Discovering the Pattern, Discerning the Potential

The role of the library in unraveling the cat’s cradle of activity data
Objectives

• To consider data strategies that can enable the library to reap maximum value from its data, both for its patrons and for the institution.

• To identify potential data sources within the library, elsewhere in the institution and beyond

• To consider how it could be connected together and at what level it should be stored and retained.

Health warnings

• I am a data analyst not a qualified library assessment practitioner

• My insights are predominantly from working with UK Higher Education libraries, starting with the first Jisc foray into ‘activity data’ in 2009 in the TILE project

• I am naturally enthusiastic, which is dangerous
With a little help from my friends

Principally thanks to Jisc for funding projects in this space since 2009. Whilst many UK universities have been partners in the journey led by Jisc, particular mention should go to

- University of Manchester (2011 Activity Data project lead),
- University of Westminster (2013 LMS Change project lead)
- University of Huddersfield, always UK library analytics trailblazers

Finally, a North American perspective has hopefully been assimilated thanks to inputs from

- Ted Dodds at Cornell University
- David Gift at Michigan State University
- Joe Zucca at the University of Pennsylvania
- Apereo and Kuali Foundation partners
The role of the library in unraveling the cat’s cradle of activity data

1 - Introduction

2 - New world disorder
   Out there
   Inside Higher Education
   Within the Library

3 - Core Questions
   Question 1 - What data should we collect? 
   Question 2 - Where might that data come from? 
   Question 3 - How will we store it? 
   Question 4 - Why are we doing this? 
   Question 5 - When should we start? 

4 - Concluding Remarks
Data is king
Data analysis in the second decade of the 21st century has been overshadowed by challenges of big data

Global response

- 2010 - *The Economist* published its ‘Special report on managing information: Data, data Everywhere’
- 2011 - McKinsey Global Institute published the landmark report ‘Big Data: The next frontier for innovation, competition and productivity’

Education response

- 2012 – The Department of Education focused the data debate on the learner experience in its report on ‘Enhancing Teaching and Learning Through Educational Data Mining and Learning Analytics’
- 2012 - Analytics featured in the ‘Top 10 Issues’ highlighted annually for EDUCAUSE by IT leaders in US colleges and higher education
Sanity check

• Is this just another IT fad, driven by highly capitalized adventurers and global purveyors of online shopping and social media?
• Are the dark arts of analytics and the exploitation of personal data best left to the commercial world and to the marketing arms of our institutions?
• Or is there something here that uniquely addresses real operational and strategic issues in the increasingly performance-driven and customer-focused business of higher education?

So what’s new?

Attitude
• Impact of Amazon and other retailers
• Consumer attitudes linked to loyalty cards and social networking
• Visibility of large scale transactional, personal and environmental data at play in the operation of almost any consumer service

Technology
• The impact of connectivity, declining hardware costs and the cloud
• Developments in underlying math and software tools

Necessity
• Austerity has been a catalyst, but the opportunities presented by analytics are equally of interest ‘come rain come shine’
Inside Higher Education

2012 - Louis Soares (Center for American Progress) emphasized the tie up between analytics and the institutional imperatives of

- Cost
- Quality
- Knowing the customer
- Consumer agency

Breakthrough areas
At corporate level, digital marketing has offered most significant breakthroughs. Relating to library activity, student success and research metrics have seized the moment.

- Learning Analytics - The heightened debate has been particularly evident in the pursuit of learning analytics, with its ultimate focus on student retention and success and how that might be supported in near real time through the monitoring and understanding of vital signs, what Purdue has called ‘Course Signals’.
- Research Altmetrics - The research community and its corporate manifestations are also alive to the potential of how wider and deeper data accumulation can fuel researcher success and researcher reputation and measures thereof.
Vision or Snake Oil?
The vision of the learning analytics and research metrics communities has been powerfully articulated without inhibition. For example, in 2010 ‘Altmetrics: A Manifesto’ proposed

“These new forms reflect and transmit scholarly impact:

- That dog-eared (but un-cited) article that used to live on a shelf now lives in – where we can see and count it.
- That hallway conversation about a recent finding has moved to blogs and social networks – now, we can listen in.
- The local genomics dataset has moved to an online repository – now, we can track it.

This diverse group of activities forms a composite trace of impact far richer than any available before.”
Within the Library

Thanks partly to the controlling forces of systems vendors, as well as concern for reason, reputation and rights, library analytics have remained

- Fuelled by library and library community data
- Focused on processes – such as collection management
- Constrained by application silos – such as ILS or the gate system

At least that is the impression conveyed by the typical ILS management dashboard, even in the acclaimed new generation library service platforms.

It is also the impression given of library analytics standing in the 2012 ECAR (the EDUCAUSE Center for Applied Research) survey of member institutions
Stepping up to the plate (ECAR findings)

Analytics in Higher Education
Benefits, Barriers, Progress, and Recommendations
The Library Ecosystem

The model developed by Jisc’s 2013 LMS Change project illustrates how the library platform and therefore library analytics need to be understood in the context of powerful and independently mediated classes of services.
Core challenges – data
So … there is an intensification of the intelligence potentially available to businesses and institutions arising from digitally mediated interactions.

However, the scale challenges our ability to accumulate (store), to analyze (process) and to assimilate (present) using traditional means.
Core challenges – audiences
Doug Chow’s diagram further illustrates the challenges of velocity and volume in terms of the response expected by different audiences:

Feedback Loops
Scale & Speed

Learner
Teacher
Manager
Policymaker

Faster

Larger

©Doug Clow http://dougclow.wordpress.com
Core Questions
The principal concern of this paper is achieve clarity on some underlying issues

1. What?
2. Where?
3. How?
4. Why?
5. When?
Core Questions

The principal concern of this paper is achieve clarity on some underlying issues

1. What data should we collect?
2. Where might that data come from?
3. How should we store it?
4. Why are we doing it?
5. When should we start?
Question 1 – What data should we collect?

We have two conflicting options
- To collect whatever data we can and let it tell its story
- To collect data specific to areas of interest or known to be useful

A third option is to let our transaction systems determine our capability

Let’s collect activity data broadly and explore what it has to tell us

Who knows what we might find?

Not all activity data represent sound indicators so we need to do some analysis

We could waste a lot of time working with insignificant data

I want to use the available activity data to improve retention and achievement

This can work for our students right now

Strategic Choices
Reconciling the Options
Conversations with data

If we choose the more expansive approach we should note the underlying and escalating challenge of scale – the Analysis Gap.

However there is a compelling argument in favor of this approach if analytics are to contribute to our service design taking account of known audiences and emerging trends.

Clearly there is a judgment call about which events are significant – as the celebrated historian E.H. Carr argued, not every cake burning represents history in any useful sense.

Tony Hirst of the Open University has been a persistent advocate for the belief that data has stories to tell. His blog post on a conversation with the data in Isle of Wight parking meter transaction logs exemplifies his approach. 
A Conversation With Data - Car Parking Meter Data

This data conversation uses the car parking meter data obtained via an FOI request to the Isle of Wight Council for Pay and Display ticket machine transaction data from the ticket machines in the River Road Car Park, Yarmouth, Isle of Wight, for the financial year 2012-13.

The data includes an identifier for the ticket machine that issued the ticket, the time the ticket was issued, the tariff band (i.e. the nominal ticket value), and the amount of cash paid for the ticket.

The following conversation is a predominantly visual one, where questions are asked of the data and responses provided in a graphical form - as charts - that then need interpreting.

Several lines of questioning naturally arise:

- when are car parks actually used, based on ticket purchases?
- are different ticket types purchased at different times of day or different days of the week?
- do customers ever pay more than they need to when purchasing a ticket?

A note on the format of this data conversation

This conversation with data has been created within an interactive IPython Notebook, using the pandas data wrangling library and the ggplot2 graphics library.

For more information, contact tony.hirst@gmail.com
Question 2 – Where might that data come from?

There are four broad patterns for sourcing and mixing data:
- Work within the library service platform (ILS)
- Work across the range of library systems and instruments
- The library wishes to combine with external data
- External units wish to combine with library data

The necessary data likely will come from a range of systems, including:
- Additional category information relating to patrons (perhaps from registry, course management, research or HR systems)
- Further transaction data enriching patterns of activity (such as assignment submission and grading).
Data weaving
The underlying challenge is how to weave that data together, a task deeply dependent on compatibility of identifiers (IDs for people and also for business objects such as organization units and courses).

- Ted Dodds, CIO at Cornell, in Jisc’s 2012 case study, stressed that an essential foundation of analytics is to identify systems and data elements of record, which requires authority and effort at a corporate level.
- Joe Zucca emphasized the importance of authoritative IDs as exemplified in Penn’s MetriDoc development, which can reach out to authoritative sources to transform digital object identifiers into citation elements or exchange user IDs for demographic markers.
Consequence – the Library as a key player

The work of the library at the University of Huddersfield on early warning indicators in the Impact Data Project illustrates the importance and value of synchronizing IDs, enabling atomic student level course and library activity data to be analyzed together.

Finding - If you do not use the library as an undergraduate, you are over 7.19 times more likely to drop out of your degree. Whilst there is no suggestion of direct cause and effect, there is a highly actionable ‘early warning’.

![Huddersfield Early Warning Indicators](image_url)
Qualifying the ‘where’ question
So, specifically, who has what data?

<table>
<thead>
<tr>
<th>Library systems</th>
<th>Course Management systems</th>
</tr>
</thead>
<tbody>
<tr>
<td>Corporate systems</td>
<td>Research systems</td>
</tr>
<tr>
<td>Above Campus services</td>
<td>Out there</td>
</tr>
</tbody>
</table>

For each data source, libraries should ask

- Whether they need it
- Whether it is the canonical source
- Whether they can access it
- Whether the data is of sufficient quality to be fit for purpose
- Whether the data is the right level
- Whether it contains IDs that enable integration with other sources
<table>
<thead>
<tr>
<th>Library systems</th>
<th>Course Management systems, including LMS &amp; MOOCs</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Circulation activity</td>
<td>• Teaching Units</td>
</tr>
<tr>
<td>• Subscribed e-resource activity – through such as COUNTER, EZproxy</td>
<td>• Recommended reading</td>
</tr>
<tr>
<td>• Institutional Repository access</td>
<td>• Submissions / Assessments</td>
</tr>
<tr>
<td>• Discovery tool logs</td>
<td>• Resource access</td>
</tr>
<tr>
<td>• Gate / access system logs</td>
<td>• Participation such as posts</td>
</tr>
<tr>
<td>• Customer Services</td>
<td>• Performance Monitoring / Early Warnings</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Corporate systems</th>
<th>Research systems</th>
</tr>
</thead>
<tbody>
<tr>
<td>• People</td>
<td>• Funder / Grant References</td>
</tr>
<tr>
<td>• Organization units</td>
<td>• Publications</td>
</tr>
<tr>
<td>• Courses</td>
<td>• Licensing</td>
</tr>
<tr>
<td>• Results / Awards</td>
<td>• Research data</td>
</tr>
<tr>
<td>• Expenditure</td>
<td>• Social altmetrics</td>
</tr>
<tr>
<td>• Website analytics</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Above Campus community services</th>
<th>Out there</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Knowledge Bases, including CUFTS, DOAB, DOAJ, GOKb</td>
<td>• Government</td>
</tr>
<tr>
<td>• ID services, including ORCID, ISNI</td>
<td>• Agencies and Regulators</td>
</tr>
<tr>
<td></td>
<td>• Social tagging and citation on the web, via such as Topsy</td>
</tr>
<tr>
<td></td>
<td>• Research and Reference Tools such as Mendeley</td>
</tr>
</tbody>
</table>
**Question 3 – How should we store it?**

The options range between two poles:
- Aggregate transaction data under agreed business headings
- Store core transaction events

Data owners will be familiar with the challenges illustrated here:

<table>
<thead>
<tr>
<th>Challenge</th>
<th>Aggregate Events</th>
<th>Store Core Events</th>
<th>Opinion</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 - Storage &amp; Processing</td>
<td>Small and predictably slow growing</td>
<td>Large and unpredictably fast growing</td>
<td>Does scale matter any more?</td>
</tr>
<tr>
<td>2 – Security</td>
<td>Personal data and IDs not required</td>
<td>Even records without IDs can be sometimes traced to individuals</td>
<td>Modern businesses mitigate such risk proactively</td>
</tr>
<tr>
<td>3 - Retention</td>
<td>No requirement to retain raw data beyond its operational lifespan</td>
<td>Requires a long term retention policy (e.g. 10-15 years or even without limit)</td>
<td>Ditto</td>
</tr>
<tr>
<td>4 - Combination with other datasets</td>
<td>Only remains possible if they are at the same or higher levels of aggregation</td>
<td>Maximizes the potential subject to compatible IDs at transaction level</td>
<td>This is a significant impediment</td>
</tr>
<tr>
<td>5 - Flexibility to address questions</td>
<td>Does what it says on the tin</td>
<td>Offers flexibility at the level of time stamp, the individual and the business object</td>
<td>Ditto</td>
</tr>
</tbody>
</table>
I threw it all away ...  
Consider the questions you can and cannot answer if you aggregate analytical datasets at different levels.

For example, imagine aggregation of activities or transactions on the basis highlighted here:

• Timestamp
  o Date Range – e.g. August 2014

• Transaction
  o Transaction Type – e.g. Loan

• Resource
  o Resource Type – Monograph

• Person
  o Person Characteristic – Year 1 Full Time Student
Counting the cost ...

Consider whether such aggregated data could support explorations such as:

- Resource utilization of particular student types on Art History courses
- Take up of resources purchased / subscribed by the Economics faculty across other disciplines
- Comparative patterns of resource utilization for students graduating with Distinction over the past 5 years
- Incidence of help desk enquiries at different times of day across the academic year

As Marcum and Schonfeld observe in their 2014 Ithaka S+R ‘Driving with Data’ briefing, the nature of COUNTER statistics exemplifies the shortcomings of aggregation:

“Activity counts, even when applied to collections, can no longer be accepted as telling the whole story. For example, usage data that are aggregated at the COUNTER-compliant level do not show breakdowns by traffic source, which is vital for understanding discovery patterns, or by user type, which would allow libraries to distinguish teaching from research uses.”
**Question 4 – Why are we doing it?**

Given clear options on **What** data we might collect, **Where** we might sources it and **How** we might store it (aggregate or not?), we’re in a better position to consider some specific motivations.

In 2013 the Jisc LAMP project interviewed collections and service managers to identify user stories – what libraries want to do with analytics data and why.

We gathered 96 stories and then took a vote in which 49 stories were headlined 11 voting libraries. Of these, 19 stories might be classified as ‘mission’ stories, indicating strategic purposes:

<table>
<thead>
<tr>
<th>Cat</th>
<th>Description</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coll’n AN</td>
<td>to inform collection management policy</td>
<td>17</td>
</tr>
<tr>
<td>Coll’n K</td>
<td>to inform dialogue with faculties</td>
<td>17</td>
</tr>
<tr>
<td>Coll’n M</td>
<td>to deliver better / best value for money</td>
<td>14</td>
</tr>
<tr>
<td>Mission D</td>
<td>to impact student measures of satisfaction, such as NSS</td>
<td>13</td>
</tr>
<tr>
<td>Coll’n L</td>
<td>to promote evidence led intelligent decision making</td>
<td>13</td>
</tr>
<tr>
<td>T&amp;L AA</td>
<td>to develop indicators regarding success / progression / retention</td>
<td>13</td>
</tr>
<tr>
<td>Mission C</td>
<td>to demonstrate value added to users</td>
<td>12</td>
</tr>
<tr>
<td>Service V</td>
<td>to target skills development (knowledge)</td>
<td>12</td>
</tr>
<tr>
<td>T&amp;L AB</td>
<td>to target and assess academic skills interventions</td>
<td>12</td>
</tr>
<tr>
<td>Mission A</td>
<td>to connect the library with the university mission</td>
<td>11</td>
</tr>
<tr>
<td>Mission F</td>
<td>to inform / justify library policy and decisions as evidence led</td>
<td>11</td>
</tr>
<tr>
<td>Mission G</td>
<td>to engage stakeholders in productive dialogue</td>
<td>11</td>
</tr>
<tr>
<td>Service Y</td>
<td>to understand access patterns through such as portals, discovery layer</td>
<td>11</td>
</tr>
<tr>
<td>Service Z</td>
<td>to validate service developments and implementations</td>
<td>10</td>
</tr>
<tr>
<td>T&amp;L AS</td>
<td>to deliver support and skills interventions at the point of need</td>
<td>9</td>
</tr>
<tr>
<td>T&amp;L AD</td>
<td>to position library teaching effort</td>
<td>9</td>
</tr>
<tr>
<td>Recom AE</td>
<td>to improve resource selection</td>
<td>9</td>
</tr>
<tr>
<td>Service X</td>
<td>to identify combinations of factors linked to dissatisfaction</td>
<td>8</td>
</tr>
<tr>
<td>T&amp;L AQ</td>
<td>to track the learner journeys</td>
<td>8</td>
</tr>
</tbody>
</table>
The remaining 30 might be classified as ‘activity’ stories, more detailed things that librarians want to do with analytics:

<table>
<thead>
<tr>
<th>Cat</th>
<th>Description</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data</td>
<td>Collect comprehensive and granular e-resource data</td>
<td>15</td>
</tr>
<tr>
<td>Data</td>
<td>Merge data from the student registry</td>
<td>14</td>
</tr>
<tr>
<td>Data</td>
<td>Map e-resource usage events to actual users</td>
<td>14</td>
</tr>
<tr>
<td>Coll’n</td>
<td>Identify how different faculties use the library and the collection</td>
<td>13</td>
</tr>
<tr>
<td>Coll’n</td>
<td>Identify least used resources / where usage is less/more than expected</td>
<td>13</td>
</tr>
<tr>
<td>Data</td>
<td>Produce compelling visualizations</td>
<td>12</td>
</tr>
<tr>
<td>Data</td>
<td>Provide a dashboard with key indicators by department</td>
<td>12</td>
</tr>
<tr>
<td>Service</td>
<td>Correlate feedback (e.g. NSS), enquiries and collection strength</td>
<td>12</td>
</tr>
<tr>
<td>T&amp;L</td>
<td>Correlate reading lists with actual usage</td>
<td>12</td>
</tr>
<tr>
<td>T&amp;L</td>
<td>Examine events by specific demographic / user groupings (e.g. Overseas)</td>
<td>12</td>
</tr>
<tr>
<td>Data</td>
<td>Merge data from multiple library systems</td>
<td>11</td>
</tr>
<tr>
<td>Data</td>
<td>Contribute to the institutional analytics effort</td>
<td>11</td>
</tr>
<tr>
<td>Coll’n</td>
<td>Understand patterns of e-resource use (like physical stock is understood)</td>
<td>11</td>
</tr>
<tr>
<td>Coll’n</td>
<td>Benchmark / improve our collection comparatively (UK and beyond)</td>
<td>11</td>
</tr>
<tr>
<td>Data</td>
<td>Merge data from IT services (e.g. PC use, Authentication)</td>
<td>10</td>
</tr>
<tr>
<td>Data</td>
<td>Access the atomic raw data whenever required</td>
<td>10</td>
</tr>
<tr>
<td>Data</td>
<td>Provide flexibility in slicing and dicing data</td>
<td>10</td>
</tr>
<tr>
<td>Coll’n</td>
<td>Inform decisions about relegation / relocation / weeding of stock</td>
<td>10</td>
</tr>
<tr>
<td>Service</td>
<td>Cohere the range of enquiry, support and survey data sources</td>
<td>10</td>
</tr>
<tr>
<td>Service</td>
<td>Analyse system use and interactions (e.g. Discovery, Website, Portal)</td>
<td>10</td>
</tr>
<tr>
<td>Service</td>
<td>Share enquiry data with other HEIs to derive more value from scale</td>
<td>10</td>
</tr>
<tr>
<td>Coll’n</td>
<td>Triangulate usage with cost and license terms</td>
<td>9</td>
</tr>
<tr>
<td>Coll’n</td>
<td>Identify value derived from investment by different faculties</td>
<td>9</td>
</tr>
<tr>
<td>Coll’n</td>
<td>Promote books to short loan / high demand / express collections</td>
<td>9</td>
</tr>
<tr>
<td>Coll’n</td>
<td>Inform decision making about library stock, space and storage</td>
<td>9</td>
</tr>
<tr>
<td>Recom</td>
<td>Promote use of content beyond the reading list</td>
<td>9</td>
</tr>
<tr>
<td>Recom</td>
<td>Surface high cost unused content for users</td>
<td>9</td>
</tr>
<tr>
<td>Data</td>
<td>Incorporate web analytics data (typically from Google)</td>
<td>8</td>
</tr>
<tr>
<td>Data</td>
<td>Review the SCONUL returns to fit the shared analytics vision</td>
<td>8</td>
</tr>
<tr>
<td>Recom</td>
<td>Link discovery, reading lists and recommendation</td>
<td>8</td>
</tr>
</tbody>
</table>
Story Categories
Each statement was categorised in terms of its broad intent:

- High level ‘mission’ statements
- Stories about the range of data required
- Managing the collection
- Improving services
- Enhancing student experience and success
- Providing recommender services

Mission Stories
The principal focus of the ‘mission’ stories was collection management (see Story AN) and its contribution to value (Story M), satisfaction (Story D) and impact (Story C).

There was also strong recognition of analytics as a tool in

- Supporting dialogue with faculty (Story K)
- Evidencing and positioning library business cases (A, F)
- Proactively enabling support activity such as patron skills development to be better designed and targeted (V, AB, AS, AD)
Activity Stories

The challenge for libraries and for systems providers is to identify what data is required to support these requirements and how it might feasibly be collected within and across systems.

- The focus on e-resources is the challenge in two of the top three activity stories (Story 38, 4, also 19) - especially linking e-resource activity to users just as we are accustomed to doing with print.
- Combining data from more than just one vendor system (2, 29, 32, 1).
- Responding to differentiations in terms of faculty (14, 9) and user demographics (33).
- Analytics relating to enquiry management and related service improvement is regarded as an important dimension (29, 48, 54)
- Whilst recognised as an opportunity (61, 62, 34), there is less emphasis on using analytics for recommendation
- Last but not least, presentation is a critical part of the challenge (8, 9)
<table>
<thead>
<tr>
<th>Analytics needs to inform</th>
<th>July 2014 Rank</th>
<th>Needs Item Data</th>
<th>Needs Person Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Student Success</td>
<td>1&lt;sup&gt;st&lt;/sup&gt;</td>
<td>Poss</td>
<td>Yes</td>
</tr>
<tr>
<td>Dialogue with faculty</td>
<td>2&lt;sup&gt;nd&lt;/sup&gt;</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>Service Improvement</td>
<td>3&lt;sup&gt;rd&lt;/sup&gt;</td>
<td>Poss</td>
<td>Poss</td>
</tr>
<tr>
<td>Benchmarking</td>
<td>4&lt;sup&gt;th&lt;/sup&gt;</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Institutional analytics</td>
<td>4&lt;sup&gt;th&lt;/sup&gt;</td>
<td></td>
<td>Poss</td>
</tr>
<tr>
<td>Investment management</td>
<td>6&lt;sup&gt;th&lt;/sup&gt;</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>Space &amp; stock management</td>
<td>6&lt;sup&gt;th&lt;/sup&gt;</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>Exploration of the unknown</td>
<td>6&lt;sup&gt;th&lt;/sup&gt;</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Personalised recommendation</td>
<td></td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>User skills programmes</td>
<td></td>
<td></td>
<td>Poss</td>
</tr>
<tr>
<td>Author impact</td>
<td></td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>Collection balance (E v. P)</td>
<td></td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>Business case development</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Decision making culture</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Question 5 – When should we start?

Management styles and organizational imperatives will suggest different responses:
• When we have decided on the library’s priorities / the strategic questions to be addressed / etc
• When the institution has established its overall analytics framework
• Just do it … now!

However, whatever your responses to our other core questions, there is a compelling argument that ‘now’ is the only tenable answer because

• Critical mass of evidence involves accumulation over multiple years
• Libraries are hamstrung by adopting non-compliant vendor systems for relatively long periods
• Not least, growing expectation in terms of user experience and business practice does not favor procrastination

The experience of the University of Huddersfield is instructive. The library accumulated around 3.9m circulation transactions over 15 years before identifying in 2009 what could be done with this wealth of data.
Concluding Remarks

Universities in general and libraries in particular have data to die for.

Much of what has been discussed is widely regarded as essential business and customer service practice.

However few academic libraries have the capacity (i.e. deployable skilled resources) to pursue all the opportunities outlined here.

So what might be the building blocks for elevating the library analytics mission, whereby the library might be a welcome influencer and institutional leader?
Building blocks

Policy
• Establish principles - so there is a clear code of practice and framework for development without legal impediment
• Activate collection - Data collection capabilities should be activated for existing systems
• Include in requirements - New systems implementations should include a requirement for accessible activity data logging

Projects
• Take an incremental approach - balance strategic development with achievable tactical projects
• Accumulate data – preserve for the long term rather than treating it as a passing project artifact
• Work on IDs – use IDs and data from systems of record, even if in addition to local library IDs

People
• Build skills in the library - Key skills need to be developed in new types of storage, analysis and visualization
• Build collaborative relationships - like-minded stakeholders (faculty, students union, institutional leaders) who will value actionable outcomes
Predictive analytics will be part of the future of Higher Ed ... and libraries need to be involved in the design of these models (Debra Gilchrist)