Migration to ONIX 3.0

Michael Gordon Smith and Cat Colwell (Australian Publishers Association)
Established by the Australian Publishers Association in 2004 for the benefit of the book industry to provide a single point of access to local availability for books in the Australian market.

1. Single search repository of Australian market titles
2. Price, availability and bibliographic information 24 hours a day, 7 days a week
3. More efficient customer service and reduced overheads
4. Market-level promotion and discovery: websites, catalogues, custom metadata feeds
5. Data from the source: all publisher provided
6. Extend capability to all publishers including the smallest
History

TitlePage created by joint investment

Joint investment by 8 publishers to build an industry P&A service - investment reimbursed over 3 years.

Data auditing reports and EDI integration

Publishers receive immediate reports upon file delivery. Integration with main EDI provider.

ONIX 3.0 development begins

New architecture to accommodate additional ONIX 3.0 data; active capacity management; UAT environment

Now

ONIX 3.0 live or in final testing for 63% of titles; delayed by Covid-19.

External 5 year license agreement

Investment in expansion of web service integration and EDI sought through 5-year license agreement

Development returns in-house

Legacy ONIX 2.1 system rebuilt using PostgreSQL; service migrated to AWS in Australia;

Transition for publishers begins

Australian publishers advised 12-month transition window
Current statistics:

- 1694 registered organisations: 152 publishers, 239 libraries, 1303 booksellers
- 1,346,925 titles
- 1M page views and 450k unique searches per month
- Average session duration: 13 minutes
Service to booksellers

- Browser searches
- API queries for key data
- API query to return a complete ONIX 3.0 record
- Full or Delta ONIX feeds
- CSV P&A feeds
- Monthly new release files
- Promotional ONIX feeds—eg new release by subject, by Australian authors etc

Browser search example
Why ONIX 3.0: the case for TitlePage and Australia

TitlePage accepted a very limited subset of ONIX 2.1, focused mainly on price & availability, playing little role in marketing or discovery. Expanding to full ONIX 3.0 offered:

<table>
<thead>
<tr>
<th></th>
<th>Handling of digital products and subscription models</th>
</tr>
</thead>
<tbody>
<tr>
<td>01</td>
<td>Block updates to allow multi-party data management</td>
</tr>
<tr>
<td>02</td>
<td>Access to all publisher collateral</td>
</tr>
<tr>
<td>03</td>
<td>Better management of sets &amp; series</td>
</tr>
<tr>
<td>04</td>
<td>Accurate market and supply detail information</td>
</tr>
<tr>
<td>05</td>
<td>Thema</td>
</tr>
<tr>
<td>06</td>
<td>...and all the other data previously discarded from ONIX 2.1</td>
</tr>
</tbody>
</table>
Our aims

- Accept, store, transmit and display all available data from publishers in any ONIX 3.0 field
- Contain the cost of development
- Maximise load and export speed
- Improve search and filter for key fields
- Allow publishers to submit either ONIX 2.1 or ONIX 3.0 during transition
- Reduce risks for publishers changing feeds
- Allow a record to be updated by publisher AND separate distributor

Illustration credit: Ainsley Seago.
Our solution: a hybrid database & S3 file repository

- UAT environment & comparison reports
- Document mapping of ONIX 2.1 -> 3.0
- Limited expansion to database structure
- AWS S3 storage of ONIX 3.0 single product records
- TitlePage ONIX 2.1 -> 3.0 converter
- TitlePage ONIX 3.0 -> 2.1 converter
- TitlePage ONIX 3.0 strict exporter
- Autoscaling database master/slave cluster
- EFS server load scaling
- API link to ONIX 3.0
Challenges & best laid plans

- Mapping between ONIX 2.1 to 3.0 for transition period
- Block updates - lack of international adoption and file update timing
- Title management for micro publishers
- Inconsistent appetite for change - Publishers, Booksellers, POS providers
- Increased variation between data from different sources
- Invalid legacy ONIX 2.1 data
- The useful display of an inconsistent but significantly increased volume of metadata
What’s Next?

- Thema
- Display
- API stock updates
- Catalogue service
- Tiered data scoring
- Merging data sources
- Diversity data