innodata®

WHITE PAPER

Content Profiling

A Methodology for Profiling
Mobile and Digital Content

Content Profiling

Content development leaders often underestimate the effort required to enrich, restructure, transform, or convert their content to support mobile and digital content development. Whether the new product is an eBook, tablet app, new website, or enhanced features on an existing platform, the content structures and enrichment must support the functionality envisioned for the product. A Content Profiling methodology brings together early in the product development cycle both content and application development best practices to ensure on budget and on time launch of mobile content.

Abstract

New digital product development is often characterized by the exploration of new, cool application features, user interfaces, and customer experiences. However it is often the case that the underlying content cannot support them. Many of the following questions should be asked:

- What are the costs to develop the right kind of content?
- What are the potential impacts on the project schedule to remediate the content?
- Is it worth removing several features (or postponing to a later phase) to meet a product launch deadline?
- Wouldn't it be best to fully understand the costs and potential trade-offs before proceeding down the path of random content development?

This paper describes our Content Profiling methodology to help you match up the current state of content with new multi-channel requirements. The end goal is to have a customized Content Readiness Assessment with a specific content transition roadmap that will increase your odds of a successful mobile content development effort.

Table of Contents

"This paper describes our Content Profiling methodology to help you match up the current state of content with new product needs."

Introduction

Not all digital "content" is equal. There are a variety of digital formats with different levels of "fidelity." For example, image-based PDFs cannot be fully text searched, whereas PDFs + Hidden Text documents are fully text searchable. If a new digital product requires full text searchable features, then an image-based PDF is not an ideal format for the source content. Even XML content can have varying degrees of rich structure. XML files can be flat, or can be highly "leveled" with nested granular structures that allow small chunks of content to be searched on, rendered, and linked to from outside and within the product. If one of the goals of the product is to allow users to navigate or search to very targeted, discrete pieces of information, then flat XML files will not support such granular display.

Beyond the format of the content source, additional enrichment may be needed to support the desired functionality. For example, an additional layer of metadata may need to be added (or the existing metadata may need to be normalized) to support advanced searching and filtering (e.g., search by jurisdiction), relevancy ranking to optimize search result sorting, or point in time functionality. A taxonomy also may need to be added (or an existing one reused) and content may need to be classified against it to support topical browsing in the product. Further, inline links and relationships may need to be created to allow users to navigate cited or related content.

It can be quite costly if dependencies between the source content and the enrichment required for new product functionality are incorrectly identified and mapped. These types of mismatches can lead to many problems including extending the product development cycle and increasing costs. Some of the key challenges can be:

- ***Product functionality not supported by content:*** Content may not have the necessary structures (e.g., leveling, inline links) or enrichments (e.g., metadata, taxonomy) to support the features and functionality defined in the product visualization (e.g., granular display, topical navigation).

- ***Issues materialize late in the project:*** Mismatches between content needs versus product features are revealed late in the project causing product launch delays and unexpected cost overruns.

- ***Online product behaves in unanticipated ways:*** Even structured content, such as XML, may have variability across the content set. These inconsistencies may lead to unexpected behaviors in the product that are experienced by customers (e.g., content not found when searched). On the surface, these may appear to be application bugs, but are actually source content issues.

- ***Content maintenance after product launch not considered:*** Typically, content in digital products is often updated, so editorial or production workflows must be adapted on an ongoing basis to support new content structure or enrichment requirements for the product.

High-Level Solution

Content Profiling is a methodology for identifying gaps between current state content formats, structures, and enrichment in relation to the required changes necessary to support desired user experiences. The methodology includes the following four steps:

- Step 1 – Gather content samples and artifacts.
- Step 2 – Analyze platform needs against content structures.
- Step 3 – Develop content delivery functionality to content structure mappings.
- Step 4 – Complete the content readiness assessment.

Solution Details

Step 1 – Gather Content Samples and Artifacts

The first step in the process is to identify the kinds of content samples and artifacts needed for the Content Profiling. Content samples should be representative of all the different content types that will be part of the content set for the new product. The sample set should not only include typical documents, but also atypical documents so the degree of variability in the content set is understood. The content samples should include all the different digital formats (e.g., PDF, XML, RTF, HTML, etc.) that may be encountered as inputs to the system. Any content architecture models, such as DTDs, XML Schemas, etc., or process documentation should also be included as artifacts with the samples.

At this point, it should be determined if there are other existing structures or enrichments that could be leveraged by the new product. Perhaps there's an existing taxonomy for another product that could be reused or a metadata database that could be leveraged. These possibilities should be included with the sample set for analysis and consideration.

Step 2 – Analyze Platform Needs Against Content Structures

An important prerequisite to the content analysis step is a high-level understanding of new platform requirements, user scenarios, and, if possible, application wireframes (expressed as an output of Solution Visualization). Without first understanding the product needs, you have no basis for determining "what's good enough" or what needs to be changed or added in the content to support the product and user experience.

This analysis step takes a top down view to examine platform needs against content structures. Starting with requirements, user scenarios, and wireframes you can determine what structures and enrichments are needed in the content to support those functionalities. In analyzing the content samples and existing structures and enrichments, you then determine what is compatible or incompatible with the product functionality as well as any additional content structures and enrichment that may be needed to make it compatible.

Typically, platforms evolve over time, and it's important to consider future functionality that may be implemented in later phases. The content architecture should be agile enough to support additional enrichment that may be needed as the product evolves.

The following figure provides three examples of platform requirements with possible existing content structures that make the current state of the content compatible or incompatible with the requirements:

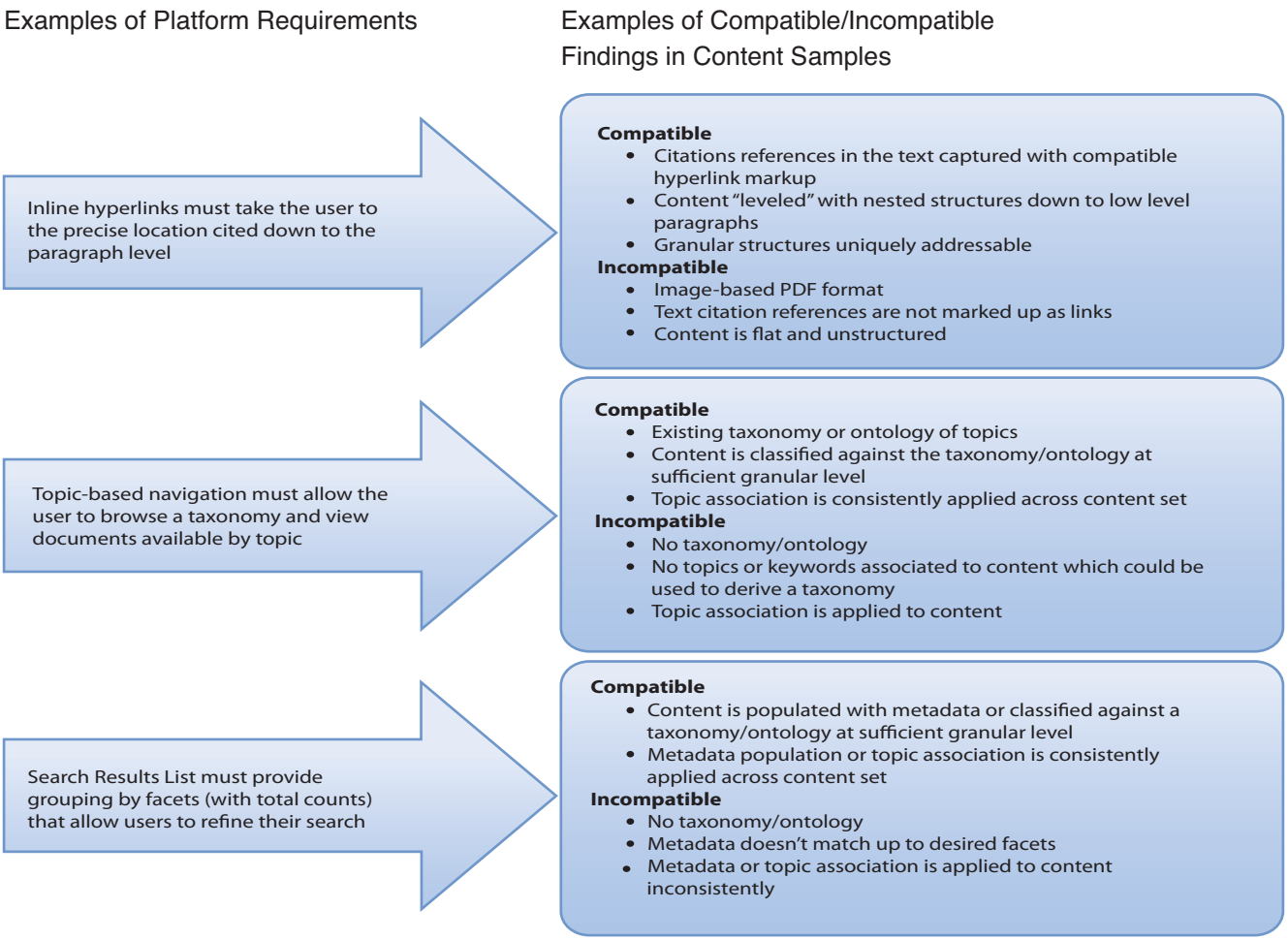

Figure 1: Product requirements and relationship to content samples

Step 3 – Develop Content Delivery to Structured Mappings

This step is intended to capture the findings of the analysis in a Content Readiness Assessment Report. Mapping is captured for product needs, requirements, application functionality, use cases, and user experiences that are dependent or impacted by content structures and enrichment. If product wireframes are available, they, too, are annotated to describe requirements for the underlying content set (e.g., required metadata). This establishes the baseline target for the future state of your product's content set and will ensure that content is ready for deployment at the same time that the application functionality is released.

An important note is that not all content requirements may be evident upfront or they may change during application development. The Content Readiness Assessment is intended as an initial step to an Agile Content Approach where the content set may need to be further adapted as the product is being developed. For example, when developing one of the application's features, the development team may determine that an additional piece of metadata needs to be added to content that wasn't anticipated in the original Content Assessment. The iterative nature of an Agile Content Approach ensures that content and application are always in sync.

The following is an example of an annotated product wireframe, highlighting specific product features and their underlying content requirements:

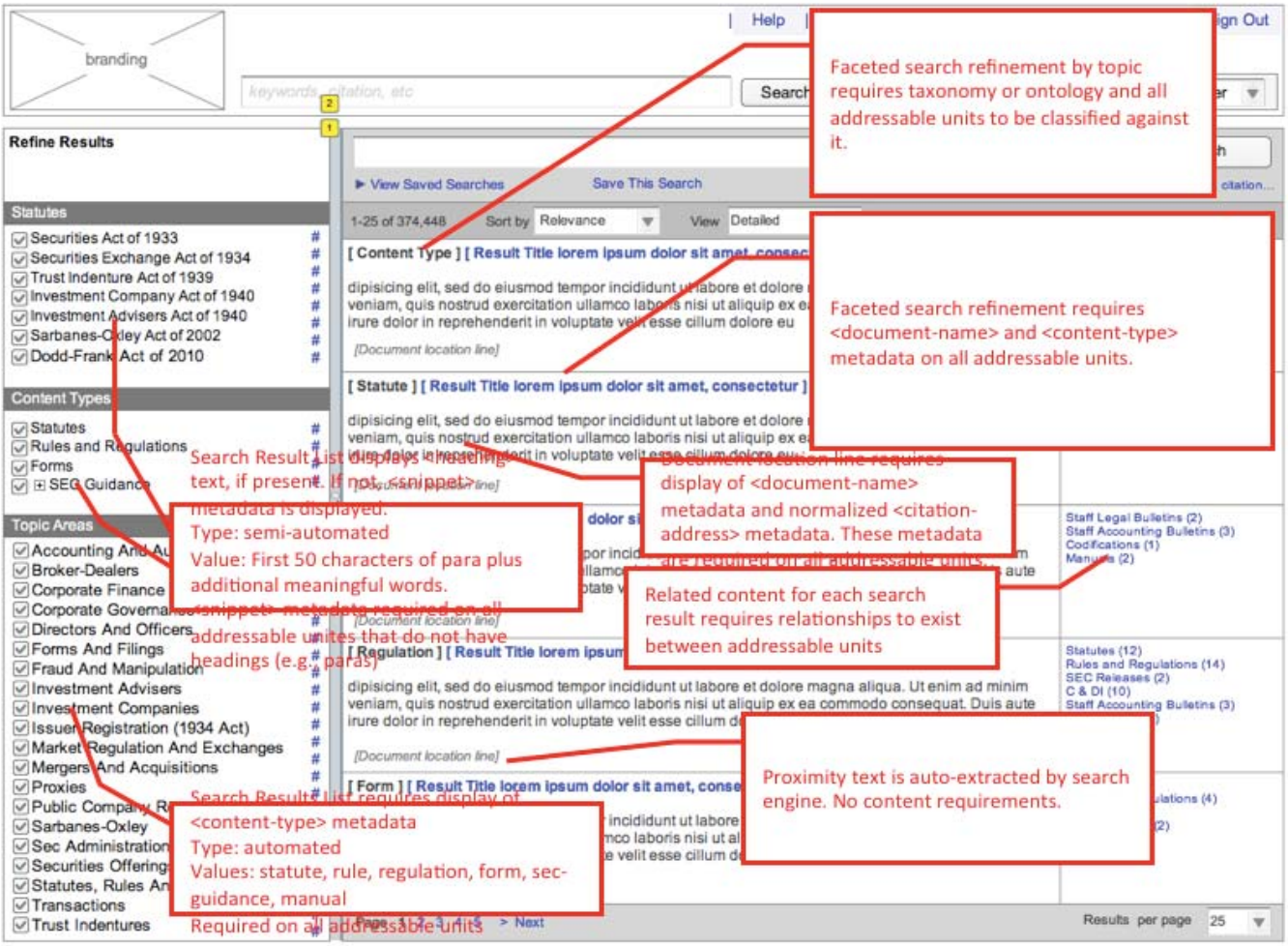

Figure 2: Annotated product wireframe

Step 4 – Develop Content Readiness Assessment

Once an initial target for the future state of the content set is developed a Content Readiness Assessment is prepared to track current state to future state. The Assessment details the gap between current state and future state of the content, with recommendations for how to implement any needed changes or additions to the content set for the product (e.g., what can be automated, what enrichment must be added manually, etc.), and a timeline for testing initial new content by the application developers.

The Content Readiness Assessment also describes any changes to content supply chain processes needed to support the future state content. If existing content will be updated or new content will be added after the initial launch on an ongoing basis, the document will capture maintenance requirements to keep the content in the product up to date.

This Content Readiness Assessment ensures that the costs of converting or enhancing the content set to meet the product needs as well as the dependencies between content and application are well known, communicated to key stakeholders, and that there are no subsequent surprises.

Content Profiling Timeline

To perform the Content Profiling, functionality mapping, and roadmap, a three-week timeline is typically required, with activities organized as follows:

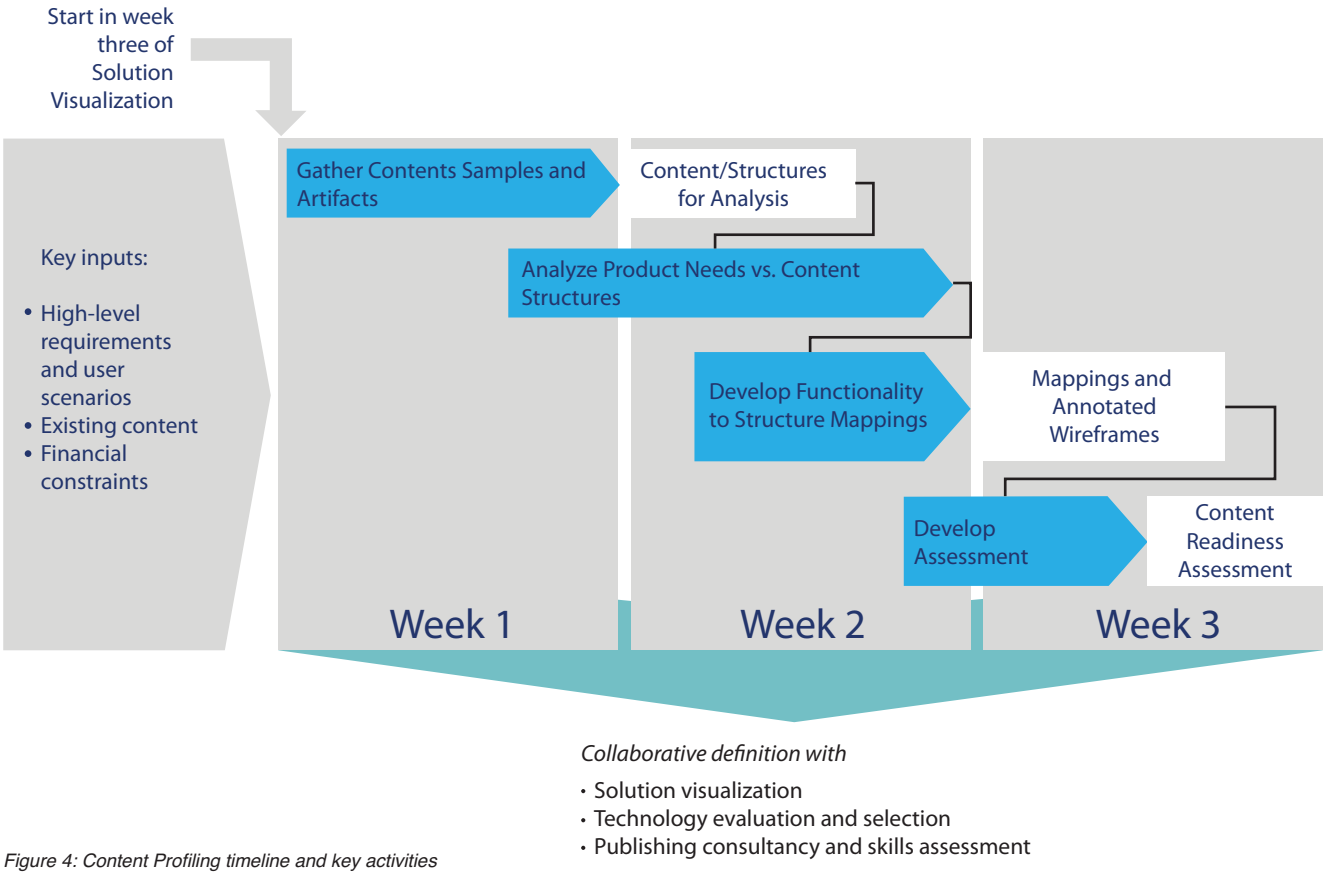

Figure 4: Content Profiling timeline and key activities

Business Benefits

A needs-based Content Profiling methodology helps to mitigate risks and ensure success by:

- ***Identifies gaps between the current state and future state content:*** Analyze sample content (e.g., documents, metadata, taxonomies, content architectures) to determine if current content structures are compatible, and if additional enrichments are required to support the application.

- ***Reuses existing enrichments where possible:*** Identify content structures already in house (e.g., existing taxonomies, metadata, and indexes) that may be repurposed and leveraged by the application.

- ***Identifies content problem areas that may be exposed by the product:*** Analyze content samples to determine if there are patterns of inconsistencies or issues that may surface in the application by different use cases (e.g., search not returning expected results because of inconsistent metadata).

- Defines processes to support ongoing updates post product launch: Analyze current processes and determine necessary changes or improvements to the process in order to support ongoing updated or new additional content that will get loaded into the product.

Summary

Migrating to the new mobile world presents many challenges for content providers. When creating new multi-channel content, it is imperative that you not only determine the functionality that will exist within the delivery platform, but you also conduct a profiling of the content to determine how content is structured to support the required functionality, how it must be enhanced, converted or transformed, and if its structure is agile enough to support functionality needs in the future.

A Digital Content Profiling methodology ensures that content will support product needs, a critical and often overlooked effort in launching on time and with success.

About Innodata

Innodata (NASDAQ: INOD) is a leading provider of business process, technology and consulting services, as well as products and solutions, that help our valued clients create, manage, use and distribute digital information. Propelled by a culture that emphasizes quality, service and innovation, we have developed a client base that includes many of the world's preeminent media, publishing and information services companies, as well as leading enterprises in information-intensive industries such as aerospace, defense, financial services, government, healthcare, high technology, insurance, intelligence, manufacturing and law.

More Information

For more information about Content Profiling, please visit www.innodata.com, call us at 201-371-8000 or contact us at solutions@innodata.com. We also encourage you to read these other papers in our *Enhancing Customer Engagement in the Post-PC Age* white paper series which you can find at www.innodata.com

- Solutions Visualization
- Agile Content Development
- Progress Release Management
- Print-to-Digital Consultancy
- Technology Blueprinting

innodata®

55 Challenger Road
Suite 202
Ridgefield Park, NJ 07660
201-371-8000
www.innodata.com