The Blueprint for a New HE Data Landscape

The Higher Education Data and Information Improvement Programme

Final Report

www.kpmg.co.uk
Foreword

Dear Colleagues,

I am pleased to present the Vision and Blueprint for the UK Higher Education Data Landscape. As the marketisation of the sector has developed, the importance and demand for more timely and accurate data has increased, both internally within Higher Education Providers (HEPs), and externally amongst Data Collectors, regulators and oversight bodies. This has strengthened the need for a change in the efficiency in the administration and collection of data.

The Blueprint for the New Data Landscape provides a strong vision. It both rationalises and simplifies data requirements through the use of standard data sets and data definitions. This creates opportunities for gaining greater value from the use of the technology being exploited by the sector.

These recommendations are also well timed as there are concurrent change programmes being undertaken by some other Data Collectors (e.g. HESA and Health Education England). This provides the welcome opportunity for the alignment of the New Landscape to a range of other important projects to provide a greater momentum for change.

For Higher Education Providers the New Landscape will mean:

- Simplification in the range of data requested by Funders and oversight bodies;
- Greater cohesion between the Data Collectors’ requirements;
- A reduction in the submission of similar data to multiple Data Collectors; and
- Improved data quality and capability that will provide benefits internally and externally.

For the users of data, the New Landscape will provide:

- Increased understanding of pre-existing data and how it can be used to a greater extent;
- More timely and accessible data such that separate data collections are not required;
- Improved data quality; and
- Reduced burden placed on the HE sector.

The New Landscape is challenging, but I am confident that it is both right and achievable. I and colleagues are conscious of the urgency with which some of these issues need resolution. For this reason the Landscape identifies benefits that can begin to be realised within 12 months, separate from other actions that form part of system and transformational changes.

The drivers for the administration of data in the sector are complex as a result of the number of independent Data Collectors and interested parties that interact with HEPs. However, this project has identified a common understanding of the issues, a willingness and commitment to overcome these difficulties to do what is right and long overdue. Whilst recognising that certain Data Collectors have competing priorities, it will be key that all of the data collectors continue to collaborate and that Higher Education Providers fully engage in the suggested improvements.

They key now is to carry the momentum forward to ensure that timely actions are delivered and the benefits realised. We all have a part to play in implementing the New Landscape and I look forward to working with you on this journey.

Yours sincerely

Professor Neil Gorman
Chair of HEDIIP Programme Board
Important notice

This Report has been prepared in accordance with the terms of our agreement with the Higher Education Statistics Agency Limited (“HESA”) dated 7 November 2014 (the “Services Contract”). Accordingly, save as set out in the Services Contract, we have (i) not verified the reliability or accuracy of any information obtained in the course of our work, or (ii) taken into account the interests, needs or circumstances of other parties (whom might read this Report).

This Report is for the benefit of HESA only, and is not suitable to be relied on by any other party wishing to acquire rights against KPMG LLP for any purpose or in any context. Any party other than HESA that obtains a copy of this Report and chooses to rely on this Report (or any part of it) does so at its own risk. To the fullest extent permitted by law, KPMG LLP does not assume any responsibility and will not accept any liability in respect of this Report to any party other than and HESA.
Executive Summary

The evolution of the sector since 2011 has led to the importance of data increasing, across institutions, students, government, funders and other key stakeholders. However, the evolution of data collections has created a lack of coherence and unnecessary complexity that needs to be addressed. This report provides a roadmap for making a stepped change in rationalising and simplifying data requirements and improving how data is collected and administered. The key now is that stakeholders remain committed to supporting the new Blueprint and do not allow timescales to drift. The task is not insignificant, but sustained and determined pursuit of the Blueprint will deliver benefits for all.

There has been a considerable body of work undertaken to develop the New Data Landscape and this has led to an extensive report. This section provides a summary report that details the key findings, proposals and benefits that the New Data Landscape will deliver. Further detail is provided in the remainder of the overall report.

1.1 The need for a New Landscape

1.1.1 HE White Paper
In June 2011 the Department for Business, Innovation and Skills (BIS) published its Higher Education White Paper, ‘Students at the Heart of the System’. This proposed specific improvements to the higher education data and information landscape by asking that: ...work be undertaken to redesign the information landscape for higher education in order to arrive at a new system that:

- Meets the needs of a wider group of users;
- Reduces the duplication that currently exists; and
- Results in timelier and more relevant data.

The broader changes that the sector is experiencing as a result of the White Paper have led to:

- Increased uncertainty in the income streams for Higher Education Providers (HEPs);
- New business models emerging in the sector;
- Further diversification in the activities undertaken by HEPs; and
- A sharper focus on the efficiency and effectiveness of HEPs’ operations.

1.1.2 Increased need for more timely and granular data within HEPs
These issues have created a shift in the importance of data to HEPs. But, the increased need and thirst for data has also heightened the focus on the issues that currently exist in the administration and collection of data across the sector:

- There are 97 different Data Collectors and over 520 different data returns that HEPs may be required to submit;
- There is no requirement for the Data Collectors to collaborate and coordinate their data needs from HEPs;
- There is duplication in the data requested by the different Data Collectors, creating unnecessary burdens for HEPs;
- The plethora of data requests has led to multiple data definitions for the same or similar data field that creates complexity for HEPs without any real gain;
- HEPs have varying levels of sophistication and capability in their data processes; and
- The need for certain data collections and the use of the data is not commonly understood by HEPs.
These matters contribute to a view that the burden of data collection is greater than it should be. Our work has identified that principally four things influence the perception of burden: 1) the physical volumes of data collected; 2) the complexity and multiplicity of data definitions; 3) the methods and number of organisations collecting data from HEPs; and 4) the extent to which the data collected is used.

1.1.3 Existing change programmes within HEPs and Data Collectors

A number of the Data Collectors and HEPs have either announced or are undertaking change programmes to improve their approach to collecting and processing data. For example:

- HESA has commenced a ‘Change in Approach to the Collection of HE Data Programme’ (CACHED) project that aims to ‘achieve relevant, more timely data, using systems that are technically future-proofed, flexible and that interface with information management systems across the sector’.

- Health Education England is undertaking a project to harmonise its data requirements across the Local Education and Training Boards.

We also found that a number of HEPs are either implementing new student records systems or are substantially improving their existing system and associated processes.

There is a groundswell of support for rationalising and improving the current data landscape. The willingness for change extends to a number of Data Collectors who are eager to see improvements to the way data is collected and administered. The existing change programmes provide a uniquely timed opportunity to increase the pace of change and deliver benefits incrementally. It is important that this opportunity is seized.

1.2 Project scope

This project has investigated and developed a Blueprint for how data should be collected by key stakeholders across the sector. A series of data principles have been developed and a governance structure that can be implemented to enhance the collective oversight of a Standard Dataset has been proposed. An overarching implementation plan has been developed along with outline costings for implementing the New Landscape. The full description of our work is provided in section two, but in summary it has encompassed:

- Close working with the HEDIIP Programme Management Office;
- Over 50 interactions with Higher Education Providers (HEPs), key stakeholders and professional bodies in the sector;
- Three stakeholder workshops;
- Survey with the sector;
- Desk based review of previous and other current HEDIIP projects; and
- Research of data landscapes in other sectors, both in the UK and internationally.

The project has focused on the collection of student data, but it is the case that the Blueprint outlined can be applied to the other forms of data collection. In scoping the data landscape project a number of assumptions were made and these are detailed in Chapter two. This project has not included the development of the standard data set, nor has it developed the financial and business case for the proposed changes. These will be necessary ‘next steps’ and as such, they are incorporated in the implementation plan provided.

A further parameter for this project has been that the Blueprint should be achievable. As such, this review was given a remit to identify a New Data Landscape within the confines of the current regulatory and organisation structures that exist.

---

1 https://www.hesa.ac.uk/component/content/article?id=3441
1.3 Key issues identified

From the stakeholder interactions, document review and the sector survey the key issues with the current landscape have been identified. The key issues, a number of which were already known, have been categorised into a small number of discrete areas, which are summarised below. A number of issues and complexities arise from the number of different organisations and government administrations that impact on HE in the UK. The diagram below illustrates this complexity. Full details of the issues identified are provided in Chapter four.

Simplified diagram of the current data landscape

The specific issues identified are summarised as follows:

- **There is no overarching governance of the Data Collectors** – Notwithstanding the various professional groups that exist in the sector, there is no overarching body or group that enables collective working across all of the Data Collectors. This means there is no group or body to whom institutions can outline issues or opportunities for improvement. It also exacerbates the problem of each collector operating to serve their own needs and interests, rather than taking a sector approach to the issue. A recent example cited was the decision to create a course directory without considering if/how the data could be obtained from a Data Collector who probably had some similar data.
- Greater alignment of Data Collector needs with HEP internal data needs – Contact with a number of HEPs identified a frustration over an apparent gap in awareness among Data Collectors concerning the difficulties that certain requests can create for HEPs, especially at a time when the level of internal data demands is increasing. There is also a commonly held view that data collections can be rationalised. Greater awareness of the operational implications for HEPs of certain data requests could enable increased alignment between the internal and external data requirements. Data demands should also be rationalised where possible.

- Absence of common agreed data standards – Each Data Collector has developed its own data standards to meet their operational needs. Whilst there is some co-operation between individual collectors, the absence of common data standards contributes to the burden on data providers and reduces the options for data sharing between collectors. Certain issues that have been identified concern the exchange and flow of data between HEPs and Data Collectors. These are summarised as follows:

  - There is not routine sharing of data between Data Collectors – Currently Data Collectors use different mechanisms to collect the data from HEPs. This creates complexity for HEPs in terms of having to service a range of different collection methods. It is also the case that with the exception of certain HESA and UCAS data, the data collected by the Data Collectors does not get shared between collectors to either reduce the collection requirements placed on HEPs or to pre-populate existing data requests.

  - There are multiple Data Collectors to whom HEPs have to provide data – The Data Collectors are not obliged to collaborate, although we found a degree of collaboration does take place. This leads to each collector, understandably, specifying data requirements that fulfil their own needs in isolation to the data that is being collected elsewhere. This has led to a variety of different, but similar, data definitions being used and the collection of data that has already been submitted to another collector. This creates complexity and inefficiency for HEPs in addition to increasing the cost of data collection for the Data Collectors.

In addition to the above representations have been made that there is an inconsistent approach to student views being obtained in respect of the data management strategies of both Data Collectors and HEPs.

### 1.4 What can be learnt from other sectors and the HE systems in other countries

The collection of complex data sets is not an issue that is unique to the UK HE Sector. Therefore to inform the identification of options for the New Landscape we considered what parallels and learning can be taken from other sectors in the UK and HE sectors around the world. Further detail on the research undertaken is provided in Chapter three.

#### 1.4.1 Other UK sectors

Examples were taken from the Health, Retail and Food sectors and reviewed. A common theme in each of the sectors reviewed is the use of a central collection body, the existence of standardised data definitions and the ability for data to be shared through the appropriate use of data sharing agreements.

#### 1.4.2 HE Data systems in other countries

We have undertaken a high level assessment of the approach to the management of student data in China, Australia, Canada and Hong Kong. This identified that each country has its own approach to Higher Education data management, but the consistent theme for the majority of the countries reviewed is that they have centralised data collections through regional or central government bodies.

The United Kingdom appears to have a unique set of circumstances due to history and evolution whereby there is a mix of central data collection agencies that are sector owned and those that form part of Government Departments. The proliferation of Data Collectors is also a factor that appears to be unique to the UK data landscape.
1.5 The building blocks for the New Landscape

Our work has found that there are four areas that will be pivotal to achieving greater cohesion in the data collections and a reduction in the burden, whether they be perceived or otherwise:

### Establishment of a collective governance function and common data principles

The establishment of a governance body that can provide collective governance and stewardship of the Standard Dataset, the Data Standards together with ownership of the Data Principles. It can act as a filter for any additional data demand that collectors wish to introduce and an opportunity for student participation. This should enable creep in data collection requirements to be reduced and enable both cross collector collaboration and a forum where HEPs and Data Collectors can come together to share understanding and resolve issues. The data principles should provide a basis for all collectors consciously working to a common cause.

**Benefits:**
- Enable a joined up approach to data requirements across the sector
- Minimise the scope for duplicate data requests to emerge
- Provide a forum for the sector to raise data collection issues
- Stewardship of common data principles will provide greater cohesion in data administration

### Development of a Standard Dataset with agreed definitions that are used by all key Data Collectors

This will require movement by some collectors on certain field definitions and will need HEPs to understand and agree on rationalising definitions. This will reduce complexity and enable a quick win for HEPs, i.e. benefits that can start to be realised within 12 months.

**Benefits:**
- Reduced complexity in the data systems and processes within HEPs
- Increased data quality

### Change in the data collection model whereby certain collectors will take the ‘standard data’ from a single body

This will avoid HEPs providing the same or very similar data multiple times and often from different parts of the institution. Over time it could also enable enhanced data analytics and intelligence.

**Benefits:**
- Alignment of reporting dates and a reduction in the number of data collection requests
- Enables HEPs to have greater control over its data submissions
- Increased scope for data analytics and business intelligence

### Enhancement of HEPs data maturity and capability

A number of institutions inputting to this review have outlined that they are on a journey with further developing their data capability and information processes. It is suggested that the shift in the nature of the HE sector is necessitating better intelligence and different and timelier information. Therefore together with the actions that Data Collectors can take to improve the data and information landscape, it is equally important that where necessary, HEPs enhance their own internal structures and processes.

**Benefits:**
- Increased understanding of how data is used and why
- Improved data processes to support HEPs information requirements
- Improved data quality
- Improved management information
We have made recommendations for how this can be enabled and the remaining sub sections describe these building blocks in more detail.

1.6 Vision and data principles

The vision for the New Data Landscape is ‘A data and information landscape for Higher Education in the UK that has effective governance and leadership, promotes data standards, rationalises data flows and maximises the value of technology and enables improved data capability’.

To underpin the vision and provide a common basis for all Data Collectors and HEPs to operate the following data principles have been developed:

- **Effective Governance:** Data collection requirements in the stakeholders and Higher Education Providers need collective governance that is free from bias to allow for effective implementation and delivery of a New Data Landscape. This governance should be delivered by representatives from HEPs and Data Collectors from across the sector, and represent sector views, with student views represented by the NUS;
- **Adherence to Principles:** These principles of information management apply to all bodies exchanging data with HE sector organisations;
- **Maximum Benefit to HE Sector as a Whole:** Information management decisions are made to provide maximum benefit to the HE sector as a whole including students;
- **Information Management is Everyone’s Business:** All key stakeholders within the HE Sector participate in information management decisions needed to accomplish business objectives;
- **Compliance with the Law:** HE information management processes comply with all relevant laws, policies, and regulations, including competition law;
- **Data is an Asset:** Data is an asset that has value to the HE Stakeholders and is managed accordingly;
- **Data is Shared:** Users have access to the data necessary to perform their duties or answer their query; therefore, data is shared across Stakeholders, where permissible and appropriate;
- **Data Trustee:** Each data element has an owner and a trustee accountable for data quality;
- **Common vocabulary and Data Definitions:** Data is defined consistently throughout the HE sector, and the definitions are understandable and available to all users;
- **Data Security:** Data is protected from unauthorised use and disclosure. In addition to the traditional aspects of national security classification, this includes, but is not limited to, protection of sensitive and proprietary information;
- **Technology Independence:** Applications are independent of specific technology choices and therefore can operate on a variety of technology platforms;
- **Responsive Change Management:** Changes to the HE information landscape are implemented in a timely manner;
- **Interoperability:** Software and hardware should conform to defined standards that promote interoperability for data, applications, and technology; and
- **Protection of Competitive Position:** Data that is viewed as compromising the competitive position of HEPs will still be collected in year, but will not be available publically, nor to other collectors (save for that data required for them to fulfil their requirements) until it is no longer viewed as commercially sensitive.

The data principles will be owned and governed by the new HE Data Governance Body.

1.7 Governance for the New Landscape

A range of governance models were openly considered during the project. These included the creation of a HE Data Regulator through government legislation, the creation of a Governance Group within a Funding Council or Government Department, through to the creation of a collective governance group. A key barrier that had to be overcome in identifying an acceptable and workable governance model was the variety of
different organisations and legal entities that have a role in collecting or providing data in the sector and the view that there will not be legislation for the HE sector in the near future.

A further complication is that a number of these bodies have statutory powers, so unless there is a governance body that has complementary statutory powers (which has been dismissed as an option), a collector with statutory powers cannot be overruled. During the project however, these bodies have committed to participate and engage with the collective governance model in pursuit of the agreed vision for the New Landscape.

This led to the selection of an ‘independent collective and consensual governance body’ that has equal representation from key Data Collectors and HEPs (including UUK and GuildHE) in addition to having representation from the National Union of Students, in order to represent the students’ interests. It is proposed that the Governance Body, termed the HE Data Governance Body, will not report to any other body or group in the sector in order to be independent, but will be physically based at HESA. This was a unanimously agreed proposal. A suggested structure for the group is as follows:

The model will rely on commitment from all key stakeholders to work collectively in accordance with the data principles and to this end, any Data Collector that is not willing to operate in this manner should not be permitted membership of the Governance Body. It is suggested that BIS are asked to support the model to encourage adherence to the data principles. The power of the sector via its representation is a critical part also. This will enable the implications of proposals and/or disagreement from certain Data Collectors to be challenged from a HEP and Student’s perspective. The group will provide a route through which HEPs can make representations, which will be important in terms of maintaining an appropriate landscape in the long term.

The Professional Statutory Reporting Bodies (PSRBs) are many in number and as such it is not feasible that all can be equally represented on the Management Board. The structure does, however, enable a PSRB group to be established as one of the ‘Consultative Groups’ which will for the first time provide an opportunity for PSRBs to discuss common data requirements. This group will initially be small, but will grow as the number of PSRBs joining the New Landscape grows.

The Governance structure will govern the Data Principles and the Standard Dataset and associated definitions in addition to the rationalisation and future development of data requirements across the sector. It could also provide good practice advice to support the development and understanding of data administration.

Further detail of the governance proposals and assessment is provided in Chapter six.
1.8 Preferred option for the New Landscape

To identify options for evaluation a modern ‘ideal’ model for the administration of data collection was identified (figure 12). This was used to identify three alternative options for the data flows in the landscape in addition to a ‘do nothing’ option.

The ‘do nothing’ option was quickly dismissed due to the shortcomings already outlined and of the three options proposed, the model below was identified as the preferred, ambitious, but deliverable option:

The key features of this model are as follows:

- To the greatest extent possible, data collections are centralised in the ‘transformed HESA’ and other Data Collectors obtain their data from HESA, via appropriate agreements. HESA will collect the Standard Dataset in the first instance, but there is scope for non-standard data to be collected on behalf of the Data Collectors. Process changes will focus on HESA becoming the single collector of the Standard HE dataset once the UCAS admissions process is complete, and other Data Collectors collecting the standard HE data from HESA.
- The operation of SLC and UCAS processes and the data flows required by these processes remains unchanged.
- The introduction of the Unique Learner Number is assumed in the model above such that data forms can be pre-populated and a student’s history understood.
- The volume of different and duplicative data returns that have to be submitted by HEPs is reduced.
- Data returns will become in-year, i.e. data will be submitted at different points in the year to create incremental data returns, rather than a single submission of data after the year end.
- The timing of data returns will be harmonised and lead to a common reporting date to HESA, such that the timing with which different Data Collectors need the data can be overcome.
- Represented in the diagram by the figures, the relationships will still exist between HEPs and the other Data Collectors and if ad-hoc or one-off survey collections are required, these could still be led by individual Data Collectors.
- The relationship between the student and the HEP is not affected by this model.
Over time the model could provide scope for two-way data exchanges between collectors (subject to adherence to relevant data sharing agreements and competition restrictions), which would enhance business intelligence that could be provided to the sector.

Further details on the operation of the new data model is provided in Chapter seven.

1.9 Impacts and benefits of the New Landscape on key stakeholders

The New Landscape will provide benefits for a range of stakeholders. These benefits and the impacts are summarised below:

1.9.1 Students

The New Landscape will provide direct benefits to students in terms of their experience of interacting with the data landscape. It will also provide indirect benefits as a result of HEPs having improved data that can contribute to the enhanced services to students in addition to freeing up resource for HEPs to repurpose in pursuit of its strategic aims. A further benefit is that a simplified data landscape makes it easier for students to understand, if they so wish, how their data moves around the system. Representations were made during the project that there is an increasing desire for students to be reassured about how their data is protected, used and by whom.

The implementation of the New Landscape and the adoption of the Principles will provide opportunities across all Data Collectors and HEPs for more consistent student participation within the data management discussions.

1.9.2 HEPs

The key changes for HEPs will be that their systems will need some amendment to enable the capture of different and/or reduced data fields, but report these centrally ‘in-year’. Some technological change will also be needed to deliver the updated ‘in-year’ student returns to HESA that will fulfil the needs of multiple Data Collectors. Interactions with a number of software suppliers of student systems has identified that the changes proposed are feasible.

HEPs will benefit through there being some rationalisation in the volume of data that is collected. They will also see a reduction in the complexity of the data collections as a result of there being common data definitions. The proposed changes should enable improved control over data as a result of more data returns being submitted from a central function ‘in year’. Data quality should also be improved where a single version of data is used by multiple collectors for multiple purposes. Finally there will be a formal body (the HE Data Landscape Management Board) to whom representations and suggestions can be made regarding the data landscape and issues impacting either positively or negatively on it.

1.9.3 Data Collectors

Data Collectors will need to amend their systems to varying degrees. Some collectors will need to amend their systems to enable a data exchange with HESA in order to reduce the data collected directly from HEPs. HESA will require the greatest level of change to enable in-year collections, enable data exchange with the other Data Collectors and administer a redesign of the student data return. UCAS and SLC processes will remain unchanged in this model.

The benefits for a number of Data Collectors will be increased data quality, increased automation in the collection of data and the ability to reduce running costs or repurpose resources to greater effect. Depending on the nature of the data sharing agreements put in place and subject to the availability of data in line with competition restrictions, the ability for greater business intelligence is increased.

1.10 Outline costs of the New Landscape

Implementation of the New Landscape will require investment by Data Collectors, HEPs and HESA. It is, however, important to note that a number of organisations are already investing in their data processes and therefore, to some degree, certain costs are already being incurred. There are costs that will be incurred in
Alongside the costs necessary to implement the New Landscape, there are also some substantial savings that can be realised. It will however be each organisation’s choice as to whether savings are ‘cashed’ or whether they are redeployed or repurposed to support other activities.

The scope of this document has not been to develop the financial case for the implementation and operation of the landscape. We have however identified some indicative, high-level estimates of the costs and savings and have found that it should deliver recurrent savings.

The survey data suggested that institutions could generate savings of around £0.5m p/a, but in our cost and saving estimates we have taken a prudent approach and reduced this to £80k p/a. On this basis savings in excess of £12.8m could be realised p/a, were all changes fully operational. We have also estimated that savings across the data collections could be in the region of £1m p/a.

Further detail on the costings is provided in Chapter nine.

1.11 High level timeline for implementing the New Data Landscape

We have provided a detailed implementation plan in Chapter eight. It is anticipated that there will be three phases to the implementation:

- The first phase will realise the new governance structure and the Standard Dataset with common definitions, which constitutes a quick win for the sector as the technology changes are not necessary to enable convergence of data definitions.
- The second phase will see the introduction of the New Landscape in full for the early adopters.
- This will then be followed by a full roll out to all remaining adopters.

Implementation is forecast to commence on 1 April 2015 and has been planned over a three-year period, to address concerns over the pace of implementation. Given the pre-existing change programmes that a number of the Data Collectors have it will be important that the implementation plan is flexed, where necessary, to align with these projects. We are also aware of some significant developments in SLC systems, which may affect the extent and timing of their engagement with the proposed new data landscape.

The diagram below provides a summary of the implementation milestones:
1.12 Next steps

The HEDIIP Programme Board should liaise with relevant stakeholders to secure agreement to the implementation plan detailed in this report. We recommend that the establishment of the governance group, should be the first task that should be implemented. Thereafter the Standard Dataset and data standards should be expedited as these can enable some quick wins for HEPs.

We are aware that various Data Collectors have their own transformation programmes ongoing. It will be important that the implementation of the landscape is aligned with these various transformations to avoid duplication.
2 Introduction

To enable a clear understanding of the reference points within this report it is useful to clarify the main factors that have influenced its development.

2.1 About HEDIIP

2.1.1 Vision

The Higher Education Data and Information Improvement Programme (HEDIIP) has been established to redesign the information landscape in order to arrive at a new system that reduces the burden on data providers and improves the quality, timeliness and accessibility of data and information about higher education (HE).

This follows the challenges set out in paragraph 6.22 of the BIS White Paper – Students at the Heart of the System which called for the HE information landscape to be redesigned “in order to arrive at a new system that meets the needs of a wider group of users; reduces the duplication that currently exists, and results in timelier and more relevant data.”

The establishment of HEDIIP has been overseen by the Regulatory Partnership Group who sponsored an initial feasibility study and a second study to define a mechanism to carry this work forward. HEDIIP is now leading a UK-wide programme of changes to build a more coherent, responsive and less burdensome information landscape.

The HEDIIP Programme Plan covers three broad themes:

- The strategy and change theme considers the broad question of what should the redesigned information landscape look like and how do we get there? As the programme evolves, this theme will focus on change activities that are required to achieve the goals of the programme.
- The standards and understanding theme promotes the standardisation of HE data and information definitions to reduce the burden involved with data supply and to improve the opportunities for data sharing and the comparability of published information. Where data and information can – or should – not be standardised, it will aim to increase the understanding of differences that exist.
- The capability and excellence theme will address the processes and capabilities associated with data and information, including management and governance issues where appropriate.

This project is part of the strategy and change theme and is therefore a key element in the Programme.

2.1.2 Rationale for the Landscape Project

The ‘A Pathway to Reform’ report for the Regulatory Partnership Group (RPG) identified the following potential benefits from redesigning the landscape:

Reducing duplication and increasing efficiency (Burden)

- Reduces collection effort through minimising duplication;
- Reduces the overhead of data transformation;
- Facilitates reuse/linking and hence reduction in the overall number of returns;
- Creates the opportunity for closer mapping between HE Provider and regulatory data exchange and business cycles;
- Increases the opportunity for automation of data collection; and

Reduced resource requirements for preparation of data offers greater opportunities to add value through analysis.

**Improving data quality**
- Better alignment with business cycles means data is checked and validated by the time it is needed;
- Standards limit the opportunity for ambiguity and misunderstanding; and
- Clear data models and increased automation reduces errors.

**Increased accessibility and opportunities for reuse**
- Use of de facto standards will make exchange of data simpler;
- Data presented in standard formats can more easily be recombined, reused and repurposed; and
- Standardised data in standardised formats more closely modelled within business processes can be accessed more readily and thus be timelier for a range of uses.

These benefits and aims, through making data more readily available and easier to comprehend, should in turn facilitate improvements to the student experience, accountability and transparency in a number of ways:
- By reducing the burden we create the opportunity for institutions to make more use of their data to improve their own performance and presentation;
- Through improving quality we increase the trust that people place in the data and therefore the confidence in the decisions that are informed by such data; and
- By making quality data more accessible we enhance the reputation of HE overall.

### 2.2 Changes in the Higher Education Sector

Changes in how the HE sector is funded, together with other implications arising from the 2011 White Paper have influenced the HE sector. There is now a more commercial approach in the way that many institutions are operating. The consequences of institutions’ reductions in income and student volumes have heightened the importance and need for high quality, timely and accurate data.

Our interactions with many HE Providers have confirmed that data is more highly valued now compared to any other time. This internal imperative for good quality data has created a greater appetite for an improved data and information landscape.

As a matter of priority HEDIIP has instigated a small number of projects alongside the landscape project, these include:
- Student Data Collection Review;
- HE Data Language;
- Subject Coding;
- Unique Learner Number (ULN); and
- Data Capability.

The Subject Coding and ULN projects aim to provide common standards, and facilitate data sharing and exchange in the areas of subject and student identification.

HEDIIP initiated the Student Data Collection Review (SDCR) to understand why so many apparently similar data collections are being carried out. The SDCR project identified that there is a common standard set of data used across the sector and identified opportunities to standardise and rationalise data collections to reduce the burden. The project also highlighted the need to progress the ULN and build data management capability, both areas in which HEDIIP is actively moving ahead. The Data Capability project aims to raise data management capability across collectors and providers; achieved through the establishment and deployment of best practice data management principles.

The New Landscape project has been defined to provide a vision and Blueprint for a new HE data and information landscape that delivers benefits across the sector and exploits the opportunities identified in the
Student Data Collection Review. The New Landscape project aims to direct the next tranche of projects and be used more broadly as a vision that all stakeholders can work towards.

2.3 Project scope

The project scope covers the UK HE data and information collection landscape and is focused on the interfaces between HE providers and Data Collectors and data flows between Data Collectors. This includes the processes of data management (in providers and collectors), data definitions and data transmission to the collectors.

The focus of this work has been on student data with the expectation that the approaches will be applicable to other HE datasets.

The Programme Board of 24th June 2014 reviewed the outcome of the Student Data Collection Review project and concluded that the major benefits of the programme would be accrued by initial focus on the larger Data Collectors. The project has therefore focused on the following HE funders, larger collectors and HE Agencies:

- Higher Education Funding Council for England (HEFCE);
- Scottish Funding Council (SFC);
- Higher Education Funding Council for Wales (HEFCW);
- Department for Employment and Learning in Northern Ireland (DELNI);
- Higher Education Statistics Agency (HESA);
- UCAS;
- Student Loans Company (SLC);
- National Health Service (NHS) (including HEE); and
- Higher Education Providers (including Higher Education Institutions and FE Colleges).

The project has considered the needs of PSRBs by engaging with a small number of representative organisations in order to ensure that the New Landscape is compatible with their requirements.

A common issue that is raised is that of data burden. This is something that the project has considered, and we have identified that there can be three drivers for data burden:

1. The complexity and breadth of the data collected;
2. The physical volumes and/or methods of data collection; and
3. The extent to which reported data is used impacts on how burdensome the requirements are judged to be by the data providers.

The development of the Blueprint seeks to make improvements in these areas, leading to a reduction in burden and duplication of data collection.
2.4 Overview of our approach

We developed a five-stage approach to this project. The five stages, shown below, were discussed and tailored as part of the initial project set-up meeting.

Figure 1: Project Approach

In undertaking this project we have consulted widely through:

- Regular meetings with the Project Management Office (PMO), Project Board, Advisory Panel and Programme Board;
- Stakeholder meetings with Data Collectors;
- Stakeholder meetings with Higher Education Providers;
- HEDIIP Workshops;
- Sector Events; and
- HE Survey.

A full list of stakeholders can be seen in Appendix A

A key part of the design and development of the New Landscape has been to clarify a number of assumptions and caveats to underpin the work. These are as follows:

Assumptions

- The scope of the Blueprint includes data required to support existing and potential HE students, with the exception of Higher Education in Further Education Colleges where this data is not currently collected by HESA;
- ‘Collect Once, Use Many Times, Used by All’ – The expectation of increased efficiency within the system is balanced against the user requirements regarding it being fit for purpose;
- Efficiency savings are sought throughout the system but these are not the primary drivers for change. Options for reducing costs via the implementation of streamlined systems or the removal of duplication would be beneficial;
- There will be resources available to implement the Blueprint;
- Roles for each organisation have been taken from publicised documentation and stakeholder interviews;
HEPs includes all institutions or businesses who are involved in the delivery of UK qualifications which are classified as HE by being in receipt of funding from a Higher Education Funding Council and/or deliver provision that is designated as being eligible for HE student finance;

There will be sufficient time to build consensus with all the relevant key stakeholders;

This project will not be required to perform a detailed assessment of the data flows as information has to be gathered via the inventory of HE collections;

The stakeholder list has been provided by the Project Team and approved by the Programme Board as being representative of the sector;

The focus of our stakeholder engagement has been with stakeholders in England, which together with Government Agencies in other countries in the UK provide a UK-wide view;

The governance of the New Data and Information Landscape is one of the key requirements to enable the landscape to generate ongoing benefits to the sector as a whole. (Data Governance is defined within our Implementation section of the report);

The views obtained from the stakeholders interviewed are representative of their areas;

We are reliant on the probity of the stakeholder comments. We have not been asked to verify the accuracy of the statements made, however where a common theme has been reported we have indicated this in the stakeholder themes; and

The diagrams produced during the development of this New Landscape have been approved by the project team as accurate for the purposes of this Blueprint.

The report is split into the following sections:

**Section 3 Current Landscape**
Describes our understanding of the current stakeholders and their data exchanges.

**Section 4 Analysis of Current Landscape**
Provides a description of the issues that we have identified.

**Section 5 The New Landscape**
A description of the building blocks for the New Landscape

**Section 6 Data Principles and Governance**
Describes the TOGAF data principles that we have developed and the new governance model.

**Section 7 Blueprint for a New Landscape**
Provides our description of the New Data and Information Landscape and its key features.

**Section 8 Implementation of a New Landscape**
Describes the roadmap for the implementation of the New Landscape.

**Section 9 Indicative Costs.**
Provides a demonstration of the cost benefits and implementation cost of the Blueprint.
Summary:

Whilst the scope of this project is similar to the profile of work that has previously been undertaken, this project aims to enable clear action and progress. The scope complements other projects and has not duplicated that work. The case for a step change in the organisation of the data landscape has never been as strong as it is currently. However, the main barriers will be to identify sufficient consensus across a sector that is fiercely independent.

As specified within our agreed scope, this project covers the UK HE student data and information collection landscape and is focused on the interfaces between HE providers and Data Collectors and data flows between Data Collectors. This includes the processes of data management (in providers and collectors), data definitions and data transmission to the collectors. Other data collections and data processing and analysis subsequently carried out by the collector is outside the scope.

It is expected that the findings will be relevant to the wider HE data collections.
3 Current Landscape

In this section of the report we categorise and document our understanding of the current data exchanges between the various stakeholders within the HE sector.

3.1 Stakeholder organisations

In our introduction we have identified the key stakeholders involved with HE Student data. In this section of the report we have grouped organisations with a specific interest in students’ data into the following categories. A full rationale for our approach is described below and individual classifications can be seen in Appendix B.

Higher Education Providers (HEPs)
Organisations that are contracted to deliver higher education and training by one of the funding and/or commissioning bodies. These include Universities, Further Education Colleges and Alternative Providers. We have also included training businesses that have had their programmes designated by the Quality Assurance Agency for Higher Education (QAA) as being eligible for student finance.

Government bodies
Government departments or organisations responsible for developing policies for education, training and skills, commissioning of training, contract management, visa and immigration (UKVI), data security or other specific services funded by central government. This would include student finance through the Student Loans Company (SLC) and NHS related collections.

Professional, Statutory and Regulatory Bodies (PSRBs)
PSRBs are an extremely diverse group of autonomous organisations. They set the standards for, and regulate the standards of entry into, particular professions.

Data Collectors
This term is used within this report to identify generic data collection activities. We have also introduced the label of ‘Key Data Collectors’ when we refer to the operations of the larger and/or more significant Data Collectors which include:

- Government bodies;
- Higher Education Statistics Agency (HESA); and
- Universities and Colleges Admission Service (UCAS).

Some of these Data Collectors are also service providers and the delivery of their services are of paramount importance to those organisations.

General public
We have grouped the requirements of general public, press, potential/current or previous students into the needs of the public as a whole. Their diverse need for access to timely high-quality information is partly covered by the Government’s expectations described in their open data agenda.

Mission groups and HE professional bodies
There are a significant number of Higher Education related professional interest groups. In general they are all aimed at providing opportunities for increasing peer development, enhancing communication and
championing the interests of their specific communities. They are important to the sector and the development of the New Landscape as they:

1. Are well respected brokers of HE opinion within the HE sector;
2. Have a common interest in data; and
3. Are influential.

Figure 2: HE Representative groups

- Mission groups: Organisations that have been developed to influence HE strategies and policies in the United Kingdom where the focus is on the needs of the membership. These organisations have direct and indirect representation on the Data Collectors.
- HE professional bodies: Normally aligned to administrative areas within a University. In general they aim to improve the operational effectiveness of their service area by enabling their members to deliver their roles more successfully. These groups are generally not represented with Data Collectors.

There is clearly some overlap between these broad categories as both are interested in developing the skills of the workforce and increasing efficiencies and effectiveness. The organisations that are recognised as representing the HE sector are accepted as being the externally focused organisations such as GuildHE and UUK. These organisations represent the sector directly or indirectly through the placing of their members onto the governing boards of the Data Collectors. The Association of Colleges (AoC) is a representative group for Further Education (FE), whilst Alternative Providers and PSRBs do not currently have a groups that represent their needs. Members of the HEI professional bodies may be present on the boards of Data Collectors but they are not representing their professional bodies.

3.2 Main stakeholders governance overviews

Effective data governance is the mechanism through which providers are able to agree a mandate for change. All of the Data Collectors are governed by their own Boards. This has evolved over time and quite understandably each collector has pursued its own priorities. To some extent the professional groups and other informal groups have played a helpful role in harmonising priorities and data needs in some areas. However, the lack of an overall governance body or group has become a significant barrier to enabling a cohesive data and information landscape.

It is the case that when the Data and Information Landscape is viewed as a whole, these governance arrangements are inconsistent in how they contribute to the governance of student and other data: some
focus on standards for data; some on standards for statistics; and others on technology or specific sector data. Coherence and systematic consistency of approach is difficult to achieve under these current arrangements due to the differing organisational priorities. This is considered further in this section.

The following table provides a summary of the major governance and sector advisory/scrutiny groups that are currently in operation across HE student data.

**Table 1: HE governance summary**

<table>
<thead>
<tr>
<th>Stakeholder group</th>
<th>Status</th>
<th>Relationship with sector</th>
<th>Chair and % of Board from Higher Education</th>
</tr>
</thead>
<tbody>
<tr>
<td>UCAS</td>
<td>Charitable company with a commercial arm (members of Company are UUK and GuildHE)</td>
<td>Shared application services, Information and Advice provision, publication of data and analysis</td>
<td>Professor Sir Steve Smith 83%</td>
</tr>
<tr>
<td>HESA</td>
<td>Charitable company with a commercial arm (members of Company are UUK and GuildHE)</td>
<td>Data Collector on behalf of statutory bodies; publication of data and information</td>
<td>Professor Simon Gaskell 95%</td>
</tr>
<tr>
<td>SLC</td>
<td>Non-Departmental Public Bodies (85% Westminster Government and 5% for each of the devolved administrations)</td>
<td>Service provider for student finance</td>
<td>Christian Brodie Chair of Council at the University of Sussex 10%</td>
</tr>
<tr>
<td>NCTL</td>
<td>Executive Agency sponsored by DfE</td>
<td>Commissioning and regulation of Teacher Training</td>
<td>Permanent Secretary n/a</td>
</tr>
<tr>
<td>NHS</td>
<td>Quasi-Autonomous Non-Governmental Organisation or Non-Departmental Public Body (NDPB)</td>
<td>Development and funding for the development of skills in the healthcare sector</td>
<td>Non HE n/a</td>
</tr>
<tr>
<td>HEFCE</td>
<td>Government Agency</td>
<td>Regulation and Funding</td>
<td>Tim Melville-Ross 45%</td>
</tr>
<tr>
<td>HEFCW</td>
<td>Government Agency</td>
<td>Regulation and Funding</td>
<td>David Allen – Former Registrar and Deputy Chief Executive, University of Exeter 63%</td>
</tr>
<tr>
<td>SFC</td>
<td>Government Agency</td>
<td>Regulation and Funding</td>
<td>Professor Alice Brown, Emeritus Professor of Politics at the University of Edinburgh 17%</td>
</tr>
<tr>
<td>DELNI</td>
<td>Department in Devolved Administration</td>
<td>Funding and Regulation</td>
<td>Non HE 0%</td>
</tr>
<tr>
<td>Jisc</td>
<td>Charitable company with a commercial arm (members of Company are UUK, GuildHE and AoC)</td>
<td>Service Provider and supports the sector with strategic change initiatives</td>
<td>Professor Martin Hall 100%</td>
</tr>
<tr>
<td>SFA</td>
<td>Government Agency</td>
<td>Regulation and Funding</td>
<td>Non HE n/a</td>
</tr>
<tr>
<td>Research Councils</td>
<td>Non-Departmental Public Bodies</td>
<td>Regulation and Funding</td>
<td>Professor Rick Rylance is Chair of the RCUK Executive Group 100%</td>
</tr>
</tbody>
</table>

Significant features of the current governance/board arrangements are:

- Most of the main stakeholders have HE representation;
Multiple audiences and differing requirements for data contribute to confusion over information requests; and

Differing needs and specific user requirements lead to different and largely uncoordinated data requests; and

Data collections are treated as mandatory as the HEPs feel that they are not free to ignore these data requests without a significant and potentially detrimental impact on their business and reputations.

Whilst most of the non-governmental organisations have direct HE representation, the influence is focused on the specific business activities and not directly on data.

### 3.2.1 Information standards

There are a number of bodies working in the data standards arena that have varying levels of influence over the data in the new HE data and information landscape. These include the Information Standards Board (ISB) for Education, Skills and Children’s Services (ESCS), the NHS Standardisation Committee for Care Information, the IMS Global Learning Consortium, Consortia Advancing Standards in Research Administration Information (CASRAI), BSI Group (BSI) and International Organisation for Standardisation (ISO). Any HE data and information governance process will need to monitor and engage with the relevant existing standards bodies as appropriate.

### 3.3 Regulation in the HE landscape in England

In March 2012 Deloitte produced the following diagram to illustrate the overlapping areas of regulation in the HE landscape. ‘Mapping the Higher Education Funding and Regulatory System in England’, Deloitte, March 2012. Whilst this work mapped the English regulation system we think that this is broadly a representative picture for the whole UK landscape in that it provides a view of the various stakeholders and the main focus of their operations and reflects the range of differing data demands.

![Figure 3: Regulation in H.E. in 2012.](image-url)
The 2012 report identified that the BIS HE White Paper and the Technical Consultation included a number of proposals that were designed to increase the efficiency and effectiveness of the funding and regulatory system. Many of these proposals have been implemented or are in the process of being implemented. However, a number of critical proposals are yet to be addressed, namely:

- The creation of a single, transparent regulatory framework that covers all institutions in the English HE system.
- An explicit remit for HEFCE to protect the interest of students, including by promoting competition where appropriate in the HE sector.

Whilst the Deloitte work focused on the regulation in England, we consider that this provides an indication of the whole UK regulatory environment as the majority of the organisations operate across the UK.

Since 2012, the regulatory environment has remained broadly consistent. Our high-level review of the current UK data landscape indicates that data governance in the UK can be represented by the following diagram. This provides a view of the silos of influence and the lack of a comprehensive data governance that joins or provides formal collaboration of data requirements between collectors and the providers.

**Figure 4: Data influence**

Within this view we have identified that business imperatives and commercial issues allow for the development of data collection strategies without the requirement for coordination or sharing of existing data, where it exists.

Each silo of governance or HE representation provides that Data Collector with a requirement for data and information for delivering its standard operations without specific consideration of other aspects of the

---


HEDIIP NEW DATA LANDSCAPE

landscape. There is a risk that the process for strategic decision making in respect of data is therefore too narrowly focused on self-interest. The Data Collectors’ issues can be summarised as:

- For HESA and UCAS, who have both made the strategic decision to grow their commercial interests alongside their charitable status for their standard services, there is the need to ensure that they do not support changes that may impact on their current or future commercial position.
- For the Data Collectors that are part of government structures, there is a mandate to deliver the operational service to the highest standard reflecting the latest changes in government policy.
- PSRBs have a diverse range of roles and legitimate requirements. Many are small standalone organisations; a number are regulated or licensed by regulatory bodies. Some are regulatory bodies that regulate by statute, Royal Charter or by other PSRBs.

Whilst the sector’s main Data Collectors have a long track record of engagement and communication on a broad range of issues, the implementation of significant changes to the status quo are not encouraged through these existing arrangements. Data Collectors do cooperate on data collection where it is deemed to be within their business interest but this is voluntary and not linked to a shared ambition that benefits the sector as a whole.

Comprehensive and consistent data governance does not currently exist across the HE landscape despite the fact that the majority of the sector agrees that data governance and data standards are an essential requirement. A key reason for this is that the Data Collectors operate independently of each other and have differences (albeit with commonalities) in their missions and purpose. The dilemma is whether it is possible to establish effective governance through non-binding consensus.

3.4 International Governance Comparators

To further assess how the issue of data governance has been addressed elsewhere, we have gained an understanding of how other Higher Education systems in the world are dealing with this issue. We have summarised their arrangements across the following three parameters which are detailed in Appendix D.

- Who owns the data collection agencies?
- Is the data collection centralised?
- Is the data collection in-year?

Our review has identified that each country has its own approach to Higher Education data management. The consistent theme for the majority of the countries reviewed is that the countries have centralised data collections through regional or central government. The United Kingdom appears to have a unique set of circumstances where the main data collections are routed through a central data collection agencies that are sector owned. The proliferation of Data Collectors is also a factor that appears to be unique to the UK data landscape. The detail of the review can be seen in Appendix D.

3.5 Further Education data governance in England

Within the FE data system the Information Authority (IA) was established in 2007 with a remit to reduce the burden of data collection on the Further Education Providers. The IA was governed by the IA Board which was chaired by an ex FE Principal. The Board had representation from Colleges and private training providers as well as FE data stakeholders such as OfSTED. The IA was housed in the Skills Funding Agency but operated as an independent organisation, its main role was to manage the FE data collection via the Individualised Learner Record. The IA was judged as a success by the majority of providers and sector representative organisations such as the Association of Colleges (AoC) and the Association of Employment and Learning Providers (AELP). Whilst the Information Authority was disbanded in 2012 during a period of significant cost reduction within the Skills Funding Agency, its role of FE data governance has been transferred to the FE data group that sits within BIS.
3.6 UK Governance comparators from other sectors

To provide an understanding of how other sectors have addressed this data governance issue we have provided a high level summary of their governance operations. More detailed descriptions are available in Appendix C but the information has been summarised below.

A review of other sector data collection for non-statutory arrangements has identified the following:

**Health**

- **Description**: Activity and Finance data in the health sector is generated by a number of publicly funded trusts, other NHS organisations such as primary care providers and private health care providers. The Health and Social Care Information Centre (HSCIC) is responsible for collating national datasets and distributing these to Monitor (the health regulator), the Department of Health and to regional centres for wider circulation. Local datasets are managed through local arrangements with bodies such as Commissioning Support Units (CSUs).

- **Governance solution** via the Health and Social Care Information Centre (HSCIC): Common Data specifications are used to gather national activity data (SUS) on a publicised schedule. An NHS Data Dictionary is in place and a set of Health Dimensions have been defined centrally to enable consistent analysis across multiple Healthcare organisations. Data quality is achieved locally through consultation with providers. Whilst this is an official responsibility the process does not include steering committees nor other formal mechanisms. Instead it forms a non-regular item on monthly/quarterly commissioning meetings between Commissioners and providers.

- **Relevance to HE sector**:
  - Multiple individual external data providers with differing systems and purposes, including both primary and secondary care. Includes GPs, Hospital Trusts, Social and mental healthcare providers;
  - While there is legislation and contractual rewards and penalties due to the symbiotic relationship between providers these are difficult to levy/enforce;
  - Similar in terms of data requirement, Patient = Student can have multiple episodes = instances of study;
  - National Repository through HSCIC (similar in role to HESA); and
  - Data Sharing Agreements are required to enable Health organisations to share information.

- **Pros**
  - Acute activity data comparable nationally through agreed data specifications and definitions;
  - Clear enhancement request process in place to manage changes;
  - Continuously improving data quality achieved from working with providers over time; and
  - Access restricted extremely securely to authorised users.

- **Cons**
  - Timeliness of data distribution due to scale of collection;
  - Local anomalies are not catered for; and
  - Datasets are not integrated centrally allowing for inconsistencies locally.

**Food standards**

- **Description**: Red Tractor Assurance is a not-for-profit organisation that is funded through membership (which is voluntary), licences and food chain support activities, such as information provision for the British Poultry Council.

- The ‘core’ schemes in six product sectors are all wholly owned by Red Tractor. In addition, a number of schemes operating in parts of the UK or more limited product ranges are recognised as ‘equivalent’ and product from farms certified to these schemes is eligible to carry the Red Tractor logo.

- **Governance solution**: Membership required submissions – governance of the individual members’ adherence is undertaken through the assessments. Red Tractor members are overseen by Assured Food
Standards (AFS) who carry out independent inspections to confirm that businesses are meeting certain standards, such as food safety, animal welfare and the environment, with any issues identified and a plan put in place to rectify. As Red Tractor has no financial or legislative powers non-conformances against a ‘key’ standard may result in suspension from the scheme.

Relevance to HE sector:
- Use of a central portal for aggregation and reporting of food standards adherence by local and national independent retailers through which data is gathered;
- Use of data for sampling and statistical analysis; and
- Regulatory submissions are undertaken through a standard mechanism via an independent body.

Pros
- Funded through membership therefore no central government cost;
- Independent of but supported by the Government; and
- Links with other assurance schemes which provides a wider coverage.

Cons
- Voluntary; and
- One of the top four supermarket chains left the scheme as they suggested that it didn’t provide sufficient value.

Retail

Description: Major retailers handle vast amounts of data, mainly through their Electronic point of sale (Epos) system, however systems such as these are driven from product catalogue/s within the business. The data contains many facets, including price and description products. The process of update takes place on a daily, or as required basis based on the procurement process.

Governance solution: There are reputational and economic drivers as there is mutual benefit in ensuring that data is passed accurately from supplier to purchaser. Data quality is not maintained through formal bodies but through informal channels. At present however development of steering groups is underway.

Relevance to HE sector:
- Multiple independent suppliers of data who own the data passed to an organisation which relies heavily on the data;
- Requirement to ensure latest version of data is captured and there is a high need for maintaining critical data dimensions; and
- Supplier may have its own software or maintain its own solution independent of the whole sector.

Pros
- Tailored to meet the needs of specific retailers;
- Not centralised; and
- Dataset determined by individual retailers.

Cons
- No central data repository and thus the ability to report on the sector performance is limited;
- Data quality standards and consistent assessment methods have not been introduced;

Whilst each of the three comparable sectors have areas in common with Higher Education we would suggest that the combination of the range of providers and their independence would suggest that the retail sector is the most comparable. The reputational and associated economic factors within the retail environment are key drivers for these organisations, in an environment where information for the public gained from consumer advice websites and social networking can have a serious impact. The Red Tractor Scheme was implemented in 2010 and the Assured Foods Standards has four main functions:

- Defining Standards;
- Ensuring conformance with the standards;
3.7 Pathway to reform

In June 2012 the Interim Regulatory Partnership Group: Project B for Redesigning the higher education data and information landscape issued the report "Pathway to reform". It suggested the implementation of a model of governance and suggested that the following actions would be needed:

- Establish governance of the information landscape as a shared responsibility;
- Draw on expertise from the IRPG member organisations, the sector and others as required for delivery of specific elements of this vision;
- Engage with the devolved administrations to address the UK aspect of the work; and
- Establish interfaces with other relevant bodies (e.g. Information Authority, Data Service, ISB for Education and Children’s Services, ISB for Health and Social Care etc.).

We agree that these recommendations remain relevant and have considered them as part of the approach to establishing effective governance for the agreed New Landscape (Section 7). Their proposed model is included in Appendix E.

3.8 Indicative policy themes from previous reports

Given the previous work referenced in section 2, we have undertaken a desk-based review of the relevant documentation from past projects and HE focused studies as part of our analysis of the current HE landscape to complement the information collected via our Stakeholder meetings.

We have used this process to identify and acknowledge any residual issues that the New Landscape will need to address alongside any information gathered during our stakeholder interviews.

The full list of documents reviewed is included in Appendix F along with the detail of our analysis where each document has been analysed to identify:

- The existing challenges/issues as documented.
- The current state at the time of publication.
- Proposals made to date.

We have then summarised the current state and the future landscape has been described from seven different perspectives, as illustrated below in figure 5 where we have provided a view of the summarised issues and the points to be considered within the New Landscape. Where we have identified that HEDIIP projects exist to mitigate the observation we have indicated with the name of the project in brackets.
3.8.1 Customer user views

Summary of Issues
- Students can find it difficult to navigate an increasingly diverse data landscape;
- The communication across data providers and collectors is sub-optimal; and
- Unique Learner Number (ULN) concept in its infancy for adoption in HE. (ULN project)

Points to consider in the New Landscape
- Improve the sector’s understanding of what data is to be collected and what it is used for; (Data inventory)
- Create rules/explanation that informs providers what information is required;
- Adopt ULN as a common identifier to improve data linking and sharing. In the longer term there is potential to replace other identifiers and only use ULN; (ULN project) and
- A definitive and comprehensive source of HE data should be developed, documented and published and maintained centrally. (Data inventory)

3.8.2 Requirements perspective

Summary of issues
- Requirements of Data Collectors are not collectively governed;
- Limited effort has been made to co-ordinate the requirements across the different data collectors; and
- The net effect of multiple requirements is increased administrative burden on HE Providers.

Points to consider in the New Landscape
- Create a list of data requirements from the Data Collectors that specifies what information they need (at logical level), what it will be used for and how often it should be provided; (Student Data Collection Review)
- Rationalise data requirements by removing duplicate requirements; and
- Put agreed data requirements list under change control, managed by collective, independent governance group/body/function.
3.8.3  Process perspective

Summary of issues
- Collection agencies are primarily focused on their needs and have not considered the collective burden on HEPs resulting from data collection in totality;
- There is not minimal alignment with HEP internal processes; and
- In some cases HEIs internal processes are not optimal, which impacts on the effort required to obtain data that is of an acceptable and suitable standard for submission to Data Collectors.

Points to consider in the New Landscape
- Create a good practice library containing tools and templates so that HEPs can learn from the experience of others across the sector; (Data Capability) and
- Work with collection agencies to define an efficient process for gathering information and sharing it amongst themselves.

3.8.4  Governance and data management

Summary of issues
- The lack of governance and data standards has led to ambiguity and inefficiency in data standards and data use; and
- In some cases there is insufficient data management control within individual HEPs, leading to variable data quality across the sector. (Data Capability)

Points to consider in the New Landscape
- Develop models of good data management practice that can be considered within a HEP; (Data Capability)
- Establish an independent central organisation that will act as the accountable body and be the single point of contact for initiating and monitoring collections; and
- Appoint data stewards and data trustees across the environments that will be managed/co-ordinated by the new organisation.

3.8.5  Organisation perspective

Summary of Issues
- There is insufficient collaboration amongst Data Collectors leading to inconsistency in the data collection arrangements; and
- Data is not consistently shared effectively amongst collection agencies.

Points to consider in the New Landscape
- Establish a shared, central data warehouse but enable individual stakeholders to have access to check that the data meets their quality requirements;
- Expand the scope of the data collections inventory to include all institutions and all types of reporting; (Data Inventory)
- Ensure that senior management at the HEPs are engaged in understanding the importance of effective data management; (Data Capability) and
- Encourage HEPs to adopt good data management practices. (Data Capability)

3.8.6  Application and data

Summary of issues
- Little automation has been applied to the data management process and in some cases it relies heavily on manual completion of spreadsheets. However, there are examples of technology being used to reduce the time and effort associated with the activity. (Data Capability)
Points to consider in the New Landscape

- HEPs do not always have a centrally co-ordinated approach for their data management system and/or oversight. Whilst we recognise that this may not be a barrier to good data management, we would suggest that it is easier to manage data through a unified approach. Whether operated centrally, locally or a combination of the two; and
- Evaluate good practice and share these with the sector. (Data Capability)

3.8.7 Data technology

Summary of Issues

- Collection agencies are not aligned in terms of their requirements for data returns;
- HEPs systems are not interoperable with those of the Data Collectors;
- Lack of data sharing between collection agencies causes more work for the HEPs; and
- The JACS coding framework is no longer fit-for-purpose. (Subject Coding)

Points to consider in the New Landscape

- Develop a data dictionary for the sector and put it under change control;
- In conjunction with the review of requirements, determine whether data collections can be rationalised;
- Design a new central data warehouse that will be the repository for use by all interested parties across the sector;
- Ensure that the new coding framework is designed in a timely fashion; and
- Allow data collection agencies to share data, thus eliminating duplication across returns from HE.

3.9 HEDIIP

HEDIIP is now formally a UK-wide structure that has provided the HE sector with an example of how data projects can be implemented through a finite structure that operates on the basis of goodwill and collaboration. Before the development of HEDIIP in 2013, Jisc funded and continues to fund strategic change-related projects, but these are agreed through their sector representation and are largely focused on technical capability and solutions. The agreed mandate for HEDIIP is more closely aligned to the reduction of burden and therefore its initial projects before the New Landscape project have been aimed at addressing individual elements of the burden. HEDIIP is hosted by HESA but retains its independence through oversight by a separate Programme Board. The Programme structure and governance is presented below.
To assess if and how the New Landscape project could benefit from the current HEDIIP arrangement we have documented the current benefits and limitations below.

**Benefits**
- Actively encourages collaboration through the inclusion of a wide range of stakeholders within its governance. The main UK stakeholders are included;
- It has strategic direction through its Programme Board along with technical and operational oversight from practitioners involved in the Advisory Panel. Members of the Advisory Panel sit on Project Boards which enhances stakeholder engagement and accountability for the projects;
- Utilises structured project management approaches for all projects; and
- It is seen by its stakeholders to be independent and it reports to an independent Board (RPG).

**Limitations**
- It has a finite timescale with current funding until July 2016;
- The scale of operations is currently small as it has a Programme Management Office of three staff including the Programme Director, which is deemed sufficient for its current scope of operations;
- Consensus was generated against a vision for HEDIIP as a programme; and
- Working through consensus can be a barrier towards implementation as the Programme Board does not have the authority to control stakeholder resourcing and prioritisation of issues.

### 3.10 Other HEDIIP projects and their relevance to New Data and Information Landscape

We have undertaken a high-level review of the status of each of the HEDIIP projects to assess how they can contribute to a more effective HE Data and Information Landscape. Our findings are summarised below and more detail can be found in Appendix G.

- **Data Capability** – Identification and dissemination of good data management practices.
- **Unique Learner Number** – To implement a consistent identifier across the HE landscape with the potential to reduce the number of unique student identifiers by the implementation of a number which is widely used in Schools and Further Education.
■ **Student Data Collection Review** Student Data Collection Review explored opportunities and barriers for the rationalisation and standardisation of student data collection in HE.

■ **Data Language** – Project to help provide an understanding of the data language that currently exists across the different Data Collectors and would support the development of data standardisation.

■ **New Subject Coding System** – The Centre for Educational Technology, Interoperability and Standards (Cetis) at the University of Bolton is undertaking the Subject Coding project to develop a replacement for the JACS system to meet the needs of a broader group of stakeholders and reflect the diverse and dynamic nature of higher education in the twenty-first century.

■ **Data Collection Inventory Review** – Aims to help Data Collectors gain a better understanding of what other collections are being undertaken and to help HE providers understand which data collections they are responding to.

---

**Summary:**

The Higher Education sector has developed and maintained a significant level of autonomy that is cherished by the main delivery providers and successfully championed by their sector bodies. The size, complexity and professionalism of the sector has resulted in a broad range of opportunities for participation and communication that is extended towards the data collection but not in a systematic or co-ordinated manner. **HE providers are represented and consulted with by the main Data Collectors and funding bodies but their input is either focused on the development of that individual business process or the impact on a single administrative function.** Whilst this autonomy and representation is seen to be a strength of the sector, it can become a potential barrier to the collective management of the data in the sector. HEDIIP has partially fulfilled this role on an interim basis but the implementation of a New Landscape would need appropriate data governance arrangements.

The review of international comparators and other sectors has identified that the UK appear to have a unique situation where there are variety of ‘non-government’ independent data collection agencies at the heart of the data collection system. A comparison of other sectors has suggested that there are examples of voluntary commercial governance arrangements that are delivered by AFS for the Red Tractor food standard.

From the documents reviewed there is a general acceptance that **data governance** is required which would seek to mitigate the various cross-cutting demands and minimise the impact of the diverse data collection requirements by a **standardisation of data specifications** and a consolidation of data collection as appropriate.

Professional bodies and Mission groups should be involved in the representation of the HEPs, however, there should also be formal opportunities for FEC and AP providers to have their needs considered.

**Documentation of data standards** and the acknowledgement of a **Standard Dataset** would provide access to a reference library of HE data for its stakeholders. As part of this process it may also be beneficial to create a repository of Data Collectors.

The findings from our review are broadly consistent with previous studies however we would stress that the effective implementation of data governance is key to the improvement of the data landscape.

As the 2011 white paper challenged the HE sector to put ‘students at the heart of the system’ we have also identified that both data Data Collectors and providers have inconsistent approaches to consulting with students.
4 Analysis of the current landscape

This section of the report provides our analysis and findings from our review of the current landscape and identifies objectives for the New Landscape.

4.1 Data flows

The following diagram describes the summarised HE landscape established via the stakeholder interviews. Figure 7 demonstrates the apparent connection between students and HEPs (primary sources of student data capture). Whilst we have indicated the direction of data flows, we have not provided a view that incorporates the timing of data flows due to complexity.

Our diagram has only included three HEIs (as an example); in reality there are approximately 160. We have attempted to provide a simplified view of the typical data flows and activities by HEPs to illustrate the duplication that is taking place.

Figure 7: Simplified diagram of the current landscape and indicative data exchanges for a typical HE Provider

Note: Lines are representative of data exchanges between the stakeholders. However, these lines of connection can be misleading as for example:

- HEPs are not homogenous organisations, even amongst the Higher Education Institutions (HEIs) there are a significant number of differences which impact on their perception of the current landscape such as:
  - They are of vastly different sizes ranging from greater than a billion pounds turnover (mostly HEIs) to less than a million (mainly FECs and APs);
  - The wide range of provider types such as HEIs where some are research-based, some are more focused on teaching, through to FECs and private providers;
  - HEPs have bespoke internal data management arrangements, some have a centralised approach to data processing whilst others have distributed approaches and several have more than one student record system;
- The majority of the HEIs utilise student record systems from one of the main providers, but approximately 20 have their own bespoke systems;
- Whilst all of the HEPs have similar returns to the statutory Data Collectors (HESA and SLC and also use UCAS for their FT undergraduate admissions process) data exchanges with other organisations such as PSRBs may not be comprehensively tracked or managed centrally;
- The PSRBs are the least likely of the Data Collectors to have consistent data exchanges with HEPs due to the variety and breadth of their interests.

Each Data Collector has evolved their collections to fulfil their own purpose. This ongoing evolution has contributed to increased complexity of the landscape as the collectors are not mandated to evaluate and minimise the impact of their separate requirements on the HEPs, or collaborate with each other such that there is a coordinated data request.

The next view of the landscape has been extended to show the data exchanges with a broader external audience.

Figure 8: Indicative simplified high level data and information flows in the current landscape
Initial observations from this Landscape:

The complexity is mainly driven by:

- Diverse range of Data Collectors and HEPs individual data collection arrangements;
- HEPs are asked to submit very similar data multiple times, often due to the timing of data collections by different collectors;
- Similar data collected by different Data Collectors at differing levels of granularity;
- Lack of clarity on the authoritative sources of data as similar data is collected at different times and used for a diverse range of purposes; and
- Data standards are applied in isolation to each separate collection.

Concerns from providers are that:

- The data supplied is used for purposes that it was not supplied for; and
- Too much data is being requested and there is a belief that it is not all being used.

The impact of this current landscape is to create confusion and burden within the providers as HEPs appear to struggle to efficiently manage their data to meet the demands of the Data Collectors alongside internal demand. It can result in HEPs being tasked with data reconciliations over protracted timescales (across multiple academic years) and the need to continually manage data to meet differing needs.

4.2 Data Collectors

The various Data Collectors’ data requirements are directly related to their business needs or related to those requirements of their administrative process. Whilst this is understandable under the current data collection arrangements, it can cause confusion where the data items collected have a common title but the actual data collected is defined differently. Some examples of this confusion are with the collection of course data by three of the main Data Collectors where:

Course

- UCAS – collects ‘course’ which is details for the student’s application and as such this may have a title that is attractive for marketing purposes. This will occur before the start of the academic year;
- SLC – collects ‘course’ which is details for the assessment, award and management of student finance, and as such the important detail is whether the course is eligible for loans and the fees chargeable. This information is collected before the start of the academic year; and
- HESA – collects ‘course’ which is for the purposes of statistical analysis, where this allows for an alignment to JACS subject codes that describe the subjects being studied. These are consistent with the codes used by UCAS but the HESA return does not contain the marketing title of the course and is collected after the end of the academic year.

Programme

- UCAS – A thematic curriculum of study covering the overall learning objectives the HE provider intends to deliver to students;
- QAA – An approved course of study that provides a coherent learning experience and normally leads to a qualification.

The examples above are an illustration and not an exhaustive list of areas where repetition of collection and inconsistencies exist.

4.3 Key features of HE students data collections and indicative data flows

We have used the information provided by the Data Collectors and our knowledge gained from the stakeholder interviews to map our understanding of the high level data flows between the Data Collectors, to gain an insight into how much data is currently shared across the sector.
HESA
- Historic individualised student return post-year end;
- Other non-student data returns;
- No direct student interactions;
- Receives StarJ data (application data) from UCAS in December;
- Receives HE in FE data from HEFCE;
- Provides data cubes to HEIDI;
- Provides NCTL with Initial Teacher Training data;
- Provides data to statutory customers; and
- Provides data under contract to third party organisations.

Future plans include, transforming the main post-year data collection into an in-year data collection.

SLC
- Course details and fees collected prior to the start of the year;
- Student details and interactions for access to student finance;
- Daily transactions with HEPs regarding students’ status;
- Different arrangements for the different constituent parts of the UK;
- Provides data to the Funding Councils;
- Has the functionality to receive student details for SLC from UCAS, but this optional link is not operative;
- Direct financial transactions with students and HEPs; and,
- Matches data with HMRC for validating payment threshold.

The SLC has recently allocated £139 million to redevelop its IT systems, although this project has not analysed the aims of this investment. We understand that this change programme is significant and ongoing. As a result, it will affect the pace and extent to which SLC can implement parts of the new landscape.

UCAS
- Course details collected from HEPs prior to start of the year;
- Applicant data collected from prospective students;
- Provides data for advice to prospective students regarding their applications;
- Provides data and data services under contract to 3rd parties;
- Provides NCTL with initial teacher training data;
- Provides data to the Funding Councils;
- Has the functionality to transfer student details for SLC but this optional link is not operative;
- Provides StarJ extract to HESA and individual HEPs; and,
- Direct and daily interaction with schools, HEPs and applicants.

UCAS is currently transforming all of its internal systems to provide a more robust and expandable infrastructure.

PSRBs
- Very diverse arrangements for data collection, with no oversight body;
- Direct student interactions;
- Some receive data from HESA;
- Direct interactions with academic departments in HEPs that align to the disciplines of the PSRBs; and
- In-year interactions with HEPs that may not be data driven.
Funding Councils

- Receive data from HESA as a statutory customer;
- Collect non HESA data directly from HEPs sometimes as aggregated returns or surveys; and
- Collect HE in FE data and share with HESA.

Key features of other HE students data collections not linked to main Data Collectors

Examples of other HE student data collections include:

**HEE (NHS)**

- Collates its data through Local Education and Training Boards (LETBs);
- Collates data for Commissioning, Performance and Quality, Workforce, Finance, Regulation and Compliance, Recruitment and Trainee Management;
- Providers exchange data directly with LETBs which is then passed onto HEE; and
- HEE also receives UCAS and HESA data from LETBs.

HEE is currently reviewing all of its data infrastructure and has recently appointed a Chief Information Officer to oversee this work.

**NCTL (Teacher Training)**

- Receives two in-year returns from HESA;
- Uses data from the main HESA student return which is used to populate NCTL database; and
- Collects trainee teacher outcome data from HEPs in July.

**RCUK**

- Receives data from HEIs within 30 days of student starting and collects research data via an annual return, it also combines information with British Libraries thesis data; and
- Receives data from HESA as a statutory customer.

RCUK is undertaking a range of projects regarding the Interoperability research systems at RCUK systems with Jisc, the Association of Research Managers and Administrators (ARMA) and universities. This is aimed at providing a more effective data exchange with HEPs.

**UKVI (UK Visas and Immigration)**

- Receives student data from foreign (non-EU) students at the application stage from HEPs via its Sponsor Management System (SMS). The SMS is a system that UKVI requires all sponsoring organisations (of whom we understand there are 30,000 in the UK) to use. Therefore there is limited scope to tailor the system just to the needs of HEPs.
- A minimum of three times during the year HEPs have to update the SMS to identify students who have withdrawn from their courses of study.
- UKVI has its own data definitions for the data it requires.

UKVI is a member of a ‘co-regulation group’ that has been established by GuideHE and UUK on behalf of the sector to debate and discuss issues affecting UKVI and its requirements on the sector.
The diagram indicates that, under the existing arrangements, Data Collectors have cooperated (sometimes with financial incentives) to exchange data with the appropriate protocols. However, whilst it demonstrates that this is possible these exchanges could be strengthened through the establishment of a Standard Dataset that is collected. Whilst the diagram indicates a transfer between UCAS and SLC this link is currently switched off pending further discussions between these Data Collectors. This diagram represents the main flows from the perspective of HEPs with degree awarding status. A view incorporating the FECs and APs who do have degree awarding status would be more complex as it would include degree validating bodies such as awarding bodies and HEIs.

4.4 Providers data processing

Higher Education Institutions are generally described as highly successful professional businesses that are motivated by providing high quality learning to their students and/or research activities to their communities. Historically they have processed significant volumes of data but have been inconsistent with their data management approaches. As autonomous institutions they have developed a wide range of internal organisational structures for meeting the plethora of largely uncoordinated data demands. Therefore alignment and rationalisation the data demands on all of the HEPs would promote an evaluation of HEI internal data management processes that would release efficiencies in data processing and provide opportunities for enhanced data management operations and better analytical capacity. The reduction of the data demands would provide the maximum benefits for HEIs, but it is envisaged that FECs and APs would also make limited efficiency gains.

4.5 Stakeholders

Our review has included the interviewing of key stakeholders involved in HE student data preparation and collection. The initial stakeholder list was provided by the HEDIIP PMO. The full list can be seen in Appendix A. The information gathered from the interviews has been analysed and summarised below. In addition to the stakeholder interviews, weekly communication with the HEDIIP PMO and direction gained from the Project
Board a series of workshops were organised to help focus the development of the New Landscape. The feedback from the stakeholder workshops has been used to refine the Data Principles and Blueprint views. Progress with the project was also discussed with the Advisory Panel.

<table>
<thead>
<tr>
<th>EVENT</th>
<th>FOCUS</th>
</tr>
</thead>
<tbody>
<tr>
<td>HEDIIP Advisory Panel</td>
<td>23rd October 2014</td>
</tr>
<tr>
<td>HEDIIP Programme Board</td>
<td>5th November 2014</td>
</tr>
<tr>
<td>Workshop 1</td>
<td>11th December 2014</td>
</tr>
<tr>
<td>Workshop 2</td>
<td>12th December 2014</td>
</tr>
<tr>
<td>Workshop 3</td>
<td>14th January 2015</td>
</tr>
<tr>
<td>HEDIIP Advisory Panel</td>
<td>6th February 2015</td>
</tr>
<tr>
<td>HEDIIP Programme Board</td>
<td>18th February 2015</td>
</tr>
</tbody>
</table>

The following summarises the information gathered from the stakeholder interviews.

4.5.1 **Data Collectors’ issues**

- Data Collectors are willing to discuss common requirements;
- There is no clear mandate from individual governance for collaborative working, but there is a willingness to do this;
- PSRBs are a very diverse group, their data collection arrangements are bespoke and they would be attracted to a proposition that provided data if it reduced their costs of collection;
- All collectors appreciate the need for common governance; and
- All collectors have legacy system issues or constraints, some of which are in varying states of transformation.

4.5.2 **Data Provider issues**

- Meeting Freedom of Information (FOI) requests is a significant issue;
- Providers have developed systems to meet the data requirements of the main Data Collectors, but the importance of institutions’ own data needs have evolved such that there are some differences now between internal and external needs;
- Providers are seeking to enhance the data collection and exchange experiences received by students within the HE landscape. Minimising the data collection burden on individual students and speeding up student data transaction related services will benefit students;
- Many institutions do not believe that the volume of data collected is necessary.
- For England FECs currently provide HE dataset to the FE Data Service (which is part of BIS) within the Individualised Learner Return (ILR) on a monthly basis. This data is then provided to HESA on an annual basis. For Scotland and Wales and Northern Ireland HE data is collected from their FEC organisations by their funding councils which is then shared with HESA;
- Alternative Providers are required to submit a reduced HESA collection annually in certain circumstances;
- Agree that data governance is required to reduce burden; and
- Would see benefit from the main Data Collectors cooperating on data standards.

4.5.3 **Government agencies and regulators**

- Are prepared to operate their own data collections and analysis where the data is not part of the HESA return;
- Are willing to be broadly supportive of data governance; and
Are not willing to be constrained by data governance arrangements through restrictive timing or scope of data, where this compromises their statutory responsibilities.

4.6 Analysis of the stakeholder responses against the key questions

All of the stakeholders were asked to provide responses to six key questions. We have summarised the responses against each question to provide a view of the breadth of responses.

1. What are the barriers to more cohesive and singular HE data governance?
   a) The self-interests and legal structures of the individual Data Collectors;
   b) Reluctance to change;
   c) Governance would potentially be easier amongst the main Data Collectors but more difficult to incorporate all of the PSRBs due to the amorphous nature of these organisations;
   d) Funding Councils need to rapidly enact policy changes and/or respond to the requirements of government;
   e) Enforcement of governance decisions given the number of autonomous bodies that collect the data;
   f) If it is easier to operate outside of the new governance body; and
   g) Lack of real accountability to the sector.

2. Should the New Landscape reduce the costs of data processing?
   a) The majority of the stakeholders agreed that it should but there was also a request for the New Landscape to rationalise the number of data collections;
   b) Yes but the implementation of change is expensive for Data Collectors and HEPs;
   c) It needs to increase the quality of the data;
   d) The New Landscape should have targets for cost reduction; and
   e) No as any savings should be redirected to spend more on data analysis.

3. What are the most significant challenges that need to be addressed?
   a) Frustration as stakeholders are becoming disenchanted with Data Collectors;
   b) Development of a clear and shared vision;
   c) Convincing the stakeholders of the cost benefit of the proposal;
   d) PSRBs largely operate independently;
   e) Influencing governmental demands;
   f) Development of standard definitions; and
   g) Complexity and breadth of collections.

4. What do you think are the quick wins that you believe are achievable?
   a) Clarification of what data can be collected by centralised data collection;
   b) Guidelines for Data Collectors to ensure that duplication is minimised;
   c) Pragmatic standardisation;
   d) Implementing a more permanent approach than HEDIIP;
   e) Implementation of the ULN; and
   f) Agreement regarding the new HE Data and Information Landscape.

5. What is perhaps desirable but probably unachievable in terms of the Blueprint?
   a) Gaining real consensus;
   b) Consolidating and rationalisation of data collections;
   c) One single data collection repository or agency;
   d) A voluntary model but in Scotland it would need to incorporate the Further Education sector; and
6. What would enable the various data demands to be more easily fulfilled?
   a) Centralised data collections;
   b) Internal clarification of good data management at an institutional level;
   c) More clarity on mapping the current data collections and documentation of definitions to identify data standards;
   d) Rationalisation of government requirements;
   e) Full implementation of ULN.

4.7 Stakeholder survey

To supplement our understanding gained from the stakeholder interviews we have developed an on-line survey to provide additional insights into a wider range of HEI data processes and their associated costs. The responses were provided on the basis that the HEIs would remain anonymous and that the results would be summarised. A more detailed summary of the results are included in Appendix H along with a list of all of the survey questions, the responses represent 17% of the HEIs. A sample of the responses is included below:

- The most common administration model from the respondents was centralised across all student types, with 67% of the respondents identifying with this administration type. In contrast, 11% of respondents operated a wholly decentralised administration model.

- Excluding a single respondent, institutions do not believe they have significant influence over any of the Data Collectors. Of the collectors, it appears that PSBRs and UKVI are perceived by the institutions as collectors that they have the least influence over. HESA and the Funding Councils have the highest number of institutions that believe they have some or significant influence over them, this is followed by UCAS and then SLC. Key ways in which influence can be improved:
  – Improved communication, via working groups, client contacts, workshops and forums.
  – Greater transparency around the need for data and more detailed specifications.

- Results varied across Data Collectors in response to a question regarding the ease of fulfilling a Data Collector’s data request, with HESA, Funding Council and PSBRs receiving significantly more ‘hard’ ratings than ‘easy’. Of all organisations, the highest number of respondents found it hard to fulfil HESA’s requests.

- Key actions that collectors could take to reduce the burden of data collections on institutions include:
  – Consistent definitions of data;
  – Clearer guidance;
  – Collect only necessary data and explain why the data is relevant;
  – Streamline collections;
  – Recognise challenges of unique organisations; and
  – Synchronisation and coordination or requests between organisations. Do not make last minute or sudden changes to the data requirements – stability of requirements.

- Key actions that institutions could take to improve the process of compiling and submitting student data returns to the key collectors?
  – Invest in relevant IT systems and specialist staff;
  – Increase automation;
  – Increase centralisation;
  – Ensure data is actively managed;
  – Streamline processes;
  – Improve data quality; and
  – More training.

The findings from our survey have been used to guide the development of the New Landscape.
Summary of the Issues with the current Landscape

The issues raised by stakeholder groups are complex but we have summarised them below by stakeholder type.

Main Data Collectors

These are independently governed and focused on the delivery of their mandates. Whilst they are all willing to be part of discussions regarding the barriers to change there are some ‘no-go’ areas according to their business priorities.

We will need a form of governance that can be accepted by autonomous organisations but provide opportunities for a meaningful rationalisation of the HE data collection. We foresee that over time PSRBs will transition to the new data landscape. In addition we need to recognise the real difficulties and political sensitivities for the SLC and UKVI as they will require a mandate from central government before committing wholly to the New Landscape.

Higher Education Providers

Providers are seeking to have a reduction in volume of data collection and in the burden of data collection, but the variety across providers makes it unclear if they all experience ‘burden’ in a similar way. They are aware that they have significant representation across the Data Collectors but as a sector they do not clearly benefit from this as their influence is fragmented.

HE providers require access to a forum that will enable them to influence the on-going development of the HE data landscape on an equal footing with the main Data Collectors. They also require tangible actions that addresses their concerns. A key challenge for the governance will be its interaction with APs who do not have a history of collaborative working. Engagements with FECs should be with their data collection agencies and or representative bodies.

Data Governance

We have identified that the absence of data governance across the Data Collectors contributes to inefficiencies in data collection and the burden experienced by HEPs. The independence of the stakeholders requires that the governance should be non-binding but all providers can be encouraged to work towards an agreed set of principles for the benefit of the sector.

Non-binding governance would be acceptable against an agreed set of data management principles. No existing structures would be suitable in their current format. Documentation of data standards and a repository of Data Collectors would contribute to a better data management approach.

Students

There is a need to identify and target initiatives that will drive benefits to HE students.

The governance arrangements should incorporate on-going dialogue with students to get a better understanding of their needs.

Data Flows

The absence of comprehensive cross data collector data governance and the absence of a timely central data collection has resulted in the inability to consolidate data returns due to the timing of the main data collection.

The proposed development of the HESA return into an in-year data collection could provide the opportunity for consolidation of data returns.

The development of a standard data set that is collected through an in-year data collection would provide an opportunity to consolidate data returns.

Transparency of data usage

There is a concern voiced by providers that they are unclear about data being used for a purpose for which it was not intended when originally collected. This concern can be seen to be inconsistent with this concern is effectively covered by the data principles.
5 The New Landscape

This section describes the building blocks that are required to develop a New Landscape.

During the development of the New Landscape we have identified a Vision Statement for the strategy to deliver. Utilising the feedback from our stakeholders and the purposes of the development of this strategy we have created the following vision for Higher Education data collection.

‘A data and information landscape for Higher Education in the UK that has effective governance and leadership, promotes data standards, rationalises data flows, maximises the value of technology and enables improved data capability.’

The need to maintain the independence and autonomy of the sector is of paramount importance to the successful adoption of the New Landscape. There are a number of building blocks required to deliver this vision. These are summarised in the diagram below and explained further in the remaining chapters.

Figure 10: Diagrammatic representation of the New Landscape

Data Principles – Will be used to provide a shared understanding of the ambitions of the landscape and an agreed framework for the Data Collectors and HEPs across the landscape to adhere to (Section 6).

Governance – Effective governance is central to the development of the landscape. It will provide the administration of the data standards, the inventory of data collections and Data Collectors. It will also provide oversight of the adherence to the Principles (Section 6).

Data Standards – Would be provided by the development and publication of the common data specifications for the data being collected within the HE sector. The publication of the data standards would provide all Data Collectors with an opportunity to ensure that the data that they require is not already being collected before they embark on additional data collection, thus minimising duplication. It will also enable retrospective data collection to cease (Section 7).
Standard Dataset – This would be the collective name for the data items that would be part of the landscape (Section 7).

Data Flows – Describes an optimised approach for the exchange of data between Data Collectors and HEPs. The development of the data flows in the New Landscape will be strongly influenced by the data collection approaches that are adopted. The potential change of the HESA return to an in-year data collection will enhance the opportunities for rationalising the data flows (Section 7).

Capability – The need to raise data management capability across Data Collectors and HEPs (Section 8).

Implementation plan – There will be a need to agree a structured and resourced action plan for the implementation of the landscape. Given the previous work with data that has been undertaken, it is of paramount importance that the New Landscape is implemented (Section 8).

We are aware of the HESA plans for an in-year data collection which would provide opportunities for the establishment of the New Landscape. In addition the New Landscape will build on the work previously undertaken in the HEDIIP projects to support the implementation of the New Landscape.
6 Data Principles and Governance

We have developed a number of data principles to assist in the enhancement of the HE Data and Information Landscape. These principles are based on The Open Group Architecture Framework (TOGAF) but have been aligned to meet the requirements of the sector.

6.1 Description of principles

For any data and information landscape to be effective, there needs to be a set of guiding principles in place that determine how that landscape operates, what the scope of the landscape is and how organisations within the landscape should interact. For the current landscape these do not exist.

The principles set the scene for the overarching framework around the New Data and Information Landscape, and provide the framework for governing how the Blueprint will operate and the expectations of Data Collectors and HEPs. They also provide a cornerstone of the Implementation Plan, and provide clear direction for how key elements of the Blueprint, including the Standard Dataset, should be developed.

These principles have been developed from the TOGAF Data Architecture principles, and further amended following group workshop sessions. TOGAF is a methodology for developing a fully-formed enterprise architecture, comprising business, data and technology architectures. It uses standard terms and approaches, and is based on a core set of principles around the development, implementation and management of an enterprise architecture. The full principles document is in Appendix H.

For these principles, and therefore the New Data and Information Landscape itself, to be effective key stakeholders will need to agree to operate in line with these principles, and adapt their operations accordingly.

The Principles are:

- **Effective Governance**: Data collection requirements in the stakeholders and Higher Education Providers need collective governance that is free from bias to allow for effective implementation and delivery of a New Data Landscape. This governance should be delivered by representatives from HEPs and Data Collectors from across the sector, and represent sector views, with student views represented by the NUS.

- **Adherence to Principles**: These principles of information management apply to all bodies exchanging data with HE sector organisations;

- **Maximum Benefit to HE Sector as a Whole**: Information management decisions are made to provide maximum benefit to the HE sector as a whole including students;

- **Information Management is Everybody’s Business**: All key stakeholders within the HE Sector participate in information management decisions needed to accomplish business objectives;

- **Compliance with the Law**: HE information management processes comply with all relevant laws, policies, and regulations, including competition law;

- **Data is an Asset**: Data is an asset that has value to the HE Stakeholders and is managed accordingly;

- **Data is Shared**: Users have access to the data necessary to perform their duties or answer their query; therefore, data is shared across Stakeholders where permissible and appropriate;

- **Data Trustee**: Each data element has an owner and a trustee accountable for data quality;

- **Common Vocabulary and Data Definitions**: Data is defined consistently throughout the HE sector, and the definitions are understandable and available to all users;

- **Data Security**: Data is protected from unauthorised use and disclosure. In addition to the traditional aspects of national security classification, this includes, but is not limited to, protection of sensitive and proprietary information;

- **Technology Independence**: Applications are independent of specific technology choices and therefore can operate on a variety of technology platforms;
Responsive Change Management: Changes to the HE information landscape are implemented in a timely manner;

Interoperability: Software and hardware should conform to defined standards that promote interoperability for data, applications, and technology; and

Protection of Competitive Position: Data that is viewed as compromising the competitive position of HEPs will still be collected in year, but will not be available publically, nor to other collectors (save for that data required for them to fulfil their requirements) until it is no longer viewed as commercially sensitive.

6.2 Governance of the New Data Landscape

Our analysis in Section 4 identified that a common issue with the current landscape is a lack of effective governance of the landscape and data as a whole. The Principles by which the New Data and Information Landscape should operate are central to its success. All key stakeholders will need to adapt their operations to a certain extent, and agree to work together and use a common dataset and standards, if any kind of collective governance of the New Landscape is to be achievable.

Without this, there will be a rapid return to the current situation, as different organisations request extra data fields, different definitions, or other data management processes as immediate responses to their own local needs without consultation with wider sector stakeholders.

Governance of this is therefore key to the initial and continued success of the New Landscape and the Blueprint.

Ideally to deliver maximum effectiveness, governance of a landscape such as this needs to be binding on all parties involved. This would allow the governing body to take potentially controversial decisions around contentious decisions relating to the Data and Information Landscape which can be enforced.

Therefore our starting point when thinking about potential governance of the New Landscape was that it should be binding. Following a number of workshops with stakeholders from across the sector, it became clear that the sheer number of different, independent organisations involved made this unrealistic without legislation to implement it.

In the absence of such legislation to create a new governing authority for data, to govern the data and information landscape and the key stakeholders within it, there are a number of options that can be considered:

- **Collective governance.** This would see all stakeholders involved in the New Landscape sign up to a memorandum of understanding on how they would operate their data collection regimes, and agreeing to the Principles. Any changes to the Principles or Standard Dataset would need to be agreed by everyone, and signed off collectively.

- **Delegated Governance (External).** This would see the stakeholders delegate their authority around agreed areas of the New Landscape to an external body, who would arbitrate disputes or requests for change to the Landscape, Principles or Standard Dataset.

- **Delegated Governance (Internal).** This would see a similar arrangement to that above, except that the body with delegated authority would remain within the sector. It could potentially be hosted within HESA, within another sector body, or be set up as a separate organisation.

These are evaluated below.

**Collective governance** would allow every party to be represented in major decisions, which would increase the likelihood of those decisions being adhered to, and organisations operating for the ‘collective good’.

However, this approach could also be unwieldy, and the likelihood of garnering agreement on any major element would be extremely low. This would, in turn, encourage organisations to act in a more unilateral fashion as they struggled to get agreement on their proposals, which would negate the whole impact and purpose of a collective approach. For this reason we have discounted this option.
Delegated Governance (External) would mitigate the risks of collective governance, in that organisations would agree to delegate their authority over issues in this area. Decision making would be more streamlined and take less time, although a mechanism for challenging decisions would also need to be implemented.

The key problem with this option is that it would be necessary to constitute a new organisation, direct funding towards that organisation, and build up the infrastructure that new organisation requires. This would impose an unnecessary cost overhead on this option that we felt was unsustainable, hence our discounting of this option.

Delegated Governance (Internal) brings all the advantages of the previous option, but mitigates the disadvantages by siting the governance organisation within an existing organisation. The most obvious candidate to host this organisation is HESA, although it does not necessarily need to be.

In this scenario the governance organisation would be hosted by another organisation but not report to it, in much the same way HEDIIP is currently hosted by HESA. This maintains the independence of the governance body, but allows it to function and operate in the most efficient way possible.

The preferred option is to host a delegated governance body within HESA, as this takes advantage of existing structures, and allows for building on existing knowledge of data within the sector. This preference was confirmed during the course of several workshops, with participants feeling it represented the most pragmatic and workable approach.

To make this structure and modus operandi work, four things will be needed:

- **Commitment from all key stakeholders** that they are prepared to delegate some of their authority and data requirements in the sector to this governance body, and that they will abide by the decisions taken and will not act unilaterally. Consultation with the governance body, and other stakeholders is imperative if this structure is to work, and this should be outlined and agreed when stakeholders sign up.

- **Key sector bodies and HEPs should have equal representation on the Governance Body.** Those that are not represented directly should be linked in to the main body via appropriate sub-committees or other consultation mechanisms. This will allow for the widest consultation on data matters, and should avoid situations whereby unilateral decisions and actions are sprung upon other sector bodies without consultation.

- **Terms of Reference** will need to be drafted and agreed by all representatives. This will need to happen as part of the development and agreement of the exact representation, structure and scope of any governance body. Terms of Reference will include:
  - Discussion of proposed amendments to a Standard Dataset; and
  - Discussion of proposed amendments to the standard definitions that support the Standard Dataset.

- **Agreement of membership of the main body, and any sub-committees.** The balance of this between different types of sector body will be key to its success. Initial indications are that an equal balance (50% HEP, 50% broader sector representation) is preferable and most likely to see success. This approach has been arrived at following discussions with a number of organisations across the sector, including a number of HEPS. The general feeling from HEPS is that they already have a significant burden of data requests, and if that is to be reduced, and kept manageable, they would like to have a majority on this body. Furthermore, they feel that they will be better able to explain the actual impact of a particular data request, and clarify its meaning appropriately before it becomes a requirement, thereby improving the quality of the data collected. NUS should have representation on the management Board. It should also be ensured that there is UUK/GuildHE representation.

Key sector bodies such as HESA, HEFCE, UCAS, SLC, GMC, HEE, NCTL, UKVI, RCUK and others will also need to be represented directly. Other bodies could be represented by similar organisations with direct representation. Agreeing this is one of the key initial tasks of any implementation workstream.

It is anticipated that over time more PSRBs will migrate to the New Landscape. The Governance structure enables a separate PSRB group to be formed as a ‘consultative group’ enabling their views to be represented.
6.2.1 Data Governance in the New Landscape

Despite the lack of clear mandate there are instances of where UCAS shares data with HESA and opportunities for UCAS to collect and share data with the SLC. HESA through its statutory mandate will share data with a range of its statutory customers. There are also opportunities for UCAS and HESA through their commercial arms to sell the processed data collected to third parties and back to the HEPs.

Data Governance is a strategic function and can be described in many ways. However, for the purposes of this report, we suggest that a more optimal HE landscape will need to provide management of the availability, usability, integrity, and security of a Higher Education dataset. For pragmatic reasons we would suggest that this Higher Education dataset would be called the HE Standard Data and this will be discussed in Section 6.3.

Examples of Data Governance tasks would be:

- Management of the governance Infrastructure – committees, working groups, meeting management;
- Developing and refining the goals and principles – establishing a process cycle for the evaluation and review of the goals and principles;
- Communications and brokering across stakeholders – communication of the policies and roles and responsibilities;
- Promotion of the Principles – pro-actively promoting the data principles with HE stakeholders;
- Enable transparency – Establishing formal opportunities for stakeholders to be challenged about their adherence to the principles, and conversely providing stakeholders with opportunities to challenge the principles;
- Establishment and management of the HE Data Stewardship function – Data Stewardship is the operational element of the governance and the role will/is likely to include functions such as:
  - Defining the standard data – Identifying standard data, gathering and documenting definitions;
  - Defining a business process – For creation of standard data for the approved use of the data,
  - Documenting data sources – Publishing and promoting the use of standard data;
  - Maintaining the data inventory;
  - Governing the scale of data requests placed on HEPs, whereby new requests from collectors are considered by the Governing Body prior to implementation.
  - Evaluating and monitoring data quality – Fit-for purpose parameters, documenting Data Collectors quality measures;
  - Identifying and recommending protocols for sensitive data;
  - Identifying data assurance processes;
  - Monitoring the range of data collected by different collectors;
  - Collating and documenting a register of HE data collections; and
  - Promoting good practice data management.

We have set out below in figure 11 to illustrate our approach to implementing effective Data Governance of the data and information landscape which would include the documentation of all data captured within the HE landscape. Whilst we appreciate that there are several options for the implementation of governance our approach has been designed to incorporate the most successful features of the HEDIIP approach and avoid the weaknesses. Our governance arrangements are consistent with the ambitions identified within pathway to reform (Appendix E).

HEDIIP itself is not the intended vehicle for this governance approach, as HEDIIP is an interim change and project management function. As the project progresses, however, it would be sensible to review, periodically, the overall governance arrangements of the New Landscape, and HEDIIP to minimise duplication across the different areas. Depending on the future for HEDIIP and the constitution of the new Governance body, HEDIIP could be collapsed into the governance body.

The planned remit for this new governance approach will be the whole HE data landscape. However, initially the focus will be on the establishment and development of the Standard Dataset. There will be a need to build momentum with the key Data Collectors and providers whilst developing links with PSRBs and Alternative Providers to enable them to be appropriately involved in the development of the landscape.
Figure 11: Governance of the New Landscape

HE Data and Information Landscape Management Board (Strategic direction)
- It would provide direction to the function aligned to the Principles and the vision for the landscape.
- It would agree the scope of the Standard Dataset.
- It would govern data standards across the HE landscape both for existing and new requests from the Data Collectors.
- It would consist equally of 50% of the Data Collectors and 50% of HEPs. Only Data Collectors and institutions signed up to comply with the Principles would form part of the board. Instances of serious organisational noncompliance with the Principles could result in removal from the board. NUS and UUK/GuildHE should also be represented on the Management Board.
- The Chair would be from a HEP or from a complementary sector. The selection would be supported by UUK in consultation with GuildHE.
- Individuals from HEPs would only be allowed a three-year term, to allow for wider inclusion. Initially this would need to be staggered to remove the risk of mass changes.

HE Data and Information Landscape Management Office
- Would be led by Service Director, who would provide the day-to-day leadership and co-ordinate the service delivery.
- Would be the operational unit for the communication of the operations of the office in defining the Standard Dataset and documenting HE data.
- Would contain specialist staff who would perform the specific data stewardships tasks required.
- HE Data Steward
  - Would be responsible for the development and documentations of data standards;
  - Would support business case production and project management for developments sanctioned by the board.
- HE Data Secretariat
  - Would publish documentation and maintain a transparent route of decisions making for access to all stakeholders;
  - Would provide the administration for the Management Office;
  - Would develop and publish the register of Data Collectors;
  - Would consult with the sector regarding new developments that are requested to meet the implementation of the Blueprint.
■ HE Data Technical Group (technical direction)
  – The Technical advisory group would consist of technical staff from the organisations that contribute to the management board.
  – Places would be open to other stakeholders such as SFA and the Learner Record Service who may have interest in the operations of the management.

■ HE Data Consultative Groups
  – Would provide an opportunity for wider stakeholder engagement, some of the groups could be interim and others could be standing groups especially in areas such as Students, PSRBs, FECs and APs to provide an in-depth understanding of their issues.

Benefits of this approach are that by centralising the generation of data specifications it should enable the HE and Information Management Office to support Data Collectors to align themselves into the governance process and the adoption of standard definitions whilst the data specification itself is owned by each organisation separately. This would provide:

■ Opportunities for the adoption of common standards;
■ Clear identification of where apparently similar things are different;
■ Single place to find data specifications;
■ Standard layout/look/feel of data specifications;
■ Standard approaches to things like XSD structures (potential for HEP systems to consume specifications and auto-build files?).

The main challenge to this approach is the perception of individual stakeholders "losing control" of their data specifications. However, this is mitigated through having representation on the HE Data and Information Landscape Board. Other challenges are likely to be

■ Resourcing the HE data and information landscape management office;
■ Agreeing a Standard Dataset sufficiently large to enable changes to the data flows whilst retaining consensus;
■ Maintaining consensus.

6.3 Establishment of a Standard HE Dataset

6.3.1 Why it is needed

Our analysis in Section 4, and our conversations and workshops with HEPs and Data Collectors, have highlighted that one of the most fundamental issues with the current situation is that there are so many competing datasets and data definitions for HEPs to respond to.

Several HEPs have told us that the bulk of their work in preparing a major data return comes in the ‘translation’ from what is being asked for to the types of data they hold, and then combining and transforming their data to meet the requirements.

The idea behind the Standard Dataset and definitions is that it will reduce that discrepancy significantly, thereby reducing the work of HEPs in delivering their data returns.

Alongside this, it will also create a Dataset that every Data Collector in the sector will use for their core requirements around student information. The idea here is to bring confidence to the different Data Collectors that data collected anywhere in the sector will provide them with the right information to undertake their business processes without having to collect it themselves.

This will allow HEPs to focus more on the quality of their data, and less on making it fit a particular return for a particular collector. It will also make the collection of in-year data easier to manage.
This will in turn create the foundation for moving to a New Landscape, and will underpin the options we outline below. Without the Standard Dataset, the options we outline in section 7 would be much more difficult, if not impossible in some cases, to deliver.

6.3.2 Composition of the Standard Dataset

For it to be effective, the Standard Dataset needs to be sufficiently broad and deep for different Data Collectors not to need to collect their own data. This means that just covering ‘Name’, ‘Address’, ‘Gender’, ‘Date of Birth’ will not be sufficient.

We do not try to define the Standard Dataset in this document, as it will be one of the key workstreams for the Implementation Phase.

Alongside the scope of the Standard Dataset will need to be standard definitions for data items. This is potentially even more challenging than establishing the Standard Dataset. During our workshops and stakeholder meetings, we have come across a number of areas where definitions differ across different collectors for ostensibly the same data items. From different definitions for ‘student’, ‘full time’, ‘part time’, ‘course’ across different collectors, to different concepts for data items, such as ‘sex’ or ‘gender’.

Because different collectors have different business processes and different reporting requirements, agreeing these standard definitions will be a difficult but important issue. Of course, the broader the Standard Dataset is, the more definitions will be required, and the more difficult it becomes to agree them. Data Collectors are open to this idea and therefore the concept has buy-in.

This will be another key focus of the implementation workstream.

6.3.3 Governance of the Standard Dataset

The Standard Dataset will need to be governed effectively, with a clearly defined change management process. It should be governed according to the Data Principles we described above.

It is certain that individual Data Collectors will request additions, amendments and changes of definition to the different data items within the Standard Dataset that they expect HEPs to respond to.

Without effective governance that moderates and manages these changes in a central, agreed fashion, individual Data Collectors will almost certainly begin going direct to HEPs for these extra, or changed, pieces of data which will very quickly lead the sector back to the situation it is now in. Data Collections will still need to exist outside of the Standard Dataset; however we would expect this to be minimised over time.

It is therefore of utmost importance to the effective and efficient functioning of the Standard Dataset that any changes to it are handled in a coherent fashion, overseen by effective, collective governance.
7 Blueprint for a New Landscape

This section details the different options for changing data flows across the sector, to deliver the Vision for the data landscape.

In considering the options below, we looked at existing data flows in the sector, and how these could be rationalised. We looked at the separation between data and services, and how the entanglement of these two areas could provide motivation to increase data collection to allow individual organisations to provide better services.

7.1 High level design of data flows against a theoretical target operating model

Drawing on our experience of modern models for data management, in the following section we provide a view of a theoretical target operating model for data management across the HE landscape which can be used to evaluate the existing data flows. Whilst we understand that the implementation of this target operating model would not be applicable within the current HE data collection environment, we suggest that the evaluation may provide useful insights.

Figure 12: Theoretical target operating model

This simplified vision of the future would see fundamental changes in the way HE data is exchanged and used today.

Features of this theoretical view are:

- There are no legacy or transition issues as it should be viewed as a conceptual model;
- The HE data repository would exist as a transactional data cloud (a data repository that enables data exchanges) that is owned by HEPs;
- A data cloud would exist to manage the agreed transactions between the HEPs and the Data Collectors. All Higher Education students would be centrally registered within the cloud and they would have a specific data file loaded with standard data associated with them labelled with a unique identifier (this could be fulfilled by a UK ULN);
All HEPs would access the student’s file “link” to that student in the cloud and download their details. For the duration of time the student is enrolled at the institution the student’s details are “checked out”. The HEP then populates the student’s data file with updates on their standard data. They would have the ability to record and retain their non-standard data and on a regular basis the student’s data file is synchronised with the cloud;

Students would have access to the cloud to register (in the case of international or non-native students) but also to update and validate their standard data file. However, all changes would need to be validated by approved Data Collectors;

The cloud would also need the ability to deal with multiple student checkouts whilst retaining the integrity of the data;

The cloud would allow for services and web applications to be developed over the data within the cloud and the data would be accessible to approved stakeholders for data analysis purposes. Other services could include;

– Admissions;
– Student finance;
– Registrations with professional bodies;
– Visa and Immigration controls;
– Destination tracking;
– Funding and performance management.

The cloud could become self-funding; as transaction charges could be levied to Data Collectors and additional revenue could be derived for developers of new services to access the data.

Data governance would exist across this independent body and it would have a mandate from HEPs to provide binding sanctions as it would licence the access to the cloud. The Governance would be provided by a separate body that was constituted from representative HEPs. Data ownership would reside with the HEPs processing their institutional data or HEP and the individual students jointly.

The table below demonstrates the differences between the theoretical model and the current approach.

<table>
<thead>
<tr>
<th>Area of comparison</th>
<th>Features of required solution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data flows</td>
<td>A central data collection would appear to be the optimal solution. However, if we acknowledge that the specialist data requirements may make this impracticable then we would suggest that a ‘Standard Dataset’ is collected through a single data collection. This Standard Dataset would need to contain the most commonly collected elements and this would reduce some of the duplication of collection.</td>
</tr>
<tr>
<td>Governance</td>
<td>In the absence of legislation, the only way to implement governance with the range of Data Collectors in the sector is through a non-binding agreement. HEPs will have a role in ensuring that the sector’s Data Collectors comply with the consensus.</td>
</tr>
<tr>
<td>Technology</td>
<td>The consolidation of data flows does not necessarily require a cloud based solution. However, there is an opportunity through the changes planned at HESA to provide an infrastructure that optimises the possibilities of consolidation and maximises interoperability.</td>
</tr>
</tbody>
</table>

7.2 Evaluation of the Blueprint options

We require a Blueprint for a new HE data landscape that will enable our vision to be achieved,

‘A data and information landscape for Higher Education in the UK that has effective governance and leadership, promotes data standards, rationalises data flows, maximises the value of technology and enables improved data capability.’

© 2015 KPMG LLP, a UK limited liability partnership and a member firm of the KPMG network of independent member firms affiliated with KPMG International Cooperative (“KPMG International”), a Swiss entity. All rights reserved. Printed in the United Kingdom.
The Standard Dataset will become recognised as the exemplar for Higher Education data quality in that it will provide the right data in the right format at the right place and time to meet the needs of its stakeholders.

In addition we will need the Blueprint to support the following requirements:

- **Rationalisation of data collections** – a reduction in the same data being collected by different collectors; and,
- **Reduction of data collections** – where standard data can reduce the need to a specific Data Collector to develop or continue with their own collection.

Building on our diagram in section 7.1 that describes a theoretical landscape for Higher Education data collection, we now consider a number of options along a spectrum of highly centralised data collection, where the collection of data is completely distinct from the services provided using that data, to a completely federated system where data is pulled from HEPs when required instead of being pushed by HEPs when requested.

In addition to the different options we considered was a ‘do nothing’ option, which was discounted early on following consultation with a range of stakeholders. In arriving at the three options outlined below, we spent a lot of time in stakeholder meetings and workshops discussing and refining different options, and discounting others as unworkable based on a wide range of stakeholder feedback around the constraints of existing organisations’ business models and transformation programmes they were already undertaking.

The options outlined in this chapter have all had at least some degree of stakeholder acceptance, although some options have garnered more support than others.

Central to all the options, and any variations on them, is the establishment of an agreed Standard Dataset. Currently there is significant overlap between the different data collections carried out by providers, both in terms of data fields they ask for as part of the collection, but also in terms of the definitions they use for each data field. In addition we would expect that any Data Collector would continue to interface directly with prospective, current or ex-students in connection with their legitimate business operations.

As discussed above, this is a fundamental issue that simply must be addressed to move to a system where in-year data collection is possible and data quality, and the understanding of what that data represents, is improved significantly.

Without a Standard Dataset, that everyone can use, the proliferation of data collections and data field definitions will continue, and our options below are predicated on the successful agreement of this Standard Dataset.

The Standard Dataset will require significant input from all stakeholders to define and agree its scope, along with the data standards that will sit alongside it. In the implementation plan in section 8 we outline the steps to agreeing the Standard Dataset and provide a suggestion below for how this could be approached.

A principle for the Standard Dataset is that it represents data that is reported to multiple Data Collectors by a large proportion of the sector. To this end many stakeholders have agreed that the current HESA Student Return should form the basis for determining the Standard Dataset. Work should be undertaken to map the HESA Student return field and definitions to the returns of the other Data Collectors to identify either common data fields or fields that are similar. Through the New Data Landscape Management Board, agreement should be reached collectively with the key Data Collectors over the content of the Standard Dataset and the data definitions.

When evaluating each option, we used several criteria:

- Impact on students and applicants;
- Ease of implementation;
- Cost of implementation;
- Potential impact on Data Collectors;
- Potential impact on HEPs;
- Ease of operation;
Stakeholder acceptance.

We examine the three options identified through workshops and stakeholder interviews against these criteria. Each option would benefit from the completion of the existing HEDIIP projects but their outcomes should be refined to ensure they are completely aligned to the requirements of the New Landscape. We would also expect that the Funding Councils support the transferring of the HE in FE data on a more timely basis to enable there to be a single and comprehensive repository of HE data in the future.

7.3 Option 1

The following diagram portrays a possible ‘future’ data environment for students, providers and strategic users of data. The principle behind this model is that it tries to deliver a complete and comprehensive data repository that is separate from the services provided using that data.

**How will it work?**

This potential future landscape envisages the separation of data collection from services. Thus the collection of data necessary for admissions would be carried out by the data collection service, while the provision of the admissions service would be carried out by one or more admissions services providers with direct access to the data held by the HE Data Collector.

Under this model, all university undergraduate and postgraduate accredited courses would see application data being collected by the Data Collector, but processed by the application service provider(s). The integration of the Unique Learner Number (ULN) will facilitate easier integration of students who have come through the UK education system into the HE sector by maintaining the unique identifier for learners.

Practically this would mean a single portal, managed by the Data Collector that would collect all the necessary data for applications and pass that over to the application processors.
Students would go to a single web portal to apply for any accredited taught course in a UK university at undergraduate or postgraduate level. This would be run and managed by the Data Collection organisation, and capture all the information necessary to process an application. The Data Collection organisation would also provide the relevant support to people filling in the form.

This data capture process would be managed by the sector Data Collector, but no application processing would be carried out. As the data is captured, it would be made available to the appropriate application process provider (of which there may be many, and which could be a HEP), who would then make contact with the applicant and manage their application from there, and any interactions with the HEP that were necessary.

Once the applicant has enrolled as a student, the HEP will collect and manage student data in the normal way, but will only submit it to the HE Data Collector, who will make it available to other Data Collectors as appropriate and required for their business processes.

The key here will be agreeing the standard data that needs to be collected up front, and subsequent data that the application processors may need to collect, and where the boundary lies.

### 7.3.2 Challenges

This model would clearly necessitate significant change within the sector around how applications are handled and in the technology used by the Collectors and HEPs. There would need to be agreement on a single platform for applicants to use to input their data, and some form of support for that.

There would also need to be agreement that international and postgraduate taught students would apply via this route, rather than direct to particular HEPs, although the latter may be possible also.

While these are not inconsiderable barriers, there are also significant potential advantages to this landscape for the applications and admissions processes.

### 7.3.3 Advantages

The key advantage would be that this model would force standardisation in data collections for all student and other data capture, which will reduce burden significantly and improve management information for Providers. It may also provide the potential for the applications process to be opened up to competition, allowing other providers to take advantage of the applications data and offering different value added services to Providers.

Other advantages include the fact that HEPs will only have to submit the standard data to one provider; this submission would be via one technical route, and the data flows themselves would be less complex and more standardised; thereby reducing processing errors.

As data standardisation and technology improved, a ‘pull’ model could be adopted, whereby the HE Data Collector pulls data from the HEP instead of waiting for the HEP to push it to them.

For the other processes in the landscape, once the data has been collected, it can be used and processed by the relevant agencies. The key element of success for this aspect of the landscape is the agreement of a Standard Dataset that provides all relevant fields for each agency or organisation so that they do not need to be primary Data Collectors but can focus instead on delivering their value add services.

### 7.3.4 Evaluation and impact on different stakeholders

Using the evaluation method described above, we have evaluated this option as follows:
### Impact on applicants and students

During the transition to this new model, the impact on applicants is potentially huge. If all went smoothly, home/EU students should not notice any discernible difference, as their interactions would be similar, just with a different organisation behind the scenes. All international applications could also flow through this landscape.

Indeed, with a single portal it could be possible to process the application and student loan request as part of the same process.

However, this is only true for the initial data collection. Following that, they would be in contact with the provider of application services who work with a particular university. In theory, there could be several of these, leaving students to deal with different organisations to process their application to different HEPs.

This is not an ideal situation and one that is likely to add confusion and complexity to an already stressful time for applicants.

For students at university, the only potential impact would be HEPs reassigning staff, freed up from the administrative and data manipulation tasks required of the current environment, to student-facing tasks that improve the student experience.

### Ease of implementation

This option would be extremely complex to implement. The changes required at organisational level across the application and admissions element of the sector are significant, as are the changes required to most of the processes around application and admissions.

While the process for Home/EU undergraduate applications would remain largely the same, responsibility for operating the initial application process, including data capture and applicant support, would move from UCAS to the Data Collector. This would require moving all the staff who handle that part of the process to the Data Collector, with all the associated TUPE implications, and potential for disruption to the service.

For international and taught postgraduate applicants, the application route varies depending on the HEP they are applying to. Some may apply directly, some may apply via a variety of agencies and other private organisations. Whichever route applicants currently take, it is almost certain that this option will remove it and replace it with the single application route for the data as described above.

This option would require a significant change to the role and scope of HESA.

In both these instances, there is an expectation that organisations who currently collect application data cease doing so directly.

### Cost of implementation

The cost of implementing this solution is likely to be extremely high. Given the complexity outlined above, significant legal and transformation expertise is likely to be required centrally to manage and co-ordinate the transformation across the different organisations involved, which will increase the costs significantly.

This is further complicated by the number of different transformation programmes being undertaken by different organisations at the present time.

### Potential impact on Data Collectors

Taught postgraduate and international applications would also move into this new service, with significant implications and/or opportunities for the private organisations who manage those services. Currently there are a number of different private organisations who collect data on international and postgraduate applicants, and manage the process of identifying potential applicants, assessing their suitability and putting them into the application process. These organisations would not want to give up the commercial advantages that have access to these data offers, making this option difficult to implement from this aspect.

For those organisations involved in collection of applicant data to facilitate their application, the impact is extremely significant, as outlined above. For these organisations, their business would be completely changed. For the rest, the main impact would be a significant dilution of their direct data collection role with HEPs, as this would be handled by the data collection organisation, and the data made available to them. This will, however, reduce burden and complexity for the HEP.
Potential impact on HEPs

The impact on the HEPs would be negligible for Home/EU undergraduate applications; they would simply be dealing with a different organisation. For international and postgraduate taught applications, however, the change would be bigger. Currently different HEPs (and even different faculties within HEPs) have different processes for international applicants, and applicants for postgraduate taught courses. At least the initial application element of this process would be moved to the new Data Collector, although the subsequent application and admission service would still be carried out by these service providers.

A benefit to HEPs would be the reduction in the number of Data Collectors to whom they would need to submit the same or similar data. The central Data Collector would handle the relationship with the HEPs, and the data handoff to the data service providers described above (applications, student finance, etc).

Ease of operation

Once implemented, this option would be relatively easy to operate. The central Data Collector would manage all relationships with applicants, HEPs and data service providers, and would be the ‘Hub’ of the data and information landscape.

Much of the data collection would be web based, and probably automated, which would reduce errors and improve overall data quality. It would also provide an element of flexibility and future proofing.

Stakeholder acceptance

The level of stakeholder acceptance of this option is extremely low at present. The organisations involved in the enquiry/admissions data collections have indicated that the impact on their business model and commercial interests would be very significant. They have indicated that they would not be prepared to implement this model.

While there has been some support for this model from HEPs, it has been very limited, and most of those we have spoken to have indicated that they would not be prepared to make the significant changes to the enquiry/admissions process for international and taught postgrad students that this option envisages.

7.3.5 Summary

While this option does have significant advantages, not least that the collection of data would be separated from the services provided with that data, given the complexity, cost and lack of stakeholder acceptance, we do not recommend it represents a viable way forward at this juncture.

Given the level of change required to move to this landscape, it has been positioned as something that could be considered at a later date. However, it is useful to generate debate about what is desirable and achievable, and for examining different roles within the sector for current organisations.

7.4 Option 2

The diagram below envisages a situation where HEPs have improved their own data management processes and technology, based on the Standard Dataset, allowing individual Data Collectors to pull data from HEPs as they require it. This is different from Option 1 in that it does not envisage a central data collection organisation and does not seek to alter any of the current processes for collecting or processing applicant and student data.
7.4.1 How will it work?

This option envisages the continuation of the existing system, with a Standard Dataset agreed and technology solutions implemented that allowed for automatic transfer of data to those organisations that required it, within their timescales.

With a Standard Dataset in place, and standard definitions around the data being captured, the requirement to carry out lots of different transformations for all the individual data collections is removed somewhat.

This would allow the HEPs to have some certainty around their data returns, and improve their internal data management processes to focus on this one data model. This in turn would make it easier to develop a technology model that allows collectors to collect when they wanted, automatically. This may remove the need for focusing on a particular ‘return date’ and would allow collectors to meet their own timelines and requirements. However, appropriate protocols would need to exist over the data collection periods.

The focus then becomes maintaining control over the Standard Dataset and definitions such that HEPs are not being expected to deliver improved technology as well as multiple data returns.

In this model, HEPs would need to implement a more continuous and consistent data collection, cleaning and transformation process. They would also need to implement some form of system, firewalled off from their core infrastructure, into which Data Collectors could gain access to pull the standard data whenever they required for their business process.

7.4.2 Challenges

The key challenges around this option are those of security and multiplicity of relationships for HEPs.

Security is a key concern with this option. Allowing Data Collectors access to HEP systems to pull data as and when they require it means that roles and permissions need to be set up and agreed, and some form of identity management needs to be in place in the HEP to facilitate this. Given that student data will be accessible to external organisations, HEPs will need to be very sure of the access model before they open it up.
Relationships are another key area of concern for HEPs. They see the number of relationships that they have to maintain with different Data Collectors as a burden, and one that they would like to see reduced. This option does not envisage any reduction in that number, although it does envisage a reduction in the complexity of those relationships.

Capability as this approach would require all Data Collectors including PSRBs and HEPs to be able to develop and maintain the appropriate infrastructure. It is also unlikely that HEPs would accept that Data Collectors could collect data when they wanted and therefore return dates would still need to exist.

7.4.3 Benefits

The benefits of this option centre around the need for HEPs to make sure their data is clean and complete on an ongoing basis. Given the implementation of the Standard Dataset across all of the options, the implementation of a pull method would relieve HEPs of trying to hit different dates for different returns for different collectors.

It also delivers much more timely data, making it more useful for policymakers and funders.

7.4.4 Evaluation and impact on different stakeholders

Using the evaluation method described in section 7.2, we have evaluated this option as follows:

<table>
<thead>
<tr>
<th>CRITERION</th>
<th>COMMENTARY</th>
</tr>
</thead>
<tbody>
<tr>
<td>Impact on applicants and</td>
<td>Home/EU applicants would not notice any change in their process or who they engaged with, as that element of the process would not change.</td>
</tr>
<tr>
<td>students</td>
<td>We do not envisage significant changes in the way the international and taught postgraduate applications and admissions processes work either, but the technology changes required under this model could result in an improved experience for these constituencies.</td>
</tr>
<tr>
<td></td>
<td>The improved technology and data processing behind it will mean that applicants and students will have access to much more up-to-date information about themselves as they relate to their institution, and other organisations in the sector with whom they need to have contact, such as the Student Loans Company.</td>
</tr>
<tr>
<td>Ease of implementation</td>
<td>This option presents some technical challenges, and will require some significant changes within most HEPs in how they manage their data, which increases the complexity slightly. Currently most HEPs have multiple systems and processes handling student data. Departments, Faculties, Schools may all have their own systems and processes for this. For this option to be effective, they will need to change how they do this, such that they have one data model, one overarching process for collecting and managing student data, and one system to manage the final data return that collectors will pull. Technology wise, there will need to be significant changes to implement this model. Security protocols will need to be agreed, management of roles and permissions will need to be co-ordinated, and the Standard Dataset and definitions will be key to making this work. However, there is no need for significant organisational change, and no requirement to change the data flows themselves, thereby increasing the ease of implementation.</td>
</tr>
<tr>
<td>Cost of implementation</td>
<td>Implementation costs will be very high, when viewed collectively. However, there will not be the need for significant, centrally provided transformation support, as most of the cost will be in technology refreshes at HEP level. Ongoing costs will drop from current levels, as HEPs will no longer need to have a large number of FTE dedicated to managing the whole process of collecting data internally, cleaning it, interpreting the central Data Collectors’ intentions for a particular data requirement, and transforming various data fields to meet that requirement. However, the longer it takes to change the current culture around data collection, management and reporting in HEPs, the longer it will take to realise these benefits. From a sector perspective, the costs for the whole project will be lower than Option One, as the level of upheaval and requirement on all parties to get to grips with a completely new system will be significantly lower.</td>
</tr>
</tbody>
</table>
The impact on Data Collectors will be limited to the technology and process changes they have to make to allow them to collect data from HEPs instead of having it pushed to them by HEPs. This should not represent a huge change for Data Collectors although there will be some system rearchitecture to facilitate it.

More significant will be the common dataset, and agreeing what data fields are in that, the definitions around them, and how it should be governed. Currently there is no agreed common dataset, although the HESA return does collect data on behalf of a number of organisations.

With this option, the governance will be crucial, as each organisation is still collecting its own data, so it would be relatively simple for them to add different data fields to their collections, which would rapidly see a return to the situation we have now.

The impact on HEPs would be limited to the changes they would need to make to their internal processes and systems. That is not to say that this change would not be significant in some instances, due to the need to have data more up-to-date and more accurate, more often.

The amount of change required would vary across HEPs, depending on their own internal systems and processes. Those who already have a well-defined process for collecting, consolidating, transforming and submitting data, and who have invested in technology solutions such as data warehouse and sophisticated reporting solutions, will likely not need to change too much.

Those who still rely on very manual processes and lots of manual databases, Excel or Access applications to get their data in a good enough state to submit will need to enact significant programmes of change.

HEPs would still have to manage relationships with the individual Data Collectors, but the implementation and effective governance of a Standard Dataset should mitigate some of the issues that they currently experience, which would reduce burden.

Once implemented, this option would be relatively easy to operate. HEPs would have a data reporting infrastructure that would be available to external users to come and get the data they required, on a roles and permissions based authentication infrastructure.

With the Standard Dataset, and HEPs moving towards an automated data warehouse type regime, overall data quality should improve.

It is important to note, however, that it would be easier for individual collectors incrementally to increase their own data collection compared to the Standard Dataset, leaving HEPs potentially facing a similar situation to the current one, but with far greater expectations on data management and frequency of collection. The investment in this option is not insignificant, and flexibility and control does not increase, as there are still multiple collections of the same data happening across the sector.

The level of stakeholder acceptance of this option is relatively low. HEPs have indicated that they feel uncomfortable with a security model where organisations can extract data from their IT environment. This concern can be mitigated with a high quality security framework, however concerns do remain.

HEPs did acknowledge that a significant advantage of this option was that they would be able to go at their own speed. This does bring with it a risk that little to no transformation happens in some HEPs, and the status quo remains.

However, HEPs also acknowledged that the burden of process transformation and cost rested on them, and they were not very comfortable with this, especially as some of the key issues with the current situation are not addressed.

On the Data Collectors’ side, there was widespread acknowledgement that this type of environment could work and would lead to improvements. However, there was also a feeling that it would represent a missed opportunity to carry on doing the same things with the same volume of data collections, and just put in place better technology to make it easier.

The general feeling was that this represents a ‘dead end’ and that a more comprehensive change was desirable.
7.4.5 Summary

This option would be feasible to implement, and would represent a step forward from where we are now. However, given the relatively low level of stakeholder acceptance, cost of implementation and the relative paucity of ambition with this option, it is not an option that we are recommending.

7.5 Option 3

The diagram below represents a situation where there is a central collector for the Standard Dataset, in the form of a transformed HESA, to whom all HEPs submit their data. This data would then be made available from HESA to other collectors as appropriate and required by their business processes.

Data Governance / Data Standards / Standard Dataset

7.5.1 How will it work?

This landscape is similar to the current one in many respects, but envisages a shift towards a much more ‘collect once, use many times’ model.

Under this landscape, the application process is carried out in exactly the same manner as currently, with the same factors in place. Once the application cycle is complete, relevant data held by UCAS is transferred to HESA, who will be the collectors of all HE data that sits within the Standard Dataset. The only addition to this is the addition of the ULN. This will be added to the process at the start, either by UCAS, other organisations who handle applications, or via the Learner Records Service records if the applicant already has a ULN. The exact process for this is being identified in the ULN project under HEDIIP.

From this point on in the student journey, nobody else collects HE data which is part of the agreed Standard Dataset. All data from the Standard Dataset is collected and held by HESA, and accessible by all organisations in the sector who require it, subject to appropriate access permissions, data sharing agreement and data protection controls. The data could be pushed to these organisations at appropriate intervals, or they could
have direct access to it, depending on technology, security and privacy constraints. Data Collectors who are receiving the standard data from HESA would be expected to liaise directly with the appropriate HEPs as required.

UCAS could theoretically share data with the SLC, and this could simplify and streamline the process for specific elements of the application data, whilst enabling UCAS, Student Finance and the other Data Collectors to maintain ownership of their non-standard data collections and the associated business processes. SLC have in the past implemented a facility whereby a student finance applicant can populate part of their application form from their UCAS application. Potential future variants on the approach will be considered by SLC and UCAS as part of their major change programmes.

The Student finance business process operations are complex and the devolved government arrangements have resulted in different arrangements across the UK. We envisage that this proposal could eventually provide SLC with access to the standard data that can be exchanged whilst the student is studying in an HEP i.e. their enrolment, attendance and achievement data. This is not an objective that could be committed to for the foreseeable future. The timing of any potential infrastructure changes would need to be aligned to the Data Collector’s requirements and priorities.

National leadership organisations such as HEE and NCTL could, once the admissions cycle is complete and data is transferred to HESA, access the Standard Dataset to confirm students’ details and status, whilst the Funding Councils would have earlier access to data depending on the decisions made about the Standard Dataset and the incremental nature of the data collection.

The scope of the Standard Dataset is crucial to the effectiveness of this option. Unless it covers the bulk of data interactions between HEPs and collectors, then the risk of this option being too close to the current situation to deliver benefits is very real. For this option to work effectively, the significant proportion of HEPs data submissions should be to the transformed HESA and their data interactions with other organisations should be extremely limited.

HESA will need to transform the way they manage collections, moving towards more regular data collections, and replace their dataset with the Standard Dataset. Their technology will also need to be updated, allowing them to provide appropriate services to a range of different collectors, to whom they will be passing data. This transformation of HESA data collection is in part being evaluated as part of their CACHED project to develop the Phase 1 Business Case to transform the HESA data collection.

Some HEPs have also expressed concern that this option could lead to a ‘student records system in the cloud’. This is not the intention of this approach, and would be almost impossible to achieve if that were the aim. HEPs will maintain responsibility for updating student data until students have graduated.

Should a student return to Higher Education to study for another qualification, their prior qualifications will be verified by their former institution, and their current one will assume responsibility for updating their record. For UK students the ULN should tie the student’s record back to prior qualifications accessible via the Personal Learning Record, non-UK would need additional validation. The responsibility for managing a student record will always sit with their current institution.

Data Collectors who are utilising the Standard Dataset from the transformed HESA will need to be supplied with the source data on feeds that enable the data to be passed through to the Data Collector against an agreed schedule or as collected. The transformed HESA would then have access to reuse the data as appropriate. Data Collectors will also need to continue to collect data that is not part of the standard data directly from the learners or appropriate third parties.

The fundamental requirement for this landscape to be effective is for the Standard Dataset to provide sufficient data to individual organisations for them to carry out their function. Common census and reporting deadlines will also be promoted through this approach but not mandated as the frequency of the collection may allow for the Data Collectors to receive standard data on a more timely basis to meet their business needs.

6 https://cached.hesa.ac.uk/
requirements. Agreement on this Standard Dataset will be the biggest challenge of implementing the New Landscape, but will provide the biggest benefit.

It should be noted that this option does not preclude PSRBs and other Data Collectors having direct relationships with HEPs and students. Direct relationships between these bodies will always need to be in place, and that should continue. Indeed, if the Standard Dataset allows Data Collectors to release some staff time from the collection and amalgamation of data from over 150 HEPs, they may well be able to improve the quality of these relationships.

It should also be noted that we do not envisage that every single item of data that is collected from HEPs will flow through the Transformed HESA only. It is likely that there will still be surveys and other smaller data collections that will need to happen. However, our recommendation is that all requests for data go first through the governance body, so that others are aware this is happening, and so a consideration can be made as to whether to add this to the Standard Dataset.

There would be an opportunity for the transformed HESA to collect specific niche data for a Data Collector which does not form part of the Standard Dataset. This collection would be on a commercial basis and would need to be consistent with the Data Principles.

7.5.2 Challenges
The key challenge of this option is the difficulty of gaining agreement from all collectors that they will no longer collect any data from the Standard Dataset, which will instead be collected by HESA and passed on and/or access granted to the data.

The other main challenge will be the transparency of how funding agencies use the data they receive from HESA to make funding decisions. Should there be an issue with this process, and a HEP receives less funding than it believes is due, there needs to be sufficient transparency for that decision, and the data on which it was based, to be examined in detail. The use of standard data definitions will help to reduce the current ambiguity in certain funding rules.

7.5.3 Benefits
The key benefit, and something the sector is aiming to move towards, will be an easier implementation of in-year data collections, which enables the reduction in the volume of data collected and would support the rationalisation of data collections. HEPs have indicated to us that with more certainty around dataset and definitions, and with a reduced number of collectors asking for data, they would be more able to facilitate in-year data collections, which will form the foundation for a move to more regular data collections as the New Landscape stabilises and becomes business as usual. This will be a key factor in persuading other key Data Collectors to come on board. Additional benefits could arise from the provision of more timely data. These include initiatives provided to HEPs through Jisc led initiatives for the implementation of other data integration and re-use services.

In addition, some elements of the HESA return could be pre-populated by UCAS/SLC data (subject to appropriate data sharing agreements), removing this burden from HEPs.

Specific benefits are described in section 7.7

7.5.4 Evaluation and impact on different stakeholders
Using the evaluation method described in section 7.2, we have evaluated this option as follows:
### Impact on applicants and students

Home/EU applicants would not notice any change in their process or who they engaged with, as that element of the process would not change. We do not envisage significant changes in the way the international and taught postgraduate applications and admissions processes work either, but the technology changes required under this model could result in an improved experience for these constituencies.

Following discussions with relevant parties, there are challenges around this in terms of affecting different organisations’ commercial models and competitive position. Appropriate safeguards on use of the data before its commercial sensitivity has expired need to be implemented as part of any detailed design phase. The key element of this is HEPs not wishing to share their in-year recruitment data with other HEPs, as this is commercially sensitive data that could affect their competitive position. From a Competition Law perspective, UUK have already indicated that care needs to be taken in this area, and this will have to be factored in to any final design.

### Ease of implementation

This option requires both technical and process changes. The technical changes do not need to occur in the initial phases, as that period will be about defining and agreeing the Standard Dataset and the data definitions. Benefits can flow quickly once this has been agreed.

If in-year collections are to become the norm, then technical changes will be required to facilitate this, as data warehousing and reporting technologies will need to be implemented for this type of collection to work effectively and efficiently.

Process changes will focus on HESA becoming the single collector of the Standard HE dataset once the UCAS admissions process is complete, and other Data Collectors collecting the standard HE data from HESA, with the exception of SLC and UKVI.

The complexity involved in achieving that should not be under-estimated; not least on an organisational trust level. For this option to work effectively, other organisations need to trust HESA to collect the data effectively, and pass it on to them. Agreeing this, along with the relevant dates for data to be collected, will be a significant undertaking as will putting in place appropriate data sharing agreements.

Most organisations have asked for access to the raw data as part of this process, and how that is facilitated would also need to be agreed at detailed design stage.

### Cost of implementation

Implementation costs are likely to be high, but lower than Option 1. The costs will come from increasing HESA’s capacity to handle all the necessary data collections once it is the single HE Data Collector, and the significant time investment necessary from all key stakeholders to agree the Standard Dataset and definitions.

In comparison to Option 2, the costs will also be lower for this option. There will be less of a requirement on HEPs to transform their technical and security infrastructures to allow multiple Data Collectors to access data as there is in Option 2.

From a HEP perspective, there will be a cost of moving to the new, single dataset, but this should be offset by the reduction in the number of different data submissions they have to make to different collectors, and the concomitant reduction in the number of different data relationships they need to maintain.

Overall for the sector, the cost of running this Option long term will be lower than the current scenario as it is more efficient, reduces the burden on HEPs to provide lots of different types of data to many different organisations, and reduces the number of data exchange relationships that need to be maintained. It also delivers a benefit to students if the student finance application and the HE application process can be harmonised.

### Potential impact on Data Collectors

The impact on Data Collectors will be significant, whichever part of the sector they are in. HESA will have to transform the way it carries out data collections, and gear up for an increase in workload, technology provision, and ability to handle and process queries.

Other Data Collectors will have to agree the Standard Dataset they want to be collected, and that HESA can collect it on their behalf.

All Data Collectors will also need to agree how the data is passed on, the dates on which certain data needs to be collected, and how individual organisations can have access to the raw data.

Again, the scope of the Standard Dataset is critical here, as is the timing of the data collections. The specialist data that collectors need to carry on collecting should be cut down to a minimum, and collectors will need to align this specialist data as far as possible with the Standard Dataset in terms of format and structure so that reconciliation issues are kept to a minimum.
### Potential impact on HEPs

The impact on HEPs would be, by and large, positive. They would not need to maintain multiple data relationships with multiple organisations, nor would they need to submit different sets of data, with different data definitions, at different times. Transparency of data and its use is key to HEPs signing up for this option. The reduction in their burden with this option is potentially immense, but they do not want to be in a position where a funder makes funding decisions based on erroneous interpretation of the data that has been passed from a HEP to a central Data Collector and thence to the funder. As we have built in the possibility for funders to access the raw data, we believe this risk is mitigated. Detailed design will look at this further, however.

### Ease of operation

Once implemented, this option would be relatively easy to operate. The data flows are relatively simple, and standard HE dataset is collected by a single collector, for all students on HEP systems. This option also offers flexibility to make further changes in the future, as requirements change.

UCAS continues to operate the applications and admissions process, and the SLC continues to operate the Student finance application. This should speed up the overall process and deliver a better service to students.

### Stakeholder acceptance

Of the three options we have looked at, this one has had the highest level of stakeholder acceptance.

Questions remain around exact definitions, dates of collection, data formats, and technology. However, most of the organisations we have spoken to felt that this option represented the best compromise between being bold and not trying to change too much, too fast.

One requirement on organisations for this option to be successful is for them to agree that HESA can collect data on their behalf and pass it through to them. This will need to be negotiated and agreed alongside the agreement of the Standard Dataset and definitions and data sharing controls.

### 7.5.5 Summary

This is our preferred option. There are significant challenges in the detail of when the data is collected, for whom, and how that data is passed on to them. It will not be easy to implement given the already significant challenge of agreeing the Standard Dataset and data standards.

The recommended Blueprint to move towards immediately requires significantly less organisational change, no direct changes in remit, and less potential impact on core processes within the sector.
## 7.6 PESTLE Analysis of preferred option

<table>
<thead>
<tr>
<th>Political</th>
<th>Economic</th>
<th>Social</th>
<th>Technological</th>
<th>Legal</th>
<th>Environmental</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Strengths</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Need for change acknowledged by most stakeholders</td>
<td>Delivers benefits across the sector compared to other options, and the status quo</td>
<td>Delivers benefits to all stakeholders</td>
<td>Pushes reform of technology to more modern underpinnings</td>
<td>Does not require legislation to implement</td>
<td>Reduces inefficiency and waste across the sector</td>
</tr>
<tr>
<td>Most stakeholders favour this option</td>
<td>Less costly to implement than other options</td>
<td>Demonstrably reduces inefficiencies, thereby improving the perception of the sector</td>
<td>Does not require massive technical change or system replacements</td>
<td>Does not require organisations to change their current commercial activities or their primary business focus</td>
<td></td>
</tr>
<tr>
<td><strong>Weaknesses</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Requires cooperation on behalf of many organisations</td>
<td>Cost implication for transformation to in-year collections</td>
<td>Governance model will not prevent disagreements around requested changes to the Standard Dataset</td>
<td>Requires standardisation of data formats across a number of different providers</td>
<td>No legal basis for governance of the Standard Dataset, thereby making it non-compulsory, with no formal sanction for non-compliance</td>
<td></td>
</tr>
<tr>
<td>Requires some organisations to change their processes and hand some activities over to HESA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Opportunities</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Greater coherence, accuracy and relevance of student data</td>
<td>Greater consolidation of processes and data collection results in ongoing savings to stakeholders</td>
<td>Promotes closer working between key stakeholders</td>
<td>Offers alternative data collection approaches (push vs. pull)</td>
<td>Ever greater use of technology reduces the need for paper and manual processing</td>
<td></td>
</tr>
<tr>
<td>More timely and relevant data allows governments and funders to make more informed decisions on HE policy</td>
<td></td>
<td>More streamlined student management processes, improving the perception of the sector</td>
<td>With greater standardisation, so comes the opportunity for greater automation</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Potential for a streamlined SLC and UCAS application processes for students</td>
<td>As data is collected and held centrally, greater use of Business Intelligence and Data Analytics tools will deliver greater insight into the sector</td>
<td></td>
<td></td>
</tr>
<tr>
<td>POLITICAL</td>
<td>ECONOMIC</td>
<td>SOCIAL</td>
<td>TECHNOLOGICAL</td>
<td>LEGAL</td>
<td>ENVIRONMENTAL</td>
</tr>
<tr>
<td>-----------</td>
<td>----------</td>
<td>--------</td>
<td>---------------</td>
<td>-------</td>
<td>---------------</td>
</tr>
<tr>
<td>Threats</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>■ The need for such broad agreement across so many independent organisations</td>
<td>■ If funding is not available for data transformation (HESA), the landscape cannot be realised</td>
<td>■ The number of different systems and technologies in use across the sector could make it difficult to harmonise them all under one data collection regime</td>
<td>■ Organisations who do not correctly understand the preferred option may perceive it as a threat to their commercial interests. This may limit their willingness to participate</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
7.7 Value Propositions for the New Blueprint (Option 3)

7.7.1 For HEPs:

For HEPs, this Blueprint will remove a significant amount of the burden of data collection. They currently have to provide multiple data returns to different organisations, with significantly overlapping datasets, and often different data definitions; but with this Blueprint, for the Standard Dataset data, they will have one dataset with one set of data definitions to provide to one collector in one format. This will reduce the complexity and difficulty of managing their data returns, and will help greatly to improve the quality of data in the sector. We were told that even without the change in data flows whereby some collectors gather the data from a transformed HESA, the introduction of data standards for an agreed Standard Dataset will reduce the burden.

This will release staff within HEPs to deliver value added activities, and provide greater focus on service improvement through data analysis, instead of simply providing data.

The biggest consistent consumption of resource within HEPs in this area is around collecting and transforming their raw data, which is collected in certain formats for different purposes, to fit the format and structure of the collection they are submitting. With a Standard Dataset, common definitions and a clear structure around the data collection, and only one main collection to submit, HEPs will be able to reduce significantly the workload on their data processing teams.

HEPs have indicated that this could free up as many as 7 FTE from their strategic planning, data processing and data returns teams, who could be allocated to other activities that would deliver strategic benefit to the provider.

If we assume that some HEPs could release fewer FTE than others, with an average of two FTE being released, at a total average salary, including on cost, of £40,000 per annum, then that is £12.8 million worth of resource per annum from approximately 160 HEPs across the sector that can be released for other activities. This is a ballpark estimate, but deliberately prudent. Set against these savings could be some additional costs from HESA for providing a more extensive or enhanced service. These costs will clarify as the HESA transformation takes shape.

7.7.2 For Data Collectors:

Organisations that require regular data collections for Higher Education data to fulfil their business requirements will no longer have to collect the same volume of data, and will be able to concentrate on their primary business purpose.

This will allow them to provide more effective service to their end clients on the basis of more accurate and up-to-date data.

Data collection organisations also have FTE involved in dealing with HEPs and managing their collections. It is likely that the number of FTEs that could be released in these organisations will be lower. We have assumed 20 FTE, in total, from all the Data Collectors could be released from their current tasks, with an average salary of £50k including costs. This equates to a further £1 million worth of resource across the sector who can be released from their current duties.

HESA has recently announced that it is evaluating the need to revise its data collection process to include more timely data collection. We have seized this opportunity to provide a single Data Collector for a Standard Dataset. As each of the key Data Collectors profiled in section 4.3 have identified current or planned data infrastructure changes it should be easy to assimilate the requirements for the new data flows as and when the systems are refreshed. Exact timings and the scope of these infrastructure changes have not been included within this report as some of this data is deemed to be out of scope for this engagement.
7.7.3 For Students:
The most important group in Higher Education, students will experience better service, more targeted to their needs and more up-to-date, as HEPs are able to focus on utilising data more effectively instead of simply having to provide it to different collectors.

Reform of HEPs’ own data processes will reduce the number of data-processing errors that students have to correct themselves, improving their experience of Higher Education as a whole.

Improved accuracy and timeliness of data will also improve the quality of data students have at their disposal when choosing their route into and through Higher Education.

Indirect benefits for students are likely to be realised by HEPs having more resources to focus on students as HEPs would be more effective at data management.

7.7.4 For Funders
The Funding Bodies across Higher Education need timely, accurate data on which to base their funding decisions. The New Data and Information Landscape will facilitate the more effective targeting of funds to those HEPs and subject areas where they can be used most effectively, reducing waste in the sector, and driving up overall performance. More timely data will be accessible through in-year data collection.

7.8 Linking to Other HEDIIP Projects
Existing HEDIIP projects were launched to resolve deficiencies within the current landscape. We therefore outline below how these projects support the New Data Landscape that we have proposed.

7.8.1 ULN
The ULN project is looking at the feasibility and barriers to adoption of implementing the Unique Learner Number, used in Further Education and schools (excluding Scotland), into the HE landscape. All the options we have looked at have the agreement and introduction of a Standard Dataset and common data definitions within that dataset as a prerequisite for their successful implementation.

The thinking behind that is that standardisation and common definitions are key to improving data quality and extracting useful information from it. The ULN is a key element of standardisation, and will make it much easier for HEPs to identify students as they move between institutions, thereby improving data quality and allowing those organisations who need to be able to identify individual students to do so.

This project can feed in to the New Data and Information Landscape, and once the ULN has been rolled out to all HE Students, it will make collecting data about them easier due to the ease with which different records can be identified and merged with the single identifier in place.

The ULN project will need to be closely aligned to this project, to ensure that as the New Data and Information Landscape is implemented, the ULN is taken into consideration when enacting the appropriate technical and process changes.

7.8.2 New Subject Coding System
This project is replacing the current Joint Academic Coding System with one that is more fit for purpose in today’s environment.

This will support directly the creation of the Standard Dataset, as it will be a commonly used data structure for the coding of academic programmes. We anticipate that this will form the standard of the new dataset and the governance of the new subject coding systems and the new HE Standard Dataset will be closely aligned.
7.8.3 Data Capability

One area that also needs to be considered is the way that stakeholders collect and manage their data before it is passed on to different Data Collectors.

HEPs have different ways of managing the data management process, and different ways of submitting their data. Imposing one way of doing this for all HEPs will not be effective. Data Collectors have implemented a range of approaches to data collection without aligning their approaches. Engagement with PSRBs with the ambitions of the New Landscape will propose a particular challenge as it is likely that a number of them may not be willing to acknowledge the development of the New Landscape until they have been made aware of the potential cost benefits of accessing standard HE data.

Through effective Governance of the landscape Data Collectors can be encouraged to move towards a harmonisation of data collection processes and standards, which should help provide a standard framework to HEPs that allows them to take advantage of this more efficient means of data collection and processing if they wish.

The Data Capability Project is looking at how to implement a standard approach to capability improvement in Data Management. Much as Prince 2 has become the standard for Project Management, so this is the first step towards creating a Data Management approach based on standard and capability.

That is central to what this project is seeking to achieve, and will be a necessary part of implementing our Blueprint. Our recommended option will require HEPs to change the way they process and manage their data, and this needs to be done in a manner which is recognised as being effective and meeting a particular standard.

7.8.4 Student Data Collection Review

The inventory of HE data collections would need to be maintained and expanded as a key source of information to provide visibility of the data collection exchanges that are in operation across the sector. Through the data governance arrangements additional information should be gathered that includes:

- Each Data Collector should be registered and made aware of the New Landscape arrangements. They should be asked to agree to the data principles and data governance arrangements the status of their responses should be recorded;
- Each data collection recorded should be reviewed to evaluate if the data collected should be part of the standard HE dataset;
- The type of data collection (personal data items, aggregated etc.); and
- The timing of the data collection

7.8.5 HE Data Language

The original project was charged with ‘delivering a data model which maps, across the HE lifecycle, the main entities, their relationships and their use in the various sector-level data systems. The expectation is that a description of each entity will form the core of a standard lexicon for HE data and a thesaurus (a dictionary of synonyms) will provide disambiguation for data interfaces within the HE sector.’

The New Data Landscape will require this work to be extended and maintained to support the development and documentation of the HE data standards for the data utilised across the HE Landscape.

7.9 Prerequisite to Implementation

For the Blueprint to be implemented effectively, there are certain challenges to overcome. Key among these is the Standard Dataset.

The Standard Dataset is crucial to the effective functioning of the Blueprint, but also possibly the hardest challenge to overcome.
To achieve the aim of the Blueprint (a reduced number of Data Collectors for the sector) the Standard Dataset needs to be sufficiently broad to provide each organisation with the data they need to deliver their service effectively without having to resort to separate collections.

The biggest challenge here will be working with a number of different independent organisations to develop a common dataset, which meets all their needs in terms of data collection, and has a common set of definitions.

Given the importance of the Standard Dataset to the success of the Blueprint, this should be the key workstream for the Implementation phase, with the biggest focus in terms of gaining collective agreement on the way forward.

7.10 Other challenges

There are other key challenges in implementing the Blueprint, aside from gaining agreement on the Standard Dataset. These include:

7.10.1 Data Collection Timescales

Gaining agreement on appropriate data collection dates will also be a very difficult exercise. Currently, different organisations require different types of data at different times of the year. Harmonising this into a number of collections that will provide each organisation with sufficiently up-to-date information will be a particularly challenging exercise.

This will require significant conversation and discussion to agree how this can be overcome, and when organisations absolutely need data to be submitted for them to meet their own business processes and requirements.

7.10.2 HEP Software System Providers

Agreeing a common format, protocol and method for submitting data will require discussion with the key software system providers, to give them appropriate lead time to build relevant modules, and upgrade their solutions where appropriate. Discussion with universities who do not use externally supplied proprietary systems from the key players will also be required so they are aware of what work may need to be done.

This challenge will be fairly easy to overcome, as the providers will want to be compliant with the new model, and as long as they are included early in the discussions and allowed to contribute their thoughts, agreement should be achievable.

7.10.3 Governance of the Standard Dataset

This will be one of the key long-term success elements for the Blueprint. Even when the challenge of agreeing and implementing the Standard Dataset has been overcome, the ongoing maintenance of the dataset is of paramount importance. There will, of course, need to be changes to the Standard Dataset. However, these need to be handled in a strategic manner, and not just implemented in an ad hoc fashion by individual organisations. Each change should have a justification behind it, and a demonstration of how and why that extra data will be used. Otherwise there will be a proliferation of new data requests, and the situation will deteriorate to something similar to the current one. However, agreeing how to govern the Standard Dataset, and what is an acceptable process for changing it will be very difficult. There will also need to be a robust process implemented for the removal of data from the Standard Dataset.

To overcome this challenge, stakeholders will need to be convinced of the advantage to them of moving to this model, and pressure needs to be brought to bear from appropriate quarters on those who are more reluctant to sign up. The recommended model of Data Governance in section 6.2 should be responsible for governing the standard HE dataset.

7.10.4 Timescales

For the Blueprint to deliver benefits, it needs to be implemented relatively quickly (within three years). However, gaining agreement on the necessary elements to be able to implement it effectively in that time
will be very difficult. There is the very real risk that everybody signs up to it in principle at the start, but implementation drifts and drifts and nothing ever really changes significantly.

There is a significant amount of work to do, and to overcome this challenge, it needs to start quickly, and maintain momentum. This requires a detailed delivery plan with frequent milestones to evidence progress and ability to deliver.

7.10.5 Competition Considerations

As detailed above, there are Competition Law considerations around the sharing of commercially sensitive recruitment and enrolment data in-year.

Understandably, HEPs are wary of this, and also do not want to compromise their commercial position by sharing what they view as commercially sensitive data with their direct competitors when it could be disadvantageous to them to do so.

This is a challenge that can be overcome with agreement on timescales for release of different elements and types of data, and the drawing up and signing up to data sharing protocols.

7.11 Next Steps

There is a need for the New Landscape Project Board to consider the above concept and principles.

Section 8 provides an implementation plan to enable the realisation of the New Data Landscape.
8 Implementation of a New Landscape

This section provides the implementation plan to realise the New Landscape over a three-year period.

8.1 Rationale for a 3-year Roadmap

The full implementation of a New Data Landscape is likely to take up to three years to be realised due to the needs for the transformation of stakeholder and provider systems together with the lead time for designing and launching data collections. However, our Blueprint describes a number of new and existing projects that will contribute towards this agreed goal. Projects that are due to deliver within the first 12 months of this implementation are the ones that we would consider to be quick wins.

We have provided a list of summary actions that will need to be undertaken to achieve these objectives and these are described in the following table. It should be noted that we have provided a very challenging implementation plan to enable benefits to be derived as early as possible. It will be important that sufficient resource and leadership is provided to ensure that these timescales can be achieved, or adjusted accordingly.

HEDIIP currently has funding until June 2016. We suggest that it should be tasked with supporting the implementation of the New Landscape until the transition to the New Landscape data management can be fully implemented.

As we have developed the implementation plan to support the realisation of benefits to the stakeholders through the new governance structure and the Standard Dataset with common definitions, these will all be available for use in the current system, before the transformed HESA data collection goes live.
<table>
<thead>
<tr>
<th>Task Name</th>
<th>Finish</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. HEDIP Implementation plan</td>
<td>Thu 31/05/18</td>
</tr>
<tr>
<td>2. Transition</td>
<td>Mon 30/05/18</td>
</tr>
<tr>
<td>3. 0.1 Communicate the approved new landscape report</td>
<td>Mon 01/06/15</td>
</tr>
<tr>
<td>4. 0.2 Communicate approved principles to the Stakeholders</td>
<td>Wed 01/07/15</td>
</tr>
<tr>
<td>5. 0.3 Develop transition arrangement</td>
<td>Wed 01/15/15</td>
</tr>
<tr>
<td>6. 0.4 Approve transition arrangements with HEDIP</td>
<td>Mon 13/07/15</td>
</tr>
<tr>
<td>7. 0.5 Evaluate and align the progress and scope of the HESA CACHED project</td>
<td>Wed 01/17/15</td>
</tr>
<tr>
<td>8. 0.6 Evaluate the current data infrastructure plans for the large data</td>
<td>Mon 03/08/15</td>
</tr>
<tr>
<td>9. 0.7 Evaluate the potential contribution of existing stakeholder led sector</td>
<td>Mon 03/08/15</td>
</tr>
<tr>
<td>10. 0.8 Agree communications strategy</td>
<td>Fri 19/06/15</td>
</tr>
<tr>
<td>11. 0.9 Begin delivery of communication strategy</td>
<td>Mon 22/06/15</td>
</tr>
<tr>
<td>12. 1.0 Agreed Data Governance model</td>
<td>Tue 01/09/15</td>
</tr>
<tr>
<td>13. 1.1 Agree terms of reference</td>
<td>Mon 15/06/15</td>
</tr>
<tr>
<td>14. 1.2 Agree constitution and remit of governance body</td>
<td>Tue 15/06/15</td>
</tr>
<tr>
<td>15. 1.3 Develop job descriptions for management function</td>
<td>Wed 29/07/15</td>
</tr>
<tr>
<td>16. 1.4 Recruit to the new structure</td>
<td>Fri 29/08/15</td>
</tr>
<tr>
<td>17. 1.5 Formally constitute governance body</td>
<td>Tue 01/09/15</td>
</tr>
<tr>
<td>18. 2.0 Agreed Standard Dataset and Data Standard</td>
<td>Mon 04/06/18</td>
</tr>
<tr>
<td>19. 2.1 Analyse current datasets and standards and identify overlaps</td>
<td>Fri 31/07/15</td>
</tr>
<tr>
<td>20. 2.2 Develop standard dataset</td>
<td>Fri 28/08/15</td>
</tr>
<tr>
<td>21. 2.3 Develop data standards</td>
<td>Fri 30/10/15</td>
</tr>
<tr>
<td>22. 2.4 Consult on standard data and data standards</td>
<td>Fri 27/11/15</td>
</tr>
<tr>
<td>23. 2.5 Initial standard dataset agreed</td>
<td>Tue 04/03/15</td>
</tr>
<tr>
<td>24. 2.6 Initial data standards agreed</td>
<td>Tue 04/03/15</td>
</tr>
<tr>
<td>25. 2.7 Work with early adopters: to agree, implement and roll out</td>
<td>Tue 04/03/15</td>
</tr>
<tr>
<td>26. 2.8 Ongoing Implementation activity</td>
<td>Mon 03/10/16</td>
</tr>
<tr>
<td>27. 2.9 Go live of new regime for early adopting collectors</td>
<td>Fri 02/06/17</td>
</tr>
<tr>
<td>28. 2.10 Amend based on operational feedback</td>
<td>Fri 02/06/17</td>
</tr>
<tr>
<td>29. 2.11 Second Phase adopters go live</td>
<td>Fri 01/06/18</td>
</tr>
<tr>
<td>30. 2.12 Full go live</td>
<td>Mon 04/06/18</td>
</tr>
<tr>
<td>31. 3.0 Process Redesign</td>
<td>Mon 04/06/18</td>
</tr>
<tr>
<td>32. 3.1 Establish working group</td>
<td>Fri 25/09/15</td>
</tr>
<tr>
<td>33. 3.2 Agree secondclass from HEPs &amp; collectors</td>
<td>Fri 26/02/16</td>
</tr>
<tr>
<td>34. 3.3 Agree HEP responsibilities in new regime</td>
<td>Fri 26/02/16</td>
</tr>
<tr>
<td>35. 3.4 HEPs assess impact of new regime</td>
<td>Thu 25/02/16</td>
</tr>
<tr>
<td>36. 3.5 HEPs amend their data collection processes where appropriate</td>
<td>Thu 03/02/17</td>
</tr>
<tr>
<td>37. 3.6 Go live of new regime</td>
<td>Mon 04/06/18</td>
</tr>
<tr>
<td>38. 4.0 Technology Implementation</td>
<td>Mon 04/06/18</td>
</tr>
<tr>
<td>39. 4.1 Current Technology Assessment</td>
<td>Mon 04/06/18</td>
</tr>
<tr>
<td>40. 4.2 Future Technology Architecture Design for data transmission and</td>
<td>Mon 04/06/18</td>
</tr>
<tr>
<td>41. 4.3 Migration plan</td>
<td>Mon 04/06/18</td>
</tr>
<tr>
<td>42. 4.4 Design and build of appropriate modules</td>
<td>Thu 01/09/15</td>
</tr>
<tr>
<td>43. 4.5 Testing and refinement</td>
<td>Mon 02/10/17</td>
</tr>
<tr>
<td>44. 4.6 Technology Go Live</td>
<td>Mon 04/06/18</td>
</tr>
</tbody>
</table>
## 8.2 Summary Actions

<table>
<thead>
<tr>
<th>NO.</th>
<th>OBJECTIVE</th>
<th>MILESTONE</th>
<th>TARGET COMPLETION DATE</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>Transition</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.1 Communicate the approved New Landscape report</td>
<td></td>
<td>01/06/15</td>
</tr>
<tr>
<td></td>
<td>0.2 Communicate approved principles to the Stakeholders</td>
<td></td>
<td>01/07/15</td>
</tr>
<tr>
<td></td>
<td>0.3 Develop transition arrangement for the New Landscape</td>
<td></td>
<td>01/07/15</td>
</tr>
<tr>
<td></td>
<td>0.4 Approve transition arrangements with HEDIIP</td>
<td></td>
<td>13/07/15</td>
</tr>
<tr>
<td></td>
<td>0.5 Evaluate and align the progress and scope of the HESA CACHED project</td>
<td></td>
<td>01/07/15</td>
</tr>
<tr>
<td></td>
<td>0.6 Evaluate the current data infrastructure plans for the large Data Collectors</td>
<td></td>
<td>03/08/15</td>
</tr>
<tr>
<td></td>
<td>0.7 Evaluate the potential contribution of existing stakeholder-led sector-wide projects and their impact on the landscape.</td>
<td></td>
<td>03/08/15</td>
</tr>
<tr>
<td></td>
<td>0.8 Agree communications strategy</td>
<td></td>
<td>19/06/15</td>
</tr>
<tr>
<td></td>
<td>09 Begin delivery of communication strategy</td>
<td></td>
<td>22/06/15</td>
</tr>
<tr>
<td>1</td>
<td>Agreed Data Governance model</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.1 Agree terms of reference</td>
<td></td>
<td>15/06/15</td>
</tr>
<tr>
<td></td>
<td>1.2 Agree constitution and remit of governance body</td>
<td></td>
<td>16/06/15</td>
</tr>
<tr>
<td></td>
<td>1.3 Develop job descriptions for management function</td>
<td></td>
<td>29/07/15</td>
</tr>
<tr>
<td></td>
<td>1.4 Recruit to the new structure</td>
<td></td>
<td>28/08/15</td>
</tr>
<tr>
<td></td>
<td>1.5 Formally constitute governance body</td>
<td></td>
<td>01/09/15</td>
</tr>
<tr>
<td>2</td>
<td>Agreed Standard Dataset and Data standards</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2.1 Analyse current datasets and standards and identify overlaps</td>
<td></td>
<td>31/07/15</td>
</tr>
<tr>
<td></td>
<td>2.2 Develop Standard Dataset</td>
<td></td>
<td>32/08/15</td>
</tr>
<tr>
<td></td>
<td>2.3 Develop data standards</td>
<td></td>
<td>30/10/15</td>
</tr>
<tr>
<td></td>
<td>2.4 Consult on Standard Dataset and data standards</td>
<td></td>
<td>27/11/15</td>
</tr>
<tr>
<td></td>
<td>2.5 Initial standard data agreed</td>
<td></td>
<td>01/03/16</td>
</tr>
<tr>
<td></td>
<td>2.6 Initial data standards agreed</td>
<td></td>
<td>01/03/16</td>
</tr>
<tr>
<td></td>
<td>2.7 Work with early adopters to agree, implement and roll out</td>
<td></td>
<td>01/03/16</td>
</tr>
<tr>
<td></td>
<td>2.8 Ongoing implementation activity</td>
<td></td>
<td>03/10/16</td>
</tr>
<tr>
<td></td>
<td>2.9 Go live of new regime for early adopting collectors</td>
<td></td>
<td>02/06/17</td>
</tr>
<tr>
<td></td>
<td>2.10 Amend based on operational feedback</td>
<td></td>
<td>02/06/17</td>
</tr>
<tr>
<td></td>
<td>2.11 Second Phase adopters go live</td>
<td></td>
<td>01/06/18</td>
</tr>
<tr>
<td></td>
<td>2.12 Full go live</td>
<td></td>
<td>04/06/18</td>
</tr>
<tr>
<td>3</td>
<td>Process Redesign</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>3.1 Establish working group</td>
<td></td>
<td>25/09/15</td>
</tr>
</tbody>
</table>
8.3 Detailed Actions

Detailed below is the relationship between the strategic objectives, milestones (as detailed in the implementation plan) and the proposed actions for implementation.

Each milestone is broken down into sections, giving a description of the milestone, a timeline for completion (defined by Target Commencement Date/Target Completion Date with an assumed Day 0 of 01/04/15), detail of the actions involved in delivering the milestone, benefits associated with the milestone and the Lead Organisation.

We have assumed there will need to be external support of some sort to HEDIIP in the implementation, although the exact format of this is still to be decided.

8.3.1 Objective 0: Transition

This workstream will handle the transition from the phase of the project that delivered this report, to the detailed design and implementation phase of the project. These activates do not appear on the timeline map below. In this section of the report we provide more details about the actions described as milestones in section 8.2 that are required to implement the Blueprint for the New Landscape. For each of the milestones we have described:

- The description of the milestone – to aid a better understanding of the objective;
- Duration – our estimate of the time it should take to deliver the milestone;
- Actions required to deliver the milestone – and explanation of the subtask required to deliver the milestone;
- Benefits of that milestone – perceived benefits; and
- The lead for the delivery of that milestone – an indication of who will be required to deliver the milestone, HEDIIP, HESA, Governance bodies, HEPs or an External Partner.
## Objective 0

### Transition

<table>
<thead>
<tr>
<th>Project</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1 Communicate the approved New Landscape report</td>
<td>Need to make all stakeholders aware of New Landscape report and its implications.</td>
</tr>
</tbody>
</table>

### Completion date

01.06.15

### Action

- Define audience to communicate report
- Define communication medium
- Draft and send correspondence
- Attend relevant sector events

### Benefits

More coordinated and wider communication leads to higher level of buy-in.

### Lead

HEDIIP

---

## Objective 0

### Transition

<table>
<thead>
<tr>
<th>Project</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.2 Communicate approved principles to the Stakeholder</td>
<td>Need to make key stakeholders aware of approved principles and their implications.</td>
</tr>
</tbody>
</table>

### Completion date

01.75.15

### Action

- Define communication medium
- Draft and send correspondence
- Attend meetings with relevant stakeholders

### Benefits

The principles are a key element of our recommendations and need to be communicated across the sector to gain understanding and buy-in.

### Lead

HEDIIP
### Objective 0
**Transition**

<table>
<thead>
<tr>
<th>Project</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.3 Develop transition arrangement</td>
<td>Need to define the transition arrangement to implementation project before transition can begin.</td>
</tr>
</tbody>
</table>

**Completion date**

01.07.15  

**Action**

Agree stakeholders  
Agree terms of reference for new project  
Consult with the sector bodies

**Benefits**

Builds momentum early and allows organisations to see change is happening  
Builds understanding that change is coming, thereby reducing some of the resistance when it arrives

**Lead**

HEDIIP

---

### Objective 0
**Transition**

<table>
<thead>
<tr>
<th>Project</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.4 Approve transition arrangements with HEDIIP</td>
<td>Must communicate the transition arrangements with HEDIIP and gain approval in order for transition to begin.</td>
</tr>
</tbody>
</table>

**Completion date**

13.07.15  

**Action**

Amend arrangements based on feedback  
Create finalised transition arrangements  
Agree arrangements with HEDIIP  
Begin transition to implementation project

**Benefits**

Transition arrangement in place

**Lead**

HEDIIP
## Objective 0
### Transition

<table>
<thead>
<tr>
<th>Project</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5 Evaluate and align the progress and scope of the HESA CACHED Project</td>
<td>The HESA CACHED project is building a business case for the transformation of HE data collection. This project needs to align with that one.</td>
</tr>
</tbody>
</table>

**Completion date**: 01.07.15

**Action**
- Evaluate output of the CACHED project
- Assess linkages and overlap with the CACHED project
- Align CACHED and New Data Landscape Projects
- Consult stakeholders who sit across both projects to agree alignment
- Consult more broadly with the sector

**Benefits**: Aligned approach to the transformation of data collection

**Lead**: HEDIIP + HESA

<table>
<thead>
<tr>
<th>Project</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.6 Evaluate the current data infrastructure plans for large Data Collectors</td>
<td>To enable a realistic transition plan towards the new Blueprint it will be necessary to gain a better understanding of the opportunities that may be gained for alignment</td>
</tr>
</tbody>
</table>

**Completion date**: 03.08.15

**Action**
- Agree Data Collectors in scope
- Development infrastructure assessment proforma
- Disseminate proformas for population
- Meet with Data Collectors to discuss feedback
- Develop transitions proposals aligned to CACHED and discuss with Data Collectors

**Benefits**: Aligned approach to the transformation of data collection

**Lead**: HEDIIP + HESA
<table>
<thead>
<tr>
<th>Objective 0</th>
<th>Transition</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Project</strong></td>
<td><strong>Description</strong></td>
</tr>
<tr>
<td>0.7 Evaluate the potential for existing stakeholder-led sector-wide projects and their impact on the landscape</td>
<td>To minimise the resources required to implement the Blueprint it will be necessary to review the current stakeholders for existing and planned cross-sector projects that can be assimilated into the landscape or realigned to meet the ambitions of the landscape.</td>
</tr>
<tr>
<td><strong>Completion date</strong></td>
<td>03.08.15</td>
</tr>
<tr>
<td><strong>Action</strong></td>
<td>Identify stakeholders with cross-sector projects</td>
</tr>
<tr>
<td></td>
<td>Development project assessment proforma</td>
</tr>
<tr>
<td></td>
<td>Disseminate proformas for population</td>
</tr>
<tr>
<td></td>
<td>Meet with stakeholders to discuss feedback</td>
</tr>
<tr>
<td></td>
<td>Develop assimilation proposals aligned to Blueprint and discuss with stakeholders</td>
</tr>
<tr>
<td><strong>Benefits</strong></td>
<td>Align cross-sector initiatives towards the implementation of the Blueprint.</td>
</tr>
<tr>
<td><strong>Lead</strong></td>
<td>HEDIIP + HESA</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Objective 0</th>
<th>Transition</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Project</strong></td>
<td><strong>Description</strong></td>
</tr>
<tr>
<td>0.8 Agree Communication Strategy</td>
<td>In order for the change to be implemented successfully, all key stakeholders need to be aware of the ongoing momentum of the New Data Landscape Project and any actions they need to take. An effective communication strategy will ensure this. As the project develops, the communications strategy will need to evolve.</td>
</tr>
<tr>
<td><strong>Completion date</strong></td>
<td>19.06.15</td>
</tr>
<tr>
<td><strong>Action</strong></td>
<td>Agree communication audience</td>
</tr>
<tr>
<td></td>
<td>Agree delivery responsibility and content creation</td>
</tr>
<tr>
<td></td>
<td>Agree key communication dates</td>
</tr>
<tr>
<td></td>
<td>Agree key communication channels</td>
</tr>
<tr>
<td><strong>Benefits</strong></td>
<td>Effective, well-organised communications</td>
</tr>
<tr>
<td></td>
<td>System for monitoring and keeping track of communication</td>
</tr>
<tr>
<td><strong>Lead</strong></td>
<td>HEDIIP + External partner</td>
</tr>
</tbody>
</table>
### Objective 0

**Transition**

<table>
<thead>
<tr>
<th>Project</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.9 Deliver Communications Strategy</td>
<td>Ongoing communications activities</td>
</tr>
</tbody>
</table>

**Completion date**

From 22.06.15 – Ongoing

**Action**

- Regular communications and meetings with stakeholders, as per the communications strategy
- Regular updating of the communication strategy as changes are implemented, such as the governance body beginning its work

**Benefits**

- Effective, well-organised communications
- System for monitoring and keeping track of communication

**Lead**

HEDIIP + External partner + Governance body

---

**8.3.2 Objective 1: Agreed Data Governance Model**

For the New Data Landscape to be effective, there will need to be effective governance of the Standard Dataset and data definitions, and data collection in the sector as a whole. This workstream will define and set up the governance body that will oversee that.

<table>
<thead>
<tr>
<th>Objective 1</th>
<th>Agreed Data Governance Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Project</td>
<td>Description</td>
</tr>
<tr>
<td>1.1 Agree terms of reference</td>
<td>There is a need to agree the terms of reference in order to clearly define the role of the governance body.</td>
</tr>
</tbody>
</table>

**Completion date**

15.06.2015

**Action**

- Create draft ToR for Governance body, based on stakeholder feedback in developing the Blueprint
- Agree terms of reference with key stakeholders and establish system to review and check progress.

**Benefits**

- Drafting ToR provides a framework for discussion of remit of governance body
- Builds on work done in Phase I, keeping momentum high

**Lead**

HEDIIP + External Partner
**Objective 1**  
Agreed Data Governance Model

<table>
<thead>
<tr>
<th>Project</th>
<th>Description</th>
<th>Completion date</th>
<th>Action</th>
<th>Benefits</th>
<th>Lead</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.2</td>
<td>Agree constitution and remit of governance body</td>
<td></td>
<td>Using ToR as a basis, have stakeholder meetings across the sector to agree the balance between HEPs and Collectors</td>
<td>With in principle agreement to the scope of the governance body, it has a greater chance of success</td>
<td>HEDIIP + External Partner</td>
</tr>
<tr>
<td></td>
<td>Building on the ToR, and Phase I, agree the balance between HEP representation and Collector representation, who actually sits on the governance body, and its remit</td>
<td></td>
<td>Work with Stakeholders to agree who will sit on the governance body, and the wider group they will represent</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Run workshops to agree the remit of the governance body</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.3</td>
<td>Develop job descriptions for governance body management function</td>
<td>29.07.15</td>
<td>Analyse similar bodies’ management structure</td>
<td>Job descriptions built from similar bodies in other sectors will allow for quicker development, and build on best practice</td>
<td>HEDIIP</td>
</tr>
<tr>
<td></td>
<td>For the governance body to be effective, it requires effective staff to deliver its remit and carry out the requests of the sector representatives</td>
<td></td>
<td>Develop job descriptions</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Objective 1</td>
<td>Agreed Data Governance Model</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-------------</td>
<td>-------------------------------</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Project</strong></td>
<td><strong>Description</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.4 Recruit to the new structure</td>
<td>Once job descriptions have been finalised, advertise and recruit to the new structure</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Completion date</strong></td>
<td>28.08.2015</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Action</strong></td>
<td>Advertise roles in appropriate media</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Shortlist candidates</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Interview</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Offer</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Benefits</strong></td>
<td>Effective recruitment will deliver motivated staff with the credibility to be effective across the sector</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Lead</strong></td>
<td>HEDIIP + HESA</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Objective 1</th>
<th>Agreed Data Governance Model</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Project</strong></td>
<td><strong>Description</strong></td>
</tr>
<tr>
<td>1.5 Formally Constitute Governance Body</td>
<td>Once the management function is in place, the governance body can be constituted formally, and begin operation</td>
</tr>
<tr>
<td><strong>Completion date</strong></td>
<td>01.09.15</td>
</tr>
<tr>
<td><strong>Action</strong></td>
<td>Communicate to stakeholders that governance body is beginning operation</td>
</tr>
<tr>
<td></td>
<td>Hold first meeting of governance body</td>
</tr>
<tr>
<td><strong>Benefits</strong></td>
<td>With a management function in place, and the first meeting held, the governance body becomes a reality and stakeholders can see it is operating, making the change more obvious</td>
</tr>
<tr>
<td><strong>Lead</strong></td>
<td>HEDIIP + HESA</td>
</tr>
</tbody>
</table>
### 8.3.3 Objective 2: Agreed Standard Dataset and Data Standards

The key underpinning of the New Data Landscape is the Standard Dataset and Data Standards. This workstream will define the new Standard Dataset and data standards that will go along with that.

<table>
<thead>
<tr>
<th>Project</th>
<th>Description</th>
<th>Completion date</th>
<th>Action</th>
<th>Benefits</th>
<th>Lead</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1 Analyse current datasets and data standards, and identify overlap</td>
<td>Reviewing the datasets used by the different collectors, and identifying areas of commonality, as well as identifying where similar data is collected, but with different standards applied to it</td>
<td>31.07.15</td>
<td>Review the datasets and standards used by the different collectors Identify the overlaps in dataset and standards, as well as the areas of similarity Agree these with stakeholders</td>
<td>Identifying common datasets and definitions will maintain momentum of the project Areas that are already common, or similar, will provide a solid foundation to build a Standard Dataset and set of definitions that can be accepted by all stakeholders</td>
<td>HEDIIP</td>
</tr>
<tr>
<td>2.2 Develop Standard Data</td>
<td>Need to create the first draft of a Standard Dataset.</td>
<td>28.08.15</td>
<td>Build on the work done in 2.1 to identify areas where a Standard Dataset would deliver benefit Base the new Standard Dataset on the HESA Student Return Develop draft Standard Dataset</td>
<td>Initial straw man for a Standard Dataset will generate discussion and acceptance of the principle</td>
<td>HEDIIP + External partner</td>
</tr>
</tbody>
</table>
### Objective 2
**Agreed Standard Dataset and Data Standards**

<table>
<thead>
<tr>
<th>Project</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.3 Develop Data Standards</td>
<td>Development of data standards that apply to the Standard Dataset</td>
</tr>
</tbody>
</table>

**Completion date**

<table>
<thead>
<tr>
<th>Project</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Completion date</td>
<td>30.10.15</td>
</tr>
</tbody>
</table>

**Action**

- Build on 2.1 to identify where data standards are already harmonised
- Build on 2.1 to identify where data standards are partially harmonised
- While developing Standard Dataset, build the standards into it

**Benefits**

- Builds a set of data standards around the Standard Dataset as it is being developed

**Lead**

- HEDIIP + External Partner

---

<table>
<thead>
<tr>
<th>Project</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.4 Consult on standard data and data standards</td>
<td>Seek feedback and input for the data standards from all key stakeholders.</td>
</tr>
</tbody>
</table>

**Completion date**

<table>
<thead>
<tr>
<th>Project</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Completion date</td>
<td>27.11.15</td>
</tr>
</tbody>
</table>

**Action**

- Identify relevant stakeholders for consultation
- Socialise proposed data standards and Standard Dataset with stakeholders
- Consult and revise data standards and standard data following feedback

**Benefits**

- Stakeholder input will lend greater legitimacy to the Standard Dataset and the common standards

**Lead**

- HEDIIP + external partner

---

<table>
<thead>
<tr>
<th>Project</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.5 Initial Standard Dataset agreed</td>
<td>Following feedback and revision, the initial Standard Dataset will need to be agreed by data collection agencies and HEPs.</td>
</tr>
</tbody>
</table>

**Completion date**

<table>
<thead>
<tr>
<th>Project</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Completion date</td>
<td>01.03.16</td>
</tr>
</tbody>
</table>

**Action**

- Issue draft of Standard Dataset, amended after consultation for broader feedback
- Amend based on feedback

**Benefits**

- Consulting more broadly will increase acceptance of the Standard Dataset

**Lead**

- HEDIIP + external partner
### Objective 2
Agreed Standard Dataset and Data Standards

<table>
<thead>
<tr>
<th>Project</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.6 Initial data standards agreed</td>
<td>Following feedback and revision, the initial data standards will need to be agreed by data collection agencies and HEPs.</td>
</tr>
<tr>
<td>2.7 Work with early adopter agencies to agree, implement and roll out</td>
<td>HEPs have suggested that this should be phased, with the greatest benefits likely to come from UCAS, HESA and SLC being in the first wave. Work with these agencies to agree timeline and implementation plan</td>
</tr>
<tr>
<td>2.8 Ongoing implementation activity</td>
<td>Ongoing project work to monitor the implementation workstream and the readiness of Data Collectors and HEPs to begin working with it</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Action</th>
</tr>
</thead>
<tbody>
<tr>
<td>Issue draft of Standard Dataset, amended after consultation for broader feedback</td>
</tr>
<tr>
<td>Amend based on feedback</td>
</tr>
<tr>
<td>Identify timeline for implementation</td>
</tr>
<tr>
<td>Agree implementation project members from each Agency</td>
</tr>
<tr>
<td>Establish a monitoring and reporting system</td>
</tr>
<tr>
<td>Analyse necessary changes to organisational structures, processes, etc</td>
</tr>
<tr>
<td>Ongoing project management of the implementation of the new dataset and standards in the first wave collectors</td>
</tr>
<tr>
<td>Design and implement new business processes</td>
</tr>
<tr>
<td>Regular meetings to discuss issues and seek solutions</td>
</tr>
<tr>
<td>Restructure of data collection departments where appropriate</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Benefits</th>
</tr>
</thead>
<tbody>
<tr>
<td>Consulting more broadly will increase acceptance of the data standards</td>
</tr>
<tr>
<td>Support from three of the largest Data Collectors, and clear timelines for implementation, will maintain momentum and increase profile and acceptance of the project in the sector</td>
</tr>
<tr>
<td>The long lead time allows organisations to prepare effectively for the New Landscape</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Lead</th>
</tr>
</thead>
<tbody>
<tr>
<td>HEDIIP + External partner</td>
</tr>
<tr>
<td>Governance Body</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Completion date</th>
</tr>
</thead>
<tbody>
<tr>
<td>01.03.16</td>
</tr>
<tr>
<td>01.03.16</td>
</tr>
<tr>
<td>03.10.2016</td>
</tr>
<tr>
<td>Objective 2</td>
</tr>
<tr>
<td>-------------</td>
</tr>
<tr>
<td><strong>Project</strong></td>
</tr>
<tr>
<td>2.9 Go live of new regime for early adopting collectors</td>
</tr>
<tr>
<td><strong>Completion date</strong></td>
</tr>
<tr>
<td><strong>Action</strong></td>
</tr>
<tr>
<td><strong>Benefits</strong></td>
</tr>
<tr>
<td><strong>Lead</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Objective 2</th>
<th>Agreed Standard Dataset and Data Standards</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Project</strong></td>
<td><strong>Description</strong></td>
</tr>
<tr>
<td>2.10 Amend based on operational feedback</td>
<td>As the new regime beds in, feedback on particular operational issues will be analysed, and amendments made to processes, Standard Dataset and data standards</td>
</tr>
<tr>
<td><strong>Completion date</strong></td>
<td>02.06.2017</td>
</tr>
<tr>
<td><strong>Action</strong></td>
<td>Regular meetings with stakeholders to discuss issues that arise Develop responses to these issues Implement responses as appropriate</td>
</tr>
<tr>
<td><strong>Benefits</strong></td>
<td>Feedback from live operation allows improvements to be implemented</td>
</tr>
<tr>
<td><strong>Lead</strong></td>
<td>Governance Body</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Objective 2</th>
<th>Agreed Standard Dataset and Data Standards</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Project</strong></td>
<td><strong>Description</strong></td>
</tr>
<tr>
<td>2.11 Go live of new regime for second phase adopters</td>
<td>Go live for the new data collection approach for early adopter Data Collectors that is linked to a phased CACHED implementation.</td>
</tr>
<tr>
<td><strong>Timeline</strong></td>
<td>01.16.18</td>
</tr>
<tr>
<td><strong>Action</strong></td>
<td>Begin using the new Standard Dataset and standard definitions in the second phase adopters Continue to monitor issues that arise, and implement solutions where appropriate</td>
</tr>
<tr>
<td><strong>Benefits</strong></td>
<td>Feedback from live operation allows improvements to be implemented</td>
</tr>
<tr>
<td><strong>Lead</strong></td>
<td>Governance Body</td>
</tr>
</tbody>
</table>
### Objective 2
Agreed Standard Dataset and Data Standards

<table>
<thead>
<tr>
<th>Project</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.12 Go live of new regime for all Data Collectors</td>
<td>Go live for the new data collection approach for all Data Collectors that is linked to a phased CACHED implementation.</td>
</tr>
</tbody>
</table>

**Timeline**
02.06.2018

**Action**
Begin using the new Standard Dataset and standard definitions in all Data Collectors
Continue to monitor its use and identify and address issues with the New Landscape

**Benefits**
Feedback from live operation allows improvements to be implemented

**Lead**
Governance Body

### 8.3.4 Objective 3: Process Redesign

The New Data Landscape will bring about a new way of working for HEPs and Data Collectors. To maximise its effectiveness, this workstream will identify areas of focus for process redesign within HEPs and Data Collectors, and implement process redesign projects to address issues that arise.

<table>
<thead>
<tr>
<th>Objective 3</th>
<th>Process Redesign</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Project</strong></td>
<td>3.1 Establish working group</td>
</tr>
</tbody>
</table>

**Description**
The working group will see secondees from collectors and HEPs design and implement the optimum data collection processes for HEPs and Collectors of different types to follow.

**Completion date**
25.09.15

**Action**
Agree role and scope of working group
Identify stakeholders to be involved in working group
Establish group

**Benefits**
The working group will be a small body that works with the HE data landscape management office.

**Lead**
HEDIIP + External partner
### Objective 3
#### Process Redesign

<table>
<thead>
<tr>
<th>Project</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.2 Agree secondees from HEPs &amp; collectors</td>
<td>The working group will need senior HEP and Data Collector administrators to be seconded to it to allow it to deliver effectively</td>
</tr>
</tbody>
</table>

**Completion date**
26.02.16

**Action**
- Identify secondees available
- Map skills to requirements
- Agree secondees with working group and governance body

**Benefits**
Secondee with direct knowledge and experience of the data collection processes for HEPs and Collectors will help identify any issues early, and be able to deliver workable solutions

**Lead**
Governance Body + External partner

<table>
<thead>
<tr>
<th>Project</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.3 Agree HEP responsibilities in the new regime</td>
<td>The new data collection regime will require changes in how HEPs collect and distribute data to collectors. This working group is best placed to agree how those changes will impact HEPs, and how they should be implemented.</td>
</tr>
</tbody>
</table>

**Completion date**
26.02.16

**Action**
- Define HEP responsibilities under new regime
- HEPs analyse their own data collection and reporting processes

**Benefits**
A shared understanding of the new regime, and how it will affect HEPs, as well as a shared view of how to implement any changes, will increase acceptance of the new regime

**Lead**
Governance Body + External partner

<table>
<thead>
<tr>
<th>Project</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.4 HEPs assess impact of new regime</td>
<td>HEPs will perform an evaluation of the impact of the new Blueprint and develop action plans</td>
</tr>
</tbody>
</table>

**Completion date**
31.05.16

**Action**
- HEPs analyse current processes
- HEPs design and implement new processes

**Benefits**
HEPs will have redesigned processes focused on the Blueprint and operating it effectively

**Lead**
HEPs
**Objective 3**
**Process Redesign**

### Project 3.5
HEPs amend their internal data collection and transformation processes where appropriate

**Description**
HEPs will need to amend how they collect, transform and report data internally, and up to Data Collectors based on the new regime.

**Completion date**
03.02.17

**Action**
- HEPs analyse current processes
- HEPs design and implement new processes
- HEPs develop implementation plans

**Benefits**
HEPs will have redesigned processes focused on the New Landscape and operating it effectively

**Lead**
HEPs

### Project 3.6
Go live of new regime

**Description**
HEPs will need to amend how they collect, transform and report data internally, and up to Data Collectors based on the new regime.

**Completion date**
04.06.18

**Action**
HEPs deliver implementation plans

**Benefits**
HEPs will have the capability to effectively interface with new Blueprint

**Lead**
HEPs

---

### Objective 4: Technology Implementation

To maximise the efficiency of the New Data Landscape, technology changes will need to be implemented. This workstream will analyse the necessary changes and implement them across Data Collectors.

### Project 4.1
Current technology assessment

**Description**
Identify 'as is' scenario for all providers’ and collectors’ technology offering

**Completion date**
25.09.2015

**Action**
- Review and assess current infrastructure (platform and network architecture) being used in HEPs and collectors
- Identify in-flight projects
- Identify common technology standards
- Identify areas of difference

**Benefits**
Shared understanding of the current technology landscape

**Lead**
HEDIIP + External partner
### Objective 4  
#### Technology Implementation

<table>
<thead>
<tr>
<th>Project</th>
<th>Description</th>
<th>Completion date</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.2 Future Technology Architecture</td>
<td>Agree future state technology reference architecture for the New Landscape, for HEPs and Collectors</td>
<td>02.10.15</td>
</tr>
</tbody>
</table>

**Action**
- Discuss and agree the high-level areas in scope for the 'To-Be' technology solution. Ensure the key stakeholders are involved in these discussions through 'Sign-off of Scope' throughout till the end of the Transformation Programme.
- Decide and agree with relevant stakeholders the approach to link different systems and applications physically and functionally in order to work as a united system.
- Develop the proposed "To-Be" Architecture for the target Solution.

**Benefits**
- Agreed technology architecture will give suppliers, HEPs and collectors sufficient time to develop towards its implementation

**Lead**
- HEDIIP + External partner

<table>
<thead>
<tr>
<th>Project</th>
<th>Description</th>
<th>Completion date</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.3 Migration Plan</td>
<td>Develop migration plan for migration to new architecture</td>
<td>01.03.16</td>
</tr>
</tbody>
</table>

**Action**
- Analyse gap between current state and future state
- Develop migration plan to move to future state

**Benefits**
- Appropriate migration plan will allow for managed transition to new technology

**Lead**
- Governance body + External partner

<table>
<thead>
<tr>
<th>Project</th>
<th>Description</th>
<th>Completion date</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.4 Design and Build of appropriate modules</td>
<td>Module build for existing technology implementations, where appropriate, to allow new processes and data transmission standards in the New Landscape to take effect</td>
<td>01.09.16</td>
</tr>
</tbody>
</table>

**Action**
- Analyse future state architecture
- Develop appropriate modules on existing technology architectures where necessary for collectors

**Benefits**
- Purpose-built modules for existing technology implementations will allow Data Collectors to maximise the benefits of the New Landscape

**Lead**
- Governance body + External partner
### Objective 4
**Technology Implementation**

<table>
<thead>
<tr>
<th>Project</th>
<th>Description</th>
<th>Completion date</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.5 Testing and refinement</td>
<td>Testing and refinement of new modules</td>
<td>02.10.17</td>
</tr>
</tbody>
</table>

**Action**
- Refine new modules based on ongoing testing
- Implement new modules into live environment

**Benefits**
- Testing and refinement will minimise errors and problems with the new modules

**Lead**
- Governance body + External partner

<table>
<thead>
<tr>
<th>Project</th>
<th>Description</th>
<th>Completion date</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.6 Technology go live</td>
<td>Go live of new technology modules</td>
<td>04.06.18</td>
</tr>
</tbody>
</table>

**Action**
- Go live of new technology

**Benefits**
- New technology modules maximise the opportunity for success of the New Landscape

**Lead**
- Governance body + External partner
8.4 Implementation Plan Horizon Diagram

The figure below depicts the general flow of ‘events’ across a timeframe that build upon each other and converge to the delivery of the “Vision”. The diagram is broken into 4 key “deliverables” of Governance, Standards, Process redesign and Technology.
9 Indicative Costs

This section describes our estimates for the implementation of the projects included within the implementation plan. The approach undertaken to deliver the costings is also described below and they represent our best estimates using the information currently available using current industry prices. We have provided this information as an indication of the cost. Additional feasibility work will need to be undertaken to establish more precise estimates.

9.1 HESA CACHED Project

An overarching element of the strategy is the understanding that the data collection process will be radically improved through the HESA transformation. That project will look at how to collect more timely data in a fashion that works better and more efficiently for the HE sector as a whole. It will be aligned closely to this project, and many of the recommendations of this report will impact the HESA CACHED project. In this respect, it is likely that there will be some overlap in the cost estimates provided in this section, with those in the CACHED project. Once the CACHED project is at a more detailed design phase, and repetition can be clarified and adjusted as required.

9.2 Sector Data Processing Costs

We have only used a review of the costs of data processing across HEIs as it was assumed that this would provide the most robust assessment that can be modelled at this stage. Whilst the Blueprint will provide benefits for FECs and APs the bulk of the impact will be on HEIs.

The overall cost of processing data in the sector is very difficult to estimate. PA Consulting’s report of January 2009, ‘Positive Accountability: Review of the costs, benefits and burdens of accountability in the English Higher Education Sector’, estimates the cost for English universities of administering just five data returns (HESA, HESES, SLC, SHA and LSC) to be £12.8m pa. As HEDIIP have identified 97 Data Collectors, with well over 500 data collections taking place annually, we can be certain that this cost is much higher.

Our own high level survey of HEIs indicated a wide variation in the costs HEIs perceive around the processing of data returns. The lowest amount reported was £75,000 per annum, including centralised and decentralised costs. The most frequently reported range was between £650,000–£1.5m, with some outliers of £2m+.

If we assume a conservative figure of £500,000 per HEI (this being a third of the average cost reported in our survey), then the total cost to the sector is likely to be in excess of £80m pa. Depending on how HEIs categorise their data return processing activity, this number will rise or fall, but it is a reasonable estimate based on the information we have collected and the HEIs we have spoken to.

The detail behind these numbers will improve as more HEIs are spoken to and their figures derived. We had fewer than 26 responses to our survey, and spoke to an additional six HEIs. It is therefore important to only use these costs as broad estimates.

9.3 Summary of Implementation Costs

The costs are built up from different elements of the project.

Secondment of Senior Staff

We have assumed 10 senior professional services staff from different HEIs and Data Collectors will be seconded to the project for one day per week, on average. Their key role will be to agree on the Standard Dataset and the definitions for the data items. In addition they will advise on potential process changes required in HEIs and Data Collectors to maximise the benefits of the New Data Landscape.
This is central to the success of the project that we are recommending seconding them to the project for one day per week for a year. This will allow the project to develop some momentum, and have the right input from the right people at the right time.

The initial elements of this activity will be key to the ongoing implementation of the HESA CACHED Project, which is aiming, subject to approval, to use the Standard Dataset and definitions for the data items for a small number of contracted data users, including HE Funding Bodies.

It will be necessary to join these two activities together and ensure that the Standard Dataset that develops will allow HESA to deliver the common elements of the CACHED Project without destabilising the delivery of the New Data Landscape.

**Project team**

The project will need a dedicated, full-time project team to drive it forward. We have assumed a project director, two project officers and a project support officer who will be full-time on the project, liaising with stakeholders, producing documentation and arranging and delivering workshops. Initially this team would sit within HEDIIP, and then most likely move to be under the remit of the Governance body, once that is up and running.

**External change management support**

It is likely that some form of external support will be required to facilitate the change management aspects of the project with different stakeholders. This should be in the form of contractors with specific skillsets or external consultancy organisations producing specific outputs.

Either way, this will be an essential part of the project if it is to be successful, as you will be able to bring in specific expertise for specific elements of the project, and know that they will be completed to the requisite standard without placing a strain on the project team, who will have other things to do. As yet, the specific activities they will undertake are impossible to define accurately. However, based on our experience of other large transformation programmes, they are likely to be required to undertake significant change management activity (mapping current processes, future process, designing the transition) and then implementing it. The volume of days estimated has been scaled to provide the significant volume of support that will be required to deliver the blueprint within a very challenging timeline.

There will be an element of crossover with the change management support required for the HESA CACHED Project. The CACHED budget provides for a significant amount of change management support, some of which will inevitably overlap the work anticipated in this document. At this stage, neither project is sufficiently mature to be able to define exactly what, or how much, that overlap will represent, but at the detailed design stage, it will need to be defined in more detail.

**Governance team**

The Governance Team will need to be set up early, and be closely linked in with the work to establish the Standard Dataset. They will also need to deliver a functioning mechanism for governing the overall Standard Dataset, and an agreed methodology for how changes will be made to it.

This could be carried out by the Project Team in the first year, but given the importance of good governance to the effective functioning of the New Landscape, we feel that it would be better to have it functioning from the beginning. As the project grows, and the governance of the Standard Dataset becomes more complex and demanding, the Governance Team can expand to cover the roles outlined above.

As per our description of the Governance function above, we have assumed 4 FTE, with an average salary (including on costs) of £60,000.

**Technology Upgrades**

The technology in some of the Data Collectors will need to be upgraded to meet the requirements of the new Standard Dataset and data standards, as well as the new processes being implemented, and the increased size of single datasets being received by some collectors (where previously they received many, smaller datasets, now they will be receiving one, much larger one).
We have assumed that the biggest effect will be on the larger Data Collectors, with the largest investment in systems and infrastructure.

Spend on programming and developing new modules has been estimated at around £150,000 per piece of work, which will cover module design, coding, building, testing and migration. Clearly at this stage no accurate figure can be provided, and this cost could go up or down significantly depending on the scope of requirements for any technology development.

It should be noted that the technology upgrades envisaged for HESA are NOT included in this figure. The technology upgrade costs for HESA are factored into the HESA CACHED budget.

<table>
<thead>
<tr>
<th>ITEM</th>
<th>UNIT COST</th>
<th>NUMBER</th>
<th>TOTAL ANNUAL COST</th>
</tr>
</thead>
<tbody>
<tr>
<td>Secondment for senior staff seconded to the project</td>
<td>£800 per day</td>
<td>50 days @ 1 day pw 10 people</td>
<td>£400,000</td>
</tr>
<tr>
<td>Project team</td>
<td>£40,000 average per person per year</td>
<td>4</td>
<td>£160,000</td>
</tr>
<tr>
<td>External change management support</td>
<td>Average £1,000 per day per resource</td>
<td>200 approximately days</td>
<td>£200,000</td>
</tr>
<tr>
<td>Governance team</td>
<td>£60,000 average per person per year</td>
<td>4 people</td>
<td>£240,000</td>
</tr>
<tr>
<td>Technology upgrades in data collectors (Not HESA)</td>
<td>Up to £150,000 per project</td>
<td>Up to 10</td>
<td>£1,500,000</td>
</tr>
</tbody>
</table>

9.4 Assumed Savings

It has only been possible to make some broad estimates of the savings that the landscape should realise. We have estimated savings for HEIs and Data Collectors, as detailed below. Savings for APs and FEs have not been included:

- **HEIs** – HEIs responding to the survey estimated an average saving of £544k pa. We have however taken a prudent assumption that a saving of two FTEs costing £40k pa including on costs could be realised or redeployed. Upscaling this for 160 HEIs provides an annual indicative saving of £12.8m pa.

- **Data Collectors** – Due to the variety of different change programmes that are being undertaken it is very difficult to ascertain a reliable estimate of the savings that could be delivered following the implementation phase. For this reason we have assumed that across the key Data Collectors covered by this review, and assuming that the landscape is adopted as proposed, then 20 FTEs could be saved or redeployed. Using an assumption £50k per FTE including on costs this provides an indicative saving of £1m pa.

In order to ensure that the benefits are not exaggerated we have taken a prudent approach which has been to utilise a fifth of the savings that have been estimated by HEIs in the survey.
9.5 **Four-Year Analysis**

Looking at the four-year time horizon, most of the spend occurs in years 1–2, during which time benefits will flow to the sector.

Initial benefits will be begun to be realised within 12 months as the communication strategy will provide evidence of data standards and the plans for the Standard Dataset.

Significant benefits begin to be felt within 24 months, when the new regime will have been working, and the second phase adopters will begin operating.

We feel that Year 2 will not see significant reallocation benefits from either HEIs or Collectors, as they test the new regime in a more difficult scenario as more collectors come on board. However, it is likely that some reallocation of staff will occur and we have reflected that below.

Year 3 will see the final wave of adopters come on board, meaning that, in Year 3, all Standard Dataset data will flow through HESA and thence to the Collectors, meaning HEIs and Collectors will be able to reallocate more staff to other value add duties or reduce headcount. We still do not anticipate the full number we identified above being reallocated in Year 3, however. This may happen in Year 4 or later.

We have allowed for external change management support in each of the first three years. The total cost for this could go up or down, depending on how much external support is required. It is possible that some of this could be moved into project team costs, but external expertise is likely to be required to help manage the change across the sector, and to maintain momentum.

<table>
<thead>
<tr>
<th>Item</th>
<th>Year 1</th>
<th>Year 2</th>
<th>Year 3</th>
<th>Year 4</th>
<th>Total</th>
<th>Incurred by</th>
</tr>
</thead>
<tbody>
<tr>
<td>Secondment for senior staff seconded to the project</td>
<td>£50,000</td>
<td>£250,000</td>
<td>£100,000</td>
<td></td>
<td>£400,000</td>
<td>HEIs, Data Collectors</td>
</tr>
<tr>
<td>Project team</td>
<td>£160,000</td>
<td>£160,000</td>
<td>£160,000</td>
<td></td>
<td>£480,000</td>
<td>HEDIIP / Governance Body</td>
</tr>
<tr>
<td>External change management support</td>
<td>£75,000</td>
<td>£100,000</td>
<td>£25,000</td>
<td></td>
<td>£200,000</td>
<td>HEDIIP / Governance Body</td>
</tr>
<tr>
<td>Governance team</td>
<td>£30,000</td>
<td>£240,000</td>
<td>£240,000</td>
<td>£240,000</td>
<td>£750,000</td>
<td>Governance Body</td>
</tr>
<tr>
<td>Technology upgrades</td>
<td>-</td>
<td>£1,000,000</td>
<td>£500,000</td>
<td></td>
<td>£1,500,000</td>
<td>Data Collectors</td>
</tr>
<tr>
<td>Data Collector Process Change</td>
<td>-</td>
<td>£1,500,000</td>
<td>£500,000</td>
<td></td>
<td>£2,000,000</td>
<td>Data Collectors</td>
</tr>
<tr>
<td>HEI Process Change</td>
<td>-</td>
<td>£4,500,000</td>
<td>£1,500,000</td>
<td></td>
<td>£6,000,000</td>
<td>HEIs</td>
</tr>
<tr>
<td>Staff reallocation saving (HEI)*</td>
<td>-</td>
<td>-</td>
<td>£3,000,000</td>
<td>£6,000,000</td>
<td>£9,000,000</td>
<td>HEIs</td>
</tr>
<tr>
<td>Staff reallocation saving (collectors)</td>
<td>-</td>
<td>-</td>
<td>£300,000</td>
<td>£600,000</td>
<td>£900,000</td>
<td>Data Collectors</td>
</tr>
<tr>
<td>Total</td>
<td>£315,000</td>
<td>£7,750,000</td>
<td>£275,000</td>
<td>£6,360,000</td>
<td>£1,430,000</td>
<td>-</td>
</tr>
</tbody>
</table>
* The total £12.8m reallocation saving for HEIs and the £1m of Data Collector savings will not deliver fully until Year 5, once the full landscape has been in place for a year.

The table above looks at the four-year horizon for the project, over which time there is a net sector cost that comes from implementation and transition costs. From Year 5, there will be net benefits to HEIs and Data Collectors.

9.5.1 Assumptions

In developing the table above, we have built the costs on a series of assumptions, in advance of the detailed design phase of the project. These are detailed below:

- **Secondment of senior staff.** We have assumed that the secondment of staff will mainly take place in Year 2, as the project is really into the detail of defining the Standard Dataset and Definitions. There may be the need for more than 50 days of secondment support, depending on how the project develops. Years 2 and 3 may not require the full complement of seconded staff at 1 day pw, depending on how much acceptance there is across the sector for the developing Standard Dataset and definitions. If there is broad acceptance and agreement, then this number can come down. However, this will be such an important aspect of the project, due to the need for senior sector representatives to define and agree on these areas, that we have kept the commitment at a high level. The development of the Standard Dataset is a key dependency for the HESA CACHED Project, and there will need to be coordination with the CACHED Project Team. Initially the CACHED Project will only seek to serve around five contractual data users, including HE Funding Bodies, so their focus will be considerably smaller.

- **Project Team.** We have assumed the project team will be required in Years 1–3, while the main body of redesign work is undertaken. There may be a requirement to maintain some project staff beyond this date, depending on how the project is progressing.

- **External Change Management Support.** We have assumed External Change Management Support will be required in Years 1–3. The exact amount of support required will depend on a number of factors, but the number we have included, we believe, realistic, based on the amount of work that will be required. The HESA CACHED Project will also require change management support for its implementation. While there will be some differences in the type of support required (far more of the support for that project will be around process and technology changes), there is likely to be some overlap. At this stage it is not possible to identify where the overlap might occur, nor what the scale of it might be, but as the two projects move forward, communication between them will continue to be vital to ensure money is not spent twice on similar activity.

- **Governance Team.** This team will be recruited and in place by Year 2, with the full complement described above (4 people, average salary £60,000 including on costs). From then on, this is the main operational cost of the New Landscape.

- **Technology Upgrades.** We have assumed this will be split between Years 2 and 3, with the bulk of the work taking place in Year 3. As described above, the costs here will be on a per project basis, covering work such as technology architecture design, new application module coding and migration. This does not include the technology upgrades within HESA to facilitate the CACHED Project. That is separate, and costed within the CACHED Budget.

- **Data Collector Process Changes.** These changes will relate to changes in process to allow Data Collectors to take maximum advantage of the move from circa 100 separate collection relationships, to one. We have assumed an average of £10,000 per collector per year in Years 2 and 3.

- **HEI Process Changes.** These changes relate to the process changes needed in institutions to allow them to manage a single return, centrally, rather than several different returns, some of which are handled via Faculties. We have assumed an average of approximately £20,000 per year for Years 2 and 3 per institution, as their changes could well be greater and more complex than the collectors’.

- **Staff Reallocation Saving.** As described above, it is likely that the New Landscape will allow some staff in HEIs and Data Collectors to be reallocated to more value add activities. We have assumed that this will not be able to start until Year 3, although in reality savings should be possible earlier than this. This
will however be a recurrent annual saving thereafter. This saving will be in the order of the £12.8m pa for HEIs and £1m pa for Data Collectors quoted above.

- **Ongoing costs.** The ongoing costs for managing the New Landscape will be the Governance Body, whose costs should remain stable beyond Year 3, and any increased headcount at HESA to take over the handling of the data returns from HEIs. For HEIs, the ongoing cost position should be positive. The people reallocation described above will reduce their net cost of data collection.

- **HESA Process Changes.** There will be a requirement for significant support for HESA as it adapts to the provision of extra services, in-year data collections and a standard dataset. These costs have been factored in to the budget for the CACHED Project and are not included here.

**Note:** All of the above assumptions are informed, but outline in nature. They will change as requirements become more precise and market testing can be undertaken.
Appendices
## Appendix A: Stakeholder Lists

<table>
<thead>
<tr>
<th>Name</th>
<th>Organisation</th>
<th>Classification</th>
</tr>
</thead>
<tbody>
<tr>
<td>Andrew Ray</td>
<td>Deputy Director, Higher Education Analysis, BIS</td>
<td>Govt</td>
</tr>
<tr>
<td>Adrian Jones</td>
<td>Head of FE Data &amp; Analysis Unit, BIS</td>
<td>Govt</td>
</tr>
<tr>
<td>Helen Mansfield</td>
<td>Head of Workforce &amp; Development, Health Education North West London</td>
<td>PSRB</td>
</tr>
<tr>
<td>Steve Egan</td>
<td>Deputy Chief Executive and Director of Finance and Corporate Resources, HEFCE</td>
<td>Regulator</td>
</tr>
<tr>
<td>Richard Puttock</td>
<td>Head of Data and Management Information, HEFCE</td>
<td>Regulator</td>
</tr>
<tr>
<td>Bethan Owen</td>
<td>Director of Institutional Engagement, HEFCW</td>
<td>Regulator</td>
</tr>
<tr>
<td>Sian McCleave</td>
<td>Department for Employment and Learning DELNI</td>
<td>Govt</td>
</tr>
<tr>
<td>Ben Coats</td>
<td>BIS</td>
<td>Govt</td>
</tr>
<tr>
<td>Hannah Falvey</td>
<td>Head of Statistics, HEFCW</td>
<td>Regulator</td>
</tr>
<tr>
<td>Alison Alden</td>
<td>Chief Executive, HESA</td>
<td>Data Collector</td>
</tr>
<tr>
<td>Philip Gallagher</td>
<td>Operations Manager, QTS and Induction Unit, NCTL</td>
<td>HE Service</td>
</tr>
<tr>
<td>Shaun Osborne</td>
<td>Project Manager, NCTL</td>
<td>HE Service</td>
</tr>
<tr>
<td>Dr Iain Cameron</td>
<td>Head of Research Careers and Diversity, RCUK</td>
<td>HE Sector</td>
</tr>
<tr>
<td>Martin Fairbaim</td>
<td>Senior Director and Secretary to the Council, SFC</td>
<td>Regulator</td>
</tr>
<tr>
<td>Martin Smith</td>
<td>Deputy Director Funding Policy, SFC</td>
<td>Regulator</td>
</tr>
<tr>
<td>David Wallace</td>
<td>Deputy CEO &amp; Director of Strategic Development, SLC</td>
<td>Data Collector</td>
</tr>
<tr>
<td>Dave Cartwright</td>
<td>Senior Manager, Data Services, SLC</td>
<td>Data Collector</td>
</tr>
<tr>
<td>Helen Thorne</td>
<td>Director of Policy and Research, UCAS</td>
<td>Data Collector</td>
</tr>
<tr>
<td>Mike Spink</td>
<td>Data Architect, UCAS</td>
<td>Data Collector</td>
</tr>
<tr>
<td>Professor Neil Gorman</td>
<td>Chair of HEDIIP Programme Board</td>
<td>HEI</td>
</tr>
<tr>
<td>Will Spinks</td>
<td>Registrar, Secretary and Chief Operating Officer, University of Manchester</td>
<td>HEI</td>
</tr>
<tr>
<td>Mike Potter</td>
<td>Principal, Guildford College</td>
<td>FEC</td>
</tr>
<tr>
<td>Graeme Wise</td>
<td>Assistant Director of Policy, NUS</td>
<td>NUS</td>
</tr>
<tr>
<td>Julie Tam</td>
<td>Head of Analysis, UUK</td>
<td>HE sector</td>
</tr>
<tr>
<td>Mike Glover</td>
<td>Academic Registrar, Warwick University. Member of the Academic Registrars Council</td>
<td>HEI</td>
</tr>
<tr>
<td>Name</td>
<td>Organisation</td>
<td>Classification</td>
</tr>
<tr>
<td>-----------------------</td>
<td>----------------------------------------------------------------</td>
<td>-----------------------</td>
</tr>
<tr>
<td>Catherine Elliott</td>
<td>Education and Skills Senior Executive, Engineering Council</td>
<td>PSRB</td>
</tr>
<tr>
<td>John Perks</td>
<td>Head of Funding Systems Architecture, Skills Funding Agency</td>
<td>Regulator</td>
</tr>
<tr>
<td>Jonathon Dalton</td>
<td>Head of Consumer Services Planning, Learning Records Service</td>
<td>Data Collector</td>
</tr>
<tr>
<td>Kirsty White</td>
<td>Head of Planning, Research and Development, General Medical Council</td>
<td>PSRB</td>
</tr>
<tr>
<td>Neil Jacobs</td>
<td>Head of Scholarly Communications Support, Jisc</td>
<td>HE Service</td>
</tr>
<tr>
<td>Rachel Bruce</td>
<td>Deputy Chief Innovation Officer, Jisc</td>
<td>HE Service</td>
</tr>
<tr>
<td>Peter Tinson</td>
<td>Executive Director, UCISA</td>
<td>PSRB</td>
</tr>
<tr>
<td>Lucy Hodson</td>
<td>HESPA</td>
<td>Professional Body</td>
</tr>
<tr>
<td>John Cartwright</td>
<td>UCISA</td>
<td>Professional Body</td>
</tr>
<tr>
<td>Prof Simon Gaskell</td>
<td>Chair of HESA</td>
<td>Data Collector</td>
</tr>
<tr>
<td>Charlie Taylor</td>
<td>NCTL</td>
<td>Government Agency</td>
</tr>
<tr>
<td>Jo Lenaghan</td>
<td>HEE’s Director of Strategy</td>
<td>Government Agency</td>
</tr>
<tr>
<td>Mary Curnock Cook</td>
<td>UCAS</td>
<td>Data Collector</td>
</tr>
<tr>
<td>Alison Alden</td>
<td>HESA</td>
<td>Data Collector</td>
</tr>
<tr>
<td>David Wallace</td>
<td>SLC</td>
<td>Data Collector</td>
</tr>
<tr>
<td>George Shirley</td>
<td>UKVI</td>
<td>Data Collector</td>
</tr>
<tr>
<td>Neil Gorman</td>
<td>Chair of HEDIIP</td>
<td>HE Service</td>
</tr>
<tr>
<td>Andrew Ray</td>
<td>BIS</td>
<td>Data Collector</td>
</tr>
<tr>
<td>Prof Madeleine Atkins</td>
<td>HEFCE</td>
<td>Data Collector</td>
</tr>
<tr>
<td></td>
<td>BUFDG NW and Midlands Regional Meetings</td>
<td>Professional Body</td>
</tr>
<tr>
<td></td>
<td>UCISA</td>
<td>Professional Body</td>
</tr>
<tr>
<td></td>
<td>Russell Group Registrars</td>
<td>Professional Body</td>
</tr>
<tr>
<td>Elaine Drennan</td>
<td>Senior Statistician, Scottish Government</td>
<td>Government Agency</td>
</tr>
<tr>
<td>Chris Williams</td>
<td>Welsh Govt</td>
<td>Government Agency</td>
</tr>
<tr>
<td>6 Individual HEIs</td>
<td>Universities</td>
<td>HEI</td>
</tr>
<tr>
<td>Carol Cook</td>
<td>Solicitors Regulation Authority</td>
<td>PSRB</td>
</tr>
<tr>
<td>Maxine Leslie</td>
<td>Academic Accreditation Manager BCS</td>
<td>PSRB</td>
</tr>
<tr>
<td>Bob Rabone</td>
<td>BUFDG NW and Midlands Regional Meetings</td>
<td>Professional Body</td>
</tr>
<tr>
<td></td>
<td>Russell Group IT Directors</td>
<td>Professional Body</td>
</tr>
</tbody>
</table>
HEDIIP Project Management Teams

HEDIIP Programme Board

<table>
<thead>
<tr>
<th>NAME</th>
<th>TITLE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Andrew Ray</td>
<td>Deputy Director, Higher Education Analysis, BIS</td>
</tr>
<tr>
<td>Steve Egan</td>
<td>Deputy Chief Executive and Director of Finance and Corporate Resources, HEFCE</td>
</tr>
<tr>
<td>Bethan Owen</td>
<td>Director of Institutional Engagement, HEFCW</td>
</tr>
<tr>
<td>Alison Allden</td>
<td>Chief Executive, HESA</td>
</tr>
<tr>
<td>Dr Iain Cameron</td>
<td>Head of Research Careers and Diversity, RCUK</td>
</tr>
<tr>
<td>Martin Fairbairn</td>
<td>Senior Director and Secretary to the Council, SFC</td>
</tr>
<tr>
<td>David Wallace</td>
<td>Deputy CEO and Director of Strategic Development, SLC</td>
</tr>
<tr>
<td>Helen Thorne</td>
<td>Director of Policy and Research, UCAS</td>
</tr>
<tr>
<td>Professor Neil Gorman (Chair)</td>
<td>Chair of HEDIIP</td>
</tr>
<tr>
<td>Ian Child</td>
<td>Deputy Vice Chancellor, University of Chichester</td>
</tr>
<tr>
<td>Will Spinks</td>
<td>Registrar, Secretary and Chief Operating Officer, University of Manchester</td>
</tr>
<tr>
<td>Mike Potter</td>
<td>Principal, Guildford College</td>
</tr>
<tr>
<td>Graeme Wise</td>
<td>Assistant Director of Policy, NUS</td>
</tr>
<tr>
<td>Julie Tam</td>
<td>Head of Analysis, UUK</td>
</tr>
</tbody>
</table>

HEDIIP New Landscape Project Board

<table>
<thead>
<tr>
<th>NAME</th>
<th>TITLE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Andy Youell</td>
<td>Director, HEDIIP; Project Executive</td>
</tr>
<tr>
<td>Richard Puttock</td>
<td>Head of Data and Management Information, HEFCE</td>
</tr>
<tr>
<td>Dave Cartwright</td>
<td>Senior Manager, Data Services, SLC</td>
</tr>
<tr>
<td>Julie Tam</td>
<td>Head of Analysis, UUK</td>
</tr>
<tr>
<td>John Britton</td>
<td>Deputy Director of Planning, Cardiff University</td>
</tr>
<tr>
<td>Kirsty White</td>
<td>Head of Planning, Research and Development, General Medical Council</td>
</tr>
<tr>
<td>Luke Taylor</td>
<td>Assistant Director of IT Services, Bristol University; Chair of the UCISA Corporate Information Systems Group</td>
</tr>
</tbody>
</table>

HEDIIP Advisory Panel
<table>
<thead>
<tr>
<th>NAME</th>
<th>TITLE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Helen Mansfield</td>
<td>Head of Workforce &amp; Development, Health Education North West London</td>
</tr>
<tr>
<td>John Stock</td>
<td>Workforce Planning and Information Analysis Lead, HEE</td>
</tr>
<tr>
<td>Rob Smith</td>
<td>Head of Planning and Information, HEE</td>
</tr>
<tr>
<td>Richard Puttock</td>
<td>Head of Data and Management Information, HEFCE</td>
</tr>
<tr>
<td>Hannah Falvey</td>
<td>Head of Statistics, HEFCW</td>
</tr>
<tr>
<td>Andrew Horsman</td>
<td>Head of Data Management, HESA</td>
</tr>
<tr>
<td>Philip Gallagher</td>
<td>Operations Manager, QTS and Induction Unit, NCTL</td>
</tr>
<tr>
<td>Shaun Osborne</td>
<td>Project Manager, NCTL</td>
</tr>
<tr>
<td>Martin Smith</td>
<td>Deputy Director Funding Policy, SFC</td>
</tr>
<tr>
<td>Dave Cartwright</td>
<td>Senior Manager, Data Services, SLC</td>
</tr>
<tr>
<td>Mike Spink</td>
<td>Data Architect, UCAS</td>
</tr>
<tr>
<td>Simon Robshaw</td>
<td>Data Analyst, UCAS</td>
</tr>
<tr>
<td>Julie Tam</td>
<td>Head of Analysis, UUK</td>
</tr>
<tr>
<td>Mike Glover (Chair)</td>
<td>Academic Registrar, Warwick University. Member of the Academic Registrars Council</td>
</tr>
<tr>
<td>Andrew Fisher</td>
<td>Academic Registrar, London South Bank University. Member of Student Records Officers Conference Committee</td>
</tr>
<tr>
<td>Catherine Elliott</td>
<td>Education and Skills Senior Executive, Engineering Council</td>
</tr>
<tr>
<td>Christine Couper</td>
<td>Director of Strategic Planning, Greenwich University. Member of the Higher Education Strategic Planners Association Executive</td>
</tr>
<tr>
<td>Esther Williams</td>
<td>Director of Student Administration, University of Bristol. Member of the AUA</td>
</tr>
<tr>
<td>Jayne Rowley</td>
<td>Director of Business Services, Graduate Prospects</td>
</tr>
<tr>
<td>Joanna Wagstaffe</td>
<td>Director of Academic Registry, University of Law</td>
</tr>
<tr>
<td>John Britton</td>
<td>Deputy Director of Planning, Cardiff University</td>
</tr>
<tr>
<td>John Perks</td>
<td>Head of Funding Systems Architecture, Skills Funding Agency</td>
</tr>
<tr>
<td>Jonathon Dalton</td>
<td>Head of Consumer Services Planning, Learning Records Service</td>
</tr>
<tr>
<td>Kirsty White</td>
<td>Head of Planning, Research and Development, General Medical Council</td>
</tr>
<tr>
<td>Lesley Donnithorne</td>
<td>HR Manager (Systems, Information &amp; Grading), UWE Bristol. Member of UHR</td>
</tr>
<tr>
<td>Luke Taylor</td>
<td>Assistant Director of IT Services, Bristol University. Chair of the UCISA Corporate Information Systems Group</td>
</tr>
<tr>
<td>Martin Boyle</td>
<td>Director of Planning and Business Intelligence, University of Glasgow</td>
</tr>
<tr>
<td>Matthew Dean</td>
<td>Technology Manager, AoC</td>
</tr>
<tr>
<td>Neil Jacobs</td>
<td>Head of Scholarly Communications Support, Jisc</td>
</tr>
<tr>
<td>Rachel Bruce</td>
<td>Deputy Chief Innovation Officer, Jisc</td>
</tr>
<tr>
<td>NAME</td>
<td>TITLE</td>
</tr>
<tr>
<td>-------------------</td>
<td>----------------------------------------------------------------------</td>
</tr>
<tr>
<td>Nick Johnstone</td>
<td>Senior Policy Advisor, GuildHE</td>
</tr>
<tr>
<td>Olivia Roberts</td>
<td>Head of Student Services &amp; Systems, Queens University Belfast</td>
</tr>
<tr>
<td>Paul Hazell</td>
<td>Assistant Director, Research, Information and Enquiry, QAA</td>
</tr>
<tr>
<td>Peter Tinson</td>
<td>Executive Director, UCISA</td>
</tr>
<tr>
<td>Zoe Stockdale</td>
<td>Graduate Studies Manager, University of Oxford</td>
</tr>
</tbody>
</table>
Appendix B: Stakeholder Descriptions

Higher Education Providers (HEPs)

Organisations that are contracted to deliver higher education and training by one of the funding and/or commissioning bodies. We have also included training businesses that have had their programmes designated by the Quality Assurance Agency for Higher Education (QAA) as being eligible for student finance. These organisations are also responsible for the primary data collection and verification of the students. In some areas of this report these organisations have been generically referred to as Higher Education Providers. We have the following provider types:

- **HEI**: Higher Education Institutions (approximately 160);
- **FEC**: Further Education Colleges (approximately 250).

**AP**: Alternative Providers (approximately 700) – two main categories exist: designated (approved for student finance) and non-designated (includes approximately 200 QAA recognised providers for UKVI) government bodies

Government departments or organisations responsible for developing policies for education, training and skills, commissioning of training, contract management, visa and immigration (UKVI), data security or other specific services funded by central government. For the purposes of this report Funding Councils, Student Loans Company (SLC), NHS (including Health Education England) and Teacher Training agencies (e.g. National College for Teaching and Leadership and General Teaching Council for Wales) should also be considered to be government bodies.

Professional, statutory and regulatory bodies (PSRBs)

PSRBs are an extremely diverse group of autonomous organisations. They set the standards for, and regulate the standards of entry into, particular professions. There is a PSRB list managed by HESA as part of the data collection that underpins Unistats but this only applies to undergrad courses. They tend to have direct relationships with the faculties staff within universities.

Data Collectors

This term is used within this report to identify generic data collection activities. We have also introduced the label of ‘Key Data Collectors’ when we refer to the operations of the larger and or more significant Data Collectors which include:

- Government bodies;
- Higher Education Statistics Agency (HESA); and
- University and College Admissions Service (UCAS).

General public

We have grouped the requirements of general public, press, potential/current or previous students into the needs of the public as a whole. Their diverse need for high quality information is partly covered by the government’s expectations for open data agenda and also directly mandated in the policies for the collection and publication of information that informs student choices. Student views of the landscape have been represented by the National Union of Students (NUS).

Jisc

Jisc is not included in the diagram in Table3.4 as it is not a Data Collector. Jisc does however have a significant position with regard to its current and future impact on the data management landscape. It is a not-for-profit charity that is funded from the Higher Education Funding Councils and Further Education Funding Councils. It
is governed by a board that contains representation from the HE funders, UUK, AoC and HEPs. They deliver services across Higher Education and Further Education providers, where the FE Providers subscription is covered by the Funding Councils and the HE Providers subscriptions is part-covered by the funding bodies with subscription levied to the HEIs. They provide:

- Network and IT services via Janet;
- Digital Content;
- Advice; and
- Research and Development.

Jisc has been evolving since the changes prompted by HEFCE 2010 Wilson review which resulted in Jisc being registered as a charity in 2013. It supports two types of membership, representative and institutional. Each representative member Association of Colleges (AoC), GuildHE and Universities UK (UUK) each hold 30% of the voting rights. Every higher and further education institution across the UK has the opportunity to become an institutional member of Jisc and collectively hold 10% of the voting rights. There are currently 260 institutional members across FE and HE with 128 being HEIs.

The Board consists of representative members who each appoint one person to the board and act collectively to appoint the chair. Each representative member also sits on the funders and owners group which provides advice to our trustees and leadership team. It is not a member of the Regulatory Partnership Group, but is linked to HEDIIP through inclusion in the Advisory Panel. HEDIIP also maintains frequent contact with Jisc during the delivery of its projects. It has independent governance which appears to have no direct impact on the data collection arrangements; some of its collaborative sector promoted projects deal with issues related to projects that are being delivered within other parts of the landscape.
Appendix C: Governance Comparators from Other Sectors

Health

Description
Activity Data in the health sector is generated by a number of publicly-funded trusts and also private healthcare providers.

This data is used to both plan for expenditure regarding the coming financial year but also to track in-year activity against plan and also to ensure on a monthly basis that the data being provided is correct, and to remove any duplication or challenge incorrectly coded episodes.

The data is approximately two months behind event with a further two submissions (one each month so it fits in with a standard submission cycle). Each submission is challenged with the provider and then corrected before the next submission before finalisation of the 3rd submission. Challenge is either done by the CCG itself or through a commissioning support service provider (CSU).

Governance solution
- Contractual data submissions – As part of the annual contracting of services locally with each provider CCGs (Clinical Commissioning groups) enter into contracts to purchase services.
- Data quality – Data quality is achieved through consultation with providers. While this is an official responsibility, the process does not included steering committees nor other formal mechanisms. Instead it forms a non-regular item on monthly/quarterly commissioning meetings between CCG and provider.

Relevance to HE sector
- Multiple individual external data providers with differing systems and purposes, including both primary and secondary care. Includes GPs, Hospital Trusts, Social and Mental healthcare providers.
- While there is legislation and contractual rewards and penalties due to the symbiotic relationship between providers, these are difficult to levy/enforce.
- Similar in terms of data requirement, Patient = Student can have multiple episodes = instances.
- National Repository through HSCIC (similar in role to HESA).

Key data facets:
- Patient Data – Driven by NHS ID, but not always available.
- Episode Data – Each treatment is coded so a pathway can be maintained (national and local submission).
- Outcomes – Are tracked.
- National submissions and regulatory requirements are faced by both provider and CCG/CSU (data processors) for their area.

Food standards

Description
Red Tractor Assurance is a not-for-profit organisation which provides a consumer-recognised assurance system via visible kite marking for produce in the following sectors: beef and lamb, dairy, pigs, poultry, combinable crops and sugar beet and fresh produce, based around the supply chain from field to shelf.
The organisation is funded through membership (which is voluntary), licences and food chain support activities, such as information provision for the British Poultry Council.

Data provided is taken from assessments conducted by 6 independent certification bodies when membership is requested; this is conducted again within 18 months. Activity and welfare data is also captured from downstream areas of the supply chain including vets. Data is captured and submitted to individual union bodies manually. However, there are application developments such as I-PIG FOR Bpex which captures Real Welfare data for pigs from vets. The data is then provided to Red Tractor Assurance.

**Governance solution**
- Membership required submissions – governance of the individual members’ adherence is undertaken through the assessments with any issues identified and a plan put in place to rectify. As Red Tractor has no financial or legislative powers non-conformances against a ‘key’ standard may result in suspension from the scheme.
- Data Quality – As the data is collected specifically for Red Tractor Assurance as part supply chain BAU the process does not include steering committees nor other formal mechanisms. However, applications under development do contain some key ‘at point of entry’ validation.

**Relevance to HE sector**
- No direct legislative or financial sanctions exist.
- Data is used for both statistical and assurance purposes.
- Data is generated by members through various other bodies and submitted to the central repository of Red Tractor Assurance.
- Data is collected once in the supply chain and used many times.

**Food Standards Agency FSA**

**Description**
Food Standards Agency (FSA) is responsible for public health issues relating to food safety. The FSA takes a governance approach to food standards across the UK by overseeing all local authorities’ public safety departments who are responsible for ensuring that standards are maintained by local companies, including shops and food processors. Annually, April time, the FSA agrees targets with each local authority, due to variations in size and structure, and requires them to undertake activities to meet these targets and submit supporting data.

Local authorities upload data, which has been generated from the local system(s) on which they record data on food law enforcement activities, to LAEMS. Once uploaded to LAEMS, the local authority data are aggregated to the pre-defined categories required by the FSA, including ‘interventions’, ‘sampling’ and ‘enforcement’.

**Governance solution**
- Backed by legislation – Food standards are governed by legalisation which starts at the EU level and has been interpreted by the UK and brought into the UK legislative system.
- Data quality – all data governance is the responsibility of the local authority. There are various penalties for local authorities and also businesses which do not follow the standards or are unable to meet targets set. All submitted data is audited to ensure accuracy and assure that public health is being sufficiently protected.

**Relevance to HE sector**
- Use of a central portal for aggregation and reporting of food standards adherence by local and national independent retailers through which data is gathered;
- Use of data for sampling and statistical analysis; and
Regulatory submissions are undertaken through a standard mechanism.

**Major retailers**

*Description*

Major retailers handle vast amounts of data, mainly through their epos system; however, systems such as these are driven from product catalogue/s within the business. These are populated via the procurement process when products are ordered from their suppliers in order to sell to customers. Once a product has been procured it must be inducted into the organisation’s own product catalogue accurately and quickly. Approaches vary between different retailers for collection process; however, for the purposes of this example we have chosen one which is developing this process.

The data contains many facets, including price and description products. The process of update takes place on a daily, as required basis based on the procurement process and any solution must be available to all suppliers, both existing and new.

**Governance solution**

- Reputational and Economic Mutual Benefit – There is mutual benefit in ensuring that data is passed accurately from supplier to purchaser.
- Data Quality – data suppliers hold the master versions of all product data with whole product details passed to the purchaser.
- Data quality is not maintained through formal bodies but through informal channels at present; however, development of steering groups is underway.

**Relevance to HE sector**

- Multiple independent suppliers of data who own the data passed to an organisation which relies heavily on the data.
- Requirement to ensure latest version of data is captured and there is a high need for maintaining critical data dimensions.
- Supplier may have its own software or maintain its own solution independent of the whole.
# Appendix D: International Governance Comparators

<table>
<thead>
<tr>
<th>Country</th>
<th>Data Collection Owners</th>
<th>Data Collection Centralised?</th>
<th>Data Collection In-Year?</th>
<th>Who Provides Oversight of Arrangements?</th>
</tr>
</thead>
<tbody>
<tr>
<td>UK</td>
<td>A mixture of sector-owned and government agencies</td>
<td>No</td>
<td>Mixed depending on requirements of Data Collectors</td>
<td>Various different bodies and Government departments</td>
</tr>
<tr>
<td>Canada</td>
<td>The Government collects its own data for its Ministry of Education. The main vehicle for financial data is an Annual Financial Information Return. Enrolment data is held centrally by the provincial government. There are also provincial and national organisations (non-government) who collect data from their members for the purposes of analyses and publication. Examples of these entities would include the Council of Ontario Financial Officers, Canadian Association of University Business Officers (CAUBO), Association of Universities and Colleges of Canada, and Higher Education Quality Council of Ontario</td>
<td>No as the data collection is highly dispersed. For example, in the province of Ontario, there is one organisation that collects application data, and a separate organisation for student finance. Also, school authorities, colleges and universities data are collected by different parts of the government</td>
<td>Data collection from the institutions would be on an annual, fiscal or academic year basis. The application and student loan data would be available to governments on a real-time basis</td>
<td>The provincial ministries of education would provide the oversight of government entities and data. For the provincial and national organisations, oversight is provided by their Board of Directors which are made up of representatives of the universities, or colleges or school boards. For example, the Board of CAUBO is made up of V-P, Finance from Canadian universities</td>
</tr>
<tr>
<td>Hong Kong</td>
<td>Mixed as the JUPAS is owned by the 8 HEIs and University Grants Committee is government owned</td>
<td>Application data in Hong Kong is collected and managed by “JUPAS” – Joint University Programmes Admissions System. Other data collections are managed by University Grants</td>
<td>JUPAS is managed by a Board of Management which comprises 16 Directors, 2 from each of the 8 member institutions (viz. CityU, HKBU, LingnanU, CUHK, HKIEd, PolyU, HKUST and HKU). Its main responsibilities are to devise, establish and develop policies and strategies for the scheme, to monitor and supervise its operation, and to review from time to time any</td>
<td></td>
</tr>
<tr>
<td>------------------------------------------</td>
<td>---------------------------------------</td>
<td>----------------------------------</td>
<td>---------------------------------------------</td>
<td></td>
</tr>
<tr>
<td>Committee which is a Government body that is responsible to oversee the eight publicly funded HEs in Hong Kong.</td>
<td>Yes</td>
<td>Within each year the providers are asked to return student data four times</td>
<td>Higher Education Data Committee provides oversight. It comprises representatives from the department, Tertiary Education Quality Standards, the Australian Research Council, Universities Australia, the Australian Council for Private Education and Training, the Council of Private Higher Education and the Australian Bureau of Statistics. The Committee will work with States and Territories in development of the datamart to ensure it meets the needs of all governments.</td>
<td></td>
</tr>
</tbody>
</table>

**Australia**

The Australian government Department for Education and Training requires HE providers to submit their data via their Higher Education Information Management System (HEIMS) this is legislated by the Higher Education Support Act 2003 (HESA).
Appendix E: Extract from Pathway to Reform

The diagram below shows a model to illustrate how the establishment of a shared governance framework and the sharing of expertise on data management and information standards might be achieved.

The elements of the model are shown in different colours as follows; note that it is quite likely that one organisation may fulfil multiple roles:

- **Blue** – an authority that provides the governance for the information landscape and commissions specific activities using the expertise that already exists in the various Agencies and HE Providers involved.
- **Red** – transactional relationships with HE Providers directly supporting business operations e.g. SLC, UCAS and UKBA.
- **Purple** – relationships that cover statistical or other more general data for processing to produce reports and analysis to inform funding and/or policy decisions as well as to assess quality, support accreditation and inform stakeholders in the wider community e.g. HEFCE, UCAS and HESA.
- **Green** – an element that attempts to address the HEBRG vision of a single interface through an optimised technical infrastructure. This element is ‘optional’ in terms of the functioning of the model.
- **Grey** – The HE Providers and their information systems suppliers and developers.
- **Yellow** – The wider data user community which comprises those HE Agencies who can use data supplied indirectly from others together with the wider public including students and HE Providers in some cases.
Appendix F: Analysis of Documentation

Key documents reviewed

- Redesigning the higher education data and information landscape. Strand 1 project report;
- The HE Information Landscape – Creating and Managing a Data Model;
- HEDIIP – Programme Initiation Document;
- HEDIIP – Programme Plan to July 2015;
- An inventory of data collections – Construction of the inventory and next steps;
- Classifying subject of study – A roadmap to a new Joint Academic Coding System;
- Student Data Collection Review – Towards standardisation and rationalisation;
- The Unique Learner Number in HE – A roadmap to implementation;
- HIGHER EDUCATION – Students at the Heart of the System;
- Seizing the data opportunity – A strategy for UK data capability;
- Higher Education Information Landscape – Current scope and future vision;
- The Information Authority – Identifying and Quantifying the Provider Data Burden – Executive Summary;
- Redesigning the higher education data and information landscape – A Pathway to Reform – Executive Summary;
- Positive accountability – Review of the costs, benefits and burdens of accountability in English higher education;
- Making your data work for you – Data Quality and Efficiency in Higher Education;
- Mapping the Higher Education Funding and Regulatory System in England;
- Professional, statutory and regulatory bodies: an exploration of their engagement with higher education;
- Higher Education Review;
- Review of the non-HE regulatory landscape and its impact on HEPs;
- Understanding the information needs of users of public information about higher education;
- What is a Course?; and
- Mapping the higher education system in England. IRPG/15, Agenda item 4.

Customer user views

<table>
<thead>
<tr>
<th>What are the challenges/Issues?</th>
<th>What is the current state?</th>
<th>What has been proposed to date?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Current system focused on too narrow a range of users.</td>
<td>Applicants to higher education primarily look for information on the UCAS website or on individual university websites, rather than on Unistats, the central website where much of this information is currently published.</td>
<td>Raise the profile of the information sources currently available to show prospective students, career advisors and teachers what they offer and how they can be used.</td>
</tr>
<tr>
<td>Students can find it difficult to navigate an increasingly diverse landscape.</td>
<td>There are currently in excess of 17 million ULNs allocated but the number is only used widely in the HE sector, where its use is compulsory. HESA and UCAS have provided for the collection of ULNs, where available, for a number of years and</td>
<td>The government will explore the skills shortages in data analytics and set out clear areas for government and industry collaboration.</td>
</tr>
<tr>
<td>The communication across data providers and collectors is sub-optimal – leading to misunderstanding, frustration, lack of clarity and data ambiguity.</td>
<td></td>
<td>Publish anonymised information about the teaching qualifications, fellowships and expertise of teaching staff at all levels.</td>
</tr>
<tr>
<td>HE providers are having to put in more effort than necessary to prepare their data returns.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

117
## What are the challenges/issues?

### What is the current state?

The major student record systems in use in the HE sector all provide the ability to handle the ULN. Each University publishes the Key Information Set (KIS) on their website on a course-by-course basis.

Students should also be able to access information about class sizes. Make available, course by course, data on the type and subjects of the qualifications held by previously successful applicants.

In order to adopt the ULN a major project will be needed:
- To be overseen by HEDIIP;
- Within the context of a sector-level data model;
- Involving any organisation that is collecting data from providers with an aim to seek to phase out other identifiers at the earliest opportunity.

### What has been proposed to date?

- To accommodate the ULN, a major project will be needed.
- The project is to be overseen by HEDIIP.
- The project will be within the context of a sector-level data model.
- The project will involve any organisation collecting data from providers with an aim to seek to phase out other identifiers at the earliest opportunity.

## Requirements perspective

### What are the challenges/issues?

Multiple unco-ordinated reporting requirements by different players in the system place a substantial administrative burden on HE Providers.

[The requirement is to] redesign the information landscape for HE in order to arrive at a new system that meets the needs of a wider group of users; reduces the duplication that currently exists, and results in timelier and more relevant data.

### What is the current state?

In general, less than half of PSRBs report having a formal requirements process in place. Formal processes, reviews and consultation are far more prevalent amongst major and NHS bodies.

The process of gathering the required data to complete the corporation tax return can be costly for HEPs in internal and external resource.

UUK estimates that the costs of FOI compliance in the HE sector are around £10 million per annum.

### What has been proposed to date?

- HEBRG and HESA should undertake further analysis to identify areas where greater alignment could be sought between PSRB requirements for data and the services offered by HESA.
- Consider simplification of the calculations for corporation tax returns for HEPs.
- Review whether existing HEPs should be removed from FOI altogether.

## Process perspective

### What are the challenges/issues?

The process used by collection agencies to gather data is inefficient (creating delay and causing extra work).

A process of convergence to a common data model must be established.

The perceived burden often arises from having to respond to a number of obligations at the same time of year, or at points in the year which are not aligned with the institutions’ own internal processes.

### What is the current state?

HE Provider and regulatory data exchange and business cycles are not aligned – thus reducing the value of the data.

Gaps in effective information sharing between sector bodies places an unnecessary reporting burden on HE providers and can impede effective working between regulatory bodies.

### What has been proposed to date?

- The process for inclusion of new PSRBs could be linked to enhanced and more robust collection of PSRB data. This would expand the detail of the collections inventory and support its maintenance.
- Set up a change control process with authorised representatives of the relevant stakeholders so as to ensure that any changes to the standard [data model] entities are effectively and quickly managed with the minimum of effort and bureaucracy.

**Good practice statements agreed by HEPs**
**What are the challenges/ issues?** | **What is the current state?** | **What has been proposed to date?**
---|---|---
Timely Quality Checking – Checking data regularly, rather than just before the return deadline, therefore reducing the peak of effort at this busy time of the year;  
Managing Change – Delivering change within a controlled, managed framework  
* Supporting Documentation  
Ensuring the dedicated data team has appropriate and complete documentation;  
HESA Protocol – Having HESA protocols that support the HEPs in focusing internal views on the value of the collection and returns process.

---

**Governance and data management**

<table>
<thead>
<tr>
<th>What are the challenges/ issues?</th>
<th>What is the current state?</th>
<th>What has been proposed to date?</th>
</tr>
</thead>
</table>
| The governance regime is relatively inefficient and ineffective.  
The lack of data standards causes ambiguity and misunderstanding.  
There is a standard NHS dataset but some SHAs have varied the standard to meet their own requirements.  
Additional institutional effort is required to meet the needs of the SHAs which have varied the Standard Dataset.  
There is a view [within Data Collectors] that there are inefficiencies in practices within providers which need to be addressed – that ‘they could do more to help themselves’.  
NHS returns to the current SHAs are typically managed within the relevant faculty/subject area (thus making governance and data consistency more difficult).  
While many institutions have policies setting out how they manage the totality of PSRB engagement, the primary link is invariably at subject/departamental level (thus making governance and data consistency more difficult).  
Less than 50% of PSRBs give data specifications to providers or use templates/proforma or external data definitions. This is more prevalent with Major & NHS bodies. | Undertake a programme of work to create a more coherent set of arrangements for the collection, sharing and dissemination of data.  
Task some of the key stakeholders in information flows to develop and propose the structure, resourcing and operation of a governance model for the data and information landscape.  
Shared understanding of data collection should be promoted through publication of more detailed information and through a forum for strategic discussions between providers and collectors.  
Establish a quarterly forum for standard [data model] development and convergence and engage all RPG stakeholders in that process as well as suitable representatives from RCUK, HE Providers and their suppliers. |  

---

**Organisation perspective**

<table>
<thead>
<tr>
<th>What are the challenges/ issues?</th>
<th>What is the current state?</th>
<th>What has been proposed to date?</th>
</tr>
</thead>
</table>
| The same data is collected by multiple agencies.  
Different strategies/approaches are followed by different organisations across the sector (e.g. inconsistency in PSRB data)  
The NHS is also in the process of major change, including replacing the current Strategic Health Authorities (SHAs) with the Local Education and Training Boards (LETBs). | Data is shared between collection agencies to avoid the need for institutions to provide the same information to multiple collectors.  
[HEDIIP should ] establish relationships with the HE Information |  

---
## What are the challenges/issues?

<table>
<thead>
<tr>
<th>What are the challenges/issues?</th>
<th>What is the current state?</th>
<th>What has been proposed to date?</th>
</tr>
</thead>
</table>
| between universities, colleges and alternative providers. Gaps in effective information sharing between sector bodies places an unnecessary reporting burden on HE Providers and can impede effective working between regulatory bodies. Not all NHS bodies which institutions engage with can be identified because they are not recognised as ‘accreditors’ within the Key Information Set (KIS). There appears to be PSRBs exercising a role in higher education that are not KIS-listed. People from institutions tended to stress the complexity and burden. Individuals from other sectors, however, questioned whether the landscape was as complex as, say, the NHS or the HE sector. | HEIs engage with the PSRBs in different ways but data collection tends to be qualitative and linked to the quality assurance of programmes of study. The inventory of data collections contains over 500 records but statutory reporting is excluded. Therefore, the inventory is not complete. [Only] one alternative provider (AP) institution contributed to the collection inventory. Therefore, the inventory is not complete. Some of the forces that potentially impede rationalisation:  
- Each funding agency is distributing public money for which it is solely accountable. This means that they cannot transfer their responsibility to other agencies.  
- Each agency requires assurance that its information requirements are being met.  
- Each funding and other agency will defend and promote its territory and sovereignty.  
- The sector is itself complex and diverse. | Authority and Information Standards Board for Education, Skills and Children’s Services (ISB-ESCS) and other similar organisations such as the HEE and/or NHS ISB, as appropriate. A key aspect of successfully developing and sustaining an ongoing process to re-design the data and information landscape is ... communication amongst organisations and people. It follows that ... commitment [is required] from the top management at every stakeholder and ... amongst their staff co-ordinating data standards and terminology |

## Application and data

### What are the challenges/issues?

Although data is routinely collected and stored electronically, this is predominantly in a relatively primitive form in relation to modern methods of automated data transmission and reception that are now commonly available. This also applies to the joining and comparison with data from other sources, which is largely both manual and informal.

Data collection is too labour intensive and could be more automated.

### What is the current state?

The predominant method used by Data Collectors is via emailed document attachments. Some organisations report use of online web forms or portals for data submission, suggesting a more modern approach. Particularly for data returns, HEPs with advanced, integrated IT systems have more automated data collection processes and incur less cost.

The UK is developing its e-infrastructure, including through substantial investment in broadband, and with 24 high performance computing systems – the fourth highest share in the world – linked together by the Janet high-bandwidth network. Developments in cloud computing mean that organisations of any size can have access to shared computing mean that organisations of any size can have access to shared computing.

### What has been proposed to date?

Develop a single application portal and integrated processes for both HE and student finance applications, where data common to both applications is entered only once. Detailed data should be made publicly available so it can be analysed and re-presented in a variety of formats. [It should be] presented ever more imaginatively. New connections can be made and new patterns will be identified.

The authoritative URIs [associated with the new coding framework] should be developed and maintained as a web service for the sector.

Establish a website for the Information Landscape and publish the initial data model and Lexicon/Thesaurus using HESA tools and presentation methods.
Data perspective

**What are the challenges/issues?**
Institutions should be able to make more use of their data to improve their own performance and presentation.

A common data model must be widely accessible, easy to understand and useful to a wide range of stakeholders.

Some HE Providers still struggle to provide timely and accurate data.

Institutions cannot easily use data derived from corporate systems for all the reporting to NHS/SHAs/PSRBs because of the particular requirements of the Data Collectors.

Although each [Data Collector] has a specific need for the information it requests, it can lead to the same basic details being collected over and over.

The word ‘course’ is used extensively in sector-level data systems and HE providers can represent their courses in different ways across these different systems.

There is a strong complaint from HEPs that they are asked for essentially the same data and information by several different bodies at different times.

HE Providers cannot perform accurate benchmarking due to inconsistent data.

**What is the current state?**
There is relatively little reference to security issues (less than 10% of respondents), with some mention of the need to maintain awareness.

The QAA does not have access to all relevant national datasets. Therefore, HE providers have to constrain their reports to the following datasets:
1. National Student Survey
2. Destination of Leavers from HE
3. Non-continuation following year of entry.

JACS coding framework is in need of revision and is not meeting the requirements of some Data Collectors – most notably the NHS and Research Councils.

The existing landscape is immensely complex consisting, as it does, of hundreds of database tables containing thousands of entities and attributes and controlled by a disparate group of complex organisations.

Data entities fall into two groups. Those widely used by a broad range of stakeholders, or indeed the whole sector, and those that are specific to a particular ‘business process’ involving a small group of stakeholders. The first group of entities is termed ‘The Core’.

UCAS and SLC use the data in the context of transactions with individuals and organisations but HESA gathers data for reports on the outcome of business processes.

**What has been proposed to date?**
Major holders of student data should make more data available on their websites in a reusable format, and at more detailed levels.

Develop a data model, lexicon and thesaurus for the sector.

The different categories of student data collections should be modelled at a high-level ...and would identify common collections that could be shared by multiple organisations.

There are data collections (circa 11 out of 15 categories) that can be removed, reduced or rethought to lighten the burden on HEPs.

Two consolidated data collections could be made instead: (i) specialist PSRB data and (ii) Student Administrative data.

In developing the new [coding] framework the Higher Education Academy’s discipline areas should be considered as a starting point.

The new [coding] framework should consist of three rather than four levels.

The new [coding] framework should be based on a six-digit coding structure.

The new [coding] framework should provide a persistent URI (uniform resource identifier) for each of the entities in the classification.

Seek to ensure that the attribute ‘subject’ is defined to avoid ambiguity and confusion with other terms such as discipline and cost centre.
# Appendix G: Review of Existing HEDIIP Projects

<table>
<thead>
<tr>
<th>PROJECT NAME</th>
<th>PROGRAMME THEME</th>
<th>OVERVIEW</th>
<th>STATUS</th>
<th>ROLE AND RELEVANCE TO NEW LANDSCAPE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Student Data Collection Review</td>
<td>Capability and Excellence</td>
<td>HEDIIP manages an inventory of HE data collections. This inventory lists around 524 data collections and highlights 93 organisations who are collecting student data. The Student Data Collection Review project sought to engage with around half of these organisations in order to determine why so many collections exist and explore opportunities and barriers to standardisation and rationalisation of student data collections.</td>
<td>This project completed in July 2014 and the report was published to the HEDIIP website. Many of the findings of this report have informed the New Landscape Project and, specifically, highlighted the need for the Data Capability project.</td>
<td>The Student Data Collection Review identified the characteristics of data collections, analysed current data management capability within both Data Collectors and providers, identified opportunities that can be taken forward in other HEDIIP projects (ULN, Inventory Review, Data Language) and highlighted barriers to standardisation and rationalisation. This, combined with project recommendations, has informed the New Landscape project and has also informed the HEDIIP Programme Plan in terms of projects that will be necessary components of moving towards and operating within the New Landscape.</td>
</tr>
</tbody>
</table>
| Data Capability            | Capability and Excellence | The need to address data management was highlighted by the HEDIIP Student Data Collection Review, which made the following observations:  
- There is wide variation in capability in terms of approaches to security, data management, and technology in use within both Data Collectors and data providers.  
- There are concerns regarding provider data management and the degree to which a ‘collect once, use many’ approach is in place.  
The HEDIIP Data Capability project aims to improve data management capability by delivering: | The project initiated mid December 2014 and is due to complete May 2015. The project has published an Engagement Plan and has made progress in developing:  
- Maturity assessment  
- TO-BE states  
- Tooling  
- Governance model.  
The project has been engaging with stakeholders and held the first meeting of the Data Capability Technical Panel on 15 January.  
The project is using the HEDIIP website, | Data management capability is recognised by HEDIIP as one of the standard building blocks in the new HE data and information landscape. In order for stakeholders to effectively undertake the necessary internal changes, respond efficiently to external changes to data flows, and realise the benefits the New Landscape has to offer, they will need to understand and if necessary develop their current data management capability. Improving data capability also opens up many more opportunities beyond what operating within the New Landscape offers. |
The Unique Learner Number (ULN) project is developing a Blueprint and implementation plan for the adoption of the ULN in Higher Education. The project is engaging with a number of stakeholders in order to explore the as-is state, the desired to-be and the stages that must then be undertaken in order to address the gap. The project is also focusing on associated benefits and stakeholder concerns in order to develop a robust business case.

The project initiated in July 2014 and is due to complete this stage February 2015. The project has undertaken the first round of engagement and has produced a draft Blueprint. The Blueprint has identified a number of adoption options and the project is now consulting with stakeholders in order to determine the preferred way forward and develop an implementation plan. During stakeholder engagement the project is also revisiting benefits to identify measures and baseline figures where it is possible to do so. The project will be running a combined stakeholder workshop early February to explore the draft implementation plan and processes changed by ULN adoption. The HEDIIP Programme consists of three themes. One of these themes is termed ‘standards and understanding.’ The new Higher Education Data and Information Landscape becomes much more joined up, consistent and efficient once there is greater standardisation across HE data and information.

The adoption of the ULN across HE will provide a consistent identifier across different datasets. The New Landscape will enable and encourage more collaborative working, and a consistent identifier will both facilitate and enhance this capability. A consistent identifier may also have a key place within the Standard Dataset envisioned for the New Landscape.

| Unique Learner Number | Standards and Understanding | The Unique Learner Number (ULN) is a standard person identifier administered by the Learning Records Service (LRS). The ULN is mandated in state funded schools and FE colleges in England and Wales and is being mandated in Northern Ireland in September 2015. Currently only a limited proportion of ULNs flow through into Higher Education and a lack of a consistent identifier across the education landscape is a real barrier to efficient data processing and to coherent student-facing services. The Unique Learner Number project is developing a Blueprint and implementation plan for the adoption of the ULN in Higher Education. The project is engaging with a number of stakeholders in order to explore the as-is state, the desired to-be and the stages that must then be undertaken in order to address the gap. The project is also focusing on associated benefits and stakeholder concerns in order to develop a robust business case. | The project initiated in July 2014 and is due to complete this stage February 2015. The project has undertaken the first round of engagement and has produced a draft Blueprint. The Blueprint has identified a number of adoption options and the project is now consulting with stakeholders in order to determine the preferred way forward and develop an implementation plan. During stakeholder engagement the project is also revisiting benefits to identify measures and baseline figures where it is possible to do so. The project will be running a combined stakeholder workshop early February to explore the draft implementation plan and processes changed by ULN adoption. The HEDIIP Programme consists of three themes. One of these themes is termed ‘standards and understanding.’ The new Higher Education Data and Information Landscape becomes much more joined up, consistent and efficient once there is greater standardisation across HE data and information. The adoption of the ULN across HE will provide a consistent identifier across different datasets. The New Landscape will enable and encourage more collaborative working, and a consistent identifier will both facilitate and enhance this capability. A consistent identifier may also have a key place within the Standard Dataset envisioned for the New Landscape. |
|---|---|---|
| New Subject Coding System | Standards and Understanding | The current method of classifying subjects is based on a scheme known as the Joint Academic Coding System (JACS). JACS required a major review as: Stage 1 of the Subject Coding project completed in November 2013 and the impact assessment and requirements | As with the ULN project (above) the New Subject Coding System project falls under the ‘standards and understanding’ HEDIIP Programme theme. |
In July 2013 the report ‘Classifying subject of study, a roadmap to a new Joint Academic Coding System’ was published. This report considered the varying requirements and uses of subject coding in HE and set out options for the development of a replacement for JACS.

In June 2014 HEDIIP commissioned the Centre for Educational Technology, Interoperability and Standards (Cetis) at the University of Bolton to undertake the Subject Coding project to develop a replacement for the JACS system. The new system had to be designed to meet the needs of a broader group of stakeholders and reflect the diverse and dynamic nature of Higher Education in the twenty-first century.

The project is divided into two stages with the following deliverables:

**Stage 1**
- Impact assessment;
- Requirements definition.

**Stage 2**
- New subject coding structure;
- Populated subject coding structure;
- Recommendations for subject based analysis and text mining;
- Governance model;
- Adoption plan.

Adopting a coding system that meets the needs of all key sector stakeholders, is used consistently by all, and has the capacity to reflect the dynamic nature of HE, aligns with the New Landscape in terms of improving the accuracy and efficiency of data flows. The new coding system also aligns with the vision of rationalising data collections by ensuring that course data that is in demand can be captured within existing data collections.

Addressing data language is a recognised contributor to the increased success of a number of HEDIIP projects. The Student Data Collection Review noted that a common data language is seen as both an outcome and a prerequisite for the standardisation and rationalisation of data collections.
<table>
<thead>
<tr>
<th>PROJECT NAME</th>
<th>PROGRAMME THEME</th>
<th>OVERVIEW</th>
<th>STATUS</th>
<th>ROLE AND RELEVANCE TO NEW LANDSCAPE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Collection Inventory Review</td>
<td>Strategy and Change</td>
<td>The project will review the inventory’s contribution towards delivering programme outcomes. The review will also identify any follow-up actions, including changes to the inventory, to leverage the benefits of the inventory.</td>
<td>Project Brief being developed. The project will be initiated after the inventory has been in place for 12 months.</td>
<td>The enhanced collaboration that the New Landscape will both require and promote will also be eased considerably by an agreed standard data language. Furthermore, a standard data language can result in many positive implications in terms of data quality.</td>
</tr>
</tbody>
</table>

Data Collection Inventory Review

The Student Data Collection Review recommended that:

- A shared understanding of data collection should be promoted through publication of more detailed information.

The Data Collection Inventory already exists but in order to promote data rationalisation within the New Landscape, provided there is a demand for the inventory to continue, it should be both reviewed and housed where it can be permanently maintained and made available to stakeholders.
Appendix H: Summary of HEI Survey

This is a summary of results from the New Landscape project data survey, which 27 HEIs completed in January 2015 to identify the cost of data processing. A separate HESA specific survey was also completed in parallel to the New Landscape survey and, where possible, this summary will link the results of the two surveys and comment on any differences. The HESA survey asked high level questions to 57 institutions and there were 4 institutions who gave responses to both surveys.

The main body of the New Landscape survey focused on Data Management and Data Returns. The results of the survey are summarised below.

Data Management

Which of the following options best describes your institution's processes for data administration?

The majority of current processes are a combination of manual and automated. In line with the HESA specific survey, only a small number of systems are fully automated (12%), or fully manual (8%). In the HESA specific survey, 16% of institutions were fully automated and none were fully manual.

How do you administer student data for the following student types? (Decentralised data is wholly administered by a team which doesn't control the central student records system)

The most common administration model from the respondents was centralised across all learner types, with 67% of the respondents identifying with this administration type. In contrast, 11% of respondents operated a wholly decentralised system for all student types. The other 22% had a varied model with some student types having a centralised system and others having a decentralised system. With the exception of one institution, the split between student types was determined by either undergraduate or postgraduate students, or Home/EU students compared to Non-EU students. These results are supported by the HESA specific survey, whose respondents all claimed to have some level of centralised system for their data administration. Of these, 82.5% had a central student record system (or database) where student record data is held and governed centrally. The remaining 17.5% had a central system that was populated from various other systems such as faculty databases.

Types of student data administration methods
Does your institution use a data warehouse to support your internal business processes?

With the exception of one respondent, institutions only use data warehouses to support business processes if they also use data warehouses for data analytics or business intelligence. Almost half use a data warehouse to support data analytics or business intelligence, and 28% use data warehouses to support both data analytics or business intelligence, and business processes. 52% of institutions do not use a data warehouse to support either process.

<table>
<thead>
<tr>
<th>Support business processes?</th>
<th>Support data analytics or business intelligence?</th>
<th>Yes</th>
<th>No</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>Yes</td>
<td>28%</td>
<td>4%</td>
<td>32%</td>
</tr>
<tr>
<td>No</td>
<td>No</td>
<td>20%</td>
<td>48%</td>
<td>68%</td>
</tr>
<tr>
<td>Total</td>
<td>Total</td>
<td>48%</td>
<td>52%</td>
<td>100%</td>
</tr>
</tbody>
</table>

What influence do you believe you have over the following collectors?

Excluding a single respondent, institutions do not believe they have significant influence over any of the collectors listed. Of the collectors, it appears that PSBR and UKIV are perceived by the institutions as collectors that they have the least influence over. HESA and the Funding Council have the highest number of institutions that believe they have some or significant influence over them. This is followed by UCAS and then SLC.
Key ways in which influence can be improved:

- Improved communication, via working groups, client contacts, workshops and forums.
- Greater transparency around the need for data and more detailed requirements.

What would be your preferred method for exchanging student data with the following types of organisations?

The most commonly preferred method for exchanging student data across all Data Collectors is via Web service API, followed by XML file and then spreadsheet. The exception to this is HESA, where approximately 50% of respondents listed their preferred method as XML. This is reflected in the HESA specific survey, where almost 80% of institutions stated XML as their preferred data exchange method. CSV and spreadsheet were the least preferred methods, apart from in the instance of Funding Council, where spreadsheets were the second most popular method of exchange.
How much extra effort would it be to move from your current method to your preferred method for each of the main Data Collectors listed?

Responses showed that, across all organisations, more institutions felt it would be easy to move to their preferred method than hard to move. There is little distinction between the Data Collectors indicating that the level of effort required was generally dependent on the institution giving the response rather than the Data Collector. The majority of responses were either all little or some effort, or all significant effort. The notable exception to this is HESA, which appears to be regarded as the easiest method to move to. 50% of respondents said there would be little effort in transferring to their preferred method with HESA.

Range of estimated costs for change:

When asked to estimate the cost for the change, most respondents either did not answer the question or stated it would be difficult to estimate. Of those who did respond, the cost estimates varied drastically, with some institutions stating that the cost would be negligible compared to the maximum cost estimate of £2 million.

The key benefits of the change were listed as:

- More time efficient;
- Would allow for standardisation;
- Removed the need for third party software;
- Greater consistency;
- Improved data quality; and
- Improved data security.
Data Returns

How easily can you fulfil the following Data Collector’s data requests?

Results varied across Data Collectors, with HESA, Funding Council and PSBR receiving significantly more ‘hard’ ratings than ‘easy’. Of all organisations, the highest number of respondents found it hard to fulfil the HESA’s requests. Similar to these results, in the HESA specific survey, less than 10% of institutions said it was easy to meet the collector’s requests. SLC and UCAS appear to be the easiest Data Collectors to fulfil data requests for as they received more ‘easy’ responses than ‘hard’.

On the following scale how easy would you say it is for your organisation to undertake its HESA student record submission currently?

1 (very easy) - 6 (very hard)
Key actions collectors could take to reduce the burden of data collections on institutions:

- Consistent definitions of data;
- Clearer guidance;
- Collect only necessary data and explain why the data is relevant;
- Streamline collections;
- Recognise challenges of unique organisations; and
- Synchronisation and coordination or requests between organisations. Do not make last minute or sudden changes to the data requirements – stability of requirements.

Key actions institutions could take to improve the process of compiling and submitting of student data returns to the key collectors?

- Invest in relevant IT systems and specialist staff;
- Increase automation;
- Increase centralisation;
- Ensure data is actively managed;
- Streamline processes;
- Improve data quality; and
- More training;

Key actions that could have the greatest and most positive impact on the data burden experienced by institutions?

- Standardise data;
- Implement a ‘freeze’ on changing requirements; and
- Data should be entered once and used many times – there needs to be collaboration between organisations.

Estimate of the annual financial impact of the removal of this data burden?

Institutions also found this difficult to estimate. Some institutions estimate savings, others list additional costs of £5 million.
Appendix I: TOGAF Framework for Data Principles

Higher Education Data & Information Improvement Programme

New Landscape Project
Data Management Principles

KPMG LLP
Final
This report contains 10 pages
Contents

1 Introduction

2 Enterprise Principles
   2.1 Effective Governance

3 Architecture Principles
   3.1 Adherence to Principles
   3.2 Maximise Benefit to HEDIIP as a Whole
   3.3 Information Management is Everybody’s Business
   3.4 Compliance with Law
   3.5 Data is an Asset
   3.6 Data is Shared
   3.7 Data Trustee
   3.8 Common Vocabulary and Data Definitions
   3.9 Data Security
   3.10 Technology Independence
   3.11 Responsive Change Management
   3.12 Interoperability
   3.13 Protection of Competitive Position
1) Introduction

The principles contained within this document have been devised as part of the Higher Education Data & Information Improvement Programme (HEDIIP) New Landscape project.
The purpose of these principles is to guide the development and evolution of the new information landscape across the Higher Education sector and to describe the agreed values of the stakeholders in respect of the HE data landscape.
Data principles are one type of architecture principles, the others being Application and Technology. They all should be related to wider Enterprise principles. In this case, we should treat the HE data landscape as the Enterprise.
This document contains a number of principles, some of which are related to the wider architecture of the landscape and others that are specifically related to data management.
Each principle will be defined in a consistent way. It will include a name/title, a definition statement and associated rationale and implications.
The principles are based upon collateral from The Open Group Architecture Framework (TOGAF).

2) Enterprise Principles

2.1 Effective Governance

Statement: Data collection requirements in the stakeholders and Higher Education Providers need collective governance that is free from bias to allow for effective implementation and delivery of a New data landscape. This governance should be delivered by representatives from HEPs and Data Collectors from across the sector, and represent sector views, with student views represented by the NUS.
Rationale: A permanent body that is free from bias from HE data providers and consumers will be required to make decisions and be accountable for the implementation and ongoing running of the systems underpinning the New Landscape. There are many stakeholders associated with the forthcoming solution and achieving absolute agreement on all issues will not be feasible. An organisation with the appropriate authority is required to drive the project forward. In order to gain the respect from the various interested parties, the governing organisation should be independent.
Implications: HEDIIP will oversee the design and implementation phases of the project but it will be necessary to create a new, permanent body to monitor and manage the New Landscape once it has been deployed.

3) Architecture Principles

3.1 Adherence to Principles

Statement: These principles of information management apply to all organisations within the HE sector organisations.
Rationale: The only way we can provide a consistent and high level of quality information is if all organisations abide by the principles.
Implications:
- Without this principle, exclusions, duplications and inconsistencies would rapidly undermine the programme.
- Subsequent initiatives will not begin until they are examined for compliance with the principles.
- A conflict with a principle will be resolved by changing the framework of the initiative.

3.2 Maximise Benefit to the HE sector as a Whole

Statement: Information management decisions are made to provide maximum benefit to the HE sector as a whole including students.
Rationale: This principle embodies ‘service above self’. Decisions made from a sector-wide perspective have greater long-term value than decisions made from any particular organisational perspective. Maximum return on investment requires information management decisions to adhere to sector-wide drivers and priorities. No minority group will detract from the benefit of the whole. However, this principle will not preclude any minority group from getting its job done.

Implications:
- Achieving maximum sector-wide benefit will require changes in the way that data is collected, managed and disseminated. Technology alone will not bring about this change.
- Some organisations may have to concede their own preferences for the greater benefit of the HE Sector.
- Applications components should be shared across organisational boundaries.
- Information management initiatives should be conducted in accordance with the overall sector plan for data. Individual organisations should pursue information management initiatives which conform to the Blueprints and priorities established by the New Data Landscape project. The plan will evolve over time.
- As needs arise, priorities must be adjusted. A forum with comprehensive sector-wide representation should make these decisions; this should include opportunities for student involvement.

3.3 Information Management is Everybody’s Business

Statement: All stakeholders within the HE Sector participate in information management decisions needed to accomplish business objectives.

Rationale: Information users are the key stakeholders, or customers, in the application of technology to address a business need. In order to ensure information management is aligned with the business, all organisations within the sector must be involved in all aspects of the information environment. The business experts from across different sector organisations (collectors and providers) and the technical staff responsible for developing and sustaining the information landscape need to come together as a team to jointly define the goals and objectives of IT.

Implications:
- To operate as a team, every stakeholder, or customer, will need to accept responsibility for developing the information landscape.
- Commitment of resources will be required to implement this principle.

3.4 Compliance with Law

Statement: Higher Education information management processes comply with all relevant laws, policies, and regulations, including competition law.

Rationale: Sector policy is to abide by laws, policies, and regulations. This will not preclude business process improvements that lead to changes in policies and regulations.

Implications:
- All Data Collectors must be mindful to comply with laws, regulations, and external policies regarding the collection, retention, and management of data. For example, the Data Protection Act and Freedom of Information Act are likely to be applicable.
- It will be necessary to ensure all stakeholders have access to and understand the rules. Efficiency, need, and common sense are not the only drivers. Changes in the law and changes in regulations may drive changes in data collection processes or applications.
3.5 Data is an Asset

Statement: Data is an asset that has value to HE stakeholders and is managed accordingly.

Rationale: Data is a valuable resource; it has real, measurable value. In simple terms, the purpose of data is to aid decision-making. Accurate, timely data is critical to accurate, timely decisions. Most corporate assets are carefully managed, and data is no exception. Data is the foundation of decision-making, so it must be carefully managed to ensure that we know where it is, can rely upon its accuracy, and can obtain it when and where it is required.

Implications:
- This is one of three closely-related principles regarding data: data is an asset; data is shared; and data is easily accessible. The implication is that there is an education task to ensure that all organisations within the sector understand the relationship between value of data, sharing of data, and accessibility to data.
- Since data is an asset of value to the collectors and providers, data stewards accountable for properly managing the data must be assigned at senior levels.
- Stewards must have the authority and means to manage the data for which they are accountable.
- We must make the cultural transition from 'data ownership' thinking to 'data stewardship' thinking.
- The role of data steward is critical because obsolete, incorrect, or inconsistent data could be passed to data consumers and adversely affect decisions they make.
- Part of the role of data steward, who manages the data, is to ensure data quality. Procedures must be developed and used to prevent and correct errors in the information and to improve those processes that produce flawed information. Data quality will need to be measured and steps taken to improve data quality: it is probable that policy and procedures will need to be developed for this as well.
- A forum with comprehensive sector-wide representation should decide on process changes suggested by data stewards.

3.6 Data is Shared

Statement: Users have access to the data necessary to perform their duties or answer their query; therefore, data is shared appropriately across organisational boundaries, where permissible.

Rationale: Timely access to accurate data is essential to improving the quality and efficiency of sector decision-making. It is less costly to maintain timely, accurate data in a single application, and then share it, than it is to maintain duplicated data in multiple applications. Within Data Collectors, there is a wealth of data, but it is stored in hundreds of discrete databases. The speed of data collection, creation, transfer, and assimilation is driven by the ability of the organisation to efficiently share these islands of data across the organisation.

Shared data will result in improved decisions since there will be fewer (ultimately one virtual) sources of more accurate and timely managed data for all decision-making. Electronically shared data will result in increased efficiency when existing data entities can be used, without re-keying, to create new entities.

Implications:
- This is one of three closely-related principles regarding data: data is an asset; data is shared; and data is easily accessible. The implication is that there is an education task to ensure that all organisations within the sector understand the relationship between value of data, sharing of data, and accessibility to data.
- To enable data sharing, providers and collectors must develop and abide by a common set of policies, procedures, and standards governing data management and access for both the short and the long term.
- In the short term, organisations must invest in software capable of migrating legacy system data into an environment capable of sharing data in a standardised way, based on modern, open standards.
- The sector will also need to develop, and agree upon, standard data models, data elements, and other metadata that defines these sharing standards, and develop a repository system for storing this metadata to make it accessible.
- In the long term, as legacy systems are replaced, organisations must agree and adopt common data access policies and guidelines for new application developers to ensure that data in new applications...
remains available to be shared where appropriate, and that data being shared can continue to be used by the new applications.

- For both the short term and the long term the sector must adopt common methods and tools for creating, maintaining, and accessing the data shared across the landscape.
- Data sharing will require a significant cultural change.
- This principle of data sharing will continually ‘bump up against’ the principle of data security. Under no circumstances will the data sharing principle cause confidential data to be compromised.
- Data made available for sharing will have to be relied upon by all users to execute their respective tasks. This will ensure that only the most accurate and timely data is relied upon for decision-making. Shared data will become the sector-wide ‘virtual single source’ of data.

### 3.7 Data Trustee

**Statement:** Each data element has an owner accountable for data quality.

**Rationale:** One of the benefits of an architected environment is the ability to share data across the HE landscape. As the degree of data sharing grows and organisations rely upon common information, it becomes essential that only the data trustee makes decisions about the content of data. Since data can lose its integrity when it is entered multiple times, the data trustee will have sole responsibility for data entry which eliminates redundant human effort and data storage resources.

**Note:** A trustee is different from a steward: a trustee is responsible for accuracy and currency of the data, while responsibilities of a steward may be broader and include data standardisation and definition tasks.

**Implications:**
- Real trusteeship dissolves the data ‘ownership’ issues and allows the data to be available to meet all users’ needs. This implies that a cultural change from data ‘ownership’ to data ‘trusteeship’ may be required.
- The data trustee will be responsible for meeting quality requirements levied upon the data for which the trustee is accountable.
- It is essential that the trustee has the ability to provide user confidence in the data based upon attributes such as ‘data source’.
- It is essential to identify the true source of the data in order that the data authority can be assigned this trustee responsibility. This does not mean that classified sources will be revealed nor does it mean the source will be the trustee.
- Information should be captured electronically once and immediately validated as close to the source as possible. Quality control measures must be implemented to ensure the integrity of the data.
- As a result of sharing data across the HE landscape, the trustee is accountable and responsible for the accuracy and currency of their designated data element(s) and, subsequently, must then recognise the importance of this trusteeship responsibility.

### 3.8 Common Vocabulary and Data Definitions

**Statement:** Data is defined consistently throughout the HE sector, and the definitions are understandable and available to all users.

**Rationale:** The data that will be used in the development of a central data warehouse must have a common definition throughout the landscape to enable sharing of data. A common vocabulary will facilitate communications and enable dialogue to be effective. In addition, it is required to interface systems and exchange data.

**Implications:**
- Significant additional energy and resources must be committed to this task. It is key to the success of efforts to improve the information landscape. This is separate from but related to the issue of data
element definition, which is addressed by a broad community (this is more like a common vocabulary and definition).

- HEDIIP must define a workstream to establish the initial common vocabulary for the sector. The definitions will be used consistently across the HE landscape.
- Whenever a new data definition is required, the definition effort will be co-ordinated and reconciled with the corporate ‘glossary’ of data descriptions. An independent, central data administrator will provide this co-ordination.
- Ambiguities resulting from multiple parochial definitions of data must give way to accepted sector-wide definitions and understanding.
- Multiple data standardisation initiatives need to be co-ordinated.
- Functional data administration responsibilities must be assigned.

3.9 Data Security

Statement: Data is protected from unauthorised use and disclosure. In addition to the traditional aspects of national security classification, this includes, but is not limited to, protection of sensitive and proprietary information.

Rationale: Open sharing of information and the release of information via relevant legislation must be balanced against the need to restrict the availability of classified, proprietary, and sensitive information. Existing laws and regulations require the safeguarding of national security and the privacy of data, while permitting free and open access.

Implications:

- Aggregation of data, both classified and not, will create a large target, requiring review and declassification procedures to maintain appropriate control. Data owners and/or functional users must determine whether the aggregation results in an increased classification level. The sector will need appropriate policy and procedures to handle this review and declassification. Access to information based on a need-to-know policy will force regular reviews of the body of information.
- In order to adequately provide access to open information while maintaining secure information, security needs must be identified and developed at the data level, not the application level.
- Data security safeguards can be put in place to restrict access to ‘view only’, or ‘never see’. Sensitivity labelling for access to classified, sensitive or proprietary information must be determined.
- Security must be designed into data elements from the beginning; it cannot be added later. Systems, data and technologies must be protected from unauthorised access and manipulation. Information must be safeguarded against inadvertent or unauthorised alteration, sabotage, disaster, or disclosure.

3.10 Technology Independence

Statement: Applications are independent of specific technology choices and therefore can operate on a variety of technology platforms.

Rationale: Independence of applications from the underlying technology allows applications to be developed, upgraded, and operated in the most cost-effective and timely way. Otherwise technology, which is subject to continual obsolescence and vendor dependence, becomes the driver rather than the user requirements themselves. Realising that every decision made with respect to IT makes us dependent on that technology, the intent of this principle is to ensure that application software is not dependent on specific hardware and operating systems software.

Implications:

- This principle will require standards which support portability.
■ For Commercial Off-The-Shelf (COTS) applications, there may be limited choices, as some of these applications are technology and platform-dependent.

■ Subsystem interfaces will need to be developed to enable legacy applications to interoperate with applications and operating environments developed under the HEDIIP architecture.

■ Middleware should be used to decouple applications from specific software solutions.

3.11 Responsive Change Management

Statement: Changes to the HE information landscape are implemented in a timely manner.
Rationale: If people are to be expected to work within the new HE information landscape, that environment must be responsive to their needs.
Implications:
■ HEDIIP has to develop processes for managing and implementing change that do not create delays.

■ A user who feels a need for change will need to connect with a ‘business expert’ to facilitate explanation and implementation of that need.

■ If changes are made to the environment, the architecture must be updated.

■ Adopting this principle might require additional resources.

■ This could conflict with other principles (such as Maximise Benefit to the sector as a Whole).

3.12 Interoperability

Statement: Software and hardware should conform to defined standards that promote interoperability for data, applications, and technology.
Rationale: Standards help ensure consistency, thus improving the ability to manage systems and improve user satisfaction, and protect existing IT investments, thus maximising return on investment and reducing costs. Standards for interoperability additionally help to ensure support from multiple vendors for their products.
Implications:
■ Interoperability standards and industry standards will be followed unless there is a compelling business reason to implement a non-standard solution.

■ A process for setting standards, reviewing and revising them periodically, and granting exceptions must be established.

■ In the short term, developing a sector-wide data model and associated data dictionary should be a priority for HEDIIP.

3.13 Protection of Competitive Position

Statement: Data that is viewed as compromising the competitive position of HEPs will still be collected in-year, but will not be available publicly, nor to other collectors (save for that data required for them to fulfil their requirements) until it is no longer viewed as commercially sensitive.
Rationale: Data Collectors and Higher Education Providers will need to be confident that adherence to the principles does not conflict directly with their commercial activities.
Implications:
■ Data sharing protocols will need to be drafted to ensure that there is no conflict with the commercial position of the stakeholders.

■ The governance will need to continually review the timing and range of the availability of the data against the requirements of stakeholders to seek greater competitive positions.
# Appendix J: Glossary of terms

<table>
<thead>
<tr>
<th>ABBREVIATION</th>
<th>FULL DESCRIPTION</th>
</tr>
</thead>
<tbody>
<tr>
<td>AOC</td>
<td>Association of Colleges</td>
</tr>
<tr>
<td>AP</td>
<td>Alternative Providers</td>
</tr>
<tr>
<td>API</td>
<td>Application Programming Interface</td>
</tr>
<tr>
<td>ARMA</td>
<td>Association of Research Managers and Administrators</td>
</tr>
<tr>
<td>AUA</td>
<td>Association of University Administrators</td>
</tr>
<tr>
<td>BAU</td>
<td>Business As Usual</td>
</tr>
<tr>
<td>BCS</td>
<td>British Computer Society</td>
</tr>
<tr>
<td>BIS</td>
<td>Department for Business Innovation and Skills</td>
</tr>
<tr>
<td>BSI</td>
<td>British Standards Institution</td>
</tr>
<tr>
<td>BUFDG</td>
<td>British Universities Finance Directors Group</td>
</tr>
<tr>
<td>CACHED project</td>
<td>Change in Approach to the Collection of HE Data Programme</td>
</tr>
<tr>
<td>CASRAI</td>
<td>Consortia Advancing Standards in Research Administration Information</td>
</tr>
<tr>
<td>CCG</td>
<td>Clinical Commissioning Groups</td>
</tr>
<tr>
<td>Cetis</td>
<td>Centre of Educational Technology, Interoperability and Standards</td>
</tr>
<tr>
<td>COTS</td>
<td>Commercial Off-The-Shelf</td>
</tr>
<tr>
<td>CSU</td>
<td>Clinical Support Unit</td>
</tr>
<tr>
<td>DELNI</td>
<td>Department for Employment and Learning in Northern Ireland</td>
</tr>
<tr>
<td>DIE</td>
<td>Department for Education</td>
</tr>
<tr>
<td>DH</td>
<td>Department of Health</td>
</tr>
<tr>
<td>DSG</td>
<td>Data and Statistics Group</td>
</tr>
<tr>
<td>DWP</td>
<td>Department for Work and Pensions</td>
</tr>
<tr>
<td>ENQA</td>
<td>European Association for Quality Assurance</td>
</tr>
<tr>
<td>ERD</td>
<td>Entity Relationship Diagram</td>
</tr>
<tr>
<td>ESCS</td>
<td>Education, Skills and Children’s Services</td>
</tr>
<tr>
<td>ETL</td>
<td>Extract Transform and Load</td>
</tr>
<tr>
<td>FE</td>
<td>Further Education</td>
</tr>
<tr>
<td>FEC</td>
<td>Further Education Colleges</td>
</tr>
<tr>
<td>FOI</td>
<td>Freedom Of Information</td>
</tr>
<tr>
<td>FSA</td>
<td>Food Standards Agency</td>
</tr>
<tr>
<td>GFE</td>
<td>General Further Education</td>
</tr>
<tr>
<td>GMC</td>
<td>General Medical Council</td>
</tr>
<tr>
<td>GSS</td>
<td>Government Statistical Service</td>
</tr>
<tr>
<td>HEBRG</td>
<td>Higher Education Better Regulation Group</td>
</tr>
<tr>
<td>HEDIIP</td>
<td>Higher Education Data and Information Improvement Programme</td>
</tr>
<tr>
<td>HEE</td>
<td>Health Education England</td>
</tr>
<tr>
<td>HEFCE</td>
<td>Higher Education Funding Council for England</td>
</tr>
<tr>
<td>HEFCW</td>
<td>Higher Education Funding Council for Wales</td>
</tr>
<tr>
<td>HEI</td>
<td>Higher Education Institution</td>
</tr>
<tr>
<td>HEIDI</td>
<td>Higher Education Information Database for Institutions.</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
</tr>
<tr>
<td>HEIDI 2</td>
<td>Name for the re-development of HEIDI</td>
</tr>
<tr>
<td>HEP</td>
<td>Higher Education Providers</td>
</tr>
<tr>
<td>HESA</td>
<td>Higher Education Statistics Agency</td>
</tr>
<tr>
<td>HSCIC</td>
<td>Health and Social Care Information Centre</td>
</tr>
<tr>
<td>IA</td>
<td>Information Authority</td>
</tr>
<tr>
<td>ILR</td>
<td>Individualised Learner Record</td>
</tr>
<tr>
<td>IQER</td>
<td>Integrated Quality Enhancement Review</td>
</tr>
<tr>
<td>IRPG</td>
<td>Interim Regulatory Partnership Group</td>
</tr>
<tr>
<td>ISB</td>
<td>Information Standards Board</td>
</tr>
<tr>
<td>ISO</td>
<td>International Organisation for Standardisation</td>
</tr>
<tr>
<td>ITSB</td>
<td>Information and Technology Strategy Board</td>
</tr>
<tr>
<td>JACS system</td>
<td>Joint Academic Coding System</td>
</tr>
<tr>
<td>Jisc</td>
<td>Previously known as Joint Information Systems Committee</td>
</tr>
<tr>
<td>KIS</td>
<td>Key Information Set</td>
</tr>
<tr>
<td>LEAMS</td>
<td>Local Environment Audit &amp; Management Systems</td>
</tr>
<tr>
<td>LETB</td>
<td>Local Education and Training Board</td>
</tr>
<tr>
<td>LRS</td>
<td>Learning Records Service</td>
</tr>
<tr>
<td>MI</td>
<td>Management Information</td>
</tr>
<tr>
<td>MIAP</td>
<td>Managing Information Across Partners</td>
</tr>
<tr>
<td>MOG</td>
<td>Machinery of Government</td>
</tr>
<tr>
<td>NCTL</td>
<td>National College for Teaching and Leadership</td>
</tr>
<tr>
<td>NHS</td>
<td>National Health Service</td>
</tr>
<tr>
<td>NUS</td>
<td>National Union of Students</td>
</tr>
<tr>
<td>OFFA</td>
<td>Office of Fair Access</td>
</tr>
<tr>
<td>OFQUAL</td>
<td>Office of Qualifications and Examinations Regulation</td>
</tr>
<tr>
<td>OFSTED</td>
<td>Office for Standards in Education, Children’s Services and Skills</td>
</tr>
<tr>
<td>ONS</td>
<td>Office for National Statistics</td>
</tr>
<tr>
<td>PA</td>
<td>PA Consulting</td>
</tr>
<tr>
<td>PLR</td>
<td>Personnel Learning Records</td>
</tr>
<tr>
<td>PMO</td>
<td>Project Management Office</td>
</tr>
<tr>
<td>PSBR</td>
<td>Public Sector Borrowing Requirement</td>
</tr>
<tr>
<td>PSRB</td>
<td>Professional, Statutory and Regulation Body</td>
</tr>
<tr>
<td>QAA</td>
<td>Quality Assurance Agency</td>
</tr>
<tr>
<td>QIA</td>
<td>Quality Improvement Agency</td>
</tr>
<tr>
<td>QTS</td>
<td>Qualified Teacher Status</td>
</tr>
<tr>
<td>RCUK</td>
<td>Research Councils United Kingdom</td>
</tr>
<tr>
<td>RPG</td>
<td>Regulatory Partnership Group</td>
</tr>
<tr>
<td>SDCR</td>
<td>Student Data Collection Review</td>
</tr>
<tr>
<td>SFA</td>
<td>Skills Funding Agency</td>
</tr>
<tr>
<td>SFC</td>
<td>Scottish Funding Council</td>
</tr>
<tr>
<td>SHA</td>
<td>Strategic Health Authority</td>
</tr>
<tr>
<td>SLC</td>
<td>Student Loans Company</td>
</tr>
<tr>
<td>SMS</td>
<td>Sponsor Management System</td>
</tr>
<tr>
<td>SQL</td>
<td>Structured Query Language</td>
</tr>
<tr>
<td>TDA</td>
<td>Training and Development Agency</td>
</tr>
<tr>
<td>Acronym</td>
<td>Description</td>
</tr>
<tr>
<td>---------</td>
<td>-------------</td>
</tr>
<tr>
<td>TOGAF</td>
<td>The Open Group Architecture Framework</td>
</tr>
<tr>
<td>ToR</td>
<td>Terms of Reference</td>
</tr>
<tr>
<td>UCAS</td>
<td>Universities and Colleges Admission Service</td>
</tr>
<tr>
<td>UCISA</td>
<td>Universities and Colleges Information Systems Association</td>
</tr>
<tr>
<td>UKBA</td>
<td>Also known as UKVI. UK Visas and Immigration</td>
</tr>
<tr>
<td>UKCES</td>
<td>UK Commission for Employment and Skills</td>
</tr>
<tr>
<td>UKVI</td>
<td>UK Visas and Immigration</td>
</tr>
<tr>
<td>ULN</td>
<td>Unique Learner Number</td>
</tr>
<tr>
<td>UPIN</td>
<td>Unique Provider Identification Number</td>
</tr>
<tr>
<td>URI</td>
<td>Uniform Resource Identifier</td>
</tr>
<tr>
<td>UUK</td>
<td>Universities United Kingdom</td>
</tr>
<tr>
<td>XML</td>
<td>eXtensible Markup Language</td>
</tr>
<tr>
<td>XSD</td>
<td>XML Schema Definition</td>
</tr>
</tbody>
</table>
Contact us

Andrew Bush
Director
E: Andrew.Bush@kpmg.co.uk

Michael Rowley
Partner
E: Michael.Rowley@kpmg.co.uk

© 2015 KPMG LLP, a UK limited liability partnership and a member firm of the KPMG network of independent member firms affiliated with KPMG International Cooperative ("KPMG International"), a Swiss entity. All rights reserved. Printed in the United Kingdom.

The KPMG name, logo and "cutting through complexity" are registered trademarks or trademarks of KPMG International.

Produced by Create Graphics | CRT033165