Data Wranglers: Human interpreters to help close the feedback loop

Doug Clow
The Open University
Institute of Educational Technology
Walton Hall, Milton Keynes, UK
doug.clow@open.ac.uk

ABSTRACT

Closing the feedback loop to improve learning is at the heart of good learning analytics practice. However, the quantity of data, and the range of different data sources, can make it difficult to take systematic action on that data. Previous work in the literature has emphasised the need for and value of human meaning-making in the process of interpretation of data to transform it in to actionable intelligence.

This paper describes a programme of human Data Wranglers deployed at the Open University, UK, charged with making sense of a range of data sources related to learning, analysing that data in the light of their understanding of practice in individual faculties/departments, and producing reports that summarise the key points and make actionable recommendations.

The evaluation of and experience in this programme of work strongly supports the value of human meaning-makers in the learning analytics process, and suggests that barriers to organisational change in this area can be mitigated by embedding learning analytics work within strategic contexts, and working at an appropriate level and granularity of analysis.

Categories and Subject Descriptors

J.1 [**Administrative Data Processing**] *Education*; K.3.1 [**Computer Uses in Education**] *Collaborative learning, Computer-assisted instruction (CAI), Computer-managed instruction (CMI), Distance learning.*

General Terms

Management, Design, Human Factors, Theory.

Keywords

Learning analytics, data wrangling, interpretation, meaning-making, organizational learning, systems.

1. INTRODUCTION

Learning analytics is widely seen as entailing a feedback loop, where 'actionable intelligence' [7] is produced from data about leaners and their contexts, and interventions are made with the aim of improving learning ([5, 8, 9, 12, 15]). This is at the heart of the Learning Analytics Cycle set out by Clow [10], which is intended to make "the necessity of closing the feedback loop through appropriate interventions unmistakable". How is the loop to be closed? Actionable intelligence needs to be coupled to intelligent action.

Sophisticated educational data mining tools are often hard for non-specialists to use and interpret [19], so previous learning analytics literature (see e.g. [16, 17, 20–22]) has highlighted the need for multidisciplinary teams, with a key role played by humans in interpreting the data, engaging in sense-making activities to mediate the information in ways that enable intelligent action. As Siemens [22] puts it, sense-making and social processes are important because of the complexity of the data and because "learning is a complex social activity".

How should such human sense-making efforts be implemented? To be effective beyond small-scale exploratory activity, programmes "will be institution-wide efforts" [4], with careful consideration given to how they will interact with educational systems, leaders and other stakeholders [20]. In particular, academic staff/faculty and researchers need to be supported to learn to interpret and design learning analytics [6], through the establishment of a contextual framework [14] and developing a culture of data use as part of increasing organisational capacity [16]. This is not a straightforward process: some academic staff may be resistant to what they perceive as the metrics agenda. However, engaging with learning analytics can steer the agenda towards richer conceptions of learning than a naïve quantitative view might imply [9].

To achieve sustainable transformation, it is important to support a Community of Practice [13, 23] around the use of learning analytics. Argyris and Schön [1] developed and popularised the conception of single- and double-loop learning as important factors in organisational learning. In a learning analytics context [10]:

"A learning analytics system may be used simply to attempt to achieve set goals (single-loop learning); greater value and insight will come if those goals themselves can be interrogated, challenged, and developed (double-loop learning)."

Despite this concern for institution-wide capacity-building learning analytics programmes, there are – so far – few accounts of such activity in the learning analytics literature. (With some notable exceptions, such as Signals at Purdue [2, 3] and Macfadyen & Dawson [17].)

This paper describes such a programme of activity, where a group of 'Data Wranglers' were deployed to engage in sense-making activity with learning analytics data. The immediate goal was producing reports with actionable recommendations, and the overall aim was to drive systematic improvement through single- and double-loop learning, and through the support and development of a Community of Practice at the Open University UK (OU). This included engagement with a range of stakeholders from individual academic staff to senior managers. (For a very brief early outline of earlier work on this project, see [18].)

LAK '14, March 24 - 28 2014, Indianapolis, IN, USA
Copyright is held by the owner/author(s). Publication rights licensed to ACM.
ACM 978-1-4503-2664-3/14/03…$15.00.
http://dx.doi.org/10.1145/2567574.2567603.

2. CONTEXT AND ROLE OF DATA WRANGLERS

The OU is a large distance teaching/online university, with around 250,000 students studying largely part-time, with much (but not all) of the tuition online. So the quantity of electronic data is considerable. Managing the scale of the student body requires considerable centralisation and many administrative systems and processes. This could create a gap in knowledge and practice between the data about learning and the academics who need to be in a position to act in order to improve teaching provision: a gap that could easily widen as the quantity and complexity of data increases.

To address this problem – to ensure that the data being collected was interpreted and turned in to actionable insight – pilot work started in 2010 and was expanded and developed into the Data Wranglers project from 2012.

The Data Wranglers are a group of academics who analyse data about student learning and prepare reports with actionable recommendations based on that data. There is a Data Wrangler for each of the OU's seven academic Faculties, and so far as possible the Data Wranglers are selected to have an academic background close to the Faculty they are working with.[1]

Their role is to translate the theory described above in to practice: to act as human sense-makers, facilitating action on feedback from learners, making better sense of what that feedback means and how the data can be improved (double-loop learning), and helping to develop the Community of Practice around the use of learning analytics.

The Data Wranglers work with four main data sources:

- Survey feedback data from students, gathered at the end of their course.
- Activity data from the VLE/LMS (Moodle).
- Delivery data about the mode of delivery and structure of courses (e.g. what use each course makes of online forums).
- Aggregated completion, pass rate and demographic data.

In practical terms, data from these sources is aggregated using a SAS data warehouse, and exported to a Tableau workbook for each Faculty. The Data Wranglers use these workbooks as their primary data investigation tool, and to generate some charts and visualisations, but also use the data sources directly (including SAS portals on the Data Warehouse) where appropriate, and produce their own charts in Excel.

To help make sense of the data, the Data Wranglers develop an understanding of the particular situation of the Faculty they are working with, building up relationships with key stakeholders to enable the reports to focus on areas where they can be of most value. The Data Wranglers also seek opportunities to engage with Faculty academics to support the use of this data in action, including work on individual courses and sitting on relevant committees.

The reports form the basis of an ongoing conversation with the Faculty about feedback on the learning experience. Further analysis is carried out where this is required, ranging from a simple extra table of data to a full-scale institutional research project, as resources permit. The Data Wranglers also support the delivery of Learning Design [11] activities. There is strong potential synergy between learning analytics and Learning Design. As Lockyer, Heathcote & Dawson argue [14], Learning Design can provide an account of pedagogical intent that can provide useful context for interpreting learning analytics, and learning analytics can provide particularly useful insight to underpin the process of Learning Design. So, for instance, some Data Wranglers were able to attend or even facilitate Learning Design workshops, to help the process of closing the feedback loop.

It is important to note that all the data are available to academics directly, via various dashboards and online facilities. The role of the Data Wrangler is not only to analyse the data, but to increase the familiarity of academics with the data sources, to build learning analytics capacity as part of a Community of Practice.

To illustrate this graphically, Figure 1 shows the situation before the Data Wranglers were in place: some users made some use of the data, but not extensively. In Figure 2, the Data Wrangler makes use of all data sources, and makes this available to all the users. As a result of the process, more users become familiar with the data sources, and begin to make more use of them directly, as depicted in Figure 3.

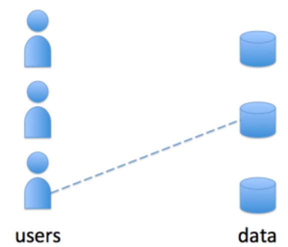

Figure 1. Situation before Data Wrangler in place: users make little use of available data sources

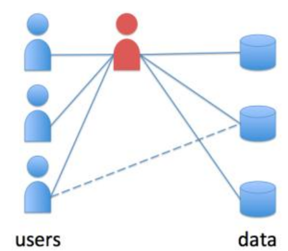

Figure 2. Situation when Data Wrangler starts work: Data Wrangler presents data from all sources to all users.

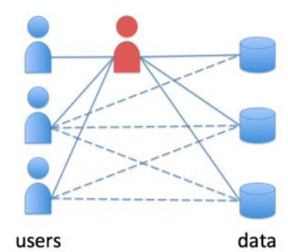

Figure 3. Situation when Data Wrangler work is mature: many users make more use of data sources directly.

The Data Wrangler activity is one component of a large, complex system of quality assurance and quality enhancement. Notably,

[1] The author of this paper is one of these Data Wranglers.

there are other processes for formal feedback and action on student feedback, completion, and pass rate data. Predictive analytics are also under development, but are not yet widely deployed.

3. EXAMPLES

Four Data Wrangling reports, of around 20-30 pages each, have been produced for each of the OU's seven Faculties[2], at roughly four-monthly intervals over the period from Spring 2012 to Summer 2013. To illustrate the role of the Data Wranglers, this section presents examples of analyses from Data Wrangler reports that resulted in changed capacity to act on learning analytics data.

The first example is usage data from the VLE/LMS. Figure 4 shows the use of various VLE/LMS components by week for one particular course. There are two troughs in overall activity, visible in the drops in Pages and Forum visits in weeks 5-8 and 11-12, which through conversation with the course team correspond to the pattern of online activity designed in to the course. Even more striking are the steep peaks for Quiz use in weeks 4, 9 and 13, which (perhaps unsurprisingly) the weeks where students are directed to complete Quizzes as part of their assessment.

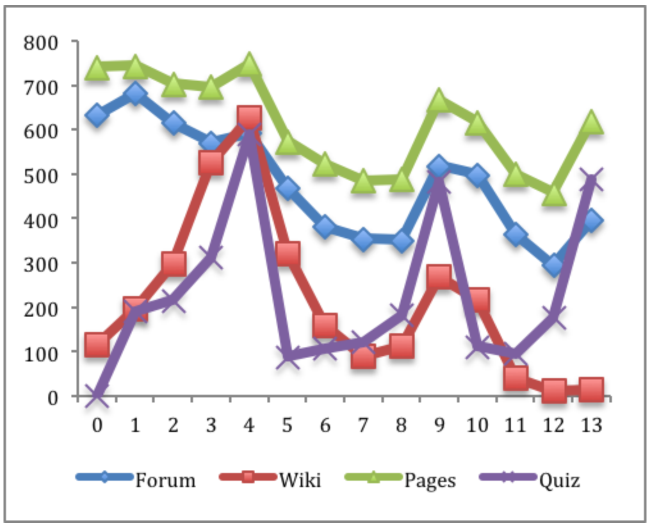

Figure 4. Unique visits to pages, forum and wiki by week for one specific course.

This provided benefit at two levels. At a high level, it gave a degree of ground-truthing to support the development of proactive student support system based on predictive modelling (still in development). At a lower level, charts such as Figure 4 drew the availability of this data to the attention of an academic who had concerns about a new cohort of students, and the Data Wrangler was able to quickly capture and present similar charts to enable them to understand the change and take appropriate action.

Other examples provided contextually-relevant data to support longstanding good practice in online learning. Two such examples are discussed below.

The first such piece of good practice concerns the importance of assessment: it is well known that students tend to ignore optional learning activities but are likely to focus on activities that are assessed. This was evident in data presented in a Data Wrangler report. As shown in Table 1, on courses that make only incidental use of Elluminate (synchronous conferencing), few students make use of it. Where the course specifies an activity, about half of students used it, but almost all (95%) students used it where the activity were referenced in the assessment. This data was in theory available to academics before the Data Wrangler process, and it is hardly a groundbreaking research finding, but it did add highly relevant weight to discussions about course design and the role of assessment.

Table 1. Usage of Elluminate broken down by course use of Elluminate for one Faculty between 2011-2012.

Course use of Elluminate	Students using Elluminate
None	18%
Informal	27%
General student support	35%
Specified activities not assessed	49%
Specified activities referenced in assessment	95%

The second good practice example comes from students' reported enjoyment of different learning activities. Figure 5 shows that many students enjoy learning through reading print, but far fewer enjoy learning through reading text online; listening to audio is a little more popular and viewing AV is almost as popular as print. However, Figure 6 shows that activities change this balance: about as many students report enjoying learning through online activities as through in-text activities, with ratings higher than for AV activities or tutor activities, and almost as high as for reading print. Again, these results provided extra, locally-valid empirical grounding for the mainstream instructional design advice: when teaching online, use less text and more activities. Charts like these were presented as part of a learning design workshop, which facilitated a very productive, detailed discussion about the nature of online learning, and helped move the focus of debate away from the subject being taught (a common preoccupation of early career teachers) and towards how it can best be learned by students.

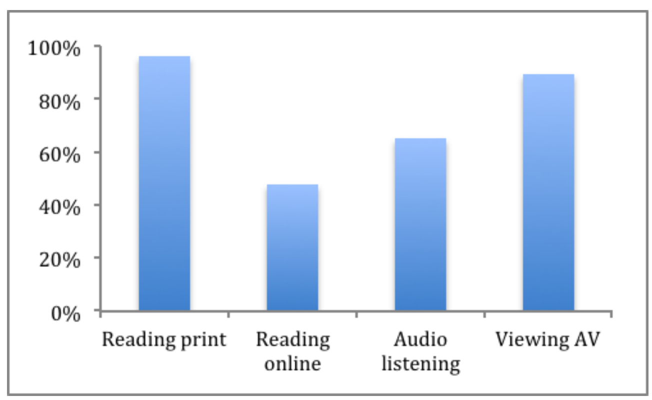

Figure 5. Proportion of students on selected courses who report that they 'enjoy learning through' different media.

One key finding from the work was a largely null one: despite considerable analytical effort, surprisingly few significant correlations could be found at the macro level between student feedback, related aspects of course design recorded in the delivery data, and VLE/LMS activity data. It may be that the data was too coarse to show effects, or that the delivery data was not accurate.

---

[2] There are some missing reports for some Faculties due to staffing issues.

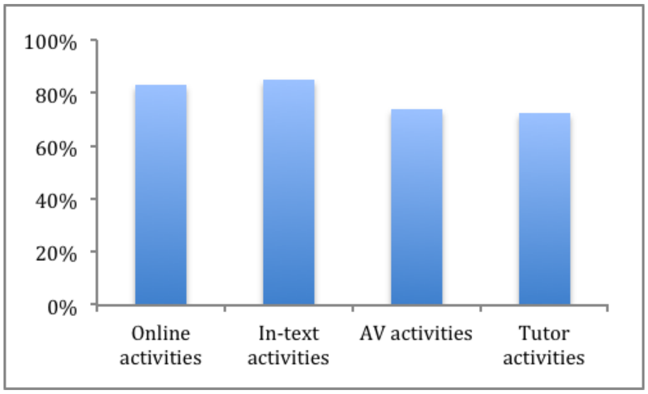

Figure 6. Proportion of students on selected courses who report that they 'enjoy learning through' different activities.

One final example illustrates the Data Wrangler process at its best. Some issues with a suite of courses emerged in one Faculty. The senior manager responsible had already engaged with their Data Wrangler through several reports, and so was able to ask for a quick bespoke report on those particular courses. The senior manager reported that this had been very helpful in the review of those courses, enabling the decisions to be made on the best available data.

4. EVALUATION

The aim of the Data Wranglers is to improve the learning experience. Disappointingly, the overall performance indicators that directly relate to this (e.g. student retention, completion, progression and feedback) have if anything worsened slightly since the project was started. However, this seems most likely to be the result of a substantial change in funding regime that took place over the same time period: the majority of student fees were previously paid via public funding, but are now paid by the student via Government-backed loans. As well as changing the profile of new students (in ways that were predicted to lead to worse outcomes), this has also required substantial changes to OU structures, systems and processes that have yet to bed in. Also, the expected timescale for an improvement is long. Course production is the main target for Data Wrangling, and OU courses, despite recent acceleration, take 1-3 years to produce, and thereafter are generally presented with only minor modifications for several years.

Encouragingly, there is good anecdotal evidence of increasing direct use of the data sources by Faculty members, although log data to support this is not available. There was also evidence in the reports of issues being raised through the reports leading to further investigation and action being taken.

An evaluation of the Data Wranglers' work was carried out and reported internally in July 2013. Feedback was obtained from most of the direct recipients of the reports in the faculties, and from other stakeholders (total N=22), by email or through face-to-face discussion if that was preferred. Feedback was also obtained from the Data Wranglers and the statisticians supporting their work.

Feedback from stakeholders was generally positive, with respondents reporting that they valued the process, and its iterative, conversational nature in particular:

"[U]seful to have an iterative phase during which queried Wrangler's interpretation"

The reports have stimulated productive reflection and discussion, with many stakeholders reporting instances where the reports have triggered or supported discussions about the development of teaching, e.g.:

"Main use of the reports has been in stimulating discussion around ways in which we can improve the students' learning, particularly online."

"Useful in supporting [course] review and curriculum planning discussions."

Another theme was a desire for more data to be included in the reports: more data sources, more fine-grained data, more historical data for comparison, and more data related to the new OU structures, systems and processes.

A less positive theme was the unevenness of the process across Faculties. The reports were all produced to the same template, and based on the same data sources, but the analysis and recommendations varied, as did the quality of the conversation between the Data Wrangler and Faculty members. Other concerns raised included the timing of the reports and which courses were included.

Perhaps the largest issue was data quality. One serious misinterpretation of VLE/LMS activity data arose during the project, and the quality and resolution of the delivery data was perceived as a serious obstacle. Until the Data Wranglers started work, this data was rarely used. It was only through engaging with this data and attempting to deploy it that the quality issues came to light. This ongoing conversation about data – between those who capture and curate it and those who can do something about it – is key to the double-loop learning aspect of the process.

This evaluation is being used to inform future Data Wrangler activity. Further reports are in progress, with improvements to the process based on the evaluation and to address stakeholder concerns. The Data Wrangler work is embedded in institutional processes, with engagement at the most senior level as well as with individual academics.

5. CONCLUSIONS

This paper has presented an account of Data Wranglers. Substantial progress has been made in establishing a Community of Practice in learning analytics, by analysing and presenting learning analytics data to academics who are in a position to take action, and through extensive engagement. Progress has also been made in improving institutional learning about the quality and interpretation of the available data, and how better data can be captured and made available.

A structurally similar project – large scale and with senior management engagement – is discussed Macfadyen & Dawson's analysis of LMS use at a large research-intensive university [17]. In that project, technical discussions swamped more meaningful change processes. In contrast, the Data Wrangler work detailed in this current paper was perhaps better placed to present data "to those involved in strategic institutional planning in ways that have the power to motivate organizational adoption and cultural change". The analysts were well embedded within the organisation to begin with, and they were able to facilitate conversations about pedagogical issues through finer-grained analyses. This project was able to pay "greater attention […] to the accessibility and presentation of analytics processes and findings so that learning analytics discoveries also have the capacity to surprise and compel, and thus motivate behavioural change" – although it is not yet a runaway success.

This approach is high cost in terms of time. The high cost has been matched by a high yield in understanding, which enabled further development. There is now an institution-wide top-down analytics strategy in place, and this is built on a bottom-up understanding of at least some of the potential of the data to improve learning.

The Data Wrangler process is not uniform. It capitalises on the individual strengths of the Data Wranglers and the key stakeholders in the Faculties. This is as expected in a capacity-building exercise: the process must start from people's existing expertise, and if capacity building is required, this expertise will of necessity be lacking.

The issues of data quality unearthed through the Data Wrangler process shows the value of sense-making activity. If it is nobody's job to make sense of the data, the risk is that the data do not make sense but nobody realises.

A bottom-up, grounded approach is necessary for sense making. However, as Macfadyen & Dawson [17] powerfully argue, organisational change is hard to achieve without meaningful engagement at the strategic, top-down level as well. The Data Wrangler process encapsulates this: the richest discussions were at an individual course level, but engagement at the levels of Faculties and the whole institution was also valuable in capacity building. As previously argued [10], students and teachers are closest to the learning experience and best placed to take rapid, appropriate action in the light of learning analytics data, but managers and policymakers are able to take action at a much greater scale of impact.

Substantial organisational change is hard. Significant effort and sensitive engagement is necessary and not always sufficient. Transformatory change is likely to take substantial amounts of time. The process can appear messy, and it is easy to focus on disappointments. However, it is only through the detailed process of engagement and dialogue between analysts, stakeholders and the data that insight and organisational change are developed.

6. ACKNOWLEDGEMENTS

The author wishes to thank all the other Data Wranglers, the statisticians, support staff and faculty members at the OU who have engaged with this work. Any errors or misrepresentations remain the author's responsibility.

7. REFERENCES

[1] Argyris, C. and Schön, D.A. 1974. *Theory In Practice: Increasing professional effectiveness*. Jossey-Bass.

[2] Arnold, K.E. 2010. Signals: Applying Academic Analytics. *Educause Quarterly*. 33, 1 (2010).

[3] Arnold, K.E. and Pistilli, M.D. 2012. Course Signals at Purdue: Using learning analytics to increase student success. *LAK12: 2nd International Conference on Learning Analytics and Knowledge* (Vancouver, BC, 2012), 2–5.

[4] Brown, M. 2012. Learning Analytics: Moving from Concept to Practice. *EDUCAUSE Learning Initiative Brief*. July (2012), 1–5.

[5] Brown, M. 2011. Learning Analytics: The Coming Third Wave. *EDUCAUSE Learning Initiative Brief*. (2011).

[6] Buckingham Shum, S. 2012. *Policy Brief: Learning Analytics*.

[7] Campbell, J.P., DeBlois, P.B. and Oblinger, D.G. 2007. Academic Analytics: A New Tool for a New Era. *Educause Review*. 42, October (2007), 40–57.

[8] Campbell, J.P. and Oblinger, D.G. 2007. Academic Analytics. *EDUCAUSE Quarterly*. October (2007).

[9] Clow, D. 2013. An overview of learning analytics. *Teaching in Higher Education*. 18, 6 (Aug. 2013), 683–695.

[10] Clow, D. 2012. The learning analytics cycle: closing the loop effectively. *LAK12: 2nd International Conference on Learning Analytics and Knowledge* (Vancouver, BC, 2012).

[11] Cross, S., Galley, R., Brasher, A. and Weller, M. 2012. *OULDI-JISC Project Evaluation Report: the impact of new curriulum design tools and approaches on institutional process and design cultures*.

[12] Johnson, L., Smith, R., Willis, H., Levine, A. and Haywood, K. 2011. *The 2011 Horizon Report*.

[13] Lave, J. and Wenger 1991. *Situated Learning: Legitimate Peripheral Participation*. Cambridge University Press.

[14] Lockyer, L., Heathcote, E. and Dawson, S. 2013. Informing Pedagogical Action: Aligning Learning Analytics With Learning Design. *American Behavioral Scientist*. 57, 10 (Mar. 2013), 1439–1459.

[15] Long, P. and Siemens, G. 2011. Penetrating the Fog: Analytics in Learning and Education. *Educause Review*. 46, 5 (2011), 31–40.

[16] Lonn, S., Krumm, A.E., Waddington, R.J. and Teasley, S.D. 2012. Bridging the gap from knowledge to action: Putting analytics in the hands of academic advisors. *LAK12: 2nd International Conference on Learning Analytics and Knowledge* (Vancouver, BC, 2012).

[17] Macfadyen, L.P. and Dawson, S. 2012. Numbers Are Not Enough. Why e-Learning Analytics Failed to Inform an Institutional Strategic Plan. *Educational Technology & Society*. 15, 3 (2012), 149–163.

[18] Powell, S. and MacNeill, S. 2012. Institutional Readiness for Analytics. *JISC CETIS Analytics Series*. 1, 8 (2012).

[19] Romero, C. and Ventura, S. 2010. Educational data mining: A review of the state of the art. *IEEE Transations on Systems, Man and Cybernetics - Part C: Applications and Reviews*. 40, 6 (2010), 601–618.

[20] Siemens, G. 2012. Learning analytics: Envisioning a research discipline and a domain of practice. *LAK12: 2nd International Conference on Learning Analytics and Knowledge* (Vancouver, BC, 2012).

[21] Siemens, G. 2013. Learning Analytics: The Emergence of a Discipline. *American Behavioral Scientist*. 57, 10 (Aug. 2013), 1380–1400.

[22] Siemens, G. 2012. *Sensemaking: Beyond analytics as a technical activity*.

[23] Wenger, E. 1998. *Communities of Practice: Learning, Meaning and Identity*. Cambridge University Press.