Innovation Showcase

New and exciting innovations in under ten minutes!
Using AI to solve the challenges of content tagging

Gerasimos Razis
Senior Product Manager, Literatum Taxonomies, Atypon
July 21, 2022
• A taxonomy is an established (whether by you, or globally) set of topics that can be used to organise content, accounts, etc.

• Enhanced content and audience discovery, onsite
  • Topic Pages
  • Topic Facets
  • Topic Search
  • Personalised recommendation and search

• Indexable taxonomical keywords contribute to SEO

• User interest profiling
Three approaches to creating a taxonomy

1. Manually create your own bespoke taxonomy
2. Employ a third party to create a bespoke taxonomy
3. Rely on publicly available taxonomies such as MeSH

Challenges

- A manual or rule-based approach is impractical on a large scale
- Bespoke fits your current needs very well, but may not fit them tomorrow
- Requires domain expert, and is costly and prone to human error
The Atypon Taxonomy

- 250k tags across 19 disciplines
- Continuously curated and expanded
- And it comes with the Atypon Auto-Tagger
The Atypon Auto-Tagger

- Consistently winning BioASQ competitions in the last 5 years
- 8% more accurate than the official solution of National Institute of Health (NIH)
- 15 publishers onboard already

“We use auto-tagger to tag articles for our ‘micro-sites’. It works really well for this purpose! We just load the content, the system tags it and then it shows up on these ‘micro-sites’!”
The Atypon Auto-Tagger

Stage 1: Training the Auto-Tagger

Tagged doc 1

Tagged doc 2

Tagged doc N

Tag1: Computer
Tag2: Education
Tag3: IT technology

Tag1: Design
Tag2: Education
Tag3: Project Management

Tag1: Computer
Tag2: Education
Tag3: IT technology
Tag4: Software engineering
Tag5: Software Quality

Auto Tagger

Trained Model
The Atypon Auto-Tagger

Stage 2: Auto-Tagging
These **product outcomes** drive **business outcomes**, such as:

- More site traffic
- Greater user experience and satisfaction
- Improved understanding of market needs
- Increased revenue for the publisher
- More time to focus on other, important aspects of your business
Subscribe to the Atypon Community Newsletter
to keep up-to-date on AI's impact on scholarly publishing

atypon.com/subscribe
Straive Data Platform – Enabling your Publishing Strategy

By Murugesh Mayandi, Chief Data Solution Officer

July 2022
Every Company Today is a Data Company

2,500,000,000,000,000,000 (2.5 quintillion) bytes of data created each day

- **Oil & Gas**: Environmental impact through geo-spatial data
- **Banking**: Reduce risk & fraud through customer data
- **Education**: Deliver personalized learning through student data
- **Health & Science**: Improve public health through pharmacovigilance & RWE data
Data as Part of Publishing Lifecycle

**Data Streams**

**Content**
Enriched asset metadata, keywords, summaries

**Social**
Social engagement scores

**Experience**
Consumption, Web analytics, Campaign analytics

**Regulatory**
Rights, Contracts

**Operational**
Publishing speed, Accept ratios,

**People**
Authors, Reviewers, Editors, Readers

**Institution**
Libraries, Universities, Funders

**Financial**
Price, Volume, RoI
What are the Possibilities of Using Data to your Advantage?

- Enriched customer and author profiling
- Granular content tagging enabling better re-useability
- Rich data models enabling easy reporting and visualization
- Alternate data indicators to predict key metrics such as customer experience, churn etc.
What are the Challenges?

**Data in Silos**
Data exists in different systems that do not talk to each other

**Governance & Stewardship**
Lack of well defined practices around data models, taxonomies, and linkages

**Data Gaps**
Many data points relevant to the business are not tracked/captured

**Data Quality**
Incoming data from different streams needs normalization before integration to derive value
How SDP can Help?

SDP is an AI-based platform with customizable extraction, enrichment, transformation, and delivery modules. It ingests unstructured data from diverse sources and transforms it into analytics-ready, integratable data for enterprises to gain actionable insights.

**Inputs**
- Public Web Sources
- Internal Documents
- Audios, Videos and Images

**Extraction Module**
- Web Content Extraction
- Entity Extraction
- Normalization Engine
- Anomaly detection

**Enrichment Module**
- Taxonomy & Classification
- Summarization
- De-Duplication
- Data quality rule build & auto recognition

**Transformation Module**
- Image Extraction
- Disambiguation & Linking
- Client QC Environment

**Data Quality Module**
- Text Extraction
- Text Classification
- Text Summarization
- Text Normalization
- Data Quality Engine

**Delivery Module**
- Image Classification
- Image Summarization
- Image Normalization
- Image Quality Engine

**Output**
- Processed Data
- Structured Data

**Hosting**
- Cloud
- On-premise

**Benefits**
- Single, integrated platform
- Platform supported by NLP, AI/ML, PDF processing
- Multiple data ingestion paths
- Ease of use for users on both ends
- Scalable
Case Study: Enabling the People Data Stream

Leading business information provider wanted an end-to-end contacts management solution for maintaining corporate leadership and contacts information for companies in North America.

<table>
<thead>
<tr>
<th>Outdated Data</th>
<th>Insufficient Coverage</th>
<th>Manual Process</th>
</tr>
</thead>
<tbody>
<tr>
<td>People data module on SDP deployed with to enable data capture and disambiguation</td>
<td>Search algorithms based on company websites and social pages</td>
<td>AI-based dedupe routines for automated fixing with exception processing by humans</td>
</tr>
<tr>
<td>Analytics driven update strategy by calibrating update frequency with expected churn based on company tiers and industry type</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Benefits Delivered

- Accuracy: 95%
- Automation: 70%
- Scale: 18M+ contacts across 3M companies in 4 months
- TAT: 100% OTD
In Summary

- Understand available data streams and develop an encompassing data strategy
- Create an integrated platform/solution plan to derive value incrementally and over a period of time
- Be an explorer in the quest for new data streams
The Power of Integration

SSP Innovation Showcase
Joe Adams, Senior Director of Sales
Integrating end-to-end solutions for improved user experience and faster research breakthroughs.
Integrated technologies lead to an integrated research ecosystem.

This is early-stage research.
What we do

Peer Review Workflows
Streamlined from calls for content to submission and integrity checks.

Conference Hubs
Focused on networking and sharing research in an easy-to-navigate platform.

Research Libraries
Multimedia libraries with robust search and user profiles for collaboration.

Automatically integrate accepted content into building conference programs or proceedings volumes.

Extend the value of conferences with interactive libraries.

@2023 Morressier. All rights reserved. Morressier Restricted.
Unlocking the potential of integration

1. Multimedia content management
2. In-platform review of all workflow stages
3. Feed conference research directly into journals
4. Identify trends in early-stage research
Partnership Funnel

Submission Workflows
Simplify pain points and pave the way to quality submissions

Research Libraries
DPC Revenues/Subs Model, corporate sales & advertising

Feed into Journals
Select the best early-stage research to submit to journals/preprint

$ OA
Generate more APC revenue
Thank you!

Get in touch:

joe.adams@morressier.com
Instant Course: Bringing Agile to Course Development

July 21, 2022
Courseware is expensive and cost prohibitive for low enrollment courses

With an average cost of $45k for a course that can easily climb to $250k or more in STEM or accounting, this makes it difficult to justify courses for upper division or small enrollment courses. As a general rule of thumb the break even point over three years for a course is as follows:

- 10,000 units: 1 color title with minimal digital components
- 25,000 units: 4 color title with minimal digital components
- 50,000 units: 4 color title with course
Lower enrollment courses often use eBooks for an online offering as a substitute to courseware

Advantages
- eBooks are a byproduct of composition – usually inexpensive to produce
- Widely accepted and available on mobile platforms
- Distributable through channels such as Amazon, VitalSource and RedShelf

Disadvantages
- You lose margin, or require higher prices by distributing through channels
- Self-distribution requires a dedicated eBook delivery mechanism
- Hard to update – either by the publisher or by a user
- Easy to pirate
- Scope and sequence not adjustable by users
- Hard to integrate graded activities and assessments
- Specialized skills are required to add video and interactivity requiring web services into the eBook package
- The eBook package itself keeps getting larger and larger
- Cannot take advantage of the features of an LMS
Why developing courseware is expensive

⚠️ Lots of starts & stops
⚠️ Very manual
⚠️ Multiple, redundant workflows
Part of the solution is Lean Content Development

Concurrent workflows with **one single source of truth**.

- Increased efficiencies & quality control
- Accelerated time to market
- Enhanced content flexibility and scalability

- No starts & stops
- Highly automated (including tech & design)
- Single workflow for all end print and digital products
- Reduced reliance on 3rd party vendors
Lowering the cost barrier to creating courseware

**AUTOMATED CONTENT TRANSFORMATION**
Take an existing epub file and ingest into the transformation tool, and automatically convert the ePUB file to component XML files

**RAPID AUTHORING & ASSEMBLY**
Collaboratively review, update, preview in real time, and assemble the course

**PUSH-BUTTON DISTRIBUTION**
Export the course in Scorm or IMS Common Cartridge at the click of a button
Content transformation: start with the ePUB file and extract the content...

```
<table>
<thead>
<tr>
<th>Filename</th>
<th>Valid?</th>
<th># Success</th>
<th># Partial matches</th>
<th># Errors</th>
<th>Accuracy %</th>
<th>HTML File</th>
</tr>
</thead>
<tbody>
<tr>
<td>navigation</td>
<td>✔</td>
<td>751</td>
<td>155</td>
<td>0</td>
<td>92%</td>
<td>navigation.html</td>
</tr>
<tr>
<td>title.xhtml</td>
<td>✔</td>
<td>17</td>
<td>2</td>
<td>0</td>
<td>94%</td>
<td>title.xhtml</td>
</tr>
<tr>
<td>copyright.xhtml</td>
<td>✔</td>
<td>24</td>
<td>2</td>
<td>0</td>
<td>95%</td>
<td>copyright.xhtml</td>
</tr>
<tr>
<td>preface.xhtml</td>
<td>✔</td>
<td>8</td>
<td>2</td>
<td>0</td>
<td>87%</td>
<td>preface.xhtml</td>
</tr>
<tr>
<td>formatted.xhtml</td>
<td>✔</td>
<td>50</td>
<td>2</td>
<td>0</td>
<td>96%</td>
<td>formatted.xhtml</td>
</tr>
<tr>
<td>features.xhtml</td>
<td>✔</td>
<td>25</td>
<td>2</td>
<td>0</td>
<td>96%</td>
<td>features.xhtml</td>
</tr>
<tr>
<td>author.xhtml</td>
<td>✔</td>
<td>6</td>
<td>2</td>
<td>0</td>
<td>83%</td>
<td>author.xhtml</td>
</tr>
<tr>
<td>edition.xhtml</td>
<td>✔</td>
<td>11</td>
<td>3</td>
<td>0</td>
<td>88%</td>
<td>edition.xhtml</td>
</tr>
</tbody>
</table>
```
Authoring & assembly step 1: review the raw course...

Accounting Refresher: Financial Statements

Four financial statements are completed near the end of the accounting cycle to summarize activity for the period and indicate the company's current financial position. The order in which the financial statements are completed is as follows:

- Income statement
- Statement of owner's equity
- Balance sheet
- Statement of cash flows

We will examine the first three financial statements here.
**Authoring & assembly step 2: add assessment...**

activity during the period. Unlike the previous two financial statements, the third line of the balance sheet header shows the final date of the period.

---

### Excel Accounting

**Question 1**

What is the formula for net income?

- Revenue + Expenses
- Liabilities + Revenue
- Expenses + Liabilities
- Owner’s Equity + Liabilities

1/2

---

[![Save answers](#) ![Grade me!](#)](#)
Authoring & assembly step 3: add videos, other assets, etc....
Push button distribution: export the Common Cartridge...
In summary, Instant Course allows...

1. A publisher to take an eBook and quickly transform it into a course.
   a. Start with the information you have
      i. eBook, test banks, supplementary material, etc.
   b. Build the course quickly and get it into students’ and professors’ hands
   c. Iterate on what works and what doesn’t with quick feedback.
   d. Use the strengths of the LMS to provide features not available in an eBook

2. In other words, you’re starting with a Minimal Viable Product (MVP), and updating as you go (This is also the essence of continuous publishing)

3. This results in a more engaging learning experience and new revenue opportunities
To Learn More
Researcher

The leading mobile discovery and discussion platform for scientists and researchers

Download for free by searching for “researcher” in your mobile app store

2022
Now more than ever, we depend on **14 million + scientists and researchers** to solve the world’s biggest challenges:
But two workflow obstacles slow the pace of progress:

<table>
<thead>
<tr>
<th>Discovery</th>
<th>“Tidal wave” of literature makes it impossible to keep up:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>3.3m articles p.a.</td>
</tr>
<tr>
<td></td>
<td>&gt;30,000 Journals</td>
</tr>
<tr>
<td></td>
<td>4% CAGR</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Discussion</th>
<th>COVID destroyed in-person conferences, hybrid now the norm:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>69%</td>
</tr>
<tr>
<td></td>
<td>Academics think lack of networking opportunities are the biggest issue with online conferences</td>
</tr>
</tbody>
</table>
Researcher has become the leading content discovery platform on mobile since launch in late 2017

8m Papers served

22,000 Content sources

“Never miss important research”
Reach and engage your followers on Researcher

Claim your journals / publisher / company profiles and get admin access on Researcher

‘The overall visual organization and ease of navigating the Analytics tab is my favourite feature so far!’

Deborah Plavin, American Society for Microbiology

‘We’ve had real success with posting a blog to the PeerJ Computer Science feed. We’re excited about making more use of that and intend to have a weekly post for both journals’

Nathaniel Gore, PeerJ
Researcher **Live**: discuss the research you discover with our global community – launched August 2021

- Free live audio events – paper presentations, Q&As, launches, lectures, casual group discussions
- Anyone with the event link can join, listen and contribute
- Live events automatically promoted to relevant users to drive dissemination and impact
- Creates unique opportunities for collaboration and networking – wherever you are and whenever you like
- Unique content to Researcher, with strong network effects
In January 2022, thought leaders in bioconjugation and DNA sequencing addressed an expert audience exclusively on Researcher Live.

Key Advances in Bioconjugation, January 2022

922 Total registrations
5,262 Event page views
440 Listeners to the recordings
405 Total attendees
Please do get in touch to learn more about partnership opportunities

Researcher

Scale Space
58 Wood Lane
London W12 7RZ
United Kingdom

Mary Sweny
Head of Business Development
mary.sweny@researcher-app.com

Use the QR code to download the app
Name: John Challice
Date: July 21, 2022

SSP Innovation Showcase
enables all publishers to turn first-party data into immediate business value.
Customer Data Platform (CDP)

- Packaged software
- Takes data from all sources
- Builds complete customer profiles
- Allows other systems access to data
- Realtime processing
A CDP for Publishers
Hum CDP’s Broad Intelligence

INSTITUTION

PERSON
Reader
Author/Reviewer
Subscriber
Anonymous

CONTENT
Article/book/chapter
Webpage
Video
Email/Newsletter
Webinar/event
Course (education)

TOPIC
Deep Intelligence: Democratizing AI

PUBLISHER SPECIFIC

- CueBERT - a world class NLP pipeline to read and tag all types of content
- Engagement scoring
- Fractional attribution modeling
- Propensity modeling
- Lookalike modeling
What makes Hum special?

<table>
<thead>
<tr>
<th>CAPABILITIES</th>
<th>APPROACH</th>
</tr>
</thead>
<tbody>
<tr>
<td>● Content Intelligence</td>
<td>● <strong>Built for humans</strong> - not just data scientists and engineers</td>
</tr>
<tr>
<td>● Segmentation</td>
<td>● <strong>Cost effective</strong> - 1/10th cost of competing solutions all-in</td>
</tr>
<tr>
<td>● Personalization</td>
<td>● <strong>Fully managed</strong> - setup, integrations + ongoing support</td>
</tr>
<tr>
<td>● Audience Explorer</td>
<td></td>
</tr>
<tr>
<td>● Content Explorer</td>
<td></td>
</tr>
</tbody>
</table>
Discover more about your content

Need help figuring out what to produce content on? Check out our recommendations below based on predictions and your current content performance.

<table>
<thead>
<tr>
<th>#</th>
<th>Keyword</th>
<th>Content Count</th>
<th>Total Views</th>
<th>Engagement Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>segmentation</td>
<td>14</td>
<td>2,354</td>
<td>97</td>
</tr>
<tr>
<td>2</td>
<td>hiring</td>
<td>32</td>
<td>2,642</td>
<td>95</td>
</tr>
<tr>
<td>3</td>
<td>millennials</td>
<td>20</td>
<td>1,463</td>
<td>92</td>
</tr>
<tr>
<td>4</td>
<td>gen-e</td>
<td>22</td>
<td>1,757</td>
<td>89</td>
</tr>
<tr>
<td>5</td>
<td>collaboration</td>
<td>8</td>
<td>935</td>
<td>88</td>
</tr>
</tbody>
</table>

About this report: Topics that have high engagement scores but low page views.
Person / Content Interactions
Audience Explorer / Segments

Image of the Audience Explorer interface with a pop-up window asking to name the segment.
Medical device security can’t be solved in healthcare: What’s ‘acceptable risk?’

Jessica Davis  December 9, 2021

Every single healthcare organization facing longstanding challenges with securing the medical device ecosystem, which will persist into the indefinite future. How then can CISOs determine “acceptable risk”?

- Personalized recommendations
- More like this (similar content)
- Others with similar interests have liked
- Topic-based browsing
- What’s new
- Most engaging
THE BIG IDEA

Makes Publishers Data-Driven
Offers Accessible, Advanced AI
Turns Data Into a Monetizable & Strategic Asset

hum
Thank You!

2022 State of Personalization Report

More about Hum for Scholarly Publishers

John@hum.works
A leading repository platform for storing, accessing, and citing research outputs visualized in the browser

info@figshare.com
visit figshare.com
Figshare and data curation

"The challenge with data sharing is now quality, not quantity."

During data deposit, personal email support is available and once submitted, a Figshare data expert will conduct a review of the metadata and a spot check of your dataset files and offer guidance, with the main aim of enhancing discoverability and reusability.
Spotlight features of Figshare+

- Publish datasets ranging from 20GB to 5TB or more
- Link your data to associated preprints, publications, grants, and other materials
- Review and checks of your metadata and description (and spot checks on a sample of your data files) by Figshare’s data review team, with the aim of making the data more discoverable and reusable

See more at [https://knowledge.figshare.com/plus](https://knowledge.figshare.com/plus)
Showcase and track the hidden impact of your research data

Portal is a place where all of the supplemental data associated with a publisher’s articles can be showcased and tracked. With a subportal for each journal (or each subject area) the data is automatically organised making it easy to find and access.

The Figshare Viewer

Viewer is an embeddable widget that allows you to preview data within your journal platform, alongside the articles, removing the need for downloading and 3rd party software.
Figshare Curation Services

- Support for individual journals
- Reviewing metadata of datasets before they get published and guiding researchers through the data sharing process.
- File and metadata checks to enhance the discoverability and reusability of data in line with the FAIR principles and support repository-specific standards.

See more at [https://knowledge.figshare.com/curation](https://knowledge.figshare.com/curation)
Thank you!

Please get in touch at: https://ds.digital-science.com/l/533712/2022-07-13/9tqdqk

Dan Valen
Head of Strategic Development
dan@figshare.com
@dvnln
- Jul 28, 2022 | Ask the Experts “Ethics in Publishing”
- Aug 9, 2022 | Early Career Webinar “Getting to the Next Level: An Exploration of Career Development for Early-Career Professionals in Scholarly Communications”
- Sep 15, 2022 | 2022 Webinar Series “The Future of Scholarly Publishing: Three Trends to Be Ready For”
- Sep 21-22, 2022 | 2022 New Directions in Scholarly Publishing “Creating the New Possible”
PROMOTE YOUR PRODUCTS, SERVICES AND CONTENT TO THE SCHOLARLY AND ACADEMIC PUBLISHING COMMUNITY!

A comprehensive resource for publishing and related services, industry freelancers, and upcoming industry events.

Support the future of scholarly communications through the power of philanthropy for all.

#sspND2022

2022 NEW DIRECTIONS IN SCHOLARLY PUBLISHING SEMINAR

Creating the New Possible
Shaping the Future from Lessons Learned

September 21-22, 2022 | Washington DC + Online

Contribute to a one-of-a-kind tool that charts positions and career paths across scholarly publishing.

Add a position description to our job data bank!