

Foreword – Rachel Bruce, Acting Director of e-Infrastructure, JISC

‘Survive or Thrive’ is an opportunity for experts across the higher education and public sectors to reflect on the changes brought about by the web and digital technologies and on how these apply to the sharing and re-use of content for research and learning. Our key question is how can we better meet the needs of users.

Significant changes in the way in which information is created, distributed and used have evolved over the past decade. These offer new ways to provide content, as shown by the Open Educational Resources movement through which content that would never have been shared before is made available on the open web. However, as reflected in the conference programme, there are not only new ways to provide content, but also emerging technologies that help to add context and meaning to resources and thereby enhance service provision; for example semantic analysis, use data and location information.

Another huge change is in the scale at which we can now work; the web allows us to work across organisations and in a global environment, which raises significant questions. When should content and the related services be delivered at a local, even a personal level, and when is operation at ‘web scale’ appropriate? What are the business and organisational models required to deliver at different levels? What underlying planning, policy and skills challenges need to be borne in mind?

We hope the conference will help to elucidate the issues and to demonstrate how a range of approaches can be used to add value to content and services, more fully exploiting assets that are produced and used as part of learning and research. For example, linked data and crowd sourcing have gained a lot of attention recently and they certainly offer new opportunities, but our discussion might focus on how these work alongside other technologies such as text mining or geo-tagging in a coherent service offer.

This discussion paper, prepared by David Kay and Paul Miller, identifies six interwoven themes, from open approaches, working at both scale and depth, to leveraging user data and real time information. Our conference discussion will be captured, adding shared understanding and identifying both opportunities and threats. As a result we will produce a position paper that will inform strategy and enable content service providers to take best advantage of technologies to meet the needs of end users.

Part 1 - Landscape

“We’re really not ready for the tsunami of digital content that’s coming down the line.”

Bill Thompson, BBC

Technological innovation often takes the form of overlapping and richly interconnected waves of change; some rapid and disruptive, others subtly gradual and yet no less transformative. In Higher Education, as elsewhere, promising innovations must be dissected and evaluated in the context of existing constraints and the requirements of learning, teaching and research. Operational realities, institutional objectives and informal community innovation frequently play an important part in complicating the neat visions put forward by proponents of the various technologies with which the sector is being enticed.

The technologies behind the Semantic Web, the fresh business models of Cloud Computing, and the community engagement demonstrated by the likes of Galaxy Zoo are as unlikely to single-handedly revolutionise Higher Education as the Virtual Learning Environment was a decade ago. By focusing too closely upon a single technology, we see the niche areas in which it has greatest impact, and miss the vast range of situations in which life continues almost unchanged. It is only by considering a *range* of technical innovations in the context of real-world usage that we begin to understand those with the greatest potential to effectively supplement the current set of complex interactions on campus, and with stakeholders elsewhere in the supply chain.

Actors and their Roles

In considering the interactions between institutions and content, it can be difficult to separate the diverse roles fulfilled by the various actors. A university, for example, *generates* content, *consumes* content, and may well also *contribute* and *channel* content. The desire of university-as-consumer to access content in an affordable manner must be juxtaposed with the desire of university-as-contributor to extract maximum value from its own contribution. A model developed by the JISC-funded TILE project¹ is useful in separating consideration of the different roles played by a single individual, organisation or institution. The model introduces the notion of *corporations*, *channels* and *clients*, and describes the interactions that may occur between each. More detail is provided in the project report, but these may be considered as;

- Corporations – the managers of assets
- Channels – the agents delivering those assets within a value added environment
- Clients - who not only consume but also contribute, engaging with multiple channels

“We need a shared value system around the goal of ‘affordable access to high

¹ <http://www.sero.co.uk/jisc-tile.html>

quality Higher Education’... and then we need to keep repeating it”

Daniel Greenstein, University of California

The current position in Higher Education

The UK Higher Education sector benefits from access to a veritable wealth of content, freely generated by various stakeholders within the sector and beyond, or procured on its behalf by the efforts of individual institutions and agencies such as JISC and the British Library.

In areas from learning resources to scholarly publishing, traditional models are being challenged by growing enthusiasm for more ‘open’ and rapid approaches to creation and distribution. The realities of shrinking budgets are aligning with a philosophical enthusiasm for greater access to the fruits of publicly funded teaching and research, driving growth in open access publications and triggering a reconsideration of the value in extensive duplication of effort, especially in the provision of introductory undergraduate course material.

Within libraries, for example, recent work by the JISC/Research Libraries UK (RLUK) Resource Discovery Taskforce² has grappled with issues such as transforming the revenue-generating creation and supply of bibliographic data. The activities of JISC’s new programme of work in this area aims to provide evidence of the manner in which any substantive change in approach ripples through the sector and the suppliers upon which institutions currently depend.

As might be expected of complex and long-lived institutions, the IT policies and infrastructure supporting the creation, delivery, use and preservation of digital content within universities are complicated, interconnected, and of widely varying utility in addressing the challenge of meeting emerging requirements. Managers face the difficult task of untangling technical, procedural and human processes in order to offload obligations that are less important than they once were, freeing resource to tackle new areas as they grow in strategic value. The very different demands of today’s service delivery and long-term preservation are often conflated, increasing cost, complexity and confusion. Apparently ‘obvious’ benefits and savings from alternative approaches such as Cloud Computing are often diluted by the difficulty of unpicking the heavily customised connections between campus systems.

“Bricks and mortar organisational practices can no longer sustain growth in Higher Education.”

Daniel Greenstein, University of California

In responding to the current economic conditions, institutional leaders may well find themselves with more latitude to respond strategically than might have been feasible in more plentiful times. Do we have the vision - and the will - to rethink the sector, its role, and its value, in order to emerge stronger than before?

² <http://rdtf.jiscinvolve.org/>

Across a range of activities, existing investment in information resources already delivers significant institutional value. By engaging strategically with emerging approaches from within the sector and beyond, there are opportunities to realise fuller value from those existing investments and to deliver real benefits to staff, students, and to the effectiveness of the whole system.

Trends evident elsewhere

"People will go around you to get your content... unless you help them."

Mike Ellis, Eduserv

In the past decade, the Web has come to dominate much of Higher Education's interaction with electronic data and information, from the interfaces of the library catalogue or Virtual Learning Environment (VLE) through to web-based interaction with distant scientific instruments. More recently, the combination of trends (AJAX, the falling costs of storage and bandwidth, commodity availability of compute resources, Open Source, increasingly ubiquitous network connectivity and a redefinition of contribution and collaboration - loosely labelled 'Web 2.0') has seen a dramatic transformation from a web of published documents *consumed* by the majority of users toward a web in which most are increasingly able to create, contribute and critique. The web experience has become richer and more engaging, and the technological and social barriers between creation and consumption have been diminished once more. Whilst mobile devices have entered consumer consciousness as an increasingly important means of both contribution and consumption, services and processes within Higher Education remain largely tied to an older paradigm reliant upon web browsers on desktops and laptops.

Continued growth in the quantity of data available, and a recognition that value may be profitably extracted from these data in the aggregate, have driven the growth of ever-larger and more far-reaching web companies. Google, Facebook, Microsoft and others increasingly operate at a level almost inconceivable only a few years ago, deploying highly automated computing solutions to process the explicit transactions of web users whilst also tracking and responding to their implicit behaviour. At web scale, the interactions of millions of users with millions of resources become map-able, meaningful, and actionable. Recent cases in which web scale companies have gone further than their users felt comfortable in making implicit connections both explicit and visible have brought privacy concerns to the fore, and aided a growing realisation as to just how powerful and far-reaching these web properties have become.

In the race to realise the full potential of the data they are gathering, web-scale companies are increasingly drawn to ideas previously associated with research into areas such as Artificial Intelligence and the Semantic Web. Microsoft and Apple have been amongst those to acquire companies that explicitly embraced Semantic Web approaches, whilst Facebook, Google and others are pursuing slightly different technological approaches to realising the

same underlying goal of describing and acting upon the implicit structure within and between resources on the Web. Everywhere we turn online, the paths we follow, the decisions we make, and the ways in which we interact become tracked and surfaced in increasingly structured and connected fashion. Semantic Web or not, the web is becoming increasingly semantic.

Emergent Themes

Through observing trends within Higher Education and beyond, it is possible to step back from the specific technologies and approaches being deployed in order to identify broader - and more sustained - themes. In the context of this event, six themes present themselves as being particularly germane. These are introduced individually below, but might briefly be characterised as follows:

- Theme 1 - The increasing adoption of 'open' approaches to the creation, sharing and consumption of diverse resources;

"There is an inevitable trend towards sharing, connecting, reuse, and the avoidance of duplicated effort."

Tom Heath, Talis

- Theme 2 - the growing importance of 'web scale' capabilities, transcending institutional (and even national) ability to provision infrastructure;
- Theme 3 - the emergence of viable means to extract value and meaning from images, audio, time-based media, and large *corpora* of unstructured text;
- Theme 4 - the explosion of 'user-generated content' online, and the recognition that crowdsourced approaches have some utility in formal research;

"Even with unlimited CPU, computers just aren't good enough ... People are good at finding the unusual, whilst computers just find what you tell them to."

Arfon Smith, Galaxy Zoo

- Theme 5 - the realisation that Amazon's profitable mining of data on user behaviours applies to a broad range of contexts, far beyond the matching of buyers to books;
- Theme 6 - a perceptible acceleration in the movement of information online, and an associated race to understand - and tap - the potential of 'real-time' data flows.

The technologies, philosophical approaches and specific projects outlined by speakers can all be considered in the light of one or more of these six themes, which should offer a means by which to evaluate and compare the arguments being made, the pros and cons of individual technologies and practices.

A path to 'intelligent content'

"If you rely solely upon the market, you will get a solution shaped by the needs of the market."

Bill Thompson, BBC

As individuals spend more of their time online *by choice*, leveraging tools inside and outside the academic sector, it becomes evermore important to recognise Higher Education's place within a broader landscape of approaches, norms, standards and systems.

There is value in stepping back from the detail of running individual services, to consider the wider implications of meeting the sector's mission in cost-effective and collaborative fashion. By designing solutions that conflate local provision of *access* with local responsibility for *preservation*, are we best serving either need? Are there alternatives that are both better *and* cheaper, and can we learn lessons from the commercial world and elsewhere without sacrificing the core values that make Higher Education special and scholarship possible?

Innovation is proceeding at a remarkable rate on the Web, and the scale at which many services now operate brings new challenges along with tremendous new opportunities. The rush to move *everything* to Web scale in order to reap similar presumed benefits should be guarded against, though. Our model of 'Core Challenges' (Model 3 - below) suggests that there are many circumstances in which it remains more effective to operate on a smaller scale.

In all of this, those responsible for guiding the sector through a period of rapid innovation and declining budgets will require open minds and wise judgement, as well as ready access to the skills and knowledge emerging from practices taking root all around.

Part 2 – The Conference Themes

As is the nature of real services and applied technologies, especially in a networked world, our six conference themes are themselves interwoven. Each theme is defined and discussed in more detail on the subsequent pages.

	Opening the fortress	Working at Web Scale	Digging Deep	Embracing UG Content	Harnessing Implicit Intentions	Approaching Real-Time
Opening the fortress	Strong Connection	Possible Connection		Strong Connection	Possible Connection	
Working at Web Scale	Possible Connection	Strong Connection	Possible Connection	Possible Connection	Strong Connection	Strong Connection
Digging Deep		Possible Connection	Strong Connection	Possible Connection	Possible Connection	
Embracing User Generated Content	Strong Connection	Possible Connection	Possible Connection	Strong Connection	Strong Connection	Strong Connection
Harnessing Implicit Intentions	Possible Connection	Strong Connection	Possible Connection	Strong Connection	Strong Connection	Possible Connection
Approaching Real-Time		Strong Connection		Strong Connection	Possible Connection	Strong Connection

Strong Connection | Possible Connection

The current and emerging application domains identified under each theme also form an interesting pattern:

Conference Theme	Application Domain	Research	Learning & Teaching	Libraries	Archives	Museums	Government	Consumer	Enterprise	Media
Opening the fortress		Strong Connection	Possible Connection	Possible Connection	Possible Connection		Strong Connection	Strong Connection		Possible Connection
Working at Web Scale		Possible Connection		Possible Connection				Strong Connection	Possible Connection	Possible Connection
Digging Deep		Strong Connection			Possible Connection	Possible Connection	Strong Connection		Possible Connection	Possible Connection
Embracing User Generated Content		Strong Connection	Strong Connection	Possible Connection	Possible Connection	Possible Connection	Possible Connection	Strong Connection		Strong Connection
Harnessing Implicit Intentions		Possible Connection	Possible Connection	Possible Connection			Possible Connection	Strong Connection		Strong Connection
Approaching Real-Time		Possible Connection						Strong Connection		Possible Connection

Current domain | Emerging domain

Theme 1 - Opening the fortress

Early interest in Open Access to the published output of scholarship has been subsumed within broader social, economic, political and technological drives toward transparency and unencumbered re-use of content (and raw data) in a growing range of application areas. Technology, licensing frameworks, and business models are becoming increasingly capable, although work remains to be done in fully integrating today's enthusiasm for openness with legacy business practices reliant upon perpetuating artificial scarcities.

Headline-grabbing initiatives to make Government data freely available have served to raise awareness of the potential for Open Data, whilst simultaneously providing a valuable pool of content relevant to teaching and research across many areas of Higher Education. The implications for data produced inside Higher Education with public funds are still to be worked through.

Unease around continued abuse of personal data in the consumer space *may* trigger an overly restrictive legislative response that ultimately overturns many of the promising advances in this area.

Current & emerging application domains

Research	L&T	Libraries	Archives	Museums	Government	Consumer	Enterprise	Media
----------	-----	-----------	----------	---------	------------	----------	------------	-------

Rich Potential

This will be of interest in settings involving;

- Engagement with Government services
- Reuse of third party material
- Protection of IPR/ revenue generation
- Scholarly publishing

Operational Examples

e-Government	data.gov, data.gov.uk
Community	Wikipedia, Galaxy Zoo, Musicbrainz
Large scale sharing	Connected Commons, Freebase, Infochimps, Amazon Public Datasets, Sunlight Foundation, Open Library

Research & Development Projects

JISC MOSAIC	Exploring utility of sharing activity data from institutional systems to support recommendation-style services.
The Depot	EDINA's deposit service for researchers without an institutional repository in which to deposit their papers.
JISC Expose	A new programme, beginning shortly, concerned with making content more widely available for use and reuse.

Connected Themes

Opening the Fortress	Working at Web Scale	Digging Deep	Embracing User Generated Content	Harnessing Implicit Intentions	Approaching Real-Time
----------------------	----------------------	--------------	----------------------------------	--------------------------------	-----------------------

Theme 2 – Working at Web Scale

Web companies such as Amazon, Google and Facebook have begun to illustrate the implications of working at scale, monetizing effectively free online offerings by mining the aggregate of millions of transactions in order to derive new value.

New models, based upon open source software, massively redundant commodity hardware and highly automated workflows alter the underlying economics and give rise to a new class of services that would previously have proved unsustainable.

The opportunity has been clearly demonstrated in the data centres of a new breed of business such as Google. Expensive, under-utilised hardware has been replaced by massively redundant clusters of commodity computing and a shift toward developing highly fault-tolerant applications.

Higher Education benefits with cost-effective access to innovative and rapidly evolving ‘consumer’ software that challenges the sector’s bespoke developments. In the area of research, the technological advances required to enable Google et al to manipulate massive data sets (Big Table, Map/Reduce, Hadoop, Cassandra, etc) are increasingly available to third parties, bringing Big Data analysis increasingly within reach of even the most modest research budget.

Current & emerging application domains

Research	L&T	Libraries	Archives	Museums	Government	Consumer	Enterprise	Media
----------	-----	-----------	----------	---------	------------	----------	------------	-------

Rich Potential

This will be of interest in settings involving;

- Massive data sets
- Distributed data sets
- Distributed processing
- Interaction with large audiences

Real World Examples

Web search	Google, Bing, Yahoo!
Commerce	Amazon, eBay
Infrastructure	Amazon Simple Storage Service (S3), Amazon Elastic Compute Cloud (EC2), Microsoft Azure, Google AppEngine, Eucalyptus
Libraries	OCLC WorldCat, e-Bird at Cornell

Research & Development Projects

Hadoop	Apache’s Hadoop project develops open-source software for reliable, scalable, distributed computing.
--------	--

Connected Themes

Opening the Fortress	Working at Web Scale	Digging Deep	Embracing User-Generated Content	Harnessing Implicit Intentions	Approaching Real-Time
----------------------	----------------------	--------------	----------------------------------	--------------------------------	-----------------------

Theme 3 – Digging Deep

Accurate and timely access to the fine grain of ‘atomic’ data, text, audio and images is highly desirable in a range of research disciplines, to government (notably in security) and potentially to media consumers. Advances in software and processing capabilities, linked to retrieval feedback mechanisms are transforming possibilities in to feasible realities.

Text Mining is about the discovery of new information, by automatic extraction from different resources and linking it together to form new facts or new hypotheses to be explored further by more conventional means of enquiry. IPR and licence issues in certain domains serve to complicate an otherwise relatively straightforward technological approach.

As the Web becomes more "visual," it is possible to go beyond traditional text and hyperlink analysis used in image collections to unlock the information stored in the image pixels, ideally using search queries and retrieval results to validate links and to create groupings (‘semantic clustering’). In the world of sound and image, further power lies in harnessing every characteristic of the medium to understand the content - analyzing the sound track using speech-to-text technology and the images on screen using video analytics.

Whilst the technologies for data, text and image mining have little in common, the issues for the academy, for curators and research groups, are very similar. All these approaches to atomic discovery and linkage differ from web search where the user is looking for something already known and identified by someone else. The challenge here is to discover unknown information, something that no one yet knows and so could not have yet written down. As this challenge shifts from invention to application, service issues come to the fore relating to the network level at which a service should be offered and the audience for whom such tools, services and the resultant possibilities of recall and re-use should be available.

Current & emerging application domains

Research	L&T	Libraries	Archives	Museums	Government	Consumer	Enterprise	Media
----------	-----	-----------	----------	---------	------------	----------	------------	-------

Rich Potential

This will be of interest in settings involving;

- Targets involving large scale with regular addition of sources
- Non-availability or unsuitable focus of man made metadata
- Value in cross referencing of atomic objects
- Thematic and enquiry based resource processing & reprocessing for new enquiries

Operational Examples

Moving image	Blinkx video search engine; digital phone applications for face recognition
Text	Clarabridge (online self service), ECHELON (classified security), PubGene (Biomedical)

Research & Development Projects

Images	Google Image Swirl; Microsoft iGroup
Text Mining	UK National Centre for Text Mining (NaCTeM)
Rights	Open Text Mining Interface and NIH's Common Journal Publishing DTD aim to enable machines to answer queries without violating publisher rights

Connected Themes

Opening the Fortress	Working at Web Scale	Digging Deep	Embracing User-Generated Content	Harnessing Implicit Intentions	Approaching Real-Time
----------------------	----------------------	--------------	----------------------------------	--------------------------------	-----------------------

Theme 4 – Embracing User Generated Content

The ability of the web to connect human beings on any topic has generated widespread recognition and expectation of web-mediated contribution, recommendation and collaboration. User contribution embraces a wide range of content generation devices – enhancement (co-creation, reuse), research data (citizen science), metadata (tagging), commentary (annotation, review), explicit recommendation (rating, lists) and discussion (network forums).

Domestic and academic consumers alike are interested in popularity, differentiated by the device of ‘users like me’ and by user recommendations mediated through such as lists, ratings, reviews and exchange forums. Network effect catalyses the interest inspired by user contributions – bringing volume (thus reliability and variety) and specialism (‘the long tail’).

Sourcing and working a crowd through the online medium can leverage wisdom (through aggregation of opinion at web scale) or resources (mobilising many hands and minds in such as citizen science). The notion of harnessing the crowd for distributed processing recognizes, ironically, that the human brain is better than a computer at tasks like pattern-recognition. The web simply yet crucially provides the globalised collaborative medium. The reach of the web also generates the dispersed ‘crowds’ valued in such as family history and ornithology.

Whilst these methods are highly compatible with research objectives and personalized learning, the academy also recognizes the dangers inherent in crowds – plagiarism, un-evidenced opinion and the diminution of originality and self-determination amongst undergraduates. Meanwhile curators may mistrust the implied challenge to the word - the purity of the canonical work, the authority of the finding aid or the official reading list.

Current & emerging application domains

Research	L&T	Libraries	Archives	Museums	Government	Consumer	Enterprise	Media
----------	-----	-----------	----------	---------	------------	----------	------------	-------

Rich Potential

This will be of interest in settings involving;

- Large scale human friendly data collection and processing challenges
- Natural communities of interest
- Genuine worth in simple contributions (such as ratings)
- Payback loops that will motivate contribution (especially early birds)
- Acceptance of open approaches to content and data

Real World Examples

Research	Galaxy Zoo invites public contributions to assist in classifying 60m galaxies
Teaching	Columbia University Digital Bridges project elicits student contribution; Oxford University First World War Poetry
Libraries	New York Public Library project opens up Voltaire’s Candide
Twitter	This all purpose crowd sourcing medium is aided by re-tweets & hash tags

Research & Development Projects

JISC DPIE	Three ‘Developing the Personalised Information Environment’ projects
JISCPress	JISCPress seeks to integrate granular reuse with user contribution

Connected Themes

Opening the Fortress	Working at Web Scale	Digging Deep	Embracing User-Generated Content	Harnessing Implicit Intentions	Approaching Real-Time
----------------------	----------------------	--------------	----------------------------------	--------------------------------	-----------------------

Theme 5 – Harnessing Implicit Intentions

Every day, our interactions with processes and organizations generate a wealth of data - Battelle's 'database of intentions', the so-called 'data exhaust' created as a byproduct of many online processes. This 'Business Intelligence' may be under utilized by some originators, but is potentially of high value to those with the interest and capacity to collect, synthesize and analyze at scale.

Amazon and a new generation of web retailers and monetized online services have built their businesses upon concentrating and mapping the data trails (both click streams and transactions) of explorers and consumers. Music fans can "scrobble" each song they listen to into the Last.fm database of listener behaviour. In return for their data (which consumers are coming to understand as an asset of mutual value) they get recommendations.

The JISC & SCONUL LMS Study (2008) emphasized the 'concentration' of users and data in web scale channels as the key to generating network effect and thereby serving learners and researchers in a Web 2.0 world. Such concentrated intelligence also provides critical mass for value added services, such as recommendations and other forms of User Generated Content.

For specialist users (such as 'the long tail' of researchers), activity data may have utility in proportion to the intelligence it carries about the users. For example, there is value in the specific intelligence uniquely available to the HE community about the users of its services, relating circulation and download transactions to the user context of areas, levels and places of stud – the DNA of the user 'academic affiliation' which can run through all activity records.

Current & emerging application domains

Research	L&T	Libraries	Archives	Museums	Government	Consumer	Enterprise	Media
----------	-----	-----------	----------	---------	------------	----------	------------	-------

Rich Potential

This will be of interest in settings involving;

- Product catalogue
- Both click streams and transactions
- Volume (establishing patterns) and breadth (characterizing the long tail) of activity
- Client demographic and / or affiliation data
- Significant client behaviours (e.g. course related in HE)

Real World Examples

Retail	Supermarkets (e.g. Dunhumby at Tesco), Loyalty cards (e.g. Nectar)
Media	Amazon.com, eMusic.com, Last.fm
Libraries	Huddersfield OPAC, University of California Digital Library, Ex Libris bX

Research & Development Projects

Mesur	'MEtrics from Scholarly Usage of Resources' has produced large-scale, longitudinal maps of the scholarly community
JISC MOSAIC	Investigated national aggregation of library and VLE resource transactions

Connected Themes

Opening the Fortress	Working at Web Scale	Digging Deep	Embracing User-Generated Content	Harnessing Implicit Intentions	Approaching Real-Time
----------------------	----------------------	--------------	----------------------------------	--------------------------------	-----------------------

Theme 6 – Approaching Real-Time

“The Real-Time Web is a paradigm based on pushing information to users as soon as it's available - instead of requiring that they or their software check a source periodically for updates. It's being implemented in social networking, search, news and elsewhere - making those experiences more like Instant Messaging and facilitating unpredictable innovations. Early benefits include increased user engagement ('flow') and decreased server loads, but these are early days. Real-time information delivery will likely become ubiquitous, a requirement for almost any website or service.” (Marshall Kirkpatrick, ReadWriteWeb)

Although most often associated with the rise of micro-blogging services such as Twitter, all of the major search engines are investing heavily in order to capture and act upon sizeable streams of data in something approaching real-time. Stock trading, sentiment analysis, traffic routing, weather services and more all benefit from systems that are capable of extracting semantics from a constant stream of information submitted by people, automated sensors, and mobile devices.

What, though, does this heavily hyped trend mean for Higher Education? Whilst scientific applications to rapidly process the outputs from sensors in the field may prove compelling, it is less clear that the sector at large really requires the ability to query, aggregate and compute at such speed.

Current & emerging application domains

Research	L&T	Libraries	Archives	Museums	Government	Consumer	Enterprise	Media
----------	-----	-----------	----------	---------	------------	----------	------------	-------

Rich Potential

This will be of interest in settings involving;

- Rapidly changing experimental conditions
- Large-scale engagement with the views/locations/attitudes of a sample population
- Personalisation on the basis of location and state
- Presence (adapting application behaviour to match device, location, context)

Real World Examples

Messaging	Twitter, Foursquare, Gowalla
Search	Twitter, Google, Bing

Research & Development Projects

PubSubHubbub	An extension to the Atom and RSS protocols, intended to support near-instant notifications.
--------------	---

Connected Themes

Opening the Fortress	Working at Web Scale	Digging Deep	Embracing User-Generated Content	Harnessing Implicit Intentions	Approaching Real-Time
----------------------	----------------------	--------------	----------------------------------	--------------------------------	-----------------------

Part 3 - Frameworks for Reflection

Consideration of our six theses suggests that the interface of technology, services and scholarly artefacts is a maelstrom, the eye of the perfect storm. No objective is fixed, no direction is linear – we are forever ‘beta’. Within this landscape, we are committed to leveraging scholarly assets to ensure the future of the academy and the health of the wider knowledge economy.

We therefore offer four frameworks or models designed to assist delegates in reflecting on the relative importance and inter-relationship of the themes and the potentially contributing technologies. They address

- Scholarly diversity – a Discipline Barometer
- Focus amidst chaos – the Sweet Spot
- Cutting to the key questions – the Core Challenges
- Deciding who should do what – an Ownership Matrix

The models are not mutually exclusive and can therefore be used as complementary lenses to focus and systematize the opportunities, the challenges and the timing in an HE context.

Model 1 – Discipline Barometer

Particular technology applications are likely to differ in fit according to discipline and level. From the perspective of each discipline and its practitioners (as opposed to the technologists), the initial broad criteria determining fit will be ‘relevance’ (does it assist scholarship?), ‘readiness’ (are practitioners ready to use it?) and ‘reward’ (will there be a return for the practitioner investment?) - though there will be others.

The following broad scholarly discipline areas, gateway criteria and levels offer an initial assessment framework that can be augmented for greater precision.

Discipline	Level	Criteria
Arts	Research	Relevance
Humanities	Postgraduate	Readiness
Social Sciences	Undergraduate	Reward
Sciences		
Medicine		

This model can also assist in identifying the requirements of research groups working in cross-discipline spaces, involving experts from diverse but potentially complementary disciplines, often in distributed settings. Those ‘marriages’ may drive specialised toolset requirements above and beyond the individual contributing disciplines.

Here we apply such criteria on a 5-point scale to the potential of image recognition and retrieval technology across three disciplines. The result suggests the potential value of such a model as part of a service innovation toolkit. Whilst there will always be individuals in any research or teaching setting who have the vision, expertise and motivation to deploy technologies against the grain, such a model provides the basis for normalizing such the dialogue.

	Technology	Application	Curation
1. Is it clearly targeted?			
2. Is there demand?			
3. Is the learning curve beneficial			
4. Is the timing right?			
5. Is it our mission?			
6. Is it good value?			
7. Can we give it sufficient attention?			

Here we apply a Red>Amber>Yellow>Green assessment to the case for local application of text indexing and mining across licensed electronic resources:

	Technology	Application	Curation
1. Is it clearly targeted?	Green	Yellow	Orange
2. Is there demand?	Yellow	Green	Orange
3. Is the learning curve beneficial	Yellow	Yellow	Yellow
4. Is the timing right?	Yellow	Yellow	Yellow
5. Is it our mission?	Red	Red	Red
6. Is it good value?	Orange	Orange	Red
7. Can we give it sufficient attention?	Yellow	Yellow	Red

The key consideration at this level of analysis, based on benchmarking techniques, is the variation across the horizontal and the vertical grading. By way of contrast, consider the very different challenge of exposing the library catalogue to external services and developers:

	Technology	Application	Curation
1. Is it clearly targeted?	Green	Green	Yellow
2. Is there demand?	Yellow	Yellow	Yellow
3. Is the learning curve beneficial	Green	Green	Green
4. Is the timing right?	Green	Green	Green
5. Is it our mission?	Green	Green	Green
6. Is it good value?	Green	Yellow	Yellow
7. Can we give it sufficient attention?	Yellow	Yellow	Yellow

Model 3 – The Core Challenges

Whenever we consider new models or new technologies for content and information management, and especially when assessing their likely fit and impact in a particular domain (e.g. a whole library, a particular collection or a research group), it is useful to have models that facilitate a first pass assessment, focusing on the core challenges ahead of piloting or implementing any given technology or service approach. We suggest those core challenges relate to scale, granularity, semantic consistency and results precision.

Scale	To be effective, does this need to be applied at Collection, Subject, System (e.g. institution) or Web-scale (e.g. National, Global) level?
Granularity	Does this apply to atomic objects (e.g. words, frames), logical / physical divisions (e.g. chapters, documents), complete Works (e.g. Books, Films) or Collections (such as archival)?
Semantic Consistency	Is descriptive consistency required; if so, within the local domain or cross-mapped between domains or is a canonical ontology implied?
Results Precision	Is the approach required to provide low or high precision results or must it provide exact recall?

These areas of enquiry represent a useful set of challenges in assessing the applicability and feasibility of any technology, which can be further specified in the key words in the following matrix (to be read horizontally).

SCALE	Collection	Subject	System	Web-scale
GRANULARITY	Atomic	Division	Work	Collection
CONSISTENCY	N/A	Local	Cross-mapped	Canonical
PRECISION	Low	Medium	High	Exact

A generic search engine requirement, for example, may be defined as follows:

SCALE	Collection	Subject	System	Web-scale
GRANULARITY	Atomic	Division	Work	Collection
CONSISTENCY	N/A	Local	Cross-mapped	Canonical
PRECISION	Low	Medium	High	Exact

Our objective in using this model is to identify the target level for the service or technology under investigation. Consider a national union catalogue:

SCALE	Collection	Subject	System	Web-scale
GRANULARITY	Atomic	Division	Work	Collection
CONSISTENCY	N/A	Local	Cross-mapped	Canonical
PRECISION	Low	Medium	High	Exact

A bioscience research group retrieval requirement may appear very differently:

SCALE	Collection	Subject	System	Web-scale
GRANULARITY	Atomic	Division	Work	Collection
CONSISTENCY	N/A	Local	Cross-mapped	Canonical
PRECISION	Low	Medium	High	Exact

These distinctions are important in clarifying the business case that might be made around deployment of a particular technology or service. In our final example, such a research group could justify applying image recognition or text mining technologies, linked to controlled vocabularies around a precise collection, an service unlikely to be justified around a national union catalogue.

Model 4 – An Ownership Matrix

Behind new service prospects lurk the realities of costs and risks, linked to changing views of process ownership, location and associated dependencies. The assets involved in scholarship demand both origination and curatorial roles, played by creators, libraries, archives, museums and data centres within and without the HE sector.

This model therefore addresses the question of ownership, of ‘Who does and should do what?’ - the ‘Leave it to Google’ (or ANO) challenge. It is designed to assist us in thinking about the roles and responsibilities, ownership and permissions that we relate to those assets – and not only the core data and intellectual works but also the associated metadata and more, which might collectively be described as the layers of the ‘content onion’.

Roles across the layers may differ according to

- Subject area - compare e-science with history
- Level - compare undergraduate studies with virtual research groups
- Curatorial domain – compare libraries and archives
- Business model – protecting IP or encouraging co-creation, perfected by local cataloguers or open to the inconsistencies of the crowd

This model therefore takes account of

- The assets that may be associated with a work or dataset (Vertical axis)
- The range of roles (i.e. responsibilities and permissions) that may be adopted for any asset, e.g. for metadata (Horizontal axis)

In the proposed model, the **assets** (the content onion) are layered as follows:

Recommend	Recommendations ranging from prescribed unit reading lists to user ratings and reviews
Activity Data	Click stream and transactional activity captured from search to loan or download
Vocabularies	Linguistic systems for linking, associating and navigating between objects – aka ontologies, thesauri, semantic networks
Expose	Making assets (e.g. metadata) accessible to search engines and to applications
Metadata	Descriptions ranging from bib records to archival finding aids to programme guides to encoded data embedded in media

Indexing	The computational means of efficient metadata, full text or digital object retrieval
Content	The scholarly works themselves, either physical entities (e.g. books) or rendered in digital form (e.g. e-books)
Datasets	Live and static datasets generated by and / or for use in research, especially in the sciences and social sciences

The **roles** are divided as follows:

Externally procured	We purchase the asset or acquire it through the scholarly network – e.g. a book or a catalogue record
Internally authored	Someone generates it within the institution – e.g. a PhD, a conference paper, a research dataset
Automatically Generated	It is generated from the knowledge or data within the IT system – e.g. Library transactions, Full text indexing
User Enhanced	We give users permission to enhance or to contribute these assets; users may be restricted to staff or to registered students
Externally shared	We share this asset with externally parties – under a license framework or without restriction

These parameters form a matrix, which can be used for collecting thoughts (putting text in the cells) or simply for assessing processes ('This is what we do') and envisioning future developments ('This is what we could do'). Our example is coded to represent the approach of an institution to its humanities collection – **We do it now**, **We are open to this**, **We are wary of that**.

	Externally Procured	Internally Authored	Automatically Generated	User Enhanced	Externally Shared
Recommend		Reading Lists		Ratings	
Activity Data					
Vocabularies					
Expose					
Metadata					
Indexing					
Content					
Datasets					