



HEDIIP NSCS Project - Governance Model



About the New Subject Coding Scheme Project

The New Subject Coding Scheme Project was commissioned by HEDIIP under the Standards and Understanding theme. The project aimed to develop a replacement for the Joint Academic Coding Scheme that met the needs of a broad group of stakeholders and reflected the diverse and dynamic nature of Higher Education in the twenty-first century. The New Subject Coding Scheme project was undertaken by the Centre for Educational Technology, Interoperability and Standards (Cetis) with partners APS Ltd and Aspire Ltd. The project undertook extensive stakeholder engagement to identify the requirements for the new coding system and developed a coding structure that aims to meet these requirements. The new coding scheme is referred to as HECoS – the Higher Education Classification of Subjects.

The project ran from May 2014 to October 2015.

The project is overseen by a Project Board made up of:

- Andy Youell, Director, HEDIIP
- Dan Cook, Head of Collections Development, HESA
- Christine Couper, Director of Strategic Planning, Greenwich University
- Hannah Falvey, Head of Statistics, HEFCW
- Lesley Donnithorne, HR Manager (Systems, Information and Grading), UWE Bristol
- Mike Spink, Data Architect, UCAS
- Paul Baron, Programme Manager, HEDIIP
- Jenni Cockram, Programme Officer, HEDIIP

Principal Authors/Editors: Lorna M. Campbell and Gill Ferrell.

Contributors: Phil Barker, Adam Cooper, Alan Paull, Wilbert Kraan.

About HEDIIP

The Higher Education Data & 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 HE.

HEDIIP is funded by the Higher Education Funding Council for England (HEFCE), the Higher Education Funding Council for Wales (HEFCW), the Scottish Funding Council (SFC) and the Department for Employment and Learning (DEL) Northern Ireland.

HEDIIP is hosted by the Higher Education Statistics Agency Ltd (HESA) which is a company limited by guarantee, registered in England at 95 Promenade Cheltenham GL50 1HZ.

Contact HEDIIP

Web: www.hediip.ac.uk

Email: info@hediip.ac.uk

Twitter: @HEDIIP

About This Report

This report is part of the work of the Higher Education Data and Information Improvement Programme (HEDIIP) New Subject Coding Scheme Project (NSCS Project). It comprises deliverable PD06, as defined in the work specification; the governance model which defines the arrangements for the on-going maintenance and development of the new subject coding scheme HECoS, (Higher Education Classification of Subjects). This report builds on the consultation responses to the draft governance plan and it has been informed by deliverables PD04 *The Higher Education Coding of Subjects (HECOS) vocabulary*, PD05 *Recommendations for Subject Based Analysis & Text Mining*, PD07 *HEDIIP NSCS Project Adoption Plan*, and *The Blueprint for a New HE Data Landscape* by KPMG.

Authorship and Status

Owner	Lorna Campbell
Principal Author/Editor	Lorna Campbell and Gill Ferrell
Contributors	Phil Barker, Adam Cooper, Wilbert Kraan, Alan Paull
This Version	Draft final for Advisory Panel review.
File Identifier	HEDIIP_NSCS_PD06_GovernanceModel_2015-11-24

Document History

Date	Person	Notes
2015-02-03	Wilbert Kraan	Draft Governance Model 0.2 sent to PMO for comment
2015-07-22	Lorna Campbell	Draft Governance Model 0.3 sent to PMO for comment
2015-08-27	Lorna Campbell	Updated following comments from PMO and input from HESA
2015-08-28	Lorna Campbell	Working draft Governance Model 0.4 sent to PMO for comment
2015-09-21	Lorna Campbell	Working draft Governance Model 0.5 harmonised with PD04 version 5.
2015-09-22	Alan Paull	Edit and proofread.
2015-09-25	Lorna Campbell	Edits accepted.
2015-10-04	Lorna Campbell	Comments from PMO accepted.
2015-10-05	Adam Cooper	Financial information from HESA pending. Minor formatting and proof amendments.
2015-10-15	Adam Cooper	Amended following comments during 2015-10-12 Project Board
2015-10-30	Adam Cooper	Amended following comments from the Advisory Panel
2015-11-24	HEDIIP PMO	Updating cross references

Contents

1. Executive Summary	5
1.1 Recommendations.....	7
2. Overview	10
3. Methodology	10
3.1 Phase 1 Methodology.....	10
3.2 Phase 2 Methodology.....	11
4. HECoS Governing Body	11
4.1 HECoS Governing Body Requirements	11
4.2 HECoS Governing Body Recommendations.....	12
5. Managing Changes to HECoS	13
5.1 Change Management Requirements.....	13
5.2 SUBSTANTIVE and ADMINISTRATIVE Changes	14
5.3 Change Management Recommendations	15
5.4 HECoS Candidate Terms Registry	16
6. Governance of Subject Based Analysis	16
6.1 Governance of Subject Group Definitions Recommendations.....	17
7. Specialist Vocabularies	17
7.1 Governing Specialist Vocabularies Options.....	17
7.2 Governing Specialist Vocabularies Recommendations	18
8. Licensing	18
8.1 HECoS Licence Requirements	19
8.2 HECoS Licence Recommendations	19
9. Web Service Management	20
9.1 HECoS Web Service Requirements	20
9.2 HECoS Web Service Recommendations	20
9.3 Proof of Concept Web Service.....	21
10. Monitoring Consistency and Accuracy of Coding	21
10.1 Monitoring Requirements	21
10.2 Monitoring Recommendations.....	21
11. Set Up and Operating Costs	22
11.1 Set-up Costs.....	22
11.2 Maintenance and Governance Costs.....	23
12. References	24
Appendix 1 - Phase 1 Governance Requirements	25
Appendix 2 - Phase 2 Consultation	26
Appendix 3 - Stakeholders Consulted	27

1. Executive Summary

During Phase 1 of the HEDIIP New Subject Coding Scheme Project, the development of a robust, impartial, independent and transparent governance model supported by an appropriate and authoritative selection of stakeholders was identified as a critical success factor for ensuring buy-in from data consumers and suppliers across the sector. The new subject coding scheme is provisionally referred to as HECoS (Higher Education Classification of Subjects). This report focuses on requirements and recommendations for the governance of HECoS.

In order to gather comprehensive requirements for the governance of HECoS, the project undertook an extensive public consultation and gathered input from a wide range of HEPs and public, regulatory and statutory bodies. Based on the requirements gathered, twenty five recommendations are outlined covering the HECoS governing body, change management procedures, subject group definitions, specialist vocabularies, licensing, web service management, monitoring consistency and accuracy, and set up and operating costs.

There is widespread agreement among stakeholders that HECoS should be sector owned and organised, and governed by an existing sector organisation that is established, trusted, respected and neutral. As HECoS will directly support the creation of the Standard Dataset, it is anticipated that both HECoS and the new HE Standard Dataset will be governed by the HE Data Governance Body as outlined in *The Blueprint for a New HE Data Landscape*. There is strong support from the majority of stakeholders that governance of HECoS should be hosted with HESA. Within the structures of the HE Data Governance Body, the Management Board will provide strategic direction for both HECoS and the HE Data and Information Landscape and the Management Office will oversee the day-to-day management of the coding scheme, collate candidate terms, monitor usage and make recommendations to the Board. A HECoS Consultative Group should be established to provide opportunities for wider stakeholder engagement.

Managing the cycle and roll out of new iterations of HECoS is one of the critical success factors for both the governance and adoption of the new subject coding scheme. The type of changes that are likely to be required to keep HECoS current and fit for purpose include both SUBSTANTIVE and ADMINISTRATIVE changes. SUBSTANTIVE changes require the authorisation of the Management Board and include proposal, acceptance and implementation of new candidate terms, deprecation of obsolete terms and changes to the definition of terms. ADMINISTRATIVE change will be undertaken by the Management Office and include clarification of definitions and scope notes, and changes to the navigation structure. Section 5.2 indicates the kind of changes which would fall into each category: SUBSTANTIVE vs ADMINISTRATIVE.

In order to provide stability of HECoS during its adoption, no SUBSTANTIVE changes should be made to the coding scheme during the first 2 years. Only ADMINISTRATIVE changes should be made during this period. The Management Office should undertake an annual review of the coding scheme to identify both SUBSTANTIVE and ADMINISTRATIVE changes that may be necessary. SUBSTANTIVE changes to term definitions, proposals for the adoption of candidate terms and the deprecation of obsolete terms, should be submitted to the HE Data and Information Landscape Management Board for consideration.

In order to allow HECoS to evolve, the NSCS Project recommends setting up a Candidate Terms Registry to enable users to submit new candidate terms for proposed inclusion in the coding scheme. This service must be kept distinct and separate from the main HECoS web service in order to preserve the integrity of the authoritative version of HECoS. Before submitting new candidate terms to the Management Board for consideration for inclusion in HECoS, the Management Office must undertake due diligence to ensure that the term is distinguishable from existing HECoS terms and evaluate the impact of its adoption.

Subject groups are widely used in reporting and analysis and were integral to the definition of JACS3. HECoS differs by separating the terms available for classifying subject of study from definitions of how data should be aggregated over related subjects. Stakeholder consultation showed that subject based analysis is an essential component of the management of Higher Education at all levels from HEPs to government. *Recommendations for Subject Based Analysis & Text Mining* (Cooper, 2015) makes the case for integrated governance of HECoS terms and the definition of subject groups which are built upon it. Governance of the Standard Cross-sector Statistical Aggregation Rules should be fully integrated with the change management process for HECoS terms, with any change being considered as SUBSTANTIVE, and therefore requiring the approval of the Management Board.

The linked data aspect of HECoS makes it feasible to define relations with other vocabularies precisely and easily from a technical point of view. However, considerable care should be taken to ensure that these relationships are appropriate and meaningful. As HECoS is designed specifically to classify subjects of study, it is not recommended that direct relationships are created with vocabularies that classify other entities such as cost centres or REF returns. However within the sphere of subjects of study, there are specific subject areas where stakeholders have expressed the need to relate HECoS to much more fine grained specialist vocabularies. In order to meet these requirements without jeopardising the integrity of HECoS, relevant sector organisations will have the option of registering their vocabularies with the Management Office. These vocabularies must meet the appropriate HECoS technical requirements and must not duplicate existing HECoS codes.

There is widespread agreement among stakeholders that HECoS must be released under an appropriate open licence in order to facilitate its adoption and use by the community. To ensure clarity and usability and in order to prevent the proliferation of customised licences, it is recommended that HECoS is released under a CC BY licence.

In order to facilitate interoperability and discoverability, HECoS should be made available as Five Star Open Data and described using the metadata format most commonly used by existing open data registries and repositories. The web service used to host HECoS must support a broad range of open formats and must provide an API that can serve different versions of the whole vocabulary set as well as versions of individual terms.

While monitoring the consistency of the Standard Data set is within the remit of the Management Board and Management Office as outlined in Section 6 of *The Blueprint for a New HE Data Landscape*, this document is concerned specifically with requirements for monitoring the quality, consistency and accuracy of the HECoS coding scheme and not with evaluating and monitoring the quality of data gathered through implementations based on HECoS. The Management Office will establish a transparent and comprehensive process to monitor the consistency and accuracy with which HECoS codes are applied during the first 2 years of adoption. After the initial 2 year adoption period, the Management Office will establish a lightweight administrative process to monitor the consistency of HECoS coding going forwards. The Management Office will ensure that HECoS structures, support and guidance are sufficient to enable a high level of consistency and accuracy of coding across the sector.

In line with the recommendations of *The Blueprint for a New HE Data Landscape* there is strong support from stakeholders to situate the HECoS governing body within the Higher Education Statistics Agency (HESA). To ensure that the governance body is viable, it is necessary to consider the indicative cost of establishing this authority, and implementing HECoS within HESA. Once HECoS has been set up and implemented across all relevant HESA systems, maintenance of the new coding scheme is likely to be comparable to the current cost of maintaining JACS. The technical systems that implement HECoS will also require maintenance in order to ensure that data can be consumed and output efficiently in an appropriate range of formats. There will also be costs associated with running the vocabulary management software chosen to host HECoS, though these may be mitigated by the adoption of an open source software solution. Financing HECoS represents a non-trivial cost that needs to be appropriately resourced.

1.1 Recommendations

1.1.1 HECoS Governing Body

Recommendation 1

HECoS will be governed by the HE Data Governance Body as outlined in *The Blueprint for a New HE Data Landscape*. 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.

Recommendation 2

Within the structures of the HE Data Governance Body, the HE Data and Information Landscape Management Board will provide strategic direction for HECoS within the context of the HE Data and Information Landscape. The Board will be composed equally of data collectors and HEPs, and chaired by a representative from a HEP or related complementary sector. Rules for tenure are outlined in *The Blueprint for a New HE Data Landscape* (KPMG, pp.51-52).

Recommendation 3

The HE Data and Information Landscape Management Office will oversee the day-to-day management of the coding scheme, collate candidate terms, monitor usage and make recommendations to the Board. Within the HE Data and Information Landscape Management Office, it is expected that the Technical Group will have the requisite technical skills to oversee the management of HECoS.

Recommendation 4

The HE Data Governance Body should publish transparent rules, regulations, operating procedures and terms of reference, and should provide clearly established communication channels to enable stakeholders to input their requirements relating to HECoS, including aggregation scheme for subject based analysis. A HECoS Consultative Group should be established under the aegis of the HE Data Consultative Groups in order to provide opportunities for wider stakeholder engagement.

1.1.2 Change Management

Recommendation 5

In order to facilitate the adoption of HECoS a stability focused approach will be taken during the first 2 years, and it is recommended that no SUBSTANTIVE changes should be made to the coding scheme during this period. However it is recognised that it may be necessary to address exceptional or unforeseen issues that arise during the adoption phase. Any issues that arise should be referred to the Board who should consider whether action should be taken and the implications for the stability of the coding scheme. At the end of this period, the Management Office should review the evidence gathered and assess the case for SUBSTANTIVE changes, for proposal to the Management Board. During the 2 year introduction period, ADMINISTRATIVE changes may be made to the coding scheme as required. It is recommended that the Management Office monitors the user experience and data quality on a frequent basis during this period, in order to quickly identify and correct confusion or misconception.

Recommendation 6

In order to strike a balance between flexibility and stability after the first 2 years, it is recommended that the HE Data and Information Landscape Management Office should undertake an annual review of the coding scheme. The annual review should consider changes to definitions, scope notes and navigation structures, the impact of adopting new candidate terms and deprecating obsolete terms, and the overall level of change to the coding scheme. In the course of the annual review, the Management Office should identify both SUBSTANTIVE and ADMINISTRATIVE changes. ADMINISTRATIVE changes to scope notes and navigation structures should be undertaken by the Management Office on an annual basis. SUBSTANTIVE changes to term definitions, proposals for the adoption of candidate terms and the deprecation of obsolete terms, should be submitted to the HE Data and Information Landscape Management Board for consideration. Before authorising any SUBSTANTIVE changes, the Management Board should consider evidence of the necessity for the proposed changes, the impact of the

proposed changes on data providers and data collectors, and agree an appropriate schedule for the changes to be implemented.

Recommendation 7

A detailed summary of all significant ADMINISTRATIVE changes should be communicated to the sector by the HE Data and Information Landscape Management Office on an annual basis, and a persistent record of all changes maintained.

Recommendation 8

A detailed summary of all SUBSTANTIVE changes, together with transparent explanations outlining the rationale for these changes, should be communicated to the sector by the HE Data and Information Landscape Management Office on behalf of the HE Data and Information Landscape Management Board on a regular basis, in line with the agreed change management schedule.

Recommendation 9

In order to allow HECoS to evolve, the NSCS Project recommends setting up a Candidate Terms Registry to enable users to submit new candidate terms for proposed inclusion in the coding scheme. The HECoS Candidate Terms Registry should be run and managed by the HE Data and Information Landscape Management Office. This service should be kept distinct and separate from the main HECoS web service in order to preserve the integrity of the authoritative version of HECoS.

1.1.3 Governance of Subject Group Definitions

Recommendation 10

Governance of the Standard Cross-sector Statistical Aggregation Rules should be fully integrated with the change management process for HECoS terms, with any change affecting statistical results being considered as SUBSTANTIVE, and therefore requiring the approval of the HE Data and Information Landscape Management Board. The Standard Aggregation Rules should be available under the same licence as HECoS and disseminated and curated as part of the same HECoS web service.

Recommendation 11

HECoS information and support services should provide access to subject group definitions which are not under the control of the HE Data Governance body but which are recognised by it. These include: ATAS subject lists, cross-reference to QAA benchmark statements, Strategically Important and Vulnerable Subject groups (HEFCE SIVS), HEFCW Academic Subject Category definitions, SFC Subject Area definitions, and others as may be deemed important. A proposed way forward for subject based analysis is outlined in *Recommendations for Subject Based Analysis & Text Mining* (Cooper, 2015, pp. 8-11) Section 3.3 addresses how these requirements can be met in different ways according to the character of the subject group definitions.

1.1.4 Governing Specialist Vocabularies

Recommendation 12

Relevant sector organisations may register specialist vocabularies under their control with the HE Data and Information Landscape Management Office. These vocabularies must meet the appropriate HECoS technical requirements and must not duplicate existing HECoS codes. Due diligence will be undertaken by the Management Office to ensure all specialist vocabularies meet these requirements. Authorisation to approve the relationship of specialist vocabularies with HECoS lies with the HE Data and Information Landscape Management Board. Once specialist vocabularies have been approved by the Management Board they can be related to HECoS by identification with one or more broader HECoS terms. It is envisaged that only a small number of HECoS related specialist vocabularies will be authorised by the HE Data and Information Landscape Management Board.

1.1.5 HECoS Licence

Recommendation 13

To ensure clarity and usability and in order to prevent the proliferation of customised licences, it is recommended that HECoS is licensed under the terms of the Creative Commons Attribution licence, CC BY 4.0.

Recommendation 14

In addition to applying an open licence to HECoS, it is recommended that the HECoS governing body register a trademark in order to protect the definitive version of the coding scheme against misrepresentation.

1.1.6 HECoS Web Service

Recommendation 15

The HECoS web service should be hosted, maintained and managed by the HE Data and Information Landscape Management Office.

Recommendation 16

HECoS should be Five Star Open Data¹. It should be available on the web under open licence as structured data in non-proprietary format, URIs should be used to denote things and the data should be linked to other data to provide context.

Recommendation 17

The system used to host HECoS should support a broad range of open formats.

Recommendation 18

In order to satisfy the version control requirements, the HECoS web service should provide an API that can serve different versions of the whole vocabulary set as well as versions of individual terms.

Recommendation 19

The HECoS web service should maintain a comprehensive history of every term used in the coding scheme, together with any changes made to each term.

Recommendation 20

In order to facilitate interoperability, and ensure that HECoS is easily discoverable, the coding scheme should be described using VoID², the metadata format most commonly used by existing open data registries and repositories, e.g. datahub.io and data.gov.uk. Other metadata formats that may be relevant include the Asset Description Metadata Schema³ and the Data Catalogue Vocabulary⁴.

Recommendation 21

Where appropriate, the HECoS web service should follow the recommendations of the Cabinet Office Standards Hub Open Standards Board⁵.

1.1.7 Monitoring Recommendations

Recommendation 22

The HE Data and Information Landscape Management Office should ensure that HECoS structures, support and guidance are sufficient to enable a high level of consistency and accuracy of coding across the sector.

¹ Five Star Open Data, <http://5stardata.info/>

² VoID, <http://www.w3.org/TR/void/>

³ Asset Description Metadata Schema, <http://www.w3.org/TR/vocab-adms/>

⁴ Data Catalogue Vocabulary, <http://www.w3.org/TR/vocab-dcat/>

⁵ Cabinet Office Standards Hub Open Standards Board, <http://standards.data.gov.uk/meeting/open-standards-board-meeting-2-24-september-2013>

Recommendation 23

The HE Data and Information Landscape Management Office should establish a transparent and comprehensive process to monitor the consistency and accuracy with which HECoS codes are applied during the first 2 years of HECoS adoption. During this period no SUBSTANTIVE changes will be made to HECoS (see Recommendation 5).

Recommendation 24

After the initial 2 year adoption period, the HE Data and Information Landscape Management Office should establish a lightweight administrative process to monitor the consistency of coding with HECoS in practice.

Recommendation 25

The HE Data and Information Landscape Management Office should gather evidence of where both SUBSTANTIVE and ADMINISTRATIVE changes to HECoS may be required. This should include identifying problems and inconsistencies in definitions and scope notes, requirements for changes to navigational structures, adoption of new candidate terms and deprecation of obsolete terms (see Recommendations 5-9).

2. Overview

During Phase 1 of the HEDIIP New Subject Coding Scheme Project, the development of a robust, impartial, independent and transparent governance model supported by an appropriate and authoritative selection of stakeholders was identified as a critical success factor for ensuring buy-in from data consumers and suppliers across the sector.

This report outlines the stakeholder consultation methodologies used to gather governance requirements and presents a comprehensive series of pragmatic recommendations based on these requirements, covering the HECoS governing body, change management procedures, licensing, web service management, aggregations governance, external specialist vocabularies, monitoring consistency and accuracy, and managing the adoption plan. These recommendations have been harmonised with the governance principles outlined in *The Blueprint for a New HE Data Landscape* (KPMG, 2015) report.

3. Methodology

In order to gather input on and requirements for the governance of the new subject coding scheme from across the sector, the project has employed a number of consultation methodologies.

3.1 Phase 1 Methodology

During Phase 1 the project team undertook face to face consultation with core sector bodies, structured telephone interviews with Professional, Statutory and Regulatory Bodies and workshops with higher education providers. During these consultations, stakeholders were asked:

The new subject coding scheme will need to be published and maintained. From your organisation's perspective what do we need to achieve in respect of governance? (Impact Assessment and Requirements Definition, Kraan and Paull, 2014, p. 44)

Stakeholders' responses were collated and synthesised in order to prioritise requirements that were subsequently presented in the *Impact Assessment and Requirements Definition report* (Kraan and Paull, 2014). See Appendix 1 for further details of the mandatory priorities relating to Governance Requirements identified in the Phase 1 report.

3.2 Phase 2 Methodology

During Phase 2, the Phase 1 interview transcripts and workshop outputs were reanalysed to collate more detailed data. In addition, a number of workshops were held to enable stakeholders, representatives of the Advisory Panel and HEPs to provide detailed input on various aspects of the development and implementation of HECoS, including governance and adoption. See Appendix 2 and 3 for further details of the Phase 2 workshops and a list of stakeholders consulted.

Following this initial phase of consultation with representatives of the Advisory Panel, HEPs and statutory bodies, a draft of the HECoS Governance Plan was circulated for public consultation from February 2015 to May 2015. A public consultation website⁶ was established to disseminate drafts of the coding scheme, the governance model and the adoption plan. Forty-three individuals left sixty-four responses relating to all aspects of HECoS on the consultation website. A HECoS Governance Model Response Form spreadsheet was also disseminated via the website, which gathered responses from ten Higher Education Providers (see Appendix 2). Formal organisational responses were also gathered from seven sector bodies and Higher Education Providers, along with numerous comments and responses from others sent via e-mail.

The project team received thirty-one email responses and numerous comments and questions during a webinar organised with Jisc, and attended by around one hundred participants. A classification exercise designed to test the coding scheme was conducted during the Student Records Officer Conference (SROC) 2015 in York, with a group of over forty participants, and at the UCAS 2015 Admissions Conference in Newport with a group of around twenty participants. All three events allowed participants to provide feedback and comments on the draft Governance Plan.

4. HECoS Governing Body

4.1 HECoS Governing Body Requirements

HECoS requires a governance model, and a governing body, that strikes the right balance between inclusiveness and authority, support and overhead, and decisiveness and agility (Impact Assessment and Requirements Definition, Kraan and Paull p. 37). These requirements match the remit of the 'independent collective and consensual governance body' envisaged in *The Blueprint for a New HE Data Landscape* (KPMG, p. 10).

Extensive consultation with the sector indicates that there is widespread agreement among stakeholders that HECoS should be sector owned and organised, and hosted by an existing organisation, or organisations, from within the Higher Education sector. There is broad consensus that a new governance model should be developed under the auspices of an existing sector organisation, and there is little or no appetite for a new entity to be established or for HECoS to be governed by a formal standards body from outwith the sector⁷. The existing sector organisation within which the governing body sits must be established, trusted, respected and neutral. It is critical that the governing body has sufficient longevity to ensure the continuity of HECoS, and that it is protected from potential abolition resulting from changes in government and government policy.

Developing a governance model that is supported by a representative cross section of stakeholders, with sufficient authority and resources to manage the evolution of HECoS was identified in the Phase 1 Impact Assessment and Requirements Definition report as a critical success factor for the new subject coding scheme (Kraan and Paull, p. 37). The governance body must be responsive to the needs of stakeholders and ensure that all requirements are considered and evaluated equally and fairly. Furthermore, the governance body must be representative of the whole higher education sector including representatives of all nations of the UK, data providers and consumers,

⁶ HECoS Public Consultation site, <https://subjectcoding.wordpress.com/>

⁷ Only one HEP raised the issue of whether there should be a role for a standards organisation such as BSI to develop the coding scheme as a formal standard.

statutory and regulatory bodies, Higher Education Providers, including university and college associations and consortia, the strategic planning community, and the student body.

As HECoS will directly support the creation of the Standard Dataset, it is anticipated that both the new subject coding scheme and the new HE Standard Dataset will be governed by the same governance structures outlined in *The Blueprint for a New HE Data Landscape* (KPMG, p.72).

The report recommends 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.' (KPMG, p. 10)

'This maintains the independence of the governance body, but allows it to function and operate in the most efficient way possible.' (KPMG, p. 49)

This is in line with the findings of the HECoS Stage 2 consultation which indicates that there is strong support from the majority of stakeholders that governance of HECoS should be situated with the Higher Education Statistics Agency (HESA).

In addition to outlining examples of governance tasks, *The Blueprint* provides an illustrative approach to implementing effective governance of the data and information landscape, whilst acknowledging that there may be other options for the implementation of governance (KPMG, p.50). In order to harmonise the recommendations of this report with the recommendations of *The Blueprint*, the same terminology has been adopted for the different tiers of governance. The roles and functions of these governance entities are outlined in Section 6.2.1 Data Governance in the New Landscape of *The Blueprint* (KPMG, pp. 51-52) and in the recommendations outlined in this report. However, it is important to recognise that, in instantiating the recommended governance structures and drawing the various strands of the HEDIIP vision together, these functional governance roles may be delivered by existing teams within HESA.

4.2 HECoS Governing Body Recommendations

Recommendation 1

HECoS will be governed by the HE Data Governance Body as outlined in *The Blueprint for a New HE Data Landscape*. 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.

Recommendation 2

Within the structures of the HE Data Governance Body, the HE Data and Information Landscape Management Board should provide strategic direction for HECoS within the context of the HE Data and Information Landscape. The Board will be composed equally of data collectors and HEPs, and chaired by a representative from an HEP or related complementary sector. Rules for tenure are outlined in *The Blueprint for a New HE Data Landscape* (KPMG, pp.51-52).

Recommendation 3

The HE Data and Information Landscape Management Office will oversee the day-to-day management of the coding scheme, collate candidate terms, monitor usage and make recommendations to the Board. Within the HE Data and Information Landscape Management Office, it is expected that the Technical Group will have the requisite technical skills to oversee the management of HECoS.

Recommendation 4

The HE Data Governance Body will publish transparent rules, regulations, operating procedures and terms of reference, and should provide clearly established communication channels to enable stakeholders to input their requirements relating to HECoS, including aggregation scheme for subject based analysis. A HECoS Consultative

Group should be established under the aegis of the HE Data Consultative Groups in order to provide opportunities for wider stakeholder engagement.

5. Managing Changes to HECoS

The two critical success factors identified for specifying and approving changes to HECoS are a transparent change management governance process and an agreed change management cycle.

5.1 Change Management Requirements

During the Phase 1 consultation stakeholders expressed mixed views regarding appropriate mechanisms for reviewing and updating HECoS, including associated aggregation scheme for subject based analysis. There was broad agreement that HECoS must be open to review and able to accommodate change, but different opinions as to how frequently the scheme should be updated or how the change management process should be moderated and facilitated. These issues were explored further during the Phase 2 consultation.

Stakeholders agreed that clear, transparent and regulated processes are required to oversee change management procedures. Furthermore, there is agreement that the governing body must ensure that HECoS is reviewed regularly in consultation with representatives of relevant core sector bodies, Higher Education Providers and professional bodies in order to add new terms and deprecate obsolete ones, and to regulate changes to the Standard Cross-sector Aggregation Rules for Subject Based Analysis (see Recommendations for Subject Based Analysis & Text Mining, Cooper, 2015).

Managing the cycle and roll out of new iterations of HECoS is one of the critical success factors for both the governance and adoption of the new subject coding scheme. The sector is receptive to a coding scheme that is reviewed and updated according to a predefined schedule, on condition that there are appropriate lead in times, and changes are not introduced mid cycle. In addition, clarity is necessary regarding where changes have been made and what the implications of these changes are for data providers and aggregators.

A range of different views were expressed by stakeholders in relation to the frequency of the change management cycle. Some stakeholders agreed that minor changes would be acceptable on an annual basis, others indicated that annual changes would be acceptable, but not desirable, while others responded that the scheme must not change annually. There was general agreement that the coding scheme should be able to accommodate minor urgent changes, but that large scale changes would not be beneficial.

While there was no clear consensus as to the frequency of the change management cycle, there was general agreement that the process should seek to minimise the subsequent burden for both data providers and collectors. Evaluating the impact of all substantive changes to the coding scheme and communicating the rationale for these changes is regarded as a critical success factor for the adoption of HECoS.

Stakeholders for whom time series maintenance and analysis is important are less likely to accept a coding scheme that changes rapidly, as this is likely to have a negative impact on their ability to maintain and analyse the data. This issue could be mitigated by recoding old data sets, but this is a costly and unpopular option. The NSCS Project report on subject based analysis (Recommendations for Subject Based Analysis & Text Mining, Cooper, 2015) notes strong views on the stability and transparent management of sector-wide aggregation rules, and these concerns apply whether or not the underlying coding scheme changes.

Contrasting views were expressed regarding the stability of HECoS during its initial phase of adoption. Some stakeholders suggested that a period of stability should be guaranteed, to enable the coding scheme to 'bed in' with no substantive changes made for the first of 3 to 5 years of HECoS adoption, while others argued that it would be beneficial to enable HECoS to evolve rapidly during this early adoption phase in order to allow a comprehensive set of terms to emerge from the sector.

5.2 SUBSTANTIVE and ADMINISTRATIVE Changes

The type of changes that are likely to be required to keep HECoS current and fit for purpose include both SUBSTANTIVE and ADMINISTRATIVE changes:

- **Proposal, acceptance and implementation of new candidate terms** - changes of this nature are likely to have a considerable impact on users and implementers and are discussed in greater detail below. Acceptance of new candidate terms is regarded as a SUBSTANTIVE change, and should require the authorisation of the HE Data and Information Landscape Management Board.
- **Deprecation of obsolete terms** - if there is evidence that terms are not being used and are unlikely to be used in the foreseeable future, they should initially be marked as Non Preferred terms before subsequently being deprecated and removed from the coding scheme. The threshold for the deprecation of obsolete terms should be proposed to the HE Data and Information Landscape Management Board by the Management Office in consultation with the sector⁸. Before deprecating specific terms, the Management Office should undertake consultation and assess usage of the term in question in order to evaluate the impact of its deprecation. Removal of obsolete terms is regarded as a SUBSTANTIVE change, and consequently should require the authorisation of the HE Data and Information Landscape Management Board.
- **Clarification of definitions and scope notes** - SUBSTANTIVE changes to the definition of terms, i.e. changes which the Management Office believes would be likely to invalidate previously-classified courses, should require the authorisation of the HE Data and Information Landscape Management Board. The necessity and impact of such changes will require consideration and evaluation, which should occur in line with the agreed HECoS change management cycle. Minor corrections and clarifications to scope notes are regarded as ADMINISTRATIVE and should be undertaken by the Management Office as part of the routine administration of HECoS.
- **Changes to the Standard Cross-sector Aggregation Rules** – sector stakeholders require the aggregation scheme for Subject Based Analysis (refer to the Recommendations for Subject Based Analysis & Text Mining report) to be stable and transparently-maintained, hence changes to the aggregation rules are SUBSTANTIVE, whether or not a change to the underlying HECoS coding scheme occurs.
- **Changes to the navigation structure** - as the structure (i.e. groups, narrow/broader relationships, related terms, etc.) of HECoS is used for navigation only, and not for coding or aggregation, changes to the navigation structure should be regarded as ADMINISTRATIVE and so be undertaken by the Management Office as part of the routine maintenance of HECoS. Changes to the navigation structure will have no impact on HECoS terms, downstream data use, or the validity of existing data.

During the establishment of governance for HECoS (including the Subject Based Analysis aggregation rules), it will be necessary to elaborate upon this account of SUBSTANTIVE vs ADMINISTRATIVE changes according to consensus, to ensure that stakeholders are content with the operationalisation of the distinction. This operationalisation should be supported by documented consultation and approvals processes, and clear terms of reference for the Management Office and the Board, separately, in respect of HECoS.

⁸ One threshold for deprecation of obsolete terms proposed during the Stage 2 consultation was any term used by two or less institutions with fewer than 100 headcounts between them.

5.3 Change Management Recommendations

Recommendation 5

In order to facilitate the adoption of HECoS a stability focused approach will be taken during the first 2 years and it is recommended that no SUBSTANTIVE changes should be made to the coding scheme during this period. However it is recognised that it may be necessary to address exceptional or unforeseen issues that arise during the adoption phase. Any issues that arise should be referred to the Board who should consider whether action should be taken, and the implications for the stability of the coding scheme. At the end of this period, the Management Office should review the evidence gathered and assess the case for SUBSTANTIVE changes, for proposal to the Management Board. During the 2 year introduction period, ADMINISTRATIVE changes may be made to the coding scheme as required. It is recommended that the Management Office monitors the user experience and data quality on a frequent basis during this period, in order to quickly identify and correct confusion or misconception.

Recommendation 6

In order to strike a balance between flexibility and stability after the first 2 years, it is recommended that the HE Data and Information Landscape Management Office should undertake an annual review of the coding scheme. The annual review should consider changes to definitions, scope notes and navigation structures, the impact of adopting new candidate terms and deprecating obsolete terms, and the overall level of change to the coding scheme. In the course of the annual review, the Management Office should identify both SUBSTANTIVE and ADMINISTRATIVE changes. ADMINISTRATIVE changes to scope notes and navigation structures should be undertaken by the Management Office on an annual basis. SUBSTANTIVE changes to term definitions, proposals for the adoption of candidate terms and the deprecation of obsolete terms, should be submitted to the HE Data and Information Landscape Management Board for consideration. Before authorising any SUBSTANTIVE changes, the Management Board should consider evidence of the necessity for the proposed changes, the impact of the proposed changes on data providers and data collectors, and agree an appropriate schedule for the changes to be implemented.

Recommendation 7

A detailed summary of all significant ADMINISTRATIVE changes should be communicated to the sector by the HE Data and Information Landscape Management Office on an annual basis, and a persistent record of all changes maintained.

Recommendation 8

A detailed summary of all SUBSTANTIVE changes, together with transparent explanations outlining the rationale for these changes, should be communicated to the sector by the HE Data and Information Landscape Management Office on behalf of the HE Data and Information Landscape Management Board on a regular basis, in line with the agreed change management schedule.

Recommendation 9

In order to allow HECoS to evolve, the NSCS Project recommends setting up a Candidate Terms Registry (see 5.4 below) to enable users to submit new candidate terms for proposed inclusion in the coding scheme. The HECoS Candidate Terms Registry should be run and managed by the HE Data and Information Landscape Management Office. This service should be kept distinct and separate from the main HECoS web service in order to preserve the integrity of the authoritative version of HECoS.

5.4 HECoS Candidate Terms Registry

As a result of consultation with core sector bodies and Higher Education providers during Phase 2, the NSCS Project recommends that it would be beneficial for the HECoS governing body to establish a channel to systematically gather new terms for the coding scheme. The HECoS Candidate Terms Registry should be run and managed by the Management Office. In order to preserve the integrity of the authoritative version of HECoS, this service should be kept distinct and separate from the main HECoS web service. A system of this nature could be managed as follows.

If a user is developing a new programme for which an appropriate term does not exist in HECoS, they should code their course using the nearest appropriate term and then consult the Candidate Terms Registry to ascertain whether the term they require has already been proposed there by another user. If the term already exists, they can indicate that they would like to use it to describe their course. If the term required by the user does not already exist in the registry, they can submit a term label, scope notes and supporting evidence.

The HE Data and Information Landscape Management Office will review all candidate terms as outlined in Recommendation 6. Before submitting new candidate terms to the HE Data and Information Landscape Management Board for consideration for inclusion in HECoS, the Management Office should undertake due diligence to ensure that the term is distinguishable from existing HECoS terms and evaluate the impact of its adoption. Candidate terms submitted to the Management Board will be considered for inclusion in HECoS on the basis of the acceptance criteria outlined in *The Higher Education Classification of Subjects (HECoS) vocabulary* (Kraan and Paull, 2015). A term may be accepted for inclusion in the coding scheme if it fills a gap in HECoS as evidenced by:

- The overloading of a HECoS code in HESA returns.
- Clearly identifiable clusters of degree programme titles and descriptions in HESA returns for which no HECoS code exists.
- Suggestions by subject matter experts, with supporting evidence from any source.

The history of the genesis of any new terms accepted for inclusion in HECoS will be preserved in a Historic Note in the HECoS web service.

6. Governance of Subject Based Analysis

Aggregation of statistics according to subject groupings is widely used in reporting and analysis and are covered in detail by a separate deliverable i.e. the *Recommendations for Subject Based Analysis & Text Mining* (Cooper, 2015). The term Subject Based Analysis (SBA) is used to refer to the analysis of data where subject of study is explicitly recorded, and for which statistics are reported according to groupings of HECoS terms, typically in published statistics. The same aggregation rules are valuable for HEP-internal statistics too, because institution-level decision-making will be more effective when sector-level data can be included in the decision-making process.

Whereas subject groups were integral to the definition of JACS3, being represented in the structure of the codes used, HECoS separates the terms available for classifying subject of study from the definition, or definitions, of how data should be aggregated over related subjects. Consequently, it is necessary to consider governance of subject group definitions separately from governance of the HECoS terms.

The Stage 2 NSCS Project consultation showed very clearly that subject based analysis is an essential component of the management of Higher Education at all levels from HEPs to government. Stakeholders were clear about the need for strong governance, the avoidance of proliferation, the need for clear and easily-found information, and the importance of usable time-series data across the JACS3-HEDIIP transition. Further information appears in *Recommendations for Subject Based Analysis & Text Mining* (Cooper, 2015), which makes the case for integrated governance of HECoS terms and the definition of subject groups built upon it.

The recommendations in PD05, which will not be duplicated here, discriminate between subject group definitions which should be formally governed, and other aspects which should be part of a centralised information and support service operated by the machinery of the HE Data Governance body during adoption (see section 13, Managing the Adoption Plan) and normal operations.

6.1 Governance of Subject Group Definitions Recommendations

Recommendation 10

Governance of the Standard Cross-sector Statistical Aggregation Rules should be fully integrated with the change management process for HECoS terms, with any change being considered as SUBSTANTIVE, and therefore requiring the approval of the HE Data and Information Landscape Management Board. The Standard Aggregation Rules should be available under the same licence (Section 9) as HECoS and disseminated and curated as part of the same HECoS web service (Section 10).

Recommendation 11

HECoS information and support services should provide access to subject group definitions which are not under the control of the HE Data Governance body but which are **recognised** by it. These include: ATAS subject lists, cross-reference to QAA benchmark statements, Strategically Important and Vulnerable Subject groups (HEFCE SIVS), HEFCW Academic Subject Category definitions, SFC Subject Area definitions, and others as may be deemed important. Amendments of the list of “recognised” subject group definitions are deemed to be ADMINISTRATIVE changes. A proposed way forward for subject based analysis is outlined in PD05 *Recommendations for Subject Based Analysis & Text Mining* (Cooper, 2015, pp. 8-11); Section 3.3 addresses how these requirements can be met in different ways according to the character of the subject group definitions.

7. Specialist Vocabularies

Balancing comprehensive detail for data analysts versus ensuring accuracy of coding and minimising effort for data providers was identified as an area of potential tension during Stage 1 of the NSCS Project (Impact Assessment and Requirements Definition, Kraan and Paull, 2014). This issue is further explored in the *Structure and Candidate Scheme* (Kraan and Paull, 2015) and *Recommendations for Subject Based Analysis & Text Mining* (Cooper, 2015). In an effort to accommodate the requirements of specialist subjects, HECoS has been designed in such a way that relations can be defined with other specialist vocabularies and data classified in other frameworks. In domains where appropriate specialist vocabularies exist, this will enable stakeholders to relate HECoS terms to more specific terms relevant to the domain; however relating any specialist vocabularies to HECoS will require careful management and governance.

7.1 Governing Specialist Vocabularies Options

During consultation, stakeholders articulated quite different opinions and requirements as to how specialist vocabularies should be accommodated within HECoS.

One option articulated during the Phase 2 consultation was for ‘specialisation hierarchies’ to be governed by relevant sector agencies or Professional, Statutory and Regulatory Bodies, e.g. the NHS might govern specialisation hierarchies relating to nursing education. This would reduce the governance overhead on the HECoS governing body and ensure that the coding scheme does not become bloated with specialist terms. Some stakeholders raised concerns with this model on the grounds that all relevant terms should be accommodated by HECoS, otherwise users will be required to code twice; once using the relevant specialisation term and once using the generic HECoS term. (It should be noted that this could be achieved programmatically.) In addition, stakeholders expressed concern that the HECoS governing body would not be able to exert sufficient influence over the sector agencies or professional bodies who manage these specialisation hierarchies.

A second option proposed was for HECoS to accommodate all specialist terms in all domains. One advantage of this approach would be that the HECoS governing body would be responsible for overseeing changes to the coding scheme across all domains, ensuring that changes are appropriate and timely. Users would then be able to choose only the specialist code that is relevant to their course or module. However some stakeholders were strongly opposed to this approach on the grounds that it could potentially destabilise HECoS as the coding scheme would potentially have to accommodate frequent changes across multiple specialised domains. Further drawbacks articulated by stakeholders are that the number of terms in HECoS would increase enormously, the coding scheme would cease to be a flat list, and managing multiple specialist vocabularies would place a significant burden on the HECoS governing body.

7.2 Governing Specialist Vocabularies Recommendations

The linked data aspect of HECoS makes it feasible to define relations with any other vocabulary precisely and easily from a technical point of view. However, considerable care should be taken to ensure that these relationships are appropriate and meaningful. As HECoS is designed specifically to classify subjects of study, it is not recommended that direct relationships are created with vocabularies that classify other entities such as cost centres or REF returns (The Higher Education Classification of Subjects (HECoS) vocabulary, Kraan and Paull, 2015, p. 42). However within the sphere of subjects of study, there are specific subject areas where stakeholders have expressed the need to relate HECoS to more fine grained specialist vocabularies, e.g. the area of work classifiers of the NHS Occupation Codes⁹. *The Higher Education Classification of Subjects (HECoS) vocabulary* outlines the principles of linking data in HECoS to data classified in other frameworks (Kraan and Paull, 2015, p. 13-14).

In order to meet stakeholder requirements without jeopardising the integrity of HECoS, relevant sector organisations should have the option of registering vocabularies under their control with the HE Data and Information Landscape Management Office. It is recommended that specialist vocabularies are reviewed and approved through the following process.

Recommendation 12

Relevant sector organisations may register specialist vocabularies under their control with the HE Data and Information Landscape Management Office. These vocabularies must meet the appropriate HECoS technical requirements and must not duplicate existing HECoS codes. Due diligence will be undertaken by the Management Office to ensure all specialist vocabularies meet these requirements. Authorisation to approve the relationship of specialist vocabularies with HECoS lies with the HE Data and Information Landscape Management Board. Once specialist vocabularies have been approved by the Management Board they can be related to HECoS by identification with one or more broader HECoS terms. It is envisaged that only a small number of HECoS related specialist vocabularies will be authorised by the HE Data and Information Landscape Management Board.

8. Licensing

There is widespread agreement among stakeholders that the new coding scheme should be released under an appropriate open licence, in order to facilitate its adoption and use by the community.

It is important to distinguish between the licence that applies to the authoritative version of HECoS maintained by the governing body, and licences that may apply to applications that implement HECoS or data generated using the coding scheme. For example, one stakeholder emphasised that the confidentiality of their HECoS coded data is critical from their perspective, however this does not preclude the actual coding scheme being released under an open licence. This report aims to articulate the licensing requirements of the authoritative version of HECoS, but does not comment on the licensing of data that uses the scheme.

⁹ Health and Social Care Information Centre, National Health Service Occupation Codes Manual, version 13.1, http://www.hscic.gov.uk/media/13060/NHS-Occupation-Code-Manual/pdf/NHS_Occupation_Code_Manual_Version_13.1.pdf

8.1 HECoS Licence Requirements

The licence must:

- Attribute HECoS to the appropriate IPR holder, e.g. the governing body or the organisation that authorises the governing body.
- Grant users a worldwide, royalty-free, non-exclusive licence to use HECoS in perpetuity.
- Enable HECoS to be used for commercial purposes, e.g. to enable commercial developers to build applications that incorporate the coding scheme.
- Allow the creation of HECoS derivatives.
- Allow users to reuse, share and reproduce HECoS in whole or in part.
- Require users who reuse, share and reproduce HECoS to attribute the IPR holder and refer to the public licence.
- Include a disclaimer of warranties and limitation or waiver of liability.

Existing licences that meet the requirements outlined above include the Creative Commons Attribution Licence (CC BY)¹⁰ and the Open Government Licence¹¹. Though both licences have their merits, use of CC BY is recommended for HECoS, as it has been adopted more widely than the Open Government Licence.

Under the terms of the CC BY 4.0 licence users are free to share, copy and redistribute the material in any medium or format, and to adapt, remix, transform, and build upon the material for any purpose, even commercially. Users must attribute the material, give appropriate credit, provide a link to the license, and indicate if changes were made. Users may not apply legal terms or technological measures that legally restrict others from doing anything the license permits. No warranties are given; the licensor offers the licensed material as-is and as available, and makes no representations or warranties of any kind concerning the material, whether express, implied, statutory, or other. The licensor will not be liable to the user on any legal theory.

8.2 HECoS Licence Recommendations

Recommendation 13

To ensure clarity and usability and in order to prevent the proliferation of customised licences, it is recommended that HECoS is licensed under the terms of the Creative Commons Attribution licence, CC BY 4.0.

Recommendation 14

In addition to applying an open licence to HECoS, it is recommended that the HECoS governing body register a trademark in order to protect the definitive version of the coding scheme against misrepresentation.

¹⁰ CC BY 4.0, <http://creativecommons.org/licenses/by/4.0/>

¹¹ Open Government Licence, <http://www.nationalarchives.gov.uk/doc/open-government-licence/version/3/>

9. Web Service Management

9.1 HECoS Web Service Requirements

A range of technical requirements for the new subject coding scheme web service were articulated by stakeholders consulted during Phase 1 and 2 of the NSCS Project. NB these requirements and recommendations relate to the management of the HECoS web service, rather than to the HECoS vocabulary itself.

- The new subject coding scheme must be made available in an open, easily reusable format that software systems can support without barriers.
- Users must be able to download the scheme.
- The HECoS web service must provide an API that is version aware.
- The HECoS governing body should record when and why subject codes and definitions are changed, alert users to recent changes, and make this information available via the web service.
- Appropriate metadata for the HECoS scheme must be maintained by the governing body.

Open formats are defined by the Open Data Handbook as follows:

‘Even if information is provided in electronic, machine-readable format, and in detail, there may be issues relating to the format of the file itself... An open format is one where the specifications for the software are available to anyone, free of charge, so that anyone can use these specifications in their own software without any limitations on reuse imposed by intellectual property rights... The benefit of open file formats is that they permit developers to produce multiple software packages and services using these formats. This then minimises the obstacles to reusing the information they contain.... The preference from the open government data perspective therefore is that information be released in open file formats which are machine-readable.’¹²

9.2 HECoS Web Service Recommendations

Recommendation 15

The HECoS web service should be hosted, maintained and managed by the HECoS Management Office.

Recommendation 16

HECoS should be Five Star Open Data¹³. It should be available on the web under open licence as structured data in non-proprietary format, URIs should be used to denote things and the data should be linked to other data to provide context.

Recommendation 17

The system used to host HECoS should support a broad range of open formats.

Recommendation 18

In order to satisfy the version control requirements, the HECoS web service should provide an API that can serve different versions of the whole vocabulary set, as well as versions of individual terms.

Recommendation 19

The HECoS web service should maintain a comprehensive history of every term used in the coding scheme, together with any changes made to each term.

¹² Open File Formats, The Open Data Handbook, <http://opendatahandbook.org/guide/en/appendices/file-formats/>

¹³ Five Star Open Data, <http://5stardata.info/>

Recommendation 20

In order to facilitate interoperability, and ensure that HECoS is easily discoverable, the coding scheme should be described using VoID¹⁴, the metadata format most commonly used by existing open data registries and repositories, e.g. datahub.io and data.gov.uk. Other metadata formats that may be relevant include the Asset Description Metadata Schema¹⁵ and the Data Catalogue Vocabulary¹⁶.

Recommendation 21

Where appropriate, the HECoS web service should follow the recommendations of the Cabinet Office Standards Hub Open Standards Board¹⁷.

9.3 Proof of Concept Web Service

For the purposes of developing and demonstrating HECoS, the NSCS Project implemented and adapted an instance of the TemaTres Controlled Vocabulary Server¹⁸. The NSCS installation of TemaTres meets all the above requirements and has been integrated with the Disqus comment hosting service to enable stakeholders and subject experts to comment on terms. Further investigation would be required in order to ascertain whether TemaTres is suitable to support the HECoS web service in the longer term.

10. Monitoring Consistency and Accuracy of Coding

10.1 Monitoring Requirements

Stakeholders consulted during the Phase 2 workshops indicated that monitoring the accuracy and consistency of subject coding was a significant priority and suggested that this should fall within the remit of the HECoS governing body, who should be empowered to take corrective action where necessary. There was some debate as to whether the HECoS governing body should also monitor or audit the consistency of coding across institutions and develop regulatory frameworks to ensure that staff have the requisite coding skills. While monitoring the consistency of the Standard Data set is within the remit of the HE Data and Information Landscape Management Board and Management Office as outlined in Section 6 of *The Blueprint for a New HE Data Landscape* (KPMG, 2015, pp. 46-53), this document is concerned specifically with requirements for monitoring the quality, consistency and accuracy of the HECoS coding scheme and not with evaluating and monitoring the quality of data gathered through implementations based on HECoS. For further information on principles and procedures for Data Governance within the new HE Data and Information Landscape refer to *The Blueprint for a New HE Data Landscape* (KPMG, 2015).

10.2 Monitoring Recommendations

Recommendation 22

The HE Data and Information Landscape Management Office should ensure that HECoS structures, support and guidance are sufficient to enable a high level of consistency and accuracy of coding across the sector.

Recommendation 23

The HE Data and Information Landscape Management Office should establish a transparent and comprehensive process to monitor the consistency and accuracy with which HECoS codes are applied during the first 2 years of HECoS adoption. During this period no SUBSTANTIVE changes will be made to HECoS (see Recommendation 5).

¹⁴ VoID, <http://www.w3.org/TR/void/>

¹⁵ Asset Description Metadata Schema, <http://www.w3.org/TR/vocab-adms/>

¹⁶ Data Catalogue Vocabulary, <http://www.w3.org/TR/vocab-dcat/>

¹⁷ Cabinet Office Standards Hub Open Standards Board, <http://standards.data.gov.uk/meeting/open-standards-board-meeting-2-24-september-2013>

¹⁸ HECoS, hosted on TemaTres Controlled Vocabulary Server, <http://ovod.net/tematres/vocab/index.php>

Recommendation 24

After the initial 2-3 year adoption period, the HE Data and Information Landscape Management Office should establish a lightweight ADMINISTRATIVE process to monitor the consistency of coding with HECoS in practice.

Recommendation 25

The HE Data and Information Landscape Management Office should gather evidence of where both SUBSTANTIVE and ADMINISTRATIVE changes to HECoS may be required. This should include identifying problems and inconsistencies in definitions and scope notes, requirements for changes to navigational structures, adoption of new candidate terms and deprecation of obsolete terms (see Recommendations 5-9).

11. Set Up and Operating Costs

In line with the recommendations of *The Blueprint for a New HE Data Landscape* (KPMG, 2015) which advises 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.’ (KPMG, p. 10)

There is strong support from stakeholders consulted during Phase 2 of the NSCS Project to situate the HECoS governing body within the Higher Education Statistics Agency (HESA).

There are clear practical advantages to the operation, maintenance, and governance of HECoS being located within HESA:

1. It has appropriate capability; it is an established higher education body with existing staff, business processes, and sector level expertise in the area of standards maintenance, data collection and management.
2. Operating HECoS along-side HESA data and statistical services offers a reduced cost option, relative to a special-purpose entity.

The *Blueprint* presents an analysis of the estimated cost, are based on information currently available using current industry prices, of implementing the New Data Landscape plan, including establishing the governance body. HECoS is part of the overall landscape and its operation, maintenance, and governance is subsumed into that estimate. Costs to establish the HECoS Service and to oversee adoption are assumed to be additional to that sum

11.1 Set-up Costs

This section is concerned only with the establishment of the technical systems, information resources, and support services for the initial adoption process: see the NSCS Project Adoption Plan report (Ferrell and Campbell, 2015). Set-up cost containment could follow two ideas, below.

Firstly, the technical systems which will comprise the web service, manage documents, and provide the web pages for HECoS are presumed to be identical to those which provide for other aspects of the HE data landscape, such as the Standard Dataset. Additionally, it is entirely possible that HESA might introduce modern systems for metadata management as part of the HESA Data Futures Programme¹⁹. Systems established for that purpose could well serve both the Standard Dataset and HECoS requirements at the cost of system configuration alone. In the long run, an integrated approach to metadata management, which encompasses critical data collection data dictionary, HECoS, the Standard Dataset definitions, etc. has clear strategic benefits. The technical system will not be a high throughput transactional system, although it should still be robust and resilient. Consequently, there would be low additional infrastructure cost, assuming cloud-based provision, which is consistent with HESA technical strategy.

¹⁹ Formerly known as CACHED, <https://www.hesa.ac.uk/datafutures>

Secondly, there should be provision in the existing HE Data and Information Improvement Programme for the creation of adoption-supporting assets, although the specific details are not yet available. That work will, naturally, reduce the additional cost required to establish HECoS.

Based on the above, which assumes software licence costs are absorbed elsewhere²⁰, configuration of technical systems, populating with HECoS, testing, and content management activities are likely to fall in the region of 6-8 person-months of effort.

11.2 Maintenance and Governance Costs

Within the Summary of Implementation Costs (KPMG, 2015, pp. 97-99), *The Blueprint* estimates the cost of establishing the Governance Team to be in the region of £240,000, on the basis of 4 full-time-equivalent (FTE) staff. For a full analysis of estimated costs please refer to *The Blueprint for a New HE Data Landscape* Section 9. Indicative Costs (KPMG, pp. 97-102).

The current cost of maintaining JACS which equates to approximately 0.8 FTE per annum²¹, excluding administrative and programming costs, and this seems to be a reasonable first estimate for HECoS. There are some differences in approach, however. Firstly, whereas JACS maintenance involved bursts of activity every 4 or 5 years, the proposed HECoS model is smoother and more reactive. By separating off the aggregation rules for subject based analysis, some of the complications which the JACS hierarchy created for coding scheme revision have been eliminated. Secondly, an approach which makes HECoS available as a web service enables a more agile approach. On the other hand, the need to maintain technical systems probably absorbs some of the savings from streamlined procedures. Overall, therefore, we suggest that maintaining HECoS will require a broadly similar level of overall resource to that currently devoted to JACS but with a different profile. It is conceivable that a more frequent series of lower-effort maintenance tasks will be easier to resource in a cost-efficient manner.

The indicative annual maintenance cost of HECoS can be estimated as being no more than 0.8 FTE (this resource is included within the 4 FTE estimated by KPMG).

²⁰ There are also Open Source Software packages available which have provisionally been identified as candidates, although a full set of requirements has not been developed to evaluate options against.

²¹ JACS is currently upgraded every 4 years. This involves 8 people at HESA and 8 at UCAS allocating 15% effort over an 18 month period. This equates to 2.4 FTE for the 8 month period, or 0.8 FTE per annum.

12. References

Alexander, K., Cyganiak, R., Hausenblas, M. and Shao, J. (2011). *Describing Linked Datasets with the Void Vocabulary*. W3C Interest Group Note 03 March 2011. Retrieved from <http://www.w3.org/TR/void/>

Cabinet Office Standards Hub (2013). *Open Standards Board - Meeting 2*. Retrieved from <http://standards.data.gov.uk/meeting/open-standards-board-meeting-2-24-september-2013>

Cooper, A., (2015). *Recommendations for Subject Based Analysis & Text Mining*. (NSCS Project Deliverable PD05)

Creative Commons. *Attribution 4.0 International*. Retrieved from <http://creativecommons.org/licenses/by/4.0/>

Dekkers, M. (2013). *Asset Description Metadata Schema (ADMS)*. W3C Working Group Note 01 August 2013. Retrieved from <http://www.w3.org/TR/vocab-adms/>

HEDIIP *New Landscape Project*. Retrieved from <http://www.hediip.ac.uk/new-landscape/>

HEDIIP *New Subject Coding Scheme Project*. Retrieved from http://www.hediip.ac.uk/subject_coding/

Campbell, L.M., Cooper, A. and Ferrell, G. (2015). *HEDIIP NSCS Project Adoption Plan*. (NSCS Project Deliverable PD07)

Five Star Open Data. Retrieved from <http://5stardata.info/>

KPMG, (2015). *The Blueprint for a New HE Data Landscape*

Kraan, W. G. and Paull, A. (2014). *New Subject Coding Scheme; Impact Assessment and Requirements Definition*.

(NSCS Project Deliverable PD01/02)

Regulatory Partnership Group. Retrieved from http://www.hediip.ac.uk/wp-content/uploads/Stage_1_report_2014-11-13.pdf

Kraan, W.G and Paull, A. (2015). *HEDIIP NSCS Structure and Candidate Scheme*.

Kraan, W.G., and Paull, A. (2015). *The Higher Education Coding of Subjects (HECOS) vocabulary*. (NSCS Project Deliverable PD04)

Maali, F. and Erickson, J. (2014). *Data Catalog Vocabulary (DCAT)*. W3C Recommendation 16 January 2014. Retrieved from <http://www.w3.org/TR/vocab-dcat/>

Open Government Licence for Public Sector Information. Retrieved from <http://www.nationalarchives.gov.uk/doc/open-government-licence/version/3/>

Appendix 1 - Phase 1 Governance Requirements

Governance requirements identified in the New Subject Coding Scheme Impact Assessment and Requirements Definition report (Kraan and Paull, 2014).

ID	Title	Description	Stakeholders	Rationale	Priority	Type	Grouping
R9	Governing and sector bodies	Governance of the NSCS shall be influenced strongly by specified sector bodies (HESA, UCAS, and others to be determined), by HEPs, representatives of Professional, Statutory and Regulatory Bodies (PSRBs) and other significant stakeholders. There shall be clear lines of responsibility, openness and transparency	HEPs, sector bodies	Provides strong sector representation on development.	Mandatory	performance	Governance
R18	Remaining static for an academic cycle	Governance of the NSCS shall enable management of the NSCS as an HE standard that shall remain static for any single specific academic annual cycle.	UCAS, HESA, HEPs	Must be stable and robust	Mandatory	constraint	Governance
R19	Facilitating annual reporting and review	Governance of the NSCS shall facilitate annual reporting and review by all stakeholders with a change implementation period of not less than 3 years, with a defined, transparent process for changes, in particular for adding and removing terms.	UCAS, HESA, HEPs	Must be stable and robust, but also capable of change	Mandatory	constraint	Governance
R29	Facilitating datasets that are fit for purpose	Governance shall facilitate the creation, maintenance and usage of authoritative data sets.	GPC, HEPs, HEE, other Sector bodies	Provides for current and wider usage in analysis via HESA, HEFCE and others.	Mandatory	performance	Uses
R47	Supporting explicit aggregations of subjects	Ways in which NSCS data is grouped (hierarchies and aggregations) shall be negotiated as part of governance and published.	HEPs	Supports use outside HEPs	Desired	constraint	Codes and structure

Appendix 2 - Phase 2 Consultation

Workshops

- NSCS Technical Workshop, HESA, Cheltenham, 11th November 2014.
- NSCS Technical Workshop, University of Loughborough, 11th December 2014.
- From concept to reality; governance and adoption of the HECoS new subject coding scheme, HESA, Cheltenham, 23rd January 2015.
- New Subject Coding Scheme (HECoS) Consultation Workshop: JACS fell down and broke his crown but what will come tumbling after? SROC 2016, Loughborough University, 4th - 6th April 2015.
- Subject Matters! A look at HECoS, the proposed new Higher Education Classification of Subjects to replace JACS, UCAS Admissions Conference, Celtic Manor Resort, Newport, 30th March 2015.

Webinars

- About the HEDIIP Subject Coding Project, JISC webinar, 23rd April 2015.

HEP Responses to HECoS Governance Model Response Form Spreadsheet

- Brunel University, London
- Imperial College, Registry
- King's College London, Planning & Analytics
- Kingston University, Planning Office
- Loughborough University
- University of Aberystwyth, Organisational Planning
- University of Brighton, Strategic Planning and Projects Division
- University of Cambridge, Academic Division
- University of Hull, Strategic Development Unit
- University of the West of England

HEP and PSRB Individual Responses

- HEFCE
- HESPA
- UCAS
- Loughborough University
- University of Aberdeen
- University of Greenwich
- University of Liverpool

Appendix 3 - Stakeholders Consulted

For a full list of stakeholders consulted during Phase 1 of the New Subject Coding Scheme Project, see Appendix 1 of the Impact Assessment and Requirements Definitions report (Kraan and Paull, 2015).

The following stakeholders were consulted during Phase 2.

Sector Bodies

- DELNI
- Enterprise Educators UK
- HEE
- HEFCE
- HEFCW
- HESA
- HESPA
- MRC
- NCTL
- QAA
- SAAS / Scottish Government
- SFA
- SLC
- UCAS

Higher Education Providers

- Aberystwyth University
- Anglia Ruskin University
- Aston University
- Bangor University
- Brunel University
- Canterbury Christ Church University
- Cardiff University
- City University London
- Conservatoire for Dance and Drama
- Coventry University
- Edge Hill University
- Edinburgh Napier University
- Harper Adams University
- Imperial College London
- Kings College London
- Lancaster University
- Leeds Trinity University
- Liverpool Hope University
- London School of Economics
- London South Bank University
- Loughborough University
- Manchester Metropolitan University

- Newcastle University
- Northumbria University Newcastle
- Nottingham Trent University
- Plymouth University
- Queen Margaret University
- Royal Holloway
- Sheffield Hallam University
- Swansea University
- University of Birmingham
- University of Brighton
- University of Bristol
- University of Buckingham
- University of Central Lancashire
- University of Chester
- University of Derby
- University of East Anglia
- University of Edinburgh
- University of Exeter
- University of Gloucestershire
- University of Greenwich
- University of Huddersfield
- University of Hull
- University of Kent
- University of Leeds
- University of London, Goldsmiths
- University of London, Queen Mary
- University of London, St Georges
- University of Manchester
- University of Northampton
- University of Oxford
- University of Portsmouth
- University of Salford
- University of Sheffield
- University of Southampton
- University of St Andrews
- University of Sunderland
- University of Surrey
- University of Sussex
- University of the West of England
- University of Wales Trinity Saint David
- University of Warwick
- University of West London
- University of West Scotland
- University of Wolverhampton
- University of Worcester
- University of York