

# THE MASHEZINE

18 MAY, 2012 | THE MUSING OF MARTIN HAWKSEY (EDTECH EXPLORER)

## Using Google Spreadsheets to dashboard project/course blog feeds #oerri

May 17, 2012 04:12PM

In the original JISC OER Rapid Innovation call[1] one of the stipulations due to the size and durations of grants is that the main reporting process is blog-based. Amber Thomas, who is the JISC Programme Manager for this strand and a keen blogger herself[2], has been a long supporter of projects adopting open practices, blogging progress as they go. Brian Kelly (UKOLN) has also an interest in this area with a some posts including Beyond Blogging as an Open Practice, What About Associated Open Usage Data?[3]

For the OERRI projects the proposal discussed at the start-up meeting was that projects adopt a taxonomy of tags to indicate keys posts (e.g. project plan, aims, outputs, nutshell etc.). For the final report projects would then compile all posts with specific tags and submit as a ms-word or pdf.

There are a number of advantages of this approach one of them, for people like me anyway, is it exposes machine readable data that can be used in a number of ways. In this post I'll show I've create a quick dashboard in Google Spreadsheets which takes a list of blog RSS feeds and filters for specific tags/categories. Whilst demonstrated this with the OERRI projects the same technique could be used in other scenarios, such as, as a way to track student blogs. As part of this solution I'll highlight some of the issues/affordances of different blogging platforms and introduce some future work to combine post content using a template structure.

Project	Dissemintation	posts	(E)R	projectplan	aims	usecase	nutshell	tag/categories	outputs	outputlist	lessons/learn	impact	grandtotal
Attribute images	<a href="http://iccommons.org">http://iccommons.org</a>	2											
bebop	<a href="http://bebop.blog">http://bebop.blog</a>	6											
Breaking Down Barriers: Building a Geosocial	<a href="http://www.oerri.org">http://www.oerri.org</a>	6											
CAMLOE	<a href="http://www.oerri.org">http://www.oerri.org</a>	1											
Improving Accessibility to Mathematics	<a href="http://www.oerri.org">http://www.oerri.org</a>	9											
Linked data approaches to OERs	<a href="http://www.oerri.org">http://www.oerri.org</a>	15											
Portfolio Commons	<a href="http://portfoliocc.org">http://portfoliocc.org</a>	5											
RedFeather	<a href="http://blogs.acs.uq.edu.au">http://blogs.acs.uq.edu.au</a>	6											
RIDL	<a href="http://www.medi.uq.edu.au">http://www.medi.uq.edu.au</a>	2											
Sharing metadata across widget stores	<a href="http://www.oerri.org">http://www.oerri.org</a>	4											
SPINOLE	<a href="http://blogs.acs.uq.edu.au">http://blogs.acs.uq.edu.au</a>	3											
SupOEROver	<a href="http://www.medi.uq.edu.au">http://www.medi.uq.edu.au</a>	2											
synite mobile	<a href="http://www.oerri.org">http://www.oerri.org</a>	8											
TRACK OER	<a href="http://cloudworks.org">http://cloudworks.org</a>	1											
Xenith	<a href="http://iccommons.org">http://iccommons.org</a>	71											

[4]  
Screenshot of OERRI post dashboard[5]

### The OERRI Project Post Directory

If you are not interested in how this spreadsheet was made and just want to grab a copy to use with your own set of projects/class blogs then just:

\*\*\* open the OERRI Project Post Directory[6] \*\*\*

File > Make a copy if you want your own editable version

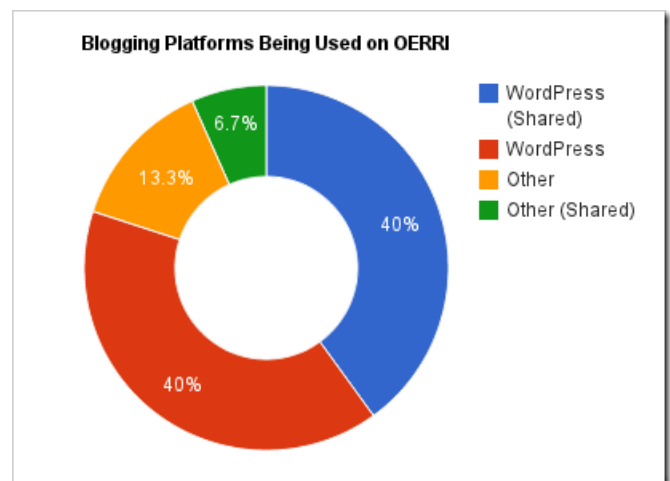
The link to the document above is the one I'll be developing

throughout the programme so feel free to bookmark the link to keep track of what the projects are doing.

The way the spreadsheet is structured is the tags/categories the script uses to filter posts is in cells D2:L2 and urls are constructed from the values in columns O-Q. The basic technique being used here is building urls that look for specific posts and returning links (made pretty with some conditional formatting).

### Blogging platforms used in OERRI

So how do we build a url to look for specific posts? With this technique it comes down to whether the blogging platform supports tag/category filtering so lets first look at the platforms being used in OERRI projects.



This chart (right) breaks down the blogging platforms. You'll see the most (12 of 15) are using WordPress in two flavours, 'shared', indicating that the blog is also a personal or team blog containing other posts not related to OERRI and 'dedicated', setup entirely for the project.

The 3 other platforms are 2 MEDEV blogs[7] and the OUs project on Cloudworks[8]. I'm not familiar with the MEDEV platform and only know a bit about cloudworks so for now I'm going to ignore these and concentrate on the WordPress blogs.

### WordPress and Tag/Category Filtering

One of the benefits of WordPress is you can can an RSS feed for almost everything by adding /feed/or ?feed=rss2 to urls (other platforms also support this, I a vague recollection of doing something similar in blogger(?)). For example, if you want a feed of all my Google Apps posts you can use [http://mashe.hawksey.info/category/google-apps/feed/\[9\]](http://mashe.hawksey.info/category/google-apps/feed/).

Even better is you can combine tags/categories with a '+' operator so if you want a feed of all my Google Apps posts that are also categorised with Twitter you can use [http://mashe.hawksey.info/category/google-apps+twitter/feed/\[10\]](http://mashe.hawksey.info/category/google-apps+twitter/feed/).

So to get the Bebop 'nutshell' categorised post as a RSS item we

can use: <http://bebop.blogs.lincoln.ac.uk/category/nutshell/feed/>  
[11]

Looking at one of the shared wordpress blogs to get the 'nutshell' from RedFeather you can use:  
<http://blogs.ecs.soton.ac.uk/oneshare/tag/redfeather+nutshell/feed/>[12]

### Using Google Spreadsheet importFeed formula to get a post url

The 'import' functions in Google Spreadsheet[13] must be my favourites and I know lots of social media professionals who use them to pull data into a spreadsheet and produce reports for clients from the data. With importFeed[14] we can go and see if a blog post under a certain category exists and then return something back, in this case the post link. For my first iteration of this spreadsheet I used the formula below:

```
if more than one
value comma
seperate

=JOIN(" ",importFeed($O3&IF(ISBLANK($P3),$Q3&"/"&$D52,$Q3&"/"&$P3&"+"&$D52)&"feed/","item url" FALSE))

base blog url   if shared tag/cat is blank then   return   No
url is (tag or category)/whatever tag you are looking for   post url   headers
else
url is (tag or category)/project tag/cat+whatever tag you are looking for
```

This works well but one of the drawback of importFeed is we can only have a maximum of 50 of them in one spreadsheet. With 15 projects and 9 tag/categories the maths doesn't add up.

To get around this I switched to Google Apps Script (macros for Google Spreadsheets I write a lot about[15]). This doesn't have an importFeed function built-in but I can do a UrlFetch[16] and Xml parse[17]. Here's the code[18] which does this (included in the template):

Note this code also uses the Cache Service[19] to improve performance and make sure I don't go over my UrlFetch quota [20].

We can call this function like other spreadsheet formula using '=fetchUrlFromRSS(aUrl)'.

### Trouble at the tagging mill

So we have a problem getting data from none WordPress blogs, which I'm quietly ignoring for now, the next problem is people not tagging/categorising posts correctly. For example, I can see Access to Math have 10 post including a 'nutshell'[21] but none of these are tagged. From a machine side there's not much I can do about this but at least from the dashboard I can spot something isn't right.

### Tags for a template

I'm sure once projects are politely reminded to tag posts they'll oblige. One incentive might be to say if posts are tagged correctly then the code above could be easily added to to not just pull post links but the full post text which could then be used to generate the projects final submission.

### Summary

So stay tuned to the **OERRI Project Post Directory**[22] spreadsheet to see if I can incorporate MEDEV and Cloudworks feeds, and also if I can create a template for final posts. Given Brian's post on usage data[23] mentioned at the beginning should I also be tracking post activity data on social networks or is that a false metric?

I'm sure there was something else but it has entirely slipped my mind ...

BTW here's the OPML file for the RSS feeds of the blogs that are live[24] (also visible here as a Google Reader bundle[25])

### QR Code

You can read this post online by scanning this barcode (or visiting <http://tinyurl.com/6sb7dtx>)



### Links

- [1] <http://tinyurl.com/86v5pkn>
- [2] <http://tinyurl.com/7aq5ugo>
- [3] <http://tinyurl.com/7n7ejmq>
- [4] <http://tinyurl.com/7b7apj7>
- [5] <http://tinyurl.com/7b7apj7>
- [6] <http://tinyurl.com/7b7apj7>
- [7] <http://tinyurl.com/8yc8nld>
- [8] <http://tinyurl.com/887kas5>
- [9] <http://tinyurl.com/7c2kfo6>
- [10] <http://tinyurl.com/7fgq9b2>
- [11] <http://tinyurl.com/7jj6cym>
- [12] <http://tinyurl.com/744tob7>
- [13] <http://tinyurl.com/7xnnoob>
- [14] <http://tinyurl.com/7nslsqh>
- [15] <http://tinyurl.com/84jc37y>
- [16] <http://tinyurl.com/88xhtqu>
- [17] <http://tinyurl.com/6v4rr9e>
- [18] <http://tinyurl.com/7yjxkta>
- [19] <http://tinyurl.com/72sf63p>
- [20] <http://tinyurl.com/8xoyehm>
- [21] <http://tinyurl.com/6nsbcx6>
- [22] <http://tinyurl.com/7b7apj7>
- [23] <http://tinyurl.com/7n7ejmq>
- [24] <http://tinyurl.com/78mpbyw>
- [25] <http://tinyurl.com/74dxrlu>

# Analytics Reconnoitre: Notes on Open Solutions in Big Data from #esym12

May 15, 2012 10:53AM

A couple of weeks ago it was Big Data Week, "a series of interconnected activities and conversations around the world across not only technology but also the commercial use case for Big Data".

**big data**<sup>[1][1][2][2]</sup> consists of data sets that grow so large and complex that they become awkward to work with using on-hand database management tools. Difficulties include capture, storage,<sup>[3][3]</sup> search, sharing, analytics,<sup>[4][4]</sup> and visualizing - BY Wikipedia[5]

In O'Reilly Radar there was a piece on Big data in Europe[6] which had Q&A from Big Data Week founder/organizer Stewart Townsend, and Carlos Somohano both of whom are big in Big Data.

Maybe I'm being naïve but I was surprised that there was no reference to what universities/research sector is doing with handling and analysing large data sets. For example at the Sanger Institute alone each of their DNA sequencers are generating 1 terabyte (1024 gigabytes) of data a day, storing over 17 petabytes (17 million gigabytes) which is doubling every year.

Those figures trip off my tongue because last week I was at the Eduserv Symposium 2012: Big Data, Big Deal?[7] which had many examples of how institutions are dealing with 'big data'. There were a couple of things I took away from this event like the prevalence of open source software as well as the number of vendors wrapping open source tools with their own systems to sell as service. Another clear message was a lack of data scientists who can turn raw data into information and knowledge.

As part of the Analytics Reconnoitre we are undertaking at JISC CETIS in this post I want to summarise some of the open source tools and 'as a service' offering in the Big Data scene.

[**Disclaimer:** I should say first I coming to this area cold. I'm not an information systems expert so what you'll see here is a very top-level view more often than not me joining the dots from things I've learned 5 minutes ago. So if you've spot anything I've got wrong or bits I'm missing let me know]

## Open source as a Service

### some of the aaS's

- CaaS - Cluster as a Service
- IaaS - Infrastructure as a Service
- SaaS - Software as a Service
- PaaS - Platform as a Service

I've already highlighted how the open source R statistical computing environment is being used as an analytics layer[8]. Open source is alive and well in other parts of the infrastructure. First up at the was Rob Anderson from Isilon Systems (division of EMC) talking about Big Data and implications for storage[9]. Rob did a great job introducing Big

Data and a couple of things I took away were the message that there is a real demand for talented 'data scientists' and getting organisations to think differently about data.

RT @mariekeguy[10]: #esym12[24][11] Rob Anderson - to sum up, big data issues & problems are mainly related to organisational issues/structures Get followers menu option

Once you've got the followers run Twitter > Combine follower sheets Method II

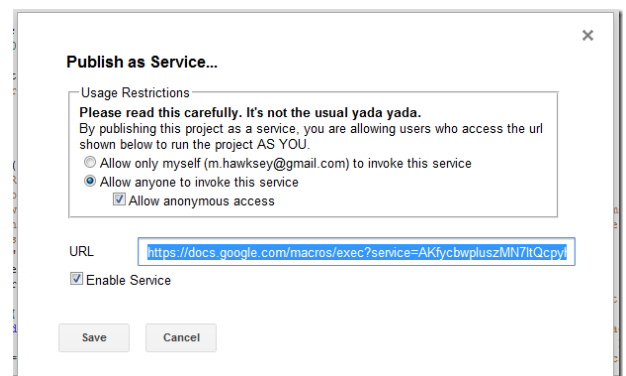
Move to the Vertices sheet and sort the data on the friends\_count column

In batches of around 250 rows select values from the id\_str column and run TAGS Advanced > Get friend IDs - this will start populating the friends\_ids column with data. For users with over 5,000 friends reselect their id\_str and rerun the menu option until the 'next\_cursor' equals 0

V	W	X	
created_at	protected	friend_ids	follow
Tue Oct 11 10:	TRUE		
Mon Nov 21 15:		{"next_cursor":0,"previous	
Tue Aug 17 11:		{"next_cursor":0,"previous	
Mon Jun 13 10:		{"next_cursor":0,"previous	
Fri Jun 03 15:0		{"next_cursor":0,"previous	
Tue Feb 14 14:		{"next_cursor":0,"previous	
Fri Jan 14 20:4	TRUE		

Next open the Script editor and open the TAGS4 file and then Run > setup.

Next select Publish > Publish as a service... and allow anyone to invoke the service anonymously. Copy the service URL and paste it into the R script downloaded earlier (also add the spreadsheet key to the R script and within your spreadsheet File > Publish to the web



Run the R script[21]! ... and fingers crossed everything works.

The files used in the SCOREProject/UKOER can be downloaded from here[22]. Changes you'll need to make are adding the output csv files to the data folder, changing references in js/gtable.js and js/wordcloud.js and the labels used in coffee/coffee.vis

So there you go. I've spent way too much of my own time on this and haven't really explained what is going on. Hopefully the

various commenting in the source code removes some of the magic (I might revisit the R code as in some ways I think it deserves a post on its own. If you have any questions or feedback leave them in the comments ;)

#### QR Code

You can read this post online by scanning this barcode (or visiting <http://tinyurl.com/c4rdqd7>)



#### Links

- [1] <http://tinyurl.com/blvwak9>
- [2] <http://tinyurl.com/2b85n5c>
- [3] <http://tinyurl.com/cb8kzhg>
- [4] <http://tinyurl.com/cxw34ns>
- [5] ERROR
- [6] <http://tinyurl.com/d88u3qe>
- [7] <http://tinyurl.com/d88u3qe>
- [8] <http://tinyurl.com/4xxex9o>
- [9] <http://tinyurl.com/cc3vkqu>
- [10] <http://tinyurl.com/79z93rp>
- [11] <http://tinyurl.com/cx3rqnb>
- [12] <http://tinyurl.com/cx3rqnb>
- [13] <http://tinyurl.com/dxza5j5>
- [14] <http://tinyurl.com/3tepy4h>
- [15] <http://tinyurl.com/6t4g2c7>
- [16] <http://tinyurl.com/728sggd>
- [17] <http://tinyurl.com/47m73oh>
- [18] <http://tinyurl.com/bqmljy2>
- [19] <http://tinyurl.com/cgrzwl8>
- [20] <http://tinyurl.com/864n2q5>
- [21] <http://tinyurl.com/bqmljy2>
- [22] <http://tinyurl.com/d9zdy5b>

## Analytics Reconnoitre: Notes on R in education and industry

May 9, 2012 06:46AM

As part of my role at JISC CETIS I've been asked to contribute to our 'Analytics Reconnoitre' which is a JISC commissioned project looking at the data and analytics landscape. One of my first tasks is to report on the broad landscape and trends in analytics service and data providers. Whilst I'm still putting this report together it's been interesting to note how one particular analytics tools, R, keeps pinging on my radar. I thought it would be useful to loosely join these together and share.

### Before R, the bigger 'data science' picture

Before I go into R there is some more scene setting required. As part of the Analytics Reconnoitre Adam Cooper (JISC CETIS) has already published *Analytics and Big Data - Reflections from the Teradata Universe Conference 2012*[1] and *Making Sense of "Analytics"*[2].

The Analytics and Big Data post is an excellent summary of the Teradata Universe event and Adam is also able to note some very useful thoughts on 'What this Means for Post-compulsory Education'. This includes identifying pathways for education to move forward with business intelligence and analytics. One of these I particularly liked was:

#### Experiment with being more analytical at craft-scale

Rather than thinking in terms of infrastructure or major initiatives, get some practical value with the infrastructure you have. Invest in someone with "data scientist" skills as master crafts-person and give them access to all data but don't neglect the value of developing apprentices and of developing wider appreciation of the capabilities and limitations of analytics.

[I'm biased towards this path because it encapsulates a lot of what I aspire to be. The craft model was one introduced to me by Joss Winn at this year's Dev8D[3] and coming for a family of craftsmen it makes me more comfortable to think I'm continuing the tradition in some way.]

Here are Adams observations and reflections on 'data science' from the same bog post:

"Data Scientist" is a term which seems to be capturing the imagination in the corporate big data and analytics community but which has not been much used in our community.

A facetious definition of data scientist is "a business analyst who lives in California". Stephen Brobst gave his distinctions between data scientist and business analyst in his talk. His characterisation of a business analyst is someone who: is interested in understanding the answers to a business question; uses BI tools with filters to generate reports. A data scientist, on the other hand, is someone who: wants to know what the question should be; embodies a combination of curiosity, data gathering skills, statistical and modelling expertise and strong communication skills. Brobst argues that the working environment for a data scientist should allow them to self-provision data, rather than having to rely on what is formally supported in the organisation, to enable them to be inquisitive and creative.

Michael Rappa from the Institute for Advanced Analytics doesn't mention curiosity but offers a similar conception of the skill-set for a data scientist in an interview in *Forbes* magazine[4]. The Guardian Data Blog has also reported on various views of what comprises a data scientist[5] in March 2012, following the Strata Conference.

While it can be a sign of hype for new terminology to be spawned, the distinctions being drawn by Brobst and others are appealing to me because they are putting space between mainstream practice of business analysis and some arguably more effective practices. As universities and

colleges move forward, we should be cautious of adopt the prevailing view from industry – the established business analyst role with a focus on reporting and descriptive statistics – and miss out on a set of more effective practices. Our lack of baked-in BI culture might actually be a benefit if it allows us to more quickly adopt the data scientist perspective alongside necessary management reporting. Furthermore, our IT environment is such that self-provisioning is more tractable.

## R in data science and in business

For those that don't know R is an open source statistical programming language[6]. If you want more background about the development of R the Information Age cover this in their piece Putting the R in analytics[7]. An important thing to note, which is covered in the story, is R was developed by two academics at University of Auckland and continues to have a very strong and active academic community supporting it. Whilst initially used as an academic tool the article highlights how it is being adopted by the business sector.

I originally picked up the Information Age post via the Revolutions blog (hosted by Revolution Analytics[8]) in the post Information Age: graduates driving industry adoption of R[9], which includes one of the following quotes from Information Age:

This popularity in academia means that R is being taught to statistics students, says Matthew Aldridge, co-founder of UK-based data analysis consultancy Mango Solutions. "We're seeing a lot of academic departments using R, versus SPSS which was what they always used to teach at university," he says. "That means a lot of students are coming out with R skills."

Finance and accounting advisory Deloitte, which uses R for various statistical analyses and to visualise data for presentations, has found this to be the case. "Many of the analytical hires coming out of school now have more experience with R than with SAS and SPSS, which was not the case years ago," says Michael Petrillo, a senior project lead at Deloitte's New York branch.

Revolutions have picked up other stories related to R in big data and analytics. Two I have bookmarked are Yes, you need more than just R for Big Data Analytics[10] in which Revolutions editor David Smith underlines that having tools like R aren't enough and a wider data science approach is needed because "*it combines the tool expertise with statistical expertise and the domain expertise required to understand the problem and the data applicable to it*".

Smith also reminds use that:

The R software is just one piece of software ecosystem — an analytics stack, if you will — of tools used to analyze Big Data. For one thing R isn't a data store in its own right: you also need a data layer where R can access structured and unstructured data for analysis. (For example, see how you can use R to extract data from Hadoop[11] in the slides from today's webinar by Antonio Piccolboni.) At the analytics layer, you need statistical algorithms that work with Big Data, like those in Revolution R Enterprise. And at the presentation layer, you need the ability to embed the results

of the analysis in reports, BI tools, or data apps.

[Revolutions also has a comprehensive list of R integrated throughout the enterprise analytics stack[12] which includes vendor integrations from IBM, Oracle, SAP and more]

The second post from Revolutions is R and Foursquare's recommendation engine[13] which is another graphic illustration of how R is being used in the business sector separately from vendor tools.

## Closing thoughts

At this point it's worth highlighting another of Adam's thoughts on directions for academia in Analytics and Big Data:

### Don't focus on IT infrastructure (or tools)

Avoid the temptation (and sales pitches) to focus on IT infrastructure as a means to get going with analytics. While good tools are necessary, they are not the right place to start.

I agree about not being blinkered by specific tools and as pointed out earlier R can only ever be just one piece of software in the ecosystem and any good data scientist will use the right tool for the job. It's interesting to see an academic tool being adopted, and arguable driving, part of the commercial sector. Will academia follow where they have led – if you see what I mean?

## QR Code

You can read this post online by scanning this barcode (or visiting <http://tinyurl.com/7xkaozn>)



## Links

- [1] <http://tinyurl.com/6tas7qs>
- [2] <http://tinyurl.com/6ptme6w>
- [3] <http://tinyurl.com/7r6eg7p>
- [4] <http://tinyurl.com/6o63xuy>
- [5] <http://tinyurl.com/7bp4sha>
- [6] <http://tinyurl.com/adcd>
- [7] <http://tinyurl.com/6pfhj4g>
- [8] <http://tinyurl.com/2av95et>
- [9] <http://tinyurl.com/6uk9qya>
- [10] <http://tinyurl.com/89kqsvc>
- [11] <http://tinyurl.com/7yyfj16>
- [12] <http://tinyurl.com/6tpx2tb>
- [13] <http://tinyurl.com/75av3dy>

# What I've starred this month: April 28, 2012

Apr 28, 2012 01:43PM

Here's some posts which have caught my attention this month:

- Study on the barriers that prevent researchers from publishing their code | Software Sustainability Institute[1] - April 27, 2012
- FOI Signals on Useful Open Data? « OUseful.Info, the blog... [2] - April 24, 2012
- MSDN Blogs[3] - April 24, 2012
- Big data in Europe - O'Reilly Radar[4] - April 23, 2012
- Exploring the Khan Academy's use of Learning Data and Learning Analytics | Emerging Education Technology[5] - April 23, 2012
- The Ed Techie: The role of respectable idiots[6] - April 20, 2012
- From Paywalls and Attention Walls to Data Disclosure Walls and Survey Walls « OUseful.Info, the blog...[7] - April 19, 2012
- John Cleese on Creativity | bavatuesdays[8] - April 13, 2012
- Evidence Framework for Innovation and Excellence in Education » Blog Archive » Big Data - Avalanche? Flood? Tsunami? What does big data mean for educators? » Evidence Framework for Innovation and Excellence in Education[9] - April 11, 2012
- Give the Chronicle Some Love | bavatuesdays[10] - April 5, 2012
- Grades 2.0: How Learning Analytics Are Changing The Teacher's Role | Edudemic[11] - April 3, 2012
- Guest Post: Being Openly Selfish and Making "OER" Work for You « UK Web Focus[12] - April 2, 2012
- Users, narcissism and control - tracking the impact of scholarly publications in the 21<sup>st</sup> century[13] - April 2, 2012
- University Funding - A Wider View « OUseful.Info, the blog...[14] - March 31, 2012
- Science of the Invisible: Putting the W in the DIK[15] - March 31, 2012
- Backchannel and notes for #GTAUK : how to get the best from G+? « DrBadgr[16] - March 31, 2012
- No MOOC Respect - CogDogBlog[17] - March 29, 2012

Automatically generated from my Diigo Starred Items[18].

## QR Code

You can read this post online by scanning this barcode (or visiting <http://tinyurl.com/83rokum>)



## Links

- [1] <http://tinyurl.com/cpqq887>
- [2] <http://tinyurl.com/7hw96kb>
- [3] <http://tinyurl.com/7ldfx77>
- [4] <http://tinyurl.com/6mmjylb>
- [5] <http://tinyurl.com/7nwb3m4>
- [6] <http://tinyurl.com/76rf2lo>
- [7] <http://tinyurl.com/753smbq>
- [8] <http://tinyurl.com/6rrb8z5>
- [9] <http://tinyurl.com/89rqhwf>
- [10] <http://tinyurl.com/7nz68jl>
- [11] <http://tinyurl.com/8xrnmd9>
- [12] <http://tinyurl.com/7w4493b>
- [13] <http://tinyurl.com/7smvfgf>
- [14] <http://tinyurl.com/7kwa54o>
- [15] <http://tinyurl.com/7e3vhwf>
- [16] <http://tinyurl.com/7mmtzwm>
- [17] <http://tinyurl.com/6lzezn2>
- [18] <http://tinyurl.com/7vokhtg>