Standardizing the discovery of digital book content: a work in progress

Michael Healy, Book Industry Study Group

Goals

- To enable digital book content to be discovered, browsed, searched and distributed in a standardized way
- To allow publishers to manage the quality and availability of the content
- To achieve these goals by defining a set of HTTP transactions between a publisher’s digital archive and the websites of syndication partners

Background

- AAP Digital Issues Working Group
- Project Brief and Preliminary Requirements were published in June 2007
- Work transferred to BISG in July 2007
- Establishment of Digital Standards Committee
- Wider industry participation: booksellers, digital repositories, search engines etc.

Participating Organizations

- AAP
- ABA
- Barnes & Noble.com
- BookNet Canada
- Bridgeport National Bindery
- CBA International
- Copyright Clearance Center
- Digital Library Federation
- FYI Solutions
- Google
- GS1 US
- Hachette Book Group USA
- Harcourt
- HarperCollins
- IDPF
- Ingram Digital Group
- John Wiley & Sons
- Knet
- LibriDigital
- LINDAwest
- McGraw-Hill
- Metacom Systems
- Microsoft
- NACS
- Offix Media
- Oxford University Press
- Pearson
- Penguin Group (USA)
- Publishing Technology
- Quality Solutions
- R.R. Bowker
- Random House
- Rosetta Solutions
- Semelde Solutions
- Simon & Schuster
- SmartBo Associates
- Wompra Publishing

Use Cases Covered by the Specification

<table>
<thead>
<tr>
<th>Use Case Reference</th>
<th>Specification Code</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>UC-01</td>
<td>TR-01</td>
<td>Get Number of Phrase Matches in Archive</td>
</tr>
<tr>
<td>UC-02</td>
<td>TR-02</td>
<td>Search Archive for Phrase</td>
</tr>
<tr>
<td>UC-03</td>
<td>TR-03</td>
<td>Search Book for Phrase</td>
</tr>
<tr>
<td>UC-04</td>
<td>TR-04</td>
<td>Enumerate Archive</td>
</tr>
<tr>
<td>UC-04.5</td>
<td>TR-04.5</td>
<td>Enumerate Book Info</td>
</tr>
<tr>
<td>UC-05</td>
<td>TR-05</td>
<td>Enumerate Book Page Images</td>
</tr>
<tr>
<td>UC-06</td>
<td>TR-06</td>
<td>Enumerate Book Title</td>
</tr>
<tr>
<td>UC-07</td>
<td>TR-07</td>
<td>Enumerate Book Text</td>
</tr>
<tr>
<td>UC-08</td>
<td>TR-08</td>
<td>Browse by Context Pages</td>
</tr>
<tr>
<td>UC-09</td>
<td>TR-09</td>
<td>Browse by Sample Pages</td>
</tr>
<tr>
<td>UC-11</td>
<td>TR-11</td>
<td>Get Full-Page Media</td>
</tr>
<tr>
<td>UC-12</td>
<td>TR-12</td>
<td>Get Thumbnail Media</td>
</tr>
<tr>
<td>UC-14</td>
<td>TR-14</td>
<td>Get Highlighted Page Media</td>
</tr>
</tbody>
</table>

The requirement ……..

…but these implementations referenced two different web service specs – that’s not scalable!
Project Participants

- **Enumerate Group**: David Mandelbaum (Barnes & Noble) - Chair
  - Paul Gore (Hachette)
  - Brian Hogue (Ingram Digital)
  - Steve Kotrch (Simon & Schuster)
  - Ricky Leung (ABA)
  - David Maralack (Microsoft)
  - Rich Rothstein (Harpercollins)
- **Get Group**: Dave Haase (Random House) – Chair
  - Neil DeYoung (Hachette)
  - Brian Hogue (Ingram Digital)
  - Steve Kotrch (Simon & Schuster)
  - Chris Odom (LibraDigital)
  - Rich Rothstein (Harpercollins)
- **Search Group**: Brian Hogue (Ingram Digital) – Chair
  - Matt Bennett (Hachette)
  - Suraj Gaurav (Microsoft)
  - Stephanie Halpern (R.R. Bowker)
  - Ricky Leung (ABA)

Current status

- Initial BISG Board approval, May 2008
- Technical specification is being finalized
- Prototyping phase – June-August 2008
- Participation of publishers, digital repositories and syndication partners
- Final specification review
- Publication – September 2008

Questions?

Michael Healy
michael@bisg.org