## SoLAR Flare UK – Morning presentations

Liveblog notes from the SoLAR Flare UK, #flareUK, held on 19 November 2012 in the Jennie Lee Building, The Open University, sponsored by JISC.

Simon Buckingham Shum welcomes everyone, on behalf of his co-chairs, Rebecca Ferguson and Doug Clow, Sheila MacNeil from JISC CETIS.

Prof Josie Taylor, Director of the Institute of Educational Technology, welcomes everyone to the building and IET. We think learning analytics is going to be very important, for a wide range of people. Wishes everyone a happy, stimulating and argumentative (in a good way!) day.

## Simon Buckingham Shum, SoLAR

This is the first national gathering of people interested in learning analytics, hopefully the first of many. Self-organising, rapid dissemination. Not much lecturing, lots of networking opportunities. This is a rapidly exploding area. In Stanford, with Roy Pea and MOOC people – this is massive there. TIME magazine cover story was MOOCs. EdX – is about big data to allow us to ask big questions about learning. The data scientist is the sexist job for the C21st. Big business intelligence companies see an exploding market in education – IBM, SAS. Big data, small data, fine granularity of trace too to understand learning.

Similar SoLAR Flares have happened in Purdue, and elsewhere. [SoLAR](#) – bringing researchers in to dialogue with senior university managers, companies, practitioners. A systems kind of dialogue. Now incorporated as a not-for-profit, with founding institutions: Athabasca, Open University, UBC, U Queensland, U Saskatchwan. Open Learning Analytics white paper.

Image and story-based activity using cards activity – [LAnoirblanc](#) – pick an image and say why it tells you something about learning analytics, post it to the Tumblr.

## Sheila MacNeill, JISC

Delighted to be co-hosting. Have a [series of briefing papers on analytics](), recently released. JISC funding business intelligence, activity data, learning analytics. These papers are a consolidation of UK and international perspectives. To help early adopters and senior managers get up to speed. Eleven papers. Four are substantial, overview papers: first is Analytics for the Whole Institution: Balancing Strategy and Tactics. Three on the context. Three on applications. They're coming out soon – timetable on [Sheila's blog]().

## Simon Buckingham Shum

Is learning analytics about the same outcomes, but higher scores? Evolutionary technology. But perhaps it could or should be a revolutionary technology – a vehicle for paradigm shift. We can pay attention to things we know are important, but couldn't observe and assess at scale. From school level, to higher ed, to HR and training.

# Lightning Presentations

## Simon Buckingham Shum

First because alphabetical! Social network diagram – a "gratuitous visualisation slide". Question is how does this help me make sense of anything? Work with Martin de Laat – tool to help you visualise your offline network. Can connect people and filter them by various analytics. Modelling the dispositions of lifelong learners. Social networks, dispositions for lifelong learning.

Q Can you tell us about your second slide?

Working with Bristol, who have a model for key dispositions for lifelong learning, generating a visualisation, not from a questionnaire as at present, but from user behaviour.

Q How would a user use that visualisation? (the SNA)

See if we have different communities on the same topic but not connected? Or a single person connecting two communities – point of vulnerability. Also have reasons for the ties as well as standard SNA. Move to semantic social networks, not just topographic ones.

Q Is there a connection to Shane Dawson's SNAPP?

They're doing forums. We're using all sorts of ties – e.g. have same interest in a learning artefact, not necessarily interacted.

**Nicola Avery**

Medicines and Healthcare Products Regulatory Agency – involved after clinical trials, licencing and life cycle management (yellow card).

Historically only do training, never had a LMS. Formed a focus group, feeding in to a new BI strategy, data assurance, performance support. Lots of people interested in data, but don't know what the right questions are. Looked at feedback from potential users.

Q Are you using analytics to improve communications from your data?

Yes. Want to explore. Have only done f2f, learning analytics a useful starting point, also on the communication side, and technology and what data exists, can we mine it. We're not mining ourselves, we're looking at what others are doing.

**Doug Clow**

A slide from me about Data Wrangling.

## Adam Cooper

Exponential random graph models. Can be nonplussed by sociograms, network density, betweenness centrality – struggling to make sense of those. Exponential is the mathematical characteristic, random because are stochastic processes. Mutuality – do you reciprocate? Transitivity – friend of my friend is my friend? Homophily – friends with people who are like you in some way. Use those small-scale phenomena to understand the whole shape. Personal experiment of JISC and CETIS teams network. Can see is quite dense from the graph; are some teams more likely to follow others? Follow own team?

Q A new approach for analysing social network diagrams?

It exists, since the mid-80s, on things like the transitivity, homophily. 1938 was first time they looked at reciprocal nature of ties. Sociology research, is a tool to overcome frustration that node-level metrics and sociograms aren't that informative.

Q Does it give you any more understanding?

Diagram is difficult to interpret. See what we can identify that might have led to that diagram – say is it arising from friend-of-a-friend or homophily answer? Would get similar diagrams, but modelling might let you see what was most likely.

Q Does it rely on manual tagging of the nodes? Is it scalable?

Weakness is that if you don't have node labels, and your network has heterogeneities, doesn't work well at all – doesn't converge.

**John Doove**

From SURF in the NL, it's like JISC. We are rash in approaching new themes – don't write reports, we give projects funds and see what questions can be answered. Seven projects funded. All user-centred.  Video of project summaries (in Dutch with English subtitles) http://bit.ly/SURF-LA-pr  and 2 videos of lessons learned http://youtu.be/9fbhR6KW53Q. One minute for every project!

Q Commonalities in projects?

Most were over-ambitious in what they could get out of the data. Said wanted to work on all the data in Blackboard … but availability is not the same as usability of data. These projects only had 7 months and about £7,000 funding. This is a multi-disciplinary thing, need educational expertise, data mining expertise, legal expertise: has to be a team.

## Cath Ellis

EBEM – a JISC project. By product is how assessment data could be part of the learning analytics dataset. What could assessment add to this? What impact does it have on students? Measure whether it does have an impact. Have harvested data assessment – always had it through marking, but not pulled together in a meaningful way. Asked the students some questions, then showed the data on results, then asked them again. Students were measurably more motivated to act on the feedback afterwards.

Q Have you used it to look at how your assessments are constructed?

Haven't yet, but want to ask how to feed that back in to assessment design. Is also tutor-facing as well.

Q Any other research in to feedback would have had the same effect?

Sure, but interested in the granularity. Motivation. Shifting attention to important areas that tend to be neglected. The qualitative data reports that this does shift their attention that way, because of the granularity.

Q How do you highlight the common problem areas?

You measure it and you show it to them.

**Rebecca Ferguson**

Social learning analytics. Usually LA is about how to improve things for learners and their environments; social is about groups, communities, networks and how to support that too. Networks and discourse. Find the knowledge-building discourse automatically: working with framework around exploratory dialogue from Neil Mercer and Karen Littleton and their team. Synchronous conference chat, 2d online conference in Elluminate/Blackboard Collaborate. Manual analysis to pick out key phrases (e.g. However, have you read, etc). We were missing a huge amount, and doing it manually was slow. We did it computationally!

Q Were the computational results different from the manual ones?

No. Found a lot more indicators we could take and apply to other situations. Still at early stages. Both methods could pick out people arriving and saying hi, saying goodbye, or problems e.g. with sound. Could also find when keynote speaker said something engaging and everyone started talking about that.

Q Did they know you were doing this before?

Was an openly available conference, transcripts are in the public domain.

Q Would it change behaviour if they knew their comments were being analysed – better discourse?

Didn't know that. It was educators, cohort like people in this room.

Q Difficulty in using this in asynchronous?

Yes, difficult, people talk in different ways. Synchronous has challenges – don't know who's talking to whom, not using conventional spelling, punctuation.

Q Environment has any impact on the kind of dialogue? Elluminate vs virtual world, kind of interactions are quite different?

Yes. Like to apply it to virtual world chat. Possibly would be even more interesting – nearly all the discussion happens in that way. In Elluminate this is background chat talking around it. In virtual worlds they are the main protagonists.

## Dai Griffiths

We think of data as having a shiny, new neutral feeling. But it has models, which determine how we act. Interested in tracking how our models of education are changing, and how those relate to the structures that manage our institutions. Distinguish models based on customer relationship management – the null hypothesis of learning analytics. As researchers, have baggage around adaptive learning, learning design patterns; the response is at last! We have the data to do this properly! Or might be, remember what happened last time when it went pear-shaped? How does the use of LA reinforce the managerial and regulatory frameworks of education at the expense of practitioners, in ways that are difficult to capture.

Q Do you have a favourite example of where analytics is reinforcing the wrong paradigm?

Talked to VP Academic for a Canadian university at an event: "We are learner-centred so we go by student grades." There's a world of interactions there that are not being paid attention to. All that work following that line is disturbing: the assumption that you can go by student grades, do your analysis by simply looking at the outcome. Big blind spot in the middle.

Q Who's scarier for teachers? Gove, or Blackboard, Inc?

I don't know! I'd add Pearson. I don't want to be a Jeremiah – there are traces we can pull out. We have to have an eye on the models. Gove's models are up front, Blackboard's aren't.

**Martin Hawksey**

MOOCs – the instructivist style, like Coursera et al, scaled up VLEs/LMSes. Also connectivist MOOCs – distributed model. Students learn in their own spaces, blogging, bookmarking, tweeting. George Siemens, Stephen Downes work. Applied elsewhere. Aggregate distributed activity, redistribute it to students – gRSShopper daily alert. Interested in extending this and applying learning analytics. Google Spreadsheets as a tool – rapidly develop dashboards. Looking at how to pull this data from open(ish) sources in to a central pot. Issues around time limited, analytically cloaked, darksocial, infrastructure messy data.

Q Tell us about the issues.

Someone who tweets and blogs will have potentially two separate identities – merging those is a challenge. Comments on posts elsewhere too. People use different userids.

Q Implied question – can we change the ways we do the cMOOCs by using an analytics tool – who is it for?

Everyone! I like spreadsheets because you can share them to the world – tutors, individuals – Sheila and I have a paper on this. Open courses, make the data open as well. It's all open-ish data.

Q Tell us abut the "ish".

Twitter has APIs that let you extract the data, but T&Cs say can't store the data in the cloud. So breaking them to do that. Not always technically available. Screenscraping resorted to sometimes where APIs unavailable.

## Jean Mutton

Student experience team at Derby. Set of indicators for staff to ID at-risk students. How do we measure student engagement? How do we share that across the institution? Looked at the data, held in seven different systems that don't talk to each other – big challenge. Not just attendance monitoring system (quite sophisticated) but the grades – but grades are too late. How can we ID students we have a hunch about, what additional support is there to improve retention, progression and achievement. Look at what indicators there might be, see what really matters to the student. Set the tutor-student conversation in context. Not just handing assignment in, but are they picking it up again? Levels of engagement.

Q How is that fed back to the student? Like Purdue Course Signals?

This is a scoping project so not done yet, but plan to design it. Idea was tutor-facing, but now clear needs to be student-facing as well.

Q These are the things that matter to the student – matter to their progression, or matter to them personally?

Important point. Wanted to find out what makes a difference to the student experience from a personal point of view. If they have a good story to tell, e.g. president of rugby club – how are we gathering that data? HE Achievement Record will have some of this indicators beyond study, e.g. volunteering, student reps.

Put information in categories – attendance monitoring, assignment submission – stuff we could get hold of. Thought of other things not currently collected.

## Jonathan San Diego

[Online slides](#) with examples. Pedagogic Planner. Design of activities can vary; users interact in different ways. Early stage examples here. Users get alerts for peer activities. Activity data, mapping different learning activities, using Diana Laurillard's conversational framework. Grainne Conole's categorisation of learning activities. Feed back in to the learning design process.

Q This is a mock-up storyboard?                    [Yes.]

## Mark Stubbs

Manchester Met. Pragmatic. Have had BI since April 2008, dashboards from 2009. Using R randomForest brute-force for learning tech review. If you're a young male and think you can leave it to the last minute, you're so wrong. The most tutors went on to a site, the worse it went for students. Didn't follow it up – needed a more systemic approach. Started building student satisfaction data collection – 10k respondents. Steering decision-making. Lots of people said they were coming this year, but didn't, in the new fees regime. Course organisation a key element.

Q More tutor on forum, worse it was? They were bad and tutor helping? Which way causation?

Hard to say! The more tutors clicked, more students did, until there was a tail-off. Only one year's worth of data, was 18k students. We're sitting on potentially really interesting data to feed back to students. Others who made this choice didn't do so well, do you want to do that? We're looking at how to feed it back. Looking at a variety of audiences – HoDs, unit leaders, looking at what their requirements are. Dashboard design.

John Doove: has a project working on this.

Q Are the top echelons listening to you?

Yes. Fortunate. Giving decision-makers the info they need – not just managers, but students.

Q Are findings big surprises? Or evidence to reinforce the message?

Some things you think are obvious in retrospect. First to second year see a big dip in satisfaction – make a big fuss for first year, not second, so makes sense.

## Annika Wolff

JISC-funded project at OU – Retain. Using different OU data to predict struggling students. Activity on VLE, student assessment data, integrate with demographic data. Train predictive modeling, find useful visualisations – prototype dashboard with traffic light low/medium/high risk. Show to lecturers who's struggling, aggregated stats to module managers to see where problems are arising. Trained and tested on historic data. Got some interesting findings. VLE data useful to integrate, most data. Assessment data useful up to a point. But combined gives best. Decision trees and SVMs – decisions trees better. Overall clicking doesn't predict outcome, but changes in behaviour in clicking highly predictive: if clicked a lot but tailed off, is an indicator that they're struggling.

Q Was the VLE data all click based?

Yes. All we have on students, we don't have data on lecture attendance, only e.g. forum, learning activities, etc.

Q Small orange segment. Would've thought that's where you want to make your interventions?

This was a snapshot in time on one course – had run model, had factors. Most of the students were doing well or badly, only a few were in the middle. Can see who they were, see what contributed to that risk factor.

Q To ignore a big red sector is ethically unsound.

Yes. The red section is where the meat of it is. Red is where bad outcome on next assessment or whole course.

Q Across a range of different modules and faculties?

Chose three courses that were typical but heavy use of the VLE. Found differences between the courses, especially in terms of the assessments (substitution). Reduction in clicking behaviour worked across all of them. Module-specific information is crucial for developing the models, because the courses are structured so different. E.g. a forum that is crucial at a particular time on a course.

## Chris Ballard

Representing a supplier. Tribal Labs. Developing and testing a predictive model, in partnership with a university. Outcome has been to develop a successful model using machine learning techniques. Supply our own student management system, so integrating this, with library, VLE. At the moment, looking at how to visualise the data to enable staff to correctly interpret it, and use that to support the students. Often a missing element in learning analytics solutions to bring together the visualisation and the action.

Q Variety in courses, institutions. How can you with a single solution cope with great diversity in the real world?

That's why it's an R&D project! Without needing a lot of machine learning experts. Don't have a clear answer at the moment, but looking at that now. Want to test our model.

Q Looking at the action data, is that feeding back?

Yes. Want to visualise the interventions, when they happened, have they led to subsequent improvement.

–