



Project: Data Language

Project Engagement Report

July 2016

Introduction

The purpose of this report is to set out the steps taken during the HEDIIP Data Language project to engage stakeholders of all kinds in the development of the new data model, and to indicate some of the key changes to the model made as a result of feedback. This is a high-level report, not an exhaustive summary of all the consultation undertaken and changes made in response. Its purpose is to give confidence to stakeholders about the depth and breadth of engagement undertaken.

Engagement Approach

The project adopted an engagement approach which was focussed on frequent and personal interaction, in meetings and seminars, with frequent iterations of the model developed in response to the feedback received.

Engagement with data collectors and data users

Our main model of engagement with data collectors and users was via in-person meetings where we discussed their current data collection and analysis needs. We held 27 meetings with data collectors and users. We also undertook about a dozen phone interviews either to follow up specialist points, or to engage individuals who had not been able to attend our meetings. Only a couple of these phone interviews engaged data collecting organisations whom we had not also spoken to face-to-face. The project had a Technical Working Group for data collectors and users which reviewed project outputs and met on two occasions. We were also able to access a number of PSRBs in workshop activity through the QAA's PSRB forum. Finally, the project engaged with HESA staff, holding two workshops and receiving feedback on technical points.

Engagement with HE Providers (HEPs)

Our main model of engagement with HEPs was through workshops. Ten workshops were held with attendance from 88 different HEPs (many HEPs sent more than one delegate, or sent delegates to more than one session in order to retain contact with the project. This was very helpful). We also engaged through sector conferences and groups, such as AUA Conference, SROC, HESPA and the London Higher HE Planners' forum. In seven cases we held meetings with individual HEPs to explore areas of particular common interest. The project had a Technical Working Group for HEPs which met separately from the data collectors' group and reviewed the project outputs. We were particularly grateful for the detailed review of project outputs undertaken by members of this group. The project website was used to publish outputs and solicit feedback. Relatively little feedback was received directly to the project website.

Engagement with Software Suppliers

The project also engaged with software suppliers to the sector, holding 5 meetings with suppliers, as well as engaging online.

Overall (considering both travel and preparation time) in-person engagement with stakeholders absorbed about a third of the project's input days. About two thirds of the project engagement time were devoted to HEPs and about a third to data collectors and users.

Engagement Timeline

A period of exploration, development and consultation ran from November 2015 through May 2016. However this period was delivered in a more structured and phased way than originally planned. The period November 2015 – January 2016 was largely devoted to stakeholder meetings with data collectors and users. A period of intense engagement mainly with HEPs followed from February 2016 – May 2016. This engagement mainly featured workshops and conference sessions, with 11 workshops and three conference sessions delivered, but

there were also highly valuable visits to a number of specific HEPs. In total 88 HEPs were engaged, many of them repeatedly. We published 5 versions of the model over this period in order to generate feedback. In June 2016 we published a version of the model for formal consultation with the sector. In general, the project experienced nothing like the expected difficulty in engaging stakeholders. HEPs took a keen interest in every aspect of the project and rapidly developed an enhanced understanding of the work being done and the possible implications.

However, whilst the iterative process was successful in securing engagement, it was not always popular with those HEP colleagues who were engaged and some HEPs expressed a wish for a more structured consultation process. The project responded to this by introducing more structure into our consultation process as the project developed, but the very tight timescales for the project would always have made it impossible to deliver a high volume of consultation if that consultation were also formally structured. It was also notable that responses to the formal consultation document, when it was published, were few in number.

Drafts of the key deliverables – PD02 (Data Model), PD03 (Glossary) and PD04 (Data Dictionary) – were published by 9 May 2016, with PD02 and PD03 published in early drafts much earlier than this. The original deadline was to circulate these to the Project Board on 20 May 2016, so timing was in line with (or even ahead of) expectations but process was not.

Despite 845 unique visitors to the consultation website, few formal responses were made. The combination of the high level of awareness/engagement with the project generated via the informal processes and a low level of response to the formal consultation implies a deliberate choice on the part of most HEPs not to respond. This can be taken as indicating that the HEPs do not object to the Common HE Student Dataset and Data Specifications. This implies acceptance rather than consensus. Most respondents chose not to contest the main features of the model, but to raise specific points of detail, or issues about implementation. The responses that were received cannot be taken as a meaningful sample of opinion in the sector, however the evidence suggests that the sector accepts the model as a basis for development in Data Futures.

Feedback received

Early feedback received on the model showed a focus on issues which were important, but outside the project remit (such as timetables for implementation of Data Futures, or challenges to the current data requirements of data collectors). This showed that it was necessary for the project to communicate more effectively about both its own particular remit, and its place within the overall HEDIIP and Data Futures Programmes where these important issues would be resolved.

In general, feedback on the model validated its appropriateness to 'standard' programmes of undergraduate and taught postgraduate provision, but raised issues about applicability to other kinds of provision:

- Postgraduate research. Many providers reported that PhD student were permitted to start their programmes at any time, so the model's use of distinct course deliveries to denote the specific times and places where a course is offered was questioned;
- Flexible programmes. Certain providers have highly flexible programmes in which students can study at variable rates or take long breaks in study without any prior arrangement. This creates challenges for the model's course session concept (or any kind of in-year return) because it is not possible to know what an individual student will be doing during any time period until that time period is over;
- Modules. Whilst some of the feedback received about the module concept merely reflected the fact that not all HEPs use this terminology to describe their own curriculum, there was also feedback about module marks, placements and mobility schemes;
- Teaching and learning data. Feedback identified that the proposed linkages between courses and their constituent modules via the course module entity was weak and could not fully reflect the complexity of course structures. At the same time the level of data about teaching and learning activities proposed for

the module delivery did not fully implement the project principles, as atomic data about teaching, learning and assessment activities was not specified. This could limit development of the model.

Outcomes

The key issues identified above were resolved as follows:

- Postgraduate research. The project engaged in depth with a number of institutions to understand processes better. This follow-up confirmed that a number of providers do indeed permit PhD students to start at any time, but once those students have started, their continuing programme is managed either on the basis of their start month, or sometimes on the basis of the academic year. We were not able to find a case where the student's continuing programme (e.g. fee invoices for second and subsequent years, dates for annual review processes, timetable for final submission) were precisely controlled by the individual student's start date. In the institutions studied, there are effectively twelve monthly start dates for the PhD programme, and the individual student is treated for most purposes as having started on one of these, even though that was not the start date. The model has a `ComencementDate` for the student registration that is separate to the course delivery start date, so can cope well with this pattern.

A related issue was that certain providers told us that although they recognised the concept of annual progress review as applied to PGR students, the timing of this process was not as rigid as for taught undergraduates. This means that there could be a significant delay after the end of any progression period before the progression review is conducted and a decision made. This is an important collection issue, because data may not exist when someone wants to collect them, and has been highlighted in the draft Data Dictionary.

- Flexible programmes. This was another area where the project engaged in more depth with specific institutions. Our conclusion is that where a student engages in a completely flexible programme where they can take long and variable breaks between periods of study without any prior agreement with the provider, then the contract to educate (and therefore the student registration) is effectively made up one module at a time. As a student pursues such a programme, there will therefore be a number of separate student registrations which each represent only one module. In earlier drafts of our model these individual student registrations were not explicitly linked (the link between them was established by the fact that all the course deliveries had the same qualification aim). Feedback identified that a more robust method of explicitly linking these associated student registrations together was needed, and the `Engagement` entity was developed for this purpose.
- Modules. There was significant feedback about the concept of module marks. Providers identified that marks data would be difficult to interpret without an understanding of mark schemes, and there is no data set identifying all the mark schemes that are (or could be) in use in the sector. Providers also identified that marks might not exist at the module level – for instance some modules are pass/fail, whilst in other cases marks are applied to assessments only, and not to modules. The project considered a more detailed approach to mapping assessment activities which would have enabled resolution of these concerns, but there is only limited current policy interest in collecting marks data, and this additional complexity did not seem justified.

Earlier drafts had separate entities for modules, mobility and placements. Feedback correctly identified that there was strong overlap between these entities, and the placement and mobility entities in particular had virtually identical attributes. The project team simplified the model in response to this feedback by creating an off-venue activity entity to hold data about an individual student's placement or mobility activity, whilst recognising that a mobility or placement scheme fits into the model as a form of module.

- Teaching learning data. As already reported, the project considered the scope for a more detailed mapping of teaching, learning and assessment data which would not only allow resolution of the marks

issue, but could also reflect actual teaching and learning activities more closely (for instance by reflecting the reality that teaching and assessment activities do not always belong to just one module, but are sometimes common between modules). This would be particularly beneficial to PSRBs, who often need to collect data about specific assessments or learning and teaching methodologies as part of their accreditation activity. The project had to balance justified feedback about these limitations of the model against the expressed desire for simplicity. In view of the fact that teaching, learning and assessment data are not likely to be critical to the first use of the model in Data Futures, it was decided to accept compromises in this area of the model.

The overall shape of the HEDIIP data model remains close to the model originally published by the project as a straw man. In particular the approach to defining curriculum data and linking people to the curriculum they are undertaking remains fundamentally unchanged. Beyond the key areas identified above, there has also been very substantial development of this model, however, and significant changes to nomenclature.

In addition, the straw man model contained no detail about attributes. Development of the attribute detail showed that the project's initial view (formed by roughly allocating existing HESA fields to places where they seemed to fit) required very substantial development. The feedback received at the level of individual attributes was extensive and is difficult to summarise adequately, but it was immensely helpful in identifying a wide range of detailed issues. A few examples include: the approach to collecting data about the proportion of teaching undertaken in the Celtic languages (where the project's initial assumptions were too simplistic), the approach to specifying subjects of specialism in initial teacher education and training (where the project's initial assumptions were too respectful of current practice and failed to secure the full level of harmonisation across UK countries) and the reporting of cost-centres for PGR students (where the project was informed that 'dummy modules' in the current HESA Student are problematic for a number of providers, and was able to develop an alternative approach).

Next Steps

The project was a necessary enabling step, but development of a data model does not produce benefits until that model is put into use. There are three follow-on actions:

- Governance of the model;
- Implementation of the model by HESA in Data Futures. A project to undertake this work is in hand;
- Continued development of the model by the Higher Education Data Landscape Steering Group.

1. Governance

Governance of the model is the responsibility of the Higher Education Data Landscape Steering Group. The Group must define a process for exercising that governance. Whilst it is clearly desirable that the model should be as stable as possible, feedback from the ISB has identified the importance of a governance model that can reach decisions rapidly in order to meet the project deadlines of users of the model (including both HESA and others). The alternative will be that users will modify the model *ad hoc* for their own purposes. The Higher Education Data Landscape Steering Group will gain substantial benefit if the strong engagement of stakeholders throughout the project can be maintained.

2. Data Futures

A project to define a data collection process and schedule in Data Futures using the data language has already been scoped. This initial use of the data language will concentrate on replacing the key existing data returns required by HESA Public Purpose customers, especially the HESA Student, AP Student and ITT returns. The project has undertaken analysis of the existing HESA returns and has a high level of confidence that the entities and attributes defined will support these requirements. However, undertaking this further work is likely to involve HESA in extensive and detailed analysis of PD04 (the Data Dictionary) and is likely to identify improvements that are required in proposed coding frames. HESA will work closely with the Higher Education Data Landscape Steering Group to ensure that these requirements are identified and implemented in a timely fashion. HESA will

also need to maintain strong engagement with stakeholders, especially HEPs, which have a strong interest in this collection work.

3. Development

The model requires further development in order to meet the needs of further groups of stakeholders. The Higher Education Data Landscape Steering Group should allocate resources and agree priorities for this development work in light of the expected value of different areas of development. Again, this work will require strong engagement from stakeholders both to ensure that the right priorities are identified for development, and to ensure the best possible outcome of the development work itself.

Appendix A: Stakeholders Engaged

The following stakeholders were directly engaged through project meetings, telephone meetings or workshops. Stakeholders who were engaged through conference sessions or purely online are not included.

A1. Higher Education Providers

Aberystwyth University
University of the Arts London
Bangor University
Bath Spa University
Birkbeck, University of London
The University of Birmingham
Birmingham City University
Bishop Grosseteste University
The University of Bradford
The University of Bristol
The University of Cambridge
Canterbury Christ Church University
Cardiff Metropolitan University
Cardiff University
The University of Central Lancashire
The City University
Coventry University
Cranfield University
University for the Creative Arts
University of Cumbria
De Montfort University
University of Derby
University of Durham
The University of East Anglia
The University of East London
The University of Edinburgh
Edinburgh Napier University
The University of Essex
University College of Estate Management
The University of Glasgow
Glasgow Caledonian University
Goldsmiths, University of London
The University of Greenwich
GSM London
Harper Adams University
Heriot-Watt University
University of Hertfordshire
The University of Huddersfield
The University of Hull
Imperial College of Science, Technology and
Medicine
Kaplan Open Learning, University of Essex
The University of Keele
The University of Kent
King's College London
Kingston University
The University of Lancaster
The University of Law
The University of Leeds
Leeds Beckett University
Leeds College of Art
The University of Leicester
The University of Lincoln
University of London
London Metropolitan University
London South Bank University
Loughborough University
The University of Manchester
The Manchester Metropolitan University
University of Newcastle-upon-Tyne
The University of Northampton
University of Northumbria at Newcastle
University of Nottingham
The Nottingham Trent University
The Open University
The University of Oxford
Oxford Brookes University
University of Plymouth
The University of Portsmouth
Queen Margaret University, Edinburgh
Queen Mary University of London
Ravensbourne
The University of Reading
The Robert Gordon University
Roehampton University
The Royal Central School of Speech & Drama
Royal College of Music
Royal Northern College of Music
The Royal Veterinary College
St Mary's University, Twickenham
The University of Sheffield
The University of Southampton
Southampton Solent University
The University of Stirling
The University of Strathclyde
The University of Sunderland
The University of Surrey
The University of Sussex
University of Ulster
University College London
University of Wales Trinity Saint David
University of the West of England, Bristol
The University of Wolverhampton
The University of York

A2 Data Collectors and Other Stakeholders

Architects Registration Board
The British Psychological Society
Department for Business, Innovation and Skills
Creative Skillset
General Medical Council
Health Education England
Higher Education Funding Council for England
Higher Education Funding Council for Wales
Information Standards Board for Educations, Skills
and Children's Services
Jisc
National College for Teaching & Leadership
NHS Business Services Authority
NHS Education for Scotland
NHS Wales
The Nursing & Midwifery Council
The Office for Fair Access
The Quality Assurance Agency for Higher Education
Scottish Government
Scottish Funding Council
Skills Funding Agency
Student Awards Agency Scotland
Student Loans Company
UCAS
UK Research Councils
Welsh Government