Single Source Production System
Print vs. Digital?

EEPG Production Network Workshop

IDM – Sofia, 12th June, 2014
Content lifecycle

Capture (Input)
- Aggregation
- Indexing
- Recognition Technologies

Store
- Repositories
- Library Services
- Storage Technologies

Manage
- Workflow
- Search & Discovery
- Sharing & Collaboration

Preserve
- Long-term Archive
- Integration with other applications
- Backup & Recovery

Deliver (Output)
- Distribution
- Security and Access Control
- Transformation Technologies
Objectives?

• Secure archiving;
• Time to market;
• Re-purposing;
• Online publishing;
• Content enhancement;
• Content updates;
• Multi-media.
ECM Framework

- Document Management
- Web Content Management
- Records Management
- Business Process Management
- Digital Assets Management
- Social Content

ECM Platform
- Meta Services
- Content Services
- Library Services

Applications
- Content Management
- Interoperability Services
Typical supply chain for print products

- Traditionally published content has been consumed in print products.
- Publishers have used digital tools optimized for production of print products.
Bottlenecks of the print products supply chain

Print oriented content production supply chain has been optimized by using PCs, popular consumer products for authoring the content such as Microsoft Word and by using affordable Desktop publishing software for producing typesetting files for print products such as Adobe InDesign.

With the print oriented content supply chain publishers end up with:

- Content stored in many proprietary formats that are closely tied to the print product;
- Content stored on computers at many locations;
- Content with final updates available in publishing format only. Transfer of content from authoring format to publishing format is typically semi-automatic and uni-directional;
- It means that publishers must use labour intensive manual effort to carry over final updates from the publishing format to the authoring format in order that authors can start updating the content for new editions from the final content published in previous editions.
Digital products as spin-off of print products

• In parallel with using digital technology for producing the content, publishers have also used digital technology for publishing the content, starting with CD-ROMs and with the success of the Web moving online.

• Finally the situation changed rapidly with the release of Apple iPad, the smartphone and other computing devices well suited to mass market digital consumption of content products.

• Publishers typically and logically started producing and publishing digital products as spin-offs of print products - i.e. the output from the production of the printed book is used for production of same content in digital format. For example as CD-ROMs, eBooks or online versions of printed books.
Pros & Cons

Advantages in using output from production of print products for production of digital products:

• Publishers have lots of legacy print products;
• Legacy products generate significant revenues, are strong brands and protect publishers' market position;
• Typesetting files for producing the print products are available (for example in Digital Asset Management systems or archives).
Pros & Cons

However, experience over several years indicates that the print-based content supply chain has serious disadvantages such as:

• For many reasons, the layout used in print products does not scale well with lower resolution and varying screen size on digital platforms and the content must therefore be broken up and presented in a new delivery format specific way.

• The print output does not provide the information that is required for exploitation of the search and browse capabilities that consumers expect to have access to in digital products.
Cost, complexity, lack of flexibility

Therefore, content supply chains are produced for each digital delivery format (eBook, online, etc.) and for each content supply chain expensive, complex, semi-automatic, product specific data processing chains are produced.

The data processing chains:

• Converts, adapts, styles the content and assets for each delivery format;

• Makes explicit content structure information and produces the information required for browsing and searching the content;

• Identifies and resolves cross references so they are 'live' in digital products.
Cost, complexity, lack of flexibility

The data processing chains:

• Are highly dependent on even small details in the layout of print product and therefore it is typically not possible to re-use scripts for updated editions/revisions of products;

• Create a multitude of files: the same content is available in many formats and resolutions and lives in many geographic places;

• Are uni-directional which means that final updates to content of digital delivery formats must manually be transferred back to the shared authoring format used for production of print and digital products.
Update cycle: print vs. digital products

Digital - and in particular online - products give reference publishers the opportunity to be in close contact with consumers and adapt and enhance the content in close cooperation with users.

The close cooperation creates the market expectation that the update cycle of digital products is shorter than the update cycle of print products.

Having a shorter update cycle of digital products compared with the update cycle of the corresponding print product does not fit well with:

• The underlying assumptions of using the content supply chain for print products for production of digital spin-off products;

• The uni-directional nature of the content supply chain.
Single source publishing?

• Develop Once, Use Many.

• Separating content from delivery format provides many business benefits, including:
  - Future Proofing Content;
  - Scalability;
  - Decrease Redundancy;
  - Shorten Time To Market;
  - Rapid Updates and Modifications.
Single source production system

• To produce and publish content products for many delivery channels an efficient single source production and publishing system combined with automated content delivery processes is a must.

• The content source is neutral to formats used by delivery channels.

• It means that from planning and all the way up to publishing the content is treated in a delivery channel neutral way.

• The content is the same no matter if it should be published in print, e-book or online format.
2014 - DPS - Streamlining Global Production Process

DPS-CMS
Content Management System

DPS-Process
- Script for online publishing
- Script for digital publishing
- Script for eBook publishing
- Script for print publishing

DPS-Reach
- Online publishing platform
- API
- eBook product
- Typesetting e.g. InDesign

Authoring XML XHTML
Proofing
Workflow
XML Search
Metadata
Publication
Content – API – Delivery channels

API for data processing is a powerful tool for:

• Centralized management of data processing scripts;
• Automated and secure formatting and export of content from CMS (via API) to delivery channel specific formats;
• Import of legacy content to CMS (via API).
Data processing & API

Multiple inputs
- DSP CMS
- file, scn, DB, misc

DPS-Processing
- Job
- Job
- Job

Scheduler

DPS-Processing SDK

Multiple outputs
- Data Processing Engine
- Mass Data Delivery
- Ad Management
- Misc
Single monitoring & managing interface

• As new delivery channels are emerging continuously, the strict separation of content from delivery and formatting of the content in delivery channel specific formats makes your content versatile enough to publish to multiple delivery channels and enables new delivery channels to be added with ease.

• Structure and streamline your publishing process. With a centralized secure XML repository, a powerful search engine, a solid workflow, Web based SaaS based hosting platform that can scale from a single user to hundreds of users (editors, project managers, system administrators).

• Make your daily job easier and enable you to focus on what you do best: successfully create and publish content products.

• Throughout the publishing process, a single user interface to create, edit, manage and publish all your content, regardless the channel.

• Through the Workflow Manager, monitor and manage the production process.
All-in-one production and publishing solution

- XML Repository combined with a powerful search feature;
- Workflow management for projects involving many people;
- Writing tools for editing the content of your products in a user friendly way (XML Editor, Online Editor and Publication Editor);
- Support unbundling and re-packaging of content, tagging and categorising content;
- Via API, integrates with data processing and with third-party software;
- Takes care of access rights, storage, versioning, security;
- It’s just as easy to create an interactive tablet publication as it is to create a print production.
Personal working environment

- User 1
  - DPS Online Editor
  - Personal space proxy

- User 2
  - DPS Online Editor
  - Personal space proxy

- User ...

Client browser

- User 1 view
- User 2 view
- User ...

Server

- XML Database

DPS CMS
Unbundling and Repackaging

• Music industry was a pioneer in unbundling content for digital consumption. However, early sales efforts weren’t successful. Fearing cannibalization, music companies restricted digital music licensing. Early music services were also not integrated with a device, which made legal downloading too cumbersome for most consumers.

• It wasn’t until music was offered the way consumers wanted it – in single tracks with an easy user interface that seamlessly linked the device and the service – that paid online music took off. Music’s experience demonstrates that breaking apart content can be a successful way of getting consumers to pay for content.

• For reference publishers – unlike music industry – it is assumed that unbundling should be combined with features for efficiently re-packaging content in new products and in the way that consumers want it.

• Spotify is in many ways a repackaging Content as a Service (CaaS) offering a large package of music for one price and with people's play lists as an alternative packaging.
Re-purposing content

- Reference publishers are migrating their offerings to keep up with the changing ways consumers access and use content. However, this doesn’t just mean providing the same content in new formats.
- It means completely rethinking how content is produced, packaged, priced, marketed and sold.
- Consumers already demand anytime, anywhere content. In the future – if not already – they will also demand “any form” content – small pieces of customized content suited to the way they live and work.
- Publishers see a real opportunity in unbundling content and repackaging them for sale and are seeking not only to maintain sales of existing product, but also to generate revenue from new ones.
- For example, a student may want to download a single chapter from a textbook to prepare for a test, while a medical professional may only want information about a particular disease and its treatment.
- Rather than risk losing a sale, publishers are increasingly willing to break apart books to meet their customers’ needs — and get paid for doing so.
De-aggregate

- Document units, fragments and assets are tagged and categorized making them easy to find
- Document units can contain links to: document fragments; and assets
- Same fragment and asset can be linked to many document units
- Preview version of assets may be imported from and shared with Digital Asset Management (DAM) system
Re-aggregate

• Publication structures contain:
  ➢ Nodes: 
  ➢ Links: 

• Links point to document units

• Same document unit can be linked to many publication structures
Documents (books) are broken down into smaller units for:

- Ease of editing
- Flexibility: several people can work on (aspects) of a book at the same time
- Flexible reuse: parts of (several) books can be combined in new publications.

For example:

- Chapters on specific grammar points + exercises
- Exercise collections etc.
Nodes & Links
Re-purposing functions

• Creates hierarchical structure of nodes.
• Nodes can contain nodes and/or be linked to documents.
• Nodes can be created, moved, copied or deleted.
• Publication can be exported with the structure plus content intact.
• Like documents, publications are versioned.
Upload legacy data

• Input: The original XML supplemented with information about:
  - How the publication is broken down in manageable units
  - The labelling of units and publication nodes

• Process: split the XML into
  - Units to be uploaded to project DB
  - An XML file representing the tree structure with links to the units to be uploaded as the publication
<book xmlns="http://www.doobook.org/xml/4.5/doobook.dtd">
  <info>
    <title>Demo Grammar</title>
    ...
  </info>
  <chapter>
    <title>I can talk about vocabulary learning</title>
    <section>
      <title>Reading</title>
      <para>When you're reading ...</para>
      ...
      <glosslist>
        <title>Glossary</title>
        <glossentry>
          <glossterm>foreign</glossterm>
          <glossdef>
            <para>from a country ...</para>
          </glossdef>
        </glossentry>
      </glosslist>
      </section>
    <section>
      <title>Speaking and listening</title>
      ...
    </section>
  </chapter>
  <chapter>
    <title>I can describe my progress and aims</title>
    ...
  </chapter>
</book>
Augmented input XML

```xml
  <chapter p:label="I can talk about vocabulary learning">
    <section p:label="Reading">
      <unit xmlns="IDMPE" name="info1" label="Doc">
        <info xmlns="http://www.docbook.org/xml/4.5/docbookx.dtd">
          <title>Reading</title>
          <para>When you're reading ...</para>
        </info>
      </unit>
      <unit xmlns="IDMPE" name="glosslist1" label="glosslist">
        <glosslist xmlns="http://www.docbook.org/xml/4.5/docbookx.dtd">
          <title>Glossary</title>
          <glossentry>
            <glossterm foreign="false">
              Smith
            </glossterm>
            <glossdef>
              from a country ...
            </glossdef>
          </glossentry>
        </glosslist>
      </unit>
    </section>
    <section p:label="Speaking and listening">
      ...
    </section>
  </chapter>
  <chapter p:label="I can describe my progress and aims">
    ...
  </chapter>
</book>
```
Tree structure

<node xmlns="http://www.docbook.org/xml/4.5/docbookx.dtd" xmlns:p="IDMPF">
  <node xmlns="" xmlns:d="http://www.docbook.org/xml/4.5/docbookx.dtd" tag="book"
    label="Deno Grammar">
    <node tag="chapter" label="I can talk about vocabulary learning">
      <node tag="section" label="Reading">
        <node document="info1" label="Doc"/>
        <node document="glosslist1" label="glosslist"/>
        <node document="gandaset1" label="gandaset"/>
        <node document="gandaset2" label="gandaset"/>
      </node>
    </node>
    <node tag="section" label="Speaking and listening">
      <node document="info2" label="Doc"/>
      <node document="glosslist2" label="glosslist"/>
      <node document="gandaset2" label="gandaset"/>
      <node document="gandaset3" label="gandaset"/>
      <node document="gandaset4" label="gandaset"/>
      <node document="gandaset5" label="gandaset"/>
      <node document="gandaset6" label="gandaset"/>
    </node>
  </node>
</node>

<node tag="chapter" label="I can describe my progress and aims">
  <node tag="section" label="Making progress">
    <node document="info4" label="Doc"/>
    <node document="glosslist3" label="glosslist"/>
    <node document="gandaset3" label="gandaset"/>
    <node document="gandaset7" label="gandaset"/>
    <node document="gandaset8" label="gandaset"/>
    <node document="gandaset9" label="gandaset"/>
  </node>
</node>

<node tag="section" label="Future goals">
  <node document="info5" label="Doc"/>
  <node document="glosslist4" label="glosslist"/>
</node>
Discoverability of content

Reference content of print products is typically:
• Built on a business-to-business (B2B) relationship;
• Consumed in a linear fashion - from first to last page in the book;
• Not rich in number of internal cross references and on links to content inside other products;
• By nature not rich in cross references;
• Not categorised and thereby not making it easy to find other related content.

In digital - and in particular online - products the situation is very different:
• Survival depends on building relationships directly with customers in a direct-to-consumer (D2C) relationship;
• Consumers are expecting that:
  ➢ Internal cross references and links to content in other digital products are only one click away and it is easy to get back again when internal cross references and links have been examined;
  ➢ Internal cross references and links to content in other products point to a relevant fragment of publications (such as chapter or section) and not to publications as a whole.
Get prepared to SEO, SMO, ...

Content of digital products is increasingly discovered by search in general purpose search engines and via cross references in social networks where links typically point to relevant fragments inside publications.

- Social networking: The number of worldwide, unique social networking users is growing rapidly. See Facebook, Twitter;
- Internet: Almost 40% of the world's total population is online.
Make content suitable

Making content of digital products accessible on Internet via general search engines and social networks requires that content follows Search Engine Optimization (SEO) recommendations such as:

• The content is broken down into document units that fits well on range of delivery channels - for example smartphones, tablets, laptops, desktops;

• Content is well-structured and structure is made available:
  ➢ To users for navigation and browsing the content;
  ➢ To general search engines so they can make the content well-indexed by trawling the content of digital products;
  ➢ To social network users so it is easy to create links to relevant content.

• Chapters and sections have clear, concise, meaningful titles

• Document units are well-categorised with metadata and metadata can be used to give users (and search engines) instant access to related content in other digital products.
Users’ expectations

• Consumers of content in digital and in particular online products are used to access content in a fragmented way.

• They expect access to cross references, links to related content in other publications.

• They assume that they can use general purpose search engines or links in blogs or social networks to particular chapters or sections in books for finding the information fragments they find relevant in a particular context they are in.

• For improved browsing experience internal cross references and links to content in other digital products are highly valued by users because the cross references are typically much more targeted than a search in general purpose search engines which searches in all indexed web content.

• Internal cross references and links to content in other digital products are also highly valued by search engines and the richness of cross references to and from given document units is an important factor for ranking content in search engines (ref. SEO).
Multiplatform consumption is today's new reality

- The digital media landscape is already well into this transition, and smartphone and tablet usage can no longer be considered a mere rounding error in digital consumer behaviour.

- According to comScore more than 1 in 3 minutes (37 percent) is now spent on smartphones and tablets compared with the PCs.

- As digital consumers become more reliant on their smartphones and tablets for everyday content consumption, it is expected that this share will rise over time and perhaps take over majority share during the course of the next year.

- Envisioning this reality places a premium not only on getting the mobile channels right, but also on proving their value in short order.
The "long tail"

• With, the constantly growing demand for content, the diversity and geographical distribution of users, reference publishers can not alone make full exploitation of their repository of unbundled content.

• They may give educators and API developers access to create entirely new versions of books and lesson plans: Educators and API developers can select chapters and section from different books by different authors, combine them with tests, study guides and other content and publishers provide these customized books through platforms that can be used by educators and students – like Spotify is giving users access to music.

• It could be considered: Content as a Service (CaaS).
Metadata management

• **System metadata**
  - Sortkey, label etc.
  -Handled by configuration

• **Customer defined metadata**
  - Simple values like Publication data, ISBN …
  -Categories like a CEFR, topics …
Costumer metadata

• Configured by administrator
  ➢ Per project, but export/import function
  ➢ Metadata values are typed: string, integer, date, closed set etc.
  ➢ The permission to edit can be restricted to user group

• Metadata are searchable
  ➢ Searches can be combined with searches on content
Configure metadata

<table>
<thead>
<tr>
<th>Category</th>
<th>Code</th>
<th>User Groups</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1</td>
<td>a1_level</td>
<td></td>
</tr>
<tr>
<td>A2</td>
<td>a2_level</td>
<td></td>
</tr>
<tr>
<td>Independent User</td>
<td>independent_user</td>
<td></td>
</tr>
<tr>
<td>B1</td>
<td>b1_level</td>
<td></td>
</tr>
<tr>
<td>B2</td>
<td>b2_level</td>
<td></td>
</tr>
<tr>
<td>Proficient User</td>
<td>proficient_user</td>
<td></td>
</tr>
<tr>
<td>C1</td>
<td>c1_level</td>
<td></td>
</tr>
<tr>
<td>C2</td>
<td>c2_level</td>
<td></td>
</tr>
<tr>
<td>Body and appearance</td>
<td>topic</td>
<td></td>
</tr>
<tr>
<td>Appearance</td>
<td>t_B_body_and-appearance</td>
<td></td>
</tr>
<tr>
<td>Body</td>
<td>t_B_body</td>
<td></td>
</tr>
<tr>
<td>Clothing</td>
<td>t_B_clothing</td>
<td></td>
</tr>
<tr>
<td>Hair</td>
<td>t_B_hair</td>
<td></td>
</tr>
<tr>
<td>Business</td>
<td>t_B_business</td>
<td></td>
</tr>
<tr>
<td>Commerce</td>
<td>t_B_commerce</td>
<td></td>
</tr>
<tr>
<td>Companies</td>
<td>t_B_companies</td>
<td></td>
</tr>
<tr>
<td>Economics</td>
<td>t_B_economics</td>
<td></td>
</tr>
<tr>
<td>Finance</td>
<td>t_B_finance</td>
<td></td>
</tr>
<tr>
<td>Products</td>
<td>t_B_products</td>
<td></td>
</tr>
<tr>
<td>Culture</td>
<td>t_B_culture</td>
<td></td>
</tr>
<tr>
<td>Film</td>
<td>t_B_film</td>
<td></td>
</tr>
</tbody>
</table>
Speaking and listening

Realia: an advice sheet

In a conversation, you can often ask another person to repeat words or phrases you don't understand, and explain the meaning to you if necessary.

Repeat new words to yourself, so you know how to pronounce them.

Look for opportunities to practise them when you're speaking. Don't be afraid to experiment with new language, and don't worry about making mistakes.

It's important to revise new vocabulary. If not, you may forget it quite quickly.
Hosting infrastructure
Growing market of digital services
Estimated revenue of Amazon Web Services from 2012 to 2015 (in billion U.S. dollars)

Revenue in billion dollars

<table>
<thead>
<tr>
<th>Year</th>
<th>Revenue</th>
</tr>
</thead>
<tbody>
<tr>
<td>2012</td>
<td>2</td>
</tr>
<tr>
<td>2013*</td>
<td>3.8</td>
</tr>
<tr>
<td>2014*</td>
<td>6.2</td>
</tr>
<tr>
<td>2015*</td>
<td>8.8</td>
</tr>
</tbody>
</table>
Business models
Free Dictionary Sites Market (Apr. 2014)

- oxforddictionaries.com: 2.5%
- dictionary.reverso.net: 4.7%
- bab.la: 5.4%
- merriam-webster.com: 6.5%
- urbandictionary.com: 6.5%
- wiktioanry.org: 9.4%
- thefreedictionary.com: 15.5%
- wordreference.com: 19%
- dictionary & thesaurus.com: 21.2%

- **Growth Nov.-Apr.**

- oxforddictionaries.com
- The Times
- macmillandictionary.com
- babla
- dictionary.cambridge.
- reverso.net
- wordreference.com
- collinsdictionary.com
- dictionary & thesauri...
- The Sun
- Wall Street Journal

- 50.00%
- 25.00%
- 0.00%
- -25.00%
- -50.00%
RTB is no longer just about remnant

Old Revenue Mix

Direct

$10+ CPM

Indirect

~$1 CPM

1

Impression Volume

100

New Revenue Mix

Direct

$10+ CPM

Programmatic

~$6 CPM

Incremental brands & budgets

~$4 CPM

Preferred access

~$1 CPM

1

Impression Volume

100

- Capture more value from unsold impressions
- Protect direct sales focused on high CPM brand campaigns
- Get transparency into the buyer, advertiser and rate
Q/A