterns inherit to the content. This will directly impact the information architecture in a re-architecture project.

However, organizations often don't have an information architect on staff. They hire IA consultants to help them with these projects.

But since architecture project can be expensive, organizations look for ways to cut corners. Performing the content audit and analysis themselves seems like a good choice, since it can be labor intensive and expensive.

If your client wants to do the audit themselves, mentor them through the process.

Your final deliverable - the new IA for the site - depends upon you having a strong understanding of the content of the site.

Mentor and guide your client through the process so you will have good inputs to your design and the output will be something that is useful and valuable for the rest of the project.

This will save time and money in the budget since you won't be filling out spreadsheets and clicking.

It builds the skills and knowledge within their organization.

The Content Audit Personality

It takes a certain type of person to successfully perform a content audit.

The characteristics of a good auditor are:

- Close attention to detail
- Deep understanding of and comfortable with the web
- Understanding of the project and how the content audit will be used
- Ability to think abstractly
- A member of the core web team
- (And maybe a bit weird)

Mentoring can be done in person or remotely.

It's always nice to be able to work directly with someone, looking over their shoulder as they start the process and guiding them along.

Often, this has to be done remotely. Send the auditor documents ahead of time and start with a call or meeting to review the overall process.

Bring lots of examples of audits and content maps from past projects.

Show how this work will fit in and influence the work that is to be done later in the project. Again, use example as a guide and tie it back to how things work on their site.

Check in regularly with your auditor to check their progress and make corrections early on.

During the initial meeting, walk through the process with the auditor together choosing items for sampling and analysis.

Have the auditor work on a small section, then review that work together. You may have to work together to "dial in" the correct level of granularity of notes that are taken, or to find the correct mixture of items in the sample.

People usually have few problems filling in the spreadsheet. Creating the content map is where they have to pull back and abstract out of the spreadsheet and is often where they run into trouble.

Making Content Map Creation Easier

The leap from a spreadsheet to tiny boxes in Visio can be hard for some people.

There is no one true way to make a content map.

Organizing the content types by the current site structure may make it easier for the auditor to work with. Be sure to step out of that structure when re-architecting though so you don't recreate what they already have.

Make everyone comfortable with the idea that the spreadsheet and content map are "stepping stones" to help you get to the next step in the process. While they can be repurposed, they are often "throw away" work. That's okay!

Use color coding and font effects sparingly. Too much makes the map hard to understand.

Be Prepared If Things Don't Go Well

There's always a chance that the person who is assigned to do the content audit just doesn't "get it."

There's a fine line between tweaking and re-tweaking the spreadsheet or map to get it just right and burning up time in the schedule trying to get someone to understand a subtle concept in the analysis.

It's possible that you have to cut your losses and just work with what you have so far.

If the auditor can't make the leap from the spreadsheet to the map, abandon the spreadsheet and just have them focus on the map. The map is often the more important piece because it is used in the mental model creation and for re-architecture. An audit spreadsheet will have to be redone as an inventory for migration anyway.

Once you've completed the audit and analysis you'll be ready to move on to the next phase of your project.

Slot content items against the tasks they support in a mental model.

Diagram a content model for database relationships.

Determine major groupings to form the basis of the site structure.

Create a descriptive metadata schema of the elements needed to describe the content.

Diagram processes, flows, and presentations of content.

Thank You!

Questions? Comments?

Chiara Fox: chiara@adaptivepath.com

Presentation available at: www.chiarafox.com