

ANALYTICS: The value in sharing

An Innovative Conversation by David Kay, Owen Stephens and Annette DeNoyer

OCTOBER 2013

In August 2013, we spoke to three leaders in library academic services at UK universities where innovative systems are long established, namely:

Martin Myhill, Assistant Director (Academic Services) E-Library Strategy at the University of Exeter;

Chris Awre, Head of Information Management at the University of Hull; and

Robert Sherratt, Head of Student Experience at the University of Hull.

Martin believes strongly in an expansive view of the role and value of analytics, based on the data that can be collected by the library or linked to it from internal and external sources. He proposes four use cases for library analytics (around which this conversation is structured) and suggests the potential value-add of sharing in each case.

Chris and Robert focus on the importance of sharing as a catalyst for developing understanding amongst library managers of the potential of analytics and of how such data might be collected and built up over time. They also pinpoint areas where meaningful scale may be achieved through a shared analytics service.

SOME BACKGROUND

The idea of shared services based on the analysis of library activity data, potentially covering more than just resource usage, is increasingly a hot topic, although certainly not new.

The aggregation and analysis of collection and activity data has long been part of the working practices of resource sharing collaborations such as OhioLINK¹, last copy services such as the UK Research Reserve², and shared licensing initiatives such as operated by SHEDL (the Scottish Higher Education Digital Library).³

So, what has changed?

In recent years, other sectors—ranging from retail to sport to financial services—have demonstrated the use of activity data to target marketing and to enhance services. General and personal patterns of human activity are tracked and analyzed using combinations of transaction data, external datasets, and even social media. We might think of this use of activity data as the ‘Amazon factor,’ whilst recognizing it as far more pervasive.

¹ <http://www.ohiolink.edu>

² <https://www.ukrr.ac.uk>

³ <http://scurll.ac.uk/what-we-do/procurement/shedl/>

At the same time, the underlying technologies for collecting, collating, analyzing, and visualizing such data on a very large scale have developed significantly. High-price monolithic warehousing approaches suited to post-hoc analysis are being superseded by near real-time agility⁴.

Not least, the Higher Education sector is warming to the value of analytics to fuel customer-facing decisions in economically constrained times, with particular focus on researcher success and on student recruitment, progression and retention, embracing learning analytics and the power of ‘early warning systems’ data.

These developments have precipitated a wide range of institutional trials and vendor developments, especially in the Learning Management Systems space. Regardless of domain, every Higher Education system vendor must incorporate an integral approach, and analytics is now a strategic focus for EDUCAUSE and for Jisc in the UK.

Against this backdrop, stage left, enters the library.

Libraries have considerable experience using analytics (previously and more prosaically known as ‘Reports’) to

⁴ Solutions include tools such as NoSQL databases, the indexing approaches offered by Lucene / Solr, and distribution frameworks such as Hadoop.

inform collection management. However, libraries typically take a cautious view of the potential of Amazon-style recommenders ('Students like you accessed this' ... 'Students who gained first class degrees also borrowed that') and of considering library data (such as resource or turnstile access) as potential early warning indicators of student progress.

In particular, library managers face two fundamental questions regarding the new wave of analytics, moving beyond collection management reporting: one regarding the personalization of services based on user activity, bringing with it a range of legal and ethical challenges; and another concerning the possibility that usage data may become much more interesting (and beneficial) when shared, not only for reasons of scale but also for informing shared service opportunities.

See our Shared Services White Paper for further details.

SERVICE PROPOSITIONS

Our conversations with Chris Awre, Martin Myhill, and Robert Sherratt explore the opportunities for shared analytics services at four levels:

- **Collections Dashboard**
- **Audience Analytics**
- **Big Data Narratives**
- **Benchmarking**

Collections Dashboard

At Exeter, Martin Myhill's baseline category for engagement with analytics is the collections dashboard. It seems simple: what we used to call reporting, we now call analytics. Data is something every library depends on, especially as weeding and demotion to secondary storage or a final copy reserve are crucial in the quest for space. However, Martin suggests we think more smartly about the range of data needed on our dashboard and also about how we use the data.

For Martin, the opportunity lies in being more directly driven by actionable data instead of calling upon data post hoc to inform periodic revisions to human-mediated policies.

Each of our conversations emphasized that the management dashboard needs to go beyond print management to include purchased and subscribed e-resources, as well as repositories. Gathering such data in a timely manner is prime shared service territory. In the UK, JUSP (the Journal Usage Statistics Portal) provides a key shared service, working on behalf of the community with the publishers to collect, validate, collate, and present. With open access repositories and perhaps digitized collections, we have a similarly distributed problem of wanting to know levels of outgoing usage (what are our students using?), as well as incoming traffic (how useful is our stuff?), so there appears scope for a JUSP-like service for a range of open web-accessible collections.

Audience Analytics

Our second consideration shifts focus from the collection to the community. As soon as 'people' come in to the frame, we need to address concerns about privacy (looking after the individual) and competitive advantage (looking after the institution).

Martin argues that sharing transaction data and patterns of use, which lead to predictive opportunities, is not a threat but rather a source of opportunity. The key is how to convert the data into effective personal recommendations, as it seems likely that such services will be more fruitful if seeded by a greater volume and wider range of data. The ultimate vision may be a personal user dashboard, including recommendations sourced both from the course reading list and also from 'users like me,' with links to access and availability.

Everyone agrees on shared service opportunities for a very different type of user data—that there is much to learn from enquiry service data in all forms, ranging across email, telephone, front of house, and social media.

At Hull, Robert Sherratt makes the connection between automated recommendations at scale and building a critical mass of social interaction between users. The value of a shared recommender service might not be in the automated recommendations themselves but rather in creating an active platform for social interactions such as ratings and comments. If so, a shared reading list platform may be the place to start building such feedback mechanisms, as reading lists seem to attract eyeballs and genuine student attention.

Everyone agrees on shared service opportunities for a very different type of user data – that there is much to learn from enquiry service data in all forms, ranging across email, telephone, front of house, and social media. Both libraries reported that new channels are expanding data collection and feedback opportunities; for example, roving library staff using tablets to take enquiries on the library floor.

Enquiry data may be a clear example of where a lone institution does not generate sufficient data to identify more than the obvious patterns, but where combining data across several institutions may reveal richer narratives. The feedback loop to enhancing online support (such as FAQs) and improving both user and staff training may be very beneficial, in addition to triggering acquisitions and informing collection policy.

Big Data Narratives

The volumes of analytical data handled by the online business giants superficially consign a library's transaction data (searches, loans, downloads, enquiries, etc.) to the Little Leagues.

Nevertheless, millions of circulation-related transactions (for example, 3+ million accumulated by a medium-sized UK academic library over 13 years⁵) without doubt contain masses of untapped user stories and actionable collection insights. These transactions also represent a scale too large for manual analysis and traditional reporting. Beyond a certain scale, big data techniques must apply; hence, our third level of analytics opportunity!

Hull's Chris Awre emphasizes how important it is for the library to have flexibility in slicing and dicing such data, rather than being constrained by fixed reports and pre-cooked analyses. Equally crucial is the ability to accumulate atomic transaction data over long periods, meaning many years, rather than losing the detail in periodic aggregations.

Likewise, Martin Myhill adopts an ambitious position regarding the value of accumulating (as opposed to aggregating) connected datasets over time. He believes 'the wisdom of the crowd' can speak through library interactions—for example, patterns relating search terms to the eventual item accessed. Such patterns may be more clearly evident at the scale offered by several universities combining their data, as already suggested in the case of enquiries. Furthermore, pure scale is not the only reason for accumulating data across libraries. There is also the benefit of variety, where lower occurrences of activity in the long tail can take on meaning.

Benchmarking

Universities in general, and libraries in particular, are culturally committed to benchmarking as a means of self-assessment and as a trigger for improvement.

Our conversations affirmed that senior managers want to make evidence-based policy, investment, and service decisions, to which end benchmarking is an important tool.

It would seem self-evident that new approaches to activity data and analytics should play strongly into benchmarking, whether or not it is explicitly a shared activity. However, Martin questions whether established library benchmarking (for example, based on the SCONUL returns in the UK) is best geared to the types of data now available and whether alternative or supplementary statistics might be introduced to do a better job.

To some extent, services such as JUSP are offering fresh approaches, though they are still focused on the same types of data. Furthermore, any shift in sector benchmarking focus would be likely to generate significantly positive ripples in terms of professional development in analytics.

Chris and Robert strongly concur that professional development is crucial for the potential of analytics to be realized. The need is not so much for technical skills as it is for building an understanding of the questions to ask of such data and recognition of the narratives they might reveal. Given a guiding hand from such entities as Jisc, RLUK and SCONUL, significant professional development benefits may be derived from collaboratively exploring new analytics opportunities in areas such as enquiry data and digital collections.

INNOVATIVE RESPONSE



We spoke to Eve-Marie Miller (Decision Center Product Manager) and Brad Jung (Vice President, Product Management) about Innovative's vision for the potential of analytics and associated shared services in academic libraries.

Librarians and vendors alike are experiencing a step change in expectations for the role of analytics in library management and customer-facing services. This shift is evident in three respects:

- The need to incorporate e-resource usage into collection management reporting;
- The opportunity to develop customer-facing services such as recommendations that draw on the raw data collected by libraries; and
- The potential for benchmarking and leveraging mutually valuable intelligence by **sharing analytics data** across collaborating libraries.

These developments resonate strongly with Innovative Decision Center product manager Eve-Marie Miller, who comes from a background in shared services at the California State Chancellor's Office and in e-resources as Director of Collection Development with EBSCO.

Collection Management

Since joining Innovative in this exciting time for analytics, Eve-Marie has been taking stock of the Decision Center product road map. She is confident that Innovative shares

It is evident from our public libraries that customers increasingly welcome personalized services, including recommendations, potentially applying more widely than books, to include databases, open access resources, and even services and events.

— Eve-Marie Miller, Decision Center Product Manager

⁵ <http://www.davevp.com/blog/archives/528>

with libraries an in-depth understanding of how transaction data can be used to inform collection development and space management, increasingly in near real-time. For Eve-Marie, the direction of travel for library reporting and analytics products is critical, recognizing that 'having established a sound analytics platform in Decision Center, I agree that the next challenge is to assist libraries in bringing together usage and trend information regarding e-resources of all types.'

Customer-Facing Services

Any service requiring storage of customer activity data for a longer term than is operationally necessary represents a major issue of principle for the library community. Nevertheless, Eve-Marie suggests that 'it is evident from our public libraries that customers increasingly welcome personalized services, including recommendations, potentially applying more widely than books, to include databases, open access resources, and even services and events.'

She observes that Innovative has significant experience and insights from the public library world to be applied appropriately in education and vice-versa. A strong example is the appropriate use of predictive analytics to make recommendations both to staff and directly to customers.

Accumulating Bigger Data

'Historic reporting approaches geared to "back of house" collection management represent only part of the emerging analytics opportunity', reflects Eve Marie, 'so we can be sure we're not holding ourselves back by limiting the scope of the Decision Center data set.' Decision Center currently covers an extended period (Circulation data for the past 3 years, Acquisitions for 5 years). However, notes Product Management VP Brad Jung, 'emerging use cases such as recommendations, suggest that we may need to revisit these boundaries, which is something that our analytics engine and storage solution allows us to do.'

Similarly the user level activity data available in Sierra is not used in Decision Center at present. Eve-Marie emphasizes that 'we are definitely thinking of positive approaches that also take care of how privacy issues, not least in cases where libraries want to combine such data into a bigger shared picture.'

Sharing Analytics

Eve-Marie recognizes that the scale offered by shared analytics would very likely enhance a recommendations approach—for example, highlighting authors and titles further in to the long tail. Furthermore, for librarians, collaborative collection development and last copy services should benefit from collection comparisons that include analytics as well as holdings.

Throughout our product strategy, we need to be enabling our library partners to interact with whomever they wish, and that includes other libraries that are not Innovative customers.

— Brad Jung, VP Product Management

Brad takes this thinking a step further in emphasizing, 'Throughout our product strategy, we need to be enabling our library partners to interact with whomever they wish, and that includes other libraries that are not Innovative customers. We need to make it easy for them to play well and to exchange data in the wider ecosystem, of which shared analytics services are a good case in point.'

Brad concludes, 'This strategy is easily said and harder to deliver in practice, but it is our committed direction of travel. Looking, for example, at the range of shared service opportunities available to UK academic libraries, it is clear that systems suppliers cannot expect a library to be working in a single consortium for all things. They might share analytics data in completely different partnerships than those with which they work on collective knowledge bases. Therefore, everything need to play in to the wider data and services ecosystem.'

RECOMMENDED READING

The following resources offer an opportunity to find out more about analytics in the areas discussed here:

- A management introduction: "*Activity Data – Delivery Benefits from the Data Deluge*"

<http://www.jisc.ac.uk/publications/reports/2012/activity-data-delivering-benefits.aspx>

- A guide to techniques and projects: "*Exploiting Activity Data in the Academic Environment*"

<http://www.activitydata.org>

- CETIS Analytics Series, including papers on *Analytics and the Whole Institution*, *Legal & Ethical*, *Teaching & Learning*, *Research*

<http://publications.cetis.ac.uk/c/analytics>

- The EDUCAUSE library 'Analytics' section

<http://www.educause.edu/library/analytics>

You can find out more about Innovative Decision Center at: <http://www.iii.com/products/decisioncenter.shtml>

For more information:

Web: www.iii.com

Email: info@iii.com

Call: 800-878-6600

 [facebook.com/innovativeinterfaces](https://www.facebook.com/innovativeinterfaces)

 twitter.com/iii_Innovative