



HEDIIP New Subject Coding System Project

Summary Report

December 2015

About this Report

This report summarises the outputs and recommendations of Stage 1 and Stage 2 of the HEDIIP Subject Coding project. In full the project has delivered the following which are published in more detail as separate documents (as outlined in Appendix B):

- Impact Assessment and Requirements Definition
- Structure and Candidate Scheme
- The Higher Education Classification of Subjects (HECoS) vocabulary
- Recommendations for Subject Based Analysis & Text Mining
- Governance Model
- Adoption Plan

The report also outlines the drivers for the initiation of the Subject Coding project and the expected benefits of adopting the new coding scheme (HECoS – Higher Education Classification of Subjects).

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

Contents

About this Report	2
About HEDIIP	2
1. About Subject Coding	4
1.1 Subject coding in Higher Education data.....	5
1.2 Problems with JACS	5
1.3 Why not JACS4?.....	5
2. The Subject Coding project.....	5
3. Requirements for a new coding scheme	6
4. New Coding Scheme – HECoS.....	6
4.1 HECoS development and characteristics.....	6
4.2 Benefits of HECoS	7
5. Implementation	8
5.1 Management and oversight	8
5.2 Subject Based Analysis.....	9
5.3 Adoption Plan and next steps.....	9
APPENDIX A	10
Subject Coding project approach	10
APPENDIX B.....	11
Subject Coding project Stage 1 and Stage 2 full reports	11
References	11

Executive summary

It is clear from the independent analysis commissioned by HEDIIP undertaken in 2013 and in the more recent studies described in this report, that the JACS approach to subject coding is no longer sustainable or fit for purpose. The issues with JACS, explored further in this report, can be summarised as follows:

- It has run out of room for the addition of newly emerging subjects because the limit of the existing coding framework has been reached
- It has become increasingly uneven with high levels of granularity in some subjects with some subjects of interest not covered at all
- It is clearly not being consistently applied
- It does not meet the needs of all of the key sector stakeholders with many data collectors taking a different approach to coding subjects

However JACS is widely understood across the HE landscape and is embedded in many different systems and processes.

After extensive consultation carried with HE providers, data collectors and processors this report describes a solution, that is in many ways an evolution of JACS, that addresses its deficiencies, provides greater flexibility for further growth and meets the needs of a wider group of stakeholders.

The solution described in the report is made up of 4 essential components:

Coding frame - the coding frame itself is a single flat list of subjects that ensures the categorisation of courses and modules is always performed at a consistent level. The list itself is in many ways an evolution of JACS3 it has been produced by reviewing each JACS3 code to ensure it has a clear definition (that distinguishes it from other subjects) and has demonstrable usage in categorising courses and modules; this has led to the elimination of subjects such as 'History by topic' and '.... not elsewhere classified'. Subject code identifiers are randomly generated to avoid the coding frame running out of space. Navigation tools will help coders locate appropriate subject codes rather than relying on the parts of the JACS3 hierarchy they are familiar with.

Standard hierarchy - standard aggregation rules, that sit above the coding frame, will enable consistent analysis of subject areas. These aggregation rules being based on the KIS/UNISTATS subject groups will bring convergence in subject categorisation across the sector. The hierarchy levels will be mapped to levels within JACS3 to provide continuity in subject analysis through the transition from JACS to the new scheme and to other schemes such as those used by QAA to improve consistency across the sector.

Governance - A governance function, described in the report, will enable the new scheme to evolve without the need for extensive reviews and overhauls that have been required to maintain JACS (i.e. moving from JACS1.7 to JACS2 and then JACS3).

Adoption - The adoption plan will co-ordinate adoption activities across HE providers and collectors and provide a range of training, mapping tools and guidance materials to assist in the transition to the new scheme.

1. About Subject Coding

1.1 Subject coding in Higher Education data

While many aspects of HE data specification can draw on standards from elsewhere, the coding of HE subjects has always been an area that requires an HE-specific solution. The profile of subjects studied at HE differs from other parts of education and research activity continues to push the boundaries of academic disciplines.

For the past two decades many HE data systems have used a system called the Joint Academic Coding System - or JACS. This system was developed jointly by HESA and UCAS in the mid-1990s by merging two existing coding systems; the Standard Classification of Academic Subjects (SCAS) and HESAcodes. Both of these systems were based on a hierarchical structure and utilised codes made up of a letter followed by three numbers.

The first operational version of JACS was published in 1999 and the system underwent major revisions in 2005 (JACS2) and 2009 (JACS3).

1.2 Problems with JACS

Although JACS is well embedded in HE systems and is very familiar to many users, there remain a number of problems with the system. The coding structure - based on letters and numbers - imposes arbitrary limits on the number of codes that can exist at each point in the hierarchy; this results in different subjects being classified at different depths in the hierarchy. It also results in some odd subject codes such as History by period (V100), History by area (V200) and History by topic (V300) which merely act as nodal points under which different subjects can sit.

Although JACS is utilised in HESA and UCAS systems, there are many data collectors for whom JACS codes do not provide the definitions they require and they therefore take different approaches to coding subjects in their data. If we are to create a truly sector-wide data language, something other than JACS3 needs to be developed.

1.3 Why not JACS4?

The limits on the JACS coding structure, combined with the number of changes that have been made since the initial implementation, mean that further development of the JACS system becomes increasingly difficult; there are diminishing numbers of unused codes in the hierarchy and they are not evenly distributed across the subject areas or across the different levels. When the JACS3 project concluded, the project team explicitly stated that the coding structure was now such a constraint that no further development in the existing structure should be undertaken; it is time for a fresh approach.

2. The Subject Coding 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 (JACS) to meet the needs of a broad group of stakeholders and reflect the diverse and dynamic nature of Higher Education in the twenty-first century. The project was undertaken by the Centre for Educational Technology, Interoperability and Standards (Cetis) with partners APS Ltd and Aspire Ltd.

The project commenced in June 2014 and was divided into two stages with the initial stage focusing on requirements gathering and the second stage focusing on development of the scheme and recommendations for governance and implementation. The project approach is outlined in more detail in Appendix A.

3. Requirements for a new coding scheme

During Stage 1 the sector was consulted on the requirements for a new coding scheme. The response to the consultation suggested that, provided a case for change could be stated clearly, it was desirable to introduce a new subject coding scheme. Forty-nine requirements were captured and these were collated and classified into design goals. The design goals were statements of desirable characteristics that defined the design space for the new coding system, and were arrived at by clustering requirements that shared a common theme and/or had close dependencies on each other. The design goals were used to guide the development and evaluation of the prototypes and were as follows:

1. Support policy implementation
2. A unified approach to support a wide group of stakeholders
3. A means of linking to data classified in other frameworks
4. The subject coding system should be simple to understand
5. Be robust
6. Be stable
7. Backwards compatibility with legacy data
8. An appropriate level of granularity (in relation to usability and meaningful data)
9. Provide comprehensive coverage of the range of subjects of study available in HE at an appropriate level of detail for its target users
10. Increased consistency of application across institutions
11. Be clearly separate conceptually from the JACS system
12. A framework with scope for evolution
13. Have a broadly consistent level of detail
14. Recognition of the perceived need for a discipline-based system
15. Have regard for the specific recommendations in the 'roadmap' about disciplines, levels and structure.

Two prototypes were developed, one a conservative evolution of JACS3, the other a more radical flat list of terms. Stage 1 recommended that both prototypes were taken forward with a view to converging them during Stage 2.

4. The new coding scheme – HECoS

4.1 HECoS development and characteristics

After the requirements gathering of Stage 1 of the Subject Coding project, two prototypes were merged into one candidate scheme, termed the Higher Education Classification of Subjects (HECoS). HECoS was hosted on a vocabulary development site, providing stakeholders with the ability to comment on each term. Through a combination of targeted solicitation of specialists, and wider publicity drives:

- Over three hundred comments were posted on specific HECoS vocabulary terms.
- Forty-three people left sixty-four messages on the comment site relating to all aspects of HECoS.
- Thirty-one email responses were also received from participants following a webinar organised by Jisc.

The resulting coding scheme has the following characteristics:

- A flat list, presented with a navigation hierarchy
- Larger than earlier prototypes but smaller than JACS3
- Focused on subjects of study rather than disciplines as the main entity for classification
- Focused on courses and modules as the main entities for classification
- Six character random numeric codes with no leading zeros; the formal identifiers for the terms are URIs
- Open licensed
- Contains related terms and non-preferred terms to aid classification

One important recommendation from the initial coding system roadmap (A roadmap to a new Joint Academic Coding System, Ferrell, 2013) was to use persistent Universal Resource Identifiers (URIs) as identifiers for terms in the new vocabulary, and to maintain such authoritative URIs in a web service for the sector. A URI is a globally unique and unambiguous identifier that can be used in any system without fear of conflicting with other identifiers. Currently HECoS uses temporary URIs until a web service on a permanent domain (HESA) can be established.

In developing the HECoS vocabulary, distinguishability was a key criterion. For a term to be included it had to:

1. Be part of JACS3 and
 - a. Have valid courses or modules coded with it, as evident in HESA returns.
 - b. Have a definition and scope that is sufficient and comprehensive to allow classification.
 - c. Be reliably distinguishable from other terms, as evident in the set of courses or modules classified with the term in HESA returns.Or:
2. Be required to fill a gap in JACS3 as evident in:
 - a. The overloading of a JACS3 code in HESA returns with too many disparate courses and modules.
 - b. Clearly identifiable clusters of course and module titles and descriptions in HESA returns for which no single JACS3 code exist.
 - c. Suggestions by subject matter experts, with supporting evidence from any source.

4.2 Benefits of HECoS

Included below are five aspects of HECoS that we expect will deliver benefits to the HE sector. The specific benefits are highlighted in bold.

1. HECoS is a flat list structure with randomly assigned codes; this will **mitigate against the limit of the coding framework being reached**.
2. HECoS focuses on the distinguishability of subjects. It is anticipated that this will **reduce inconsistent course coding**.
3. Implementation of HECoS will involve training, and resources and guidance will be available for ongoing good practice. It is anticipated that this will **reduce imprecise course and module coding**. Analysis has shown examples of modules and courses that could be coded at a greater level of detail to greater reflect the course or module being coded.
4. The move away from a hierarchical structure should encourage the practice of searching for appropriate codes rather than applying known codes or looking for codes where you would expect to find them in the hierarchy. It is anticipated that this will **improve course and module coding accuracy**. Furthermore, the project undertook a sector-wide consultation to identify codes to be included in the scheme and identify changes to existing codes; this should also improve accuracy.
5. Recommendation for a common approach to subject-based analysis using HECoS has been developed. It is anticipated that will **remove unnecessarily inconsistent subject-based analyses**.

These benefits will all contribute to improving course and module data in the HE sector. This will result in improved information for prospective students who are exploring which HE pathway to follow.

5. Implementation

5.1 Management and oversight

Implementation of HECoS and the ongoing use of HECoS requires management and oversight for a number of reasons. Initially adoption of HECoS will need to be overseen and coordinated across a number of organisations ensuring that the timing of adoption does not compromise processes integral to an organisation's operations. Secondly the ongoing use of HECoS will require management and oversight in order to manage changes to the framework, ensure consistent application, and monitor and realise benefits.

HECoS will be governed by the HE Data Governance body as outlined in the New Landscape report. This body will not report to any other body or group in the sector in order to be independent, but will be hosted by HESA.

The HE Data Landscape Management Board will provide strategic direction for HECoS. The board will be composed equally of data collectors and data providers and chaired by a representative from a HEP.

The HE Data Landscape Management Office at HESA will undertake the day-to-day management of the coding scheme, collate candidate terms, monitor usage and make recommendations to the Board. The Technical Group will have the requisite technical skills to oversee the management of HECoS.

A HECoS Consultative group could be established to provide opportunities for wider stakeholder engagement.

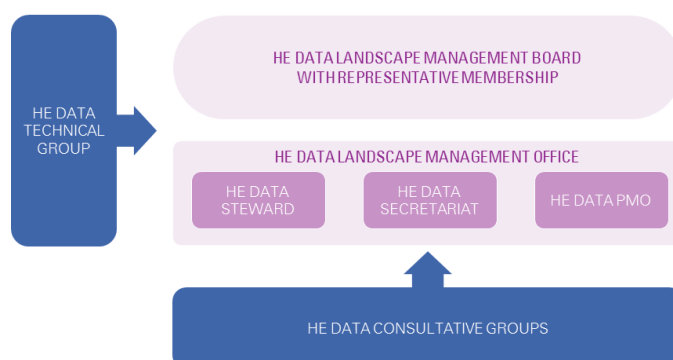


Figure 1: Governance of the New Landscape (from the New Landscape report)

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. It is however recognised that it may be necessary to address exceptional or unforeseen issues that arise during the adoption phase. A SUBSTANTIVE change requires the authorisation of the Management Board and includes proposal, acceptance and implementation of new candidate terms, deprecation of obsolete terms and changes to the definition of terms.

During the two year introduction period ADMINISTRATIVE changes may be made. ADMINISTRATIVE changes are undertaken by the Management Office and include clarification of definitions and scope notes, and changes to the navigation structure.

After the first two years it is recommended that the Management Office should undertake an annual review of the coding scheme.

In order to allow HECoS to evolve it is recommended that a Candidate Terms Registry is set up to enable users to submit new terms for proposed inclusion in the coding scheme. This should be run by the Management Office.

5.2 Subject Based Analysis

Subject Based Analysis refers to the analysis of data where subject of study is explicitly recorded, and for which statistics are reported according to groupings of HECoS terms.

HECoS is designed to work with many hierarchies by separating the terms from the way they are aggregated. This allows HECoS to fulfil more diverse functions but it also means that one widely agreed analytical hierarchy is essential to maintain comparability of analyses. The project therefore sought, through engagement with the main data collectors, a broadly supported common approach to subject-based analysis using HECoS. The project also considered the prospect of text mining.

The project made the following recommendations:

- Base a set of standard cross-sector aggregation rules on the KIS/Unistats subject groups and with apportionment for major/minor and balanced studies.
- Integrate governance of the standard aggregation rules with governance of the HECoS classification scheme under the aegis of the body recommended in the New Landscape report.
- Ensure that subject group definitions, which may be owned by individual Core Sector Bodies or government bodies, and maintained by them in addition to the standard set, are effectively disseminated alongside HECoS. These include: HEFCE SIVS, HEFCW ASCs, mapping to QAA subject benchmarks, FCO ATAS.
- Provide support to these bodies by drafting subject group rules for them as part of the adoption process, in the interest of consistency and efficiency.
- Promote a small pilot and encourage experimentation, isolated from impact on business processes, to build evidence for the value of text mining for management and administration.

5.3 Next steps

The immediate next steps will be to:

1. Agree with stakeholders an implementation timescale
2. Develop supporting materials (including mappings) to aid organisations with adoption of the new scheme.

The full Adoption Plan outlines a number of implementation tasks identified by stakeholders during the consultation process. These tasks cover the following broad areas:

- Dependency analysis
- Communications and awareness raising
- Developing standard approaches to subject-based analyses
- Creating information resources and training materials
- Establishing the management service for HECoS

APPENDIX A

Subject Coding project approach

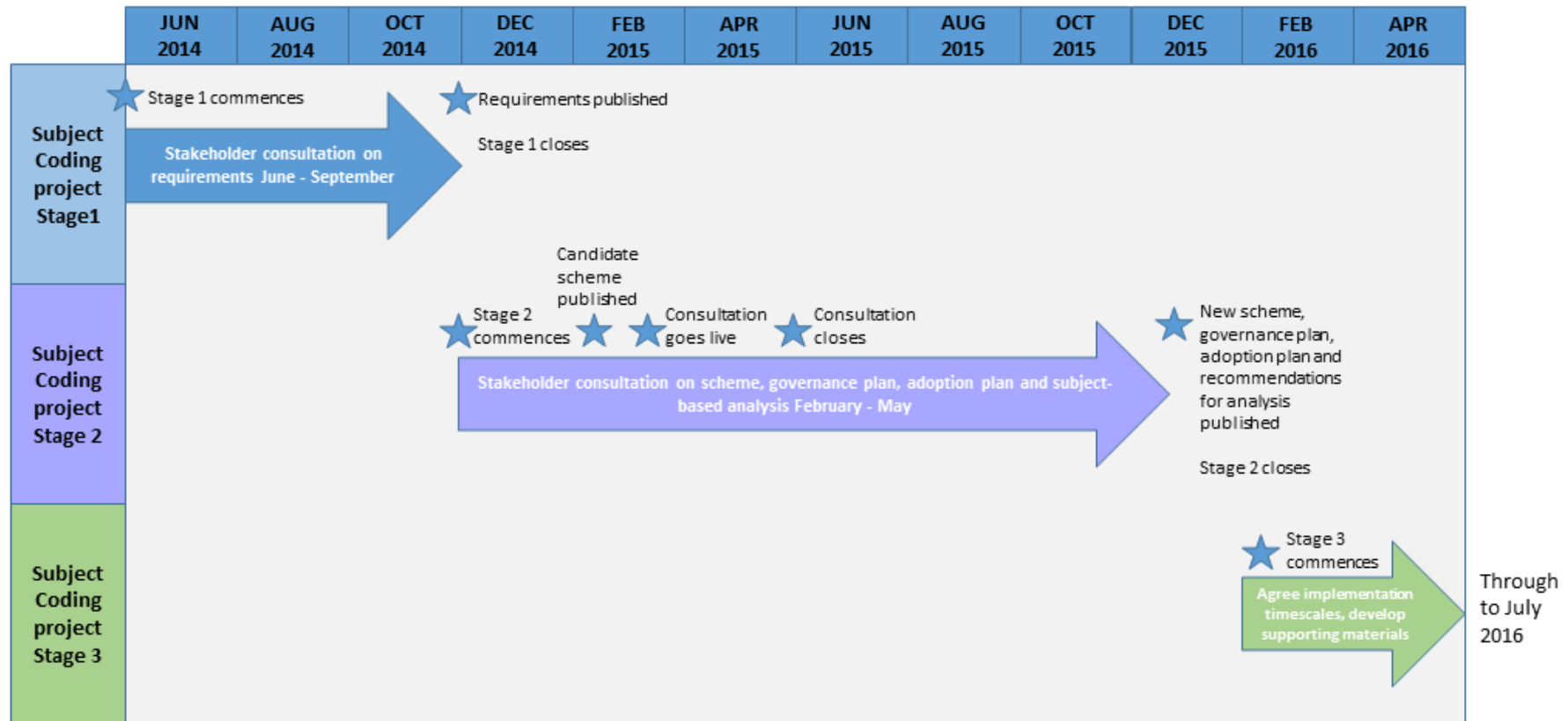


Figure 2: Project timeline

APPENDIX B

Subject Coding project Stage 1 and Stage 2 full reports

Stage 1

New Subject Coding System, Impact Assessment and Requirements Definition, November 2014

http://www.hediip.ac.uk/wp-content/uploads/Stage_1_report_2014-11-13.pdf

Stage 2

New Subject Coding System, Structure and Candidate Scheme, February 2015

http://hediip.ac.uk/wp-content/uploads/HEDIIP_NSCS_PD03_2015-02-18b.pdf

The Higher Education Classification of Subjects (HECoS) vocabulary (report), December 2015

http://www.hediip.ac.uk/wp-content/uploads/HEDIIP_NSCS_PD04_Scheme_2015-12-15.pdf

The Higher Education Classification of Subjects (HECoS) vocabulary (coding scheme), December 2015

http://www.hediip.ac.uk/wp-content/uploads/HEDIIP_NSCS_PD04_HECoS_Vocabulary_2015-12-15.xlsx

New Subject Coding System, Recommendations for Subject Based Analysis and Text Mining, December 2015

http://www.hediip.ac.uk/wp-content/uploads/HEDIIP_NSCS_PD05_SubjectAnalysis_2015-12-15.pdf

New Subject Coding System, Governance Model, December 2015

http://www.hediip.ac.uk/wp-content/uploads/HEDIIP_NSCS_PD06_GovernanceModel_2015-12-15.pdf

New Subject Coding System, Adoption Plan, December 2015

http://www.hediip.ac.uk/wp-content/uploads/HEDIIP_NSCS_PD07_AdoptionPlan_2015-12-15.pdf

References

Classifying subject of study, a roadmap to a new Joint Academic Coding System, July 2013

http://www.hediip.ac.uk/wp-content/uploads/JACS_Report_2013-07.pdf

The Blueprint for a New HE Data Landscape, May 2015

http://www.hediip.ac.uk/wp-content/uploads/HEDIIP_Data_Landscape_Report.pdf