Higher Education Data & Information Improvement Programme

End of Programme Report

July 2016

Foreword

With data and information playing an ever increasing role in the policy and processes of higher education, the establishment of the HEDIIP programme in 2013 was both timely and necessary. Duplication, inefficiencies and a lack of oversight led to data being more of a burden than an asset and a huge challenge for those who sought to understand the sector through published information.

The HEDIIP programme has worked hard to define a model for the information landscape that will reduce burden on HE providers and improve the quality, timeliness and accessibility of data and information about HE. Colleagues from across the landscape have come together to share perspectives and build a greater understanding of the problems we face and the solutions that are required. Without exception they have contributed to the programme in a way that is positive, open and honest; the strength of the outcomes reflects the quality of their contribution.

The information landscape now enters a new phase with the creation of the Data Landscape Steering Group in August 2016 which will provide oversight and leadership and with the development of a transformed approach to data collections led by HESA through the Data Futures programme. The end of the HEDIIP programme marks a major milestone in this endeavour; we have a clear vision for the future and a solid understanding of the journey ahead. Now we must work together to turn this vision into reality.

Professor Neil Gorman

Chair of the HEDIIP Programme Board

Andy Youell

Programme Director

Background to the programme

The higher education information landscape contains a myriad of data flows that support admissions, funding, policy and a range of information outputs for students and others with an interest in the sector. The volume and complexity of data flows has grown exponentially in recent decades and a lack of coordination between data collectors and HE providers has resulted in a landscape that contains high levels of unnecessary duplication and complexity, a lack of coherence in published information and a broader sense that things are out of control.

The 2011 BIS HE White Paper called for the information landscape to be redesigned in order to reduce the burden on HE providers and improve the quality, timeliness and accessibility of data and information about HE[1]. The Regulatory Partnership Group commissioned a study to unpack the issues raised and recommend a way forward[2] and in March 2013 took the decision to establish the Higher Education Data & Information Improvement Programme (HEDIIP) at HESA.

Although the original call to redesign the information landscape was made in the White Paper for England, the challenges and opportunities it presents mostly apply UK-wide. Throughout this journey the work has been undertaken on a UK-wide basis and when HEDIIP was established in 2013 it was formally defined as a UK-wide programme with proportionate funding contributions from the former Department for Employment and Learning (NI)[3], the Higher Education Funding Council for England, the Higher Education Funding Council for Wales and the Scottish Funding Councils.

Programme structure

HEDIIP has been overseen by a Programme Board made up of nominations from three broad areas: HE funders and regulators, HE service providers and HE providers and their students. The programme has also worked with a larger Advisory Panel made up of colleagues from a broad range of HE organisations and HE professional groups and bodies. A small Programme Office, based at HESA, has coordinated the planning and oversight of individual projects. A list of the Programme Board and Advisory Panel members can be found at www.hediip.ac.uk .

---

[1] Students at the Heart of the System, BIS (2011); paragraph 6.22
[2] A Pathway to Reform, RPG (2012)
[3] Responsibility for Higher Education in Northern Ireland moved from the former Department for Employment and Learning (DEL(NI)) to the newly created Department for the Economy (DfE(NI))in May 2016

A vision for the new landscape

HEDIIP has created an architecture for the information landscape that is based on four key elements, summarized in the diagram below.

Collective governance	
Standardisation	Rationalisation
Improved capabilities	
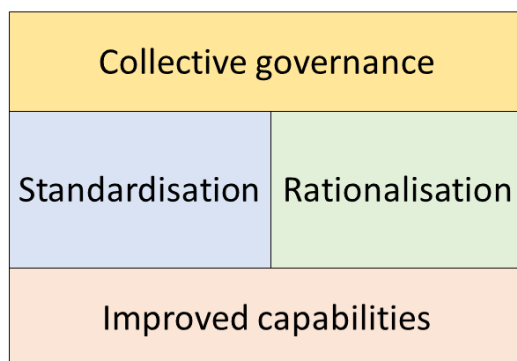
The detailed proposals for the collective governance of the information landscape have been discussed with key stakeholders and we are now moving forward to establish the HE Data Landscape Steering Group at HESA. The core objectives of the group will be a reduction in burden for HE providers and an improvement in the quality, timeliness and accessibility of data and information for all stakeholders.

It will do this by providing oversight and leadership for the information landscape and providing a forum in which data collectors and HE providers can come together to improve the collection, sharing and handling of data. Where appropriate it will challenge and make recommendations about data collection and processing issues to HE providers, data collectors and other bodies across the sector.

The group will oversee the management and adoption of a standard HE data language and promote good practice in the collection and handling of data.

HEDIIP projects

Inventory of HE data collections

One of the first projects HEDIIP undertook was to build an inventory of HE data collections. Previous studies have shed some light on the extent of HE data collections though there has never been an attempt to build a definitive list of collections and collectors. The HEDIIP inventory was created in 2013 and was published on the [HEDIIP website](). It has proved a valuable reference source for the programme as well as for HE providers who have used it to cross check against their own external reporting requirements.

The inventory underwent a major review and update in 2016 and is being handed over to HESA who will continue to maintain and publish the inventory and use it to support the Data Landscape Steering Group.

The inventory has helped HE providers build a better understanding of their external reporting obligations and has proved an invaluable resource in helping the HEDIIP programme understand the true extent of data collections across the landscape. The updated inventory has better coverage and more detailed information about individual collections, and has improved sorting and searching capabilities.

Review of Data Collections

Although the inventory of data collections lists a few of the key attributes of the various HE data collections, it did not contain the detail of what is collected, or information about why and how data is collected. This 2014 project set out to gain a greater understanding of the range of student data collections in order to establish the potential for the standardisation and rationalisation of these collections.

The project conducted interviews with 24 data collecting organisations and received survey responses from a further 19 organisations. The project identified the main drivers for data collection and analysed the specifications of the various data collections. It concluded that there is a core of student data that is common across a broad range of data collections and therefore the scope for standardisation and rationalisation is significant.

The review provided a rich understanding of the current range of student data collections and identified some of the key challenges and opportunities that HEDIIP faced. It laid a solid foundation on which the conversations about the new landscape architecture could take place.

New Landscape

The New Landscape project defined the architecture for the information landscape through extensive engagement with stakeholders and the testing and refinement of principles and models that would gain widespread support. The project built on the results of the Review of Data Collections project and on the studies that were conducted under the Regulatory Partnership Group[4] in 2011 and 2012. The New Landscape report was published in May 2015 and subsequent programme activities have focussed on establishing the key elements of the architecture in a sustainable manner.

The delivery of the New Landscape project marked the first time there has ever been a coherent and broadly-agreed vision about how the HE information landscape should operate and how that vision could be realised. It set the information landscape journey off on a new trajectory of change.

HE Data Language

The Data Language project analysed a number of existing HE student data collections to build a logical data model that can underpin a range of data collections in the future. The project worked extensively with HE providers in order to define and agree the key concepts around students and curriculum and with data collectors to ensure that the model can meet the range of requirements that drive data collections.

The project has delivered a comprehensive logical data model along with a vocabulary of terms and a data dictionary of field definitions. These elements of the data language will be managed by HESA, under the auspices of the HE Data Landscape Steering Group.

Although it might appear an obvious move, the development of a data language that describes concepts and terms in a way that meets the needs of many organisations has been one of the most significant and technically challenging projects undertaken by HEDIIP. The data language is robust yet flexible and has successfully delivered a major technical advance while carrying the support of HEDIIP stakeholders.

---

[4] http://www.hefce.ac.uk/about/unicoll/other/RPG/

Subject Coding

The studies that preceded HEDIIP identified subject coding as one of the key elements of the information landscape that required significant attention and HEDIIP pursued this strand of work from the outset. The existing Joint Academic Coding System (JACS) did not meet the needs of many stakeholders and it had run out of space for further development. HEDIIP established a detailed set of requirements and defined a coding system - the Higher Education Classification of Subjects (HECoS) - that meets these requirements and that is not constrained for future development.

This strand of work then focussed on implementation issues, defining a Common Aggregation Hierarchy for standard analyses, mappings from JACS to HECoS and setting out recommendations for the on-going management and governance of the system.

HECoS will be published and managed by HESA, on behalf of the sector, and will be overseen by the Data Landscape Steering Group.

The development of a subject coding system that can meet the needs of a broad range of stakeholders clears one of the most significant blockages to the adoption of a common data language across the landscape. The JACS system – which we think can be traced back to the early 1970s – was in desperate need of replacing.

Adoption of the Unique Learner Number

The lack of a coherent approach to student identifiers in HE has been the source of much duplication and a barrier to data sharing and more joined up services. Previous studies have highlighted the number of different student identifiers that are used across the system and recommended the adoption of the Unique Learner Number (ULN) which is gaining traction in the schools and FE sectors in England, Wales and Northern Ireland.

HEDIIP undertook a major study to define a roadmap for ULN adoption in HE and define the benefits that should flow at the various stages on this journey. The roadmap sets out three broad phases of adoption: improve the harvesting of ULNs for young people entering the system, expand the coverage of the ULN to new groups and drive coverage towards 100%. HEDIIP has been working with key stakeholders to drive forward the first phase of this journey and this will continue under the Data Landscape Steering Group.

HEDIIP has also launched a strand of work focussed specifically on building some kind of link between the ULN and the Scottish Candidate Number which is well embedded in Schools and FE in Scotland. The programme has worked with a broad group including the Scottish Qualifications Authority, the Student Awards Agency Scotland, NHS Education Scotland and the Scottish Funding Councils and government in order to define a model of interoperability

that will facilitate the delivery of the ULN benefits across the UK. The Scottish Commission on Widening Access has identified this as one of the key elements of data infrastructure required to monitor and improve access to higher education and its final report includes a recommendation to carry forward the unique learner number development in Scotland.

The adoption of the ULN in higher education will continue to be monitored and steered by the Data Landscape Steering Group.

Data Capability

The New Landscape report identified data capability as one of the main building blocks of the architecture for the information landscape and HEDIIP met with a number of the sectors professional bodies to scope and define a strand of work that would help HE providers understand and improve their standards of data management and governance. The Data Capability project developed a toolkit which included an on-line self-assessment of capability maturity. This model was tested with delegates at the 2015 Student Records Officers Conference and was fully launched in June 2015. The project then embarked on a range of activities to drive adoption of the model and build momentum for this agenda across the sector.

By the end of 2015 over 100 institutions had undertaken the self-assessment process and HEDIIP published the [analysis of the sectors data capability](#) in January 2016. The project has continued to support the sector in its adoption through workshops and other direct engagement, further development of the toolkit and by building a community of practice for data management professionals in the sector. The data capability agenda now has a high profile and some momentum across the sector.

This agenda will continue to be supported by HESA through the development and delivery of training in this area and by promoting and supporting this agenda through the Data Futures programme. The Data Landscape Steering Group will also monitor the sectors data capability as it supports the broader agendas of rationalisation of data collections and the reduction of burden.

With the ever-increasing power of information technology, and the ever-increasing desire to drive more value from data, the sectors ability to properly manage and oversee its data resources is more important, and challenging, than ever. The Data Capability project has placed this issue firmly on the agenda for many providers and has established momentum for this agenda within the community of data professionals across the sector.

Lessons learned

The HEDIIP programme has sought to change the approach to data collection and management across a complex landscape that contains many organisations and a plethora of other policy and operational initiatives. In doing this it has sought to reposition the conversation about data from being one of burden and difficulty to being one that understands data as a strategic issue that enables the development and monitoring of policy and the delivery of high quality services. In order to succeed HEDIIP has had to demonstrate progress at a technical level and shift the culture and mind-set around data to a more positive place. The main lessons and observations from this work are set out below.

HEDIIP was primarily a change programme, not a data programme

Although the title might suggest otherwise, HEDIIP was fundamentally a change programme whose aim was to change a culture and mind-set through the broad support of a new model of working across the sector. Although the objectives were framed around data and information flows, success is only realised through the broad buy-in to a model that requires stakeholders across the landscape to change both processes and behaviours. The programme needed to strike a balance between progress and credibility around the specific, often technical issues of data and information while also operating successfully at a strategic and policy level. The long-term success of the programme will depend on the extent to which it has been able to succeed at both levels.

Communication and engagement is key

With such a large range of stakeholders, communication and engagement was always going to be a significant challenge. The programme devoted a lot of effort to understanding and engaging with stakeholders. The programme achieved a high level of visibility amongst many of its key communities, especially those that are more focussed on data and information issues. The creation of a strong brand identity, and the use of a variety of communication channels were significant factors in the success of communications.

Where the programme faltered in communications it was often due to a lack of clarity around some of the more technically-focussed work and a failure to contextualise and explain messages thoroughly.

Data needs to be everybody's issue

Data and information has grown to be a significant area of specialisation in HE; there is a strong community of data specialists within providers and the sector bodies that collect and process data. Although this community has been supportive of the work of HEDIIP, the broader changes that are required to achieve the programmes aims mean that the key messages from HEDIIP need to be absorbed by stakeholders who might not normally consider themselves as having a direct interest in data issues. Although it is entirely appropriate for the data specialists to lead on the detail of technical issues, issues around the strategy, governance and management of data – both within organisations and across the landscape – are issues that all stakeholders should engage with, to some extent at least.

The HE sector is at its best when it works collaboratively

Although HEDIIP has come at a time of massive change for HE, the way in which organisations across the landscape have come together to share and collaborate and to develop collective solutions shows that HE remains a sector that understands the importance of collaborative working to develop solutions for complex problems. Throughout the programme stakeholders have engaged with an openness and a positivity that has allowed genuinely game-changing solutions to develop. The information landscape journey does not finish with the end of HEDIIP and this spirit of openness and cooperation will be needed as the various elements of the HEDIIP architecture move into implementation.

Programme objectives

The RPG study[5] that recommended the creation of HEDIIP set out the six key issues faced by the information landscape and the desired outcomes from the programme that address these issues. This analysis of issues and desired objectives has informed the creation and planning of the programme. The table below sets out what HEDIIP has achieved against these objectives.

Issue	Desired objective	Outcome
1. A lack of visibility of data collections conducted and data available in the sector	a. Better understanding by institutions of how they can use data for their own benefit	The Data Capability project has been highly successful in engaging HE providers in the data capability improvement agenda. The Toolkits produced by the project enables HE providers to link their strategic objectives, and use of data directly to data capabilities and develop action plans to realise tangible benefits. Over 100 HE providers have assessed their data management practices and there are currently 32 projects in-flight across the sector, using HEDIIP Toolkits to improve data capability. The sector has demonstrated the value it places in this initiative which will continue with the support of HESA.

[5] http://www.hediip.ac.uk/wp-content/uploads/RPG_HEDIIP_proposal_2013-03.pdf

Issue	Desired objective	Outcome
	b. Better understanding by data collectors of data available in order to reduce duplicated collections	HEDIIP has constructed and published an inventory of data collections being carried out across the sector which is openly available for data collectors and providers to interrogate. The value of the inventory was shown by a review carried out in 2016. HEDIIP has defined and, with the support of the Programme Board, set up the HE Data Landscape Steering Group and defined a set of good practice data principles. The Steering Group and principles will guide the adoption of good data collection practice across the sector. The HEDIIP vision, described above, has been a key driver in the HESA Data Futures programme. The programme with input from HEDIIP's Advisory Panel has developed a collection model that will support the rationalisation of collections. The programme is currently underway and is due to be implemented for the 2019/20 Academic Year.
2. Poor understanding of the purpose of data collections	a. Clarity for institutions on the purpose and value of data collections	HEDIIP's inventory of data collections has established, for the first time, a list of collections taking place across the sector. The inventory is openly available and has been used by some HE providers to help understand where their data is being used. The Student Data Collection Review clearly identified the purpose and value of student data collections being carried out across the sector. The findings from this work were published in 2014 and have been used to inform the development of the HEDIIP vision, described above.

Issue	Desired objective	Outcome
	b. Identification of any non-value added data collections	HEDIIP has not identified any non-value added data collections that could be eliminated as such, however through the Student Data Collection Review, it has identified that many data collectors particularly PSRBs are collecting a similar core set of data, hence the scope to which collection activity across the sector could be rationalised. Rationalisation of data collection is a being taken forward in the Data Futures programme agenda and particularly by the development of a collection model that will support the rationalisation of collections. Other sources of non-value added activity identified by HEDIIP include: • The lack of a common data language and the need to translate data for different purposes (see 3. below). • Inefficiencies in data management practices due to relatively low data capability (see 1. a. above).
3. No common data language	a. Comparable, better quality and more accessible data sets	The HEDIIP Data Language project has engaged with a wide range of stakeholders across the HE landscape to establish a Logical Data Model with a supporting data dictionary. The model and dictionary have been accepted by the stakeholders and will be taken forward by the Data Futures programme. The Data Futures programme will develop a Physical Data Model which will be openly available and will support the rationalisation of data collection activity.
	b. Reduced duplication of data collections through data sharing (facilitated by common data definitions)	The HE Data Landscape Steering Group and Advisory Panel, set up by HEDIIP, will drive the adoption of good practice data principles across the landscape. These principles underpin data sharing, managing data as an asset, common vocabulary and interoperability.

Issue	Desired objective	Outcome
4. Weak operational co-ordination between organisations including data sharing	a. Reduced duplication of data collections through more efficient data collection processes and data sharing	As described earlier HEDIIP's vision has shaped HESA's Data Futures programme. The HEDIIP architecture for landscape proposed a 'transformed' HESA collecting data from HE providers and disseminating data to data collectors, who currently collect data directly from providers. This architecture, supported by a common data language means HE providers do not have to translate their data for different purposes, is designed to lead to more efficient collection processes and promote data sharing rather than silo collection activity.
	b. Improved timeliness of data	The HEDIIP architecture vision for a 'transformed' HESA identified the requirement for in year data collection. This concept was the subject of a HESA consultation at the end of 2015 and gained widespread support from the sector. The Data Futures programme is developing the concept further by initiating the Collection Design project which will also use the Logical Data Model developed by the HEDIIP Data Language project.
5. Absence of a sector wide strategy on data	a. Sector wide strategies to promote consistency across all organisations	As previously described HEDIIP has achieved this outcome by gaining sector support to the architecture vision and data principles. The HE Data Landscape Steering Group and Advisory Panel will provide the collective governance to promote consistency. The New Subject Coding System project has developed the HE Classification of Subjects (HECoS) scheme to replace JACS. The adoption of this scheme is a sector wide strategy developed by HEDIIP that is being taken forward by HESA and UCAS with oversight from the HE Data Landscape Steering Group. Similarly HEDIIP has facilitated the adoption of the ULN as a sector strategy which will be overseen by the HE Data Landscape Steering Group supported by HESA, UCAS and the Skills Funding Agency with interest from HEE, SFC, SQA and NES amongst others.

Issue	Desired objective		Outcome
	b.	Sector wide understanding of the needs and requirements of data users	The collective governance arrangements particularly the HE Data Landscape Advisory Panel will provide a forum for data users particularly at the practitioner level to share understanding and develop the common data language for the benefit of the sector.
6. No mandate or expectations to change behaviours to drive efficiencies	a.	Data collectors to seek most efficient method of collecting data to meet their requirements	The collective governance of the HE Data Landscape Steering Group and Advisory Panel has been established to drive good practice data collection. Building on the HEDIIP vision, Data Futures is in the process of defining a collection model that will enable data collector's requirements to be rolled into HESA by adoption of the common data language.
	b.	HE institutions to adopt most efficient and effective methods of data management and governance	HEDIIPs Data Capability project, as already described, has addressed this objective and gained widespread support across the sector. The momentum of this initiative will be maintained by HESA.

Further information and contacts

Further information about the work of HEDIIP can be found at www.hediip.ac.uk

The HE Data Landscape Steering Group secretariat can be contacted at dlsg@hesa.ac.uk