



New Subject Coding Scheme

Impact Assessment and Requirements Definition



About the New Subject Coding System project

The project was commissioned by HEDIIP under the Standards and Understanding theme. This first stage was delivered by the Centre for Educational Technology, Interoperability and Standards (Cetis) at the University of Bolton. The second stage of the project is now underway and due for completion in 2015.

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

Andy Youell, Director, HEDIIP
Andrew Horsman, Head of Data Management, HESA
Dr Christine Couper, Director of Strategic Planning, Greenwich University
Hannah Falvey, Head of Statistics, HEFCW
Lesley Donnithorne, HR Manager (Systems, Information and Grading), UWE Bristol
Simon Robshaw, Data Analyst, UCAS
Paul Baron, Programme Manager, HEDIIP
Jenni Cockram, Programme Officer, HEDIIP

Principal Authors/Editors: Wilbert Kraan, Alan Paull

Contributors: Charlie Paull, Lorna Campbell, Gill Ferrell, Adam Cooper, Phil Barker

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 the New Subject Coding System project	1
About HEDIIP.....	1
Contents	2
1. Executive summary.....	4
1.1. The vision	4
1.2. The methodology.....	4
1.3. The result	4
1.4. Recommendations.....	5
2. Overview	5
3. Methodology	6
3.1. Phone interviews and site visits.....	6
3.2. Workshops.....	7
3.3. Consultation website	8
3.4. Synthesis	8
4. Initial design questions and examples.....	9
4.1. Subject vocabularies consulted	9
4.2. Hierarchies, flat lists, and how subjects are shared	10
4.3. Disciplines, subjects and degree programmes	12
4.4. How many top terms?	13
4.5. How to deal with subjects not listed	14
4.6. What should the NSCS codes look like?.....	15
4.7. Initial design features	15
5. Design goals, use cases and requirements	16
5.1. Support policy implementation	17
5.2. A unified approach to support a wide group of stakeholders.....	18
5.3. A means of linking to data classified in other frameworks.	19
5.4. The NSCS should be simple to understand.....	20
5.5. Be robust.....	20
5.6. Be stable	21
5.7. Backwards compatibility with legacy data.	21
5.8. An appropriate level of granularity (in relation to usability and meaningful data).....	22
5.9. Provide comprehensive coverage of the range of subjects of study available in HE at an appropriate level of detail for its target users	22
5.10. Increased consistency of application across institutions.....	23
5.11. Be clearly separate conceptually from the JACS system	24
5.12. A framework with scope for evolution	24
5.13. Have a broadly consistent level of detail.....	25
5.14. The perceived need for a discipline-based system.....	25
5.15. Have regard for the specific recommendations in the 'roadmap' about disciplines, levels and structure...25	25

6.	Areas of conflict, impact risks and their remediation.....	26
6.1.	Comprehensive detail for data analysts versus minimal coding effort for data providers	26
6.2.	Agility versus stability	26
6.3.	Supporting different purposes in a single scheme	27
7.	The prototypes	27
7.1.	Prototype 1	27
7.2.	Prototype 2	32
8.	Critical success factors for the new scheme.....	36
9.	Next steps	37
10.	References.....	37
	Appendices.....	39
1	Stakeholders contacted	39
2	HEDIIP partnership NSCS design principles	41
3	NSCS Benefits Map	42
4	Interview Protocol	43
5	HEP Workshop Plan	45
6	Full Requirements Matrix	47

1. Executive summary

1.1. The vision

A national subject coding scheme for higher education touches many stakeholders, because it is pivotal within and between many datasets. JACS has fulfilled that role so far, but informal feedback and formal reports have indicated that JACS3 no longer fulfils that role adequately because, as the **Classifying subject of study** report (Ferrell, 2013) noted:

- the limit of the existing coding framework has been reached;
- changes and growth in JACS' range of functions mean it is no longer consistently applied;
- it does not meet the needs of all of the key sector stakeholders;
- Higher education providers (HEPs) sometimes use JACS3 in inconsistent ways;
- incomplete and misunderstood JACS3 terminology and definitions have led to poor quality data in some instances.

There is now an opportunity to develop and introduce a completely new scheme. Most of the work done by the New Subject Coding Scheme (NSCS) project in its first phase has consisted of: consulting many parties; gathering requirements; and trying to gauge the impact of a vocabulary whose shape at this stage remains completely open.

1.2. The methodology

The project team tackled these aspects by designing a consultation process to engage many, whilst optimising the time and depth of engagement with each. It did this by posing a number of initial design questions about basic features of the proposed new scheme. These questions were illustrated with examples, and were designed as much to elicit requirements and gauge impact on stakeholders' processes and systems as they were used to steer the development of early prototypes ("straw men"). This way, requirements and use cases could be gathered at the same time as the impact of potential solutions was gauged. Not only that, the examples that came with the initial questions were collaboratively adjusted to reduce impact and meet emerging requirements through the consultation period.

1.3. The result

The result of this approach was a wealth of responses gathered from a wide range of interested parties in a short space of time. While there are clear and unsurprising tensions within the feedback, no insurmountable differences emerged. The main tensions found in the responses are the need to:

1. provide comprehensive detail for data analysts whilst requiring minimal coding effort for data suppliers;
2. enabling agility whilst ensuring stability;
3. support many purposes in a single, simple scheme.

The necessity of balancing these different interests and needs has steered the development of two early prototype schemes, which will be reduced to one prototype in the next phase of the project. In order to provide the required stability, both prototypes take JACS3 as a starting point, but differ in their structure. One prototype has multiple levels and is comprehensive, the other has a single

level, and is designed to be used with other schemes. Each prototype addresses the use cases and requirements that were gathered, but in a different way and to a different extent. Both can also accommodate complementary schemes, which means that the new subject coding scheme will function as both a vocabulary and a framework.

Managing the tensions also informed plans for the future development of an appropriate governance model for the new scheme. There is an emerging consensus that no new organisation is required, but ensuring that the governing body is representative is important. Managing consensus about the pace of change in the new coding scheme will be the most important task of the governing body. The body will also be crucial in ensuring buy-in from both data consumers and suppliers: the most critical success factor for the new scheme.

1.4. Recommendations

The response of stakeholders to the consultation of stage 1 of the NSCS suggests that, provided a case for change can be stated clearly, it is desirable to introduce a new subject coding scheme. In order to progress that development, we recommend that:

- both prototypes be developed further with an expectation to converge them into a single prototype, depending on further feedback;
- a new governance model be developed under the auspices of an existing sector organisation, with broad representation;
- a subject coding framework be developed alongside a specific core scheme.

2. Overview

The report will first outline the consultation methodology and provide an account of who was contacted and how. This will be followed by a description of the development of the initial design questions, as well as the impact they had on various stakeholders.

The requirements derived from the consultation exercise will then be considered. Because there are forty-four of these, they will be presented together as part of the broader design goals that they inform. Similar to design principles, the design goals define the design space in which the straw men and later prototypes have been developed. Two prototypes will be described in terms of how they arose from the initial questions and examples, and how they address each design goal.

Areas of conflict and impact risks will be summarised and considered separately before we end with a consideration of critical success factors and next steps. An amended benefit case for the new scheme is available in appendix 3.

3. Methodology

With many and varied stakeholders, and just three summer months to contact them all, an approach was developed that maximised engagement with each stakeholder, proportional to the degree of interest that parties were likely to have in the new coding scheme. The allocation of sector bodies to these activities was done in consultation with the HEDIIP Programme Management Office (PMO), and worked broadly as outlined in Table 1. A precise breakdown of who was consulted in what way is available in appendix 1.

Table 1 Data collection activities and stakeholders

	Workshops	Phone interviews	Meetings	Comment web site
Core sector bodies such as UCAS			√	√
Other sector bodies such as the QAA		√		√
Professional, Statutory and Regulatory Bodies such as the GMC		√		√
Higher Education providers (HEPs) including Universities	√			√
Any other interested parties	√			√

Common to all these consultation activities was the set of initial design questions and examples (see section 4), and these were developed iteratively as feedback was received, and the consultation activities progressed.

In addition to a wide range of engagement activities, the project team's ability to divide work among six team members ensured optimal flexibility when planning meetings and consultations and helped to maximise the feedback gathered. Nevertheless it proved challenging to engage with some organisations within the constraints of Stage 1. As a result, some feedback has been integrated either very late in the day, or after the consultation period officially closed.

3.1. Phone interviews and site visits

A targeted approach was taken to engaging stakeholders who depend on subject coding in their central business processes so that the team could fully understand their needs. Each stakeholder was consulted through a semi-structured interview at a face to face meeting, or if that wasn't possible, via telephone. Stakeholders also had an opportunity to provide a written submission. Half day consultations were allocated to the ten core sector bodies with complex requirements for the NSCS, with about an hour each for the others.

The interview protocol (see appendix 4) covered current use cases, future requirements, impact and governance, but with room to diverge and shift focus depending on the nature of the sector body. The initial design questions and examples were sent to the interviewees in advance, along with a letter outlining the scope of the project and the interview process.

Reports were made of these bodies' written and oral submissions, and sent back to them for any corrections or additions.

3.2. Workshops

The workshops were one day invitation events that were designed to engage with higher education providers in general, and staff with a responsibility for planning, data returns, marketing or IT system management in particular. In order to facilitate attendance by people from a range of institutions, workshops were held in southern and northern England and central Scotland. The HE provider attendees came from a range of different institutions that varied from new universities to Russell Group institutions, and large all-round universities to small, specialist ones. A complete overview of the attendees is available in appendix 1.

The agenda for each of the workshops was the same (see appendix 5) and was designed to elicit criticism of JACS3 at the beginning, in order to clear the way for focus on the NSCS later in the day. The rest of the day was divided up into a user story stage to elicit requirements, and an activity where attendees picked preferred solutions to the initial design questions and a final discussion to elicit feedback around issues such as preferred governance models.

In the event, the initial exercise had to be expanded, because people were not convinced of the business case for changing JACS3 and wanted to list what they thought was good about JACS3 as well as what they didn't like about it. In one case, the user stories exercise had to be abandoned to make room for an extensive discussion of the prioritisation of JACS3 replacement by HEDIIP. Clearly, for these stakeholders, making the business case for the NSCS will need to be a priority in stage 2 of the project. In the other workshop events, the user stories activity went ahead, and did unearth valuable requirements and insights, and the initial design questions were answered in small groups by all.

In terms of background, twenty-one of the thirty-one HE provider attendees had a role in planning or other student data management related jobs, five had an IT role and four a marketing and admissions responsibility (see Figure 1). One person had an HR role. This appears to reflect reasonably accurately how responsibility for subject coding and data reporting is apportioned in institutions, and, as such, we received responses from those who oversee the creation of the source course coding data that is ultimately used by all other stakeholders.

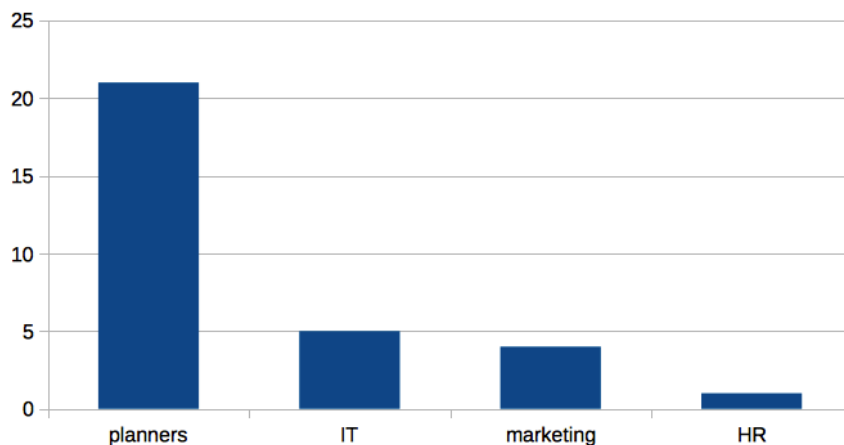


Figure 1 Workshop attendees

3.3. Consultation website

The consultation website enabled the widest possible set of interested parties to express the needs of their organisations and the likely impact of change, as well as emphasising to the sector at large that the NSCS project will operate with transparency. The content of the site reflects current thinking on the use, impact and adoption, governance and design of the NSCS. Each section is open for anyone to comment, subject to spam controls.

In terms of informing the sector, the site has garnered a total of about 3,000 views by around 670 unique visitors, very nearly all whom were in the UK. Other than the front page, viewers have primarily been interested in the Design section, which is perhaps to be expected, since this will be the area that produces the revised scheme that will affect all users. The majority of referrals to the comment site came from the HEDIIP website, with smaller numbers coming in via Twitter, the Cetus website and UCAS and HESPA mailing lists and forums.

Five of these visitors left eleven comments, all from HEPs, and all broadly in line with comments from the workshops. While it is difficult to compare a comment site of this nature with other sites, the amount of responses per unique visitor seems relatively high compared to general blogs.

3.4. Synthesis

The outcome of the various consultation activities was a combination of formally and informally stated requirements from interviews on the one hand, and discussions and votes on the initial design questions and examples on the other. This material contained a variety of types of information: use cases, user stories, desirable scheme features, governance principles, data modelling preferences and much more.

In order to present this information and draw conclusions from it, requirements were extracted and collated from the reports of the meetings and phone interview activities. These were then grouped into more abstract design principles that summarise these requirements and define the design space within which the NSCS will need to be developed.

Each initial design question was accompanied by a small number of proposed solutions, illustrated by examples, for stakeholders to choose from. They could also propose alternative solutions. During the workshops, small groups discussed the questions and picked a favoured solution. Some of the sector bodies that were interviewed and the respondents on the comments site also followed this process. These preferences have been collated and will be presented with an examination of the initial design discussions in the next section.

4. Initial design questions and examples

As it was likely to be difficult to elicit meaningful requirements and impact assessments from stakeholders about a new subject coding scheme that has not yet been developed, the project team formulated a number of initial design questions that dealt with fundamental decisions about the structure of the NSCS, along with possible solutions that were illustrated with examples. These design questions and potential solutions have been continuously re-formulated in light of feedback, and the examples formed the basis of the prototypes described later in this report (see section 7).

The current form of the design questions and examples is one of the main outcomes of the exercise, because they were developed iteratively and collaboratively throughout the consultation period. For that reason, the stories of how the initial questions and examples were developed are intermingled with the responses they elicited, and an analysis of the outcome given.

4.1. Subject vocabularies consulted

In the process of designing the initial questions and examples, many existing vocabularies were consulted. Though some of the initial examples are rooted in existing JACS3 terms, neither the questions nor the examples are straightforward adaptations of existing vocabularies. The initial questions and examples did, however, take a steer from the ideas embodied in the structure of these other subject vocabularies. The list consulted included:

- JACS3 (HESA - Higher Education Statistics Agency, 2012).
- HE Academy discipline areas (The Higher Education Academy, 2014).
- QAA subject benchmarks (QAA, 2014).
- RCUK Research Classification (Research Councils UK, n.d.).
- Elsevier's list of academic subjects (Elsevier, n.d.).
- Wikipedia list of academic disciplines and sub-disciplines ("List of academic disciplines and sub-disciplines," 2014).
- Classification of Instructional Programs (CIP 2010 and earlier) from the U.S. Department of Education (U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, & U.S. Department of Education, 2014).
- Australian Standard Classification of Education (ASCED) (Australian Bureau of Statistics, n.d.).
- Australian and New Zealand Standard Research Classification (ANZSRC) (Australian Research Council, n.d.).
- Medical Subject Headings (MeSH) (U.S. National Library of Medicine, n.d.).
- Learn Direct Classification System (LDCS).
- Sector Subject Area (SSA) classification system (Learning and Skills Council, n.d.).
- NHS Occupation Codes (Health and Social Care Information Centre, n.d.)
- Organisation for Economic Co-operation and Development (OECD) Field of Science (FOS) Classification (Organisation for Economic Co-operation and Development, 2007).
- UNESCO International Standard Classification of Education (ISCED) (UNESCO, 1997).

The HE Academy, RCUK and QAA lists provide richer views of subject groupings than are likely to be familiar to many stakeholders, and they also provide a useful indication of the optimal number of top level terms in a subject hierarchy. The issue was developed into an initial design question (see section 4). The Elsevier and Wikipedia lists give an impression of the size, depth and major groupings of subject classifications that are designed for wide audiences. The CIP 2010 provides a

view of the size and structure of a mature classification scheme for degree programmes, as well as an idea about the modularity of levels. The ASCED and ANZSRC provide good examples of subject classification schemes that are formally rigorous, comprehensive and wide in scope, but perhaps not designed for wide audiences. The OECD FOS and ISCED classifications offer a view of subject vocabularies independent of national requirements. The LDCS and SSA schemes are examples of UK-based educational schemes in current operation. The NHS Occupation Codes scheme provides a specific example of a target for mapping.

In future NSCS work, mappings to some of these vocabularies will need to be explored and developed.

4.2. Hierarchies, flat lists, and how subjects are shared

Development

In its original form, this question was intended to elicit responses about how or whether to share lower level subjects such as “Social policy and administration” between two or more higher level groupings such as “Social sciences” and “Health and social care” rather than just one. Proposed solutions included:

- A. unique subjects, where no subjects are shared at all;
- B. shared subjects, where a subject can be an equal part of more than one subject hierarchy;
- C. synonyms, where a placeholder in one hierarchy points to a “real” subject in another hierarchy;
- D. linked subjects, where two or more subjects in multiple hierarchies could be considered equivalent under certain criteria. This was later changed to a flat list (see below).

Within the first workshop, a more radical version of 4.2.D was proposed by one of the HEP planners: a flat list of subjects. To understand the connection between the earlier and later version of 4.2.D, it is necessary to consider the nature of the relations between the subjects. Under 4.2.A through C, the relations – whether hierarchical or not – are fixed. Under D, that is relaxed; depending on the purpose that the scheme is being used for, one set or another set of relations obtains. The final version of D takes that one step further; nearly all relations¹ between the subjects in the scheme depend on the purpose, and there is no limit to how many such relations there can be. That is, once subjects in the scheme are considered as terms that relate to each other conditionally – as they would be in the original 4.2.D – it becomes possible to think of a scheme where the terms are agreed and fixed in number and meaning, but the relations between them could be varied according to the needs of different stakeholders – as they would be in the later version of 4.2.D. In subsequent iterations of the workshop, the comment site and other documentation, the flat list solution was substituted for the linked subject solution as option 4.2.D: one finite set of subjects which are all at broadly the same level of granularity, with no defined hierarchy. On top of such a shared flat list, different hierarchies can be composed for specific purposes by particular sector bodies. This idea has been developed with respondents in subsequent consultation events, and is most clearly embodied in Prototype 2 (see section 7.2).

¹ The subjects in the later version of 4.2.D would still have one relation to each other that isn't dependent on purpose, and that is the fact that they all belong to the same scheme.

Response

In all three workshops, 4.2.D was the preferred option by all but one of the subgroups. At least three sector bodies –two core– preferred it, or expressed themselves open to it. There were no categorical objections to it from anyone, even if some could see drawbacks. Reasons cited for preferring the ‘flat list’ solution include:

- simplicity and ease of use because the coding scheme is no more complex than it needs to be;
- greater political neutrality because how subjects are aggregated becomes an explicit policy decision, which should reduce the incentive for strategic coding;
- greater flexibility as more granular taxonomies for specific purposes can be attached ‘underneath’ the common flat list;
- easier and more robust coding over time, because subjects do not have to be aggregated at the same time as the subjects are coded;
- it’s the subjects that matter, the structure in which they sit is secondary;
- a flat list formalises actual current practice and eliminates an appearance of precision the data can’t bear.

Reservations about option 4.2.D voiced by some include:

- the potential lack of control - not knowing what the consequences of coding decisions are because the aggregation and the associated policies might change later;
- different purposes might require different flat lists;
- may require significant system redevelopment;
- may be perceived as a radical solution and therefore more difficult to “sell” to the sector.

There were a few who were open to solutions 4.2.A and 4.2.B, or preferred them, but as many disliked them or thought they couldn’t work. Solution 4.2.A – no shared subjects – was preferred as well as disliked because it resembles JACS; familiarity versus known coding limitations. A polyhierarchy like solution 4.2.B was used internally by one sector body, but was thought too complex and potentially inflexible by others.

Analysis

At the moment, a flat list is the preferred option among the majority of stakeholders consulted. Reservations about it are few, and can be addressed: for HEPs, the control that some fear they might lose via coding strategically is a consequence of separating policy from the scheme structure. If subject aggregations and policy decisions about them are openly discussed within the sector, this fear could be addressed effectively.

The point about different purposes requiring different lists is pertinent. To some extent this can be addressed by having specialised vocabularies at different levels which can be articulated from a common flat list. This does, however, beg the question which purpose the common list should address - this issue will be dealt with in the next initial question. Also, HEFCE and HESA have mentioned the growing importance for finer detail from across the sector, which may or may not be satisfied by auxiliary schemes.

Finally, the fear that a flat list could mean a costly system redevelopment was raised by a core sector body representative who was also “a fan” of the flat list solution. Nonetheless, this impact aspect needs close monitoring in stage 2 of the project. It is possible that other considerations could still favour a hierarchical common scheme solution.

4.3. Disciplines, subjects and degree programmes

Development

This question was designed to elicit feedback about an issue raised in the **Classifying subject of study** report (Ferrell, 2013): whether a NSCS should classify subjects, disciplines or something else. The problem was that JACS, as the name implies, classifies subjects, while many stakeholders use it as if it classifies disciplines. For the purposes of the question, the discipline and subject were defined as follows:

Disciplines can be characterised as having:

- a defined approach (shared methods and research questions);
- a label that is different from the subject of study/research;
- a history as a human practice;
- an associated profession.

Examples include: Medicine and Dentistry; Linguistics; Physics, Astronomy and Astrophysics.

Subjects, on the other hand, are characterised by:

- being concepts independent of how they've been studied ;
- being studied by multiple disciplines;
- having little or no overlap with adjacent subjects (i.e. are defined in contra-distinction).

Examples include: medicine, French language, physics.

Since academics might identify more strongly and reliably with the community they operate in (discipline) than what that community studies (subject), discipline might be a more reliable category than subject. Initially, the solutions were formulated as:

- A. subject;
- B. discipline;
- C. four distinct levels: branch, discipline, subject and field.

In the light of the addition of the flat list (solution 4.2.D), the list of solutions was changed to:

- A. subject hierarchy;
- B. subject list;
- C. discipline hierarchy;
- D. discipline list;
- E. four distinct levels: branch, discipline, subject and field.

Response

In the workshops, all HEP participants preferred subject, sometimes strongly. Because of the general preference for flat lists, solution 4.3.B was most popular, and 4.3.E thought to be a variant that could be achieved by having several vocabularies, where needed. Several Professional, Statutory and Regulatory Bodies (PSRBs) noted that they effectively represented disciplines, but few thought that disciplines ought to be the main structuring concept of the NSCS. Being able to identify relevant subjects, and whether a degree programme is accredited or not were felt to be much higher priorities.

The preference for subjects mainly took the form of objections to using disciplines:

- disciplines are not as stable as subjects;
- disciplines and their boundaries can be controversial;
- given the increasing popularity of combined studies, disciplines are less relevant;
- while it's possible to impose one or more discipline structures on a list of subjects, the obverse is not so easy.

Analysis

The subject versus discipline question did prompt two further questions. One was the relation of disciplines to cost centres, with some HEP representatives arguing that they ought to be the same. Many others, including some sector bodies, noted that the relation between discipline and cost centre is quite complex in general, and several HEP staff noted that they'd struggle to align them. Some sector bodies such as HEFCW rarely used them as a result.

The other question concerned the precise level of granularity of subjects: is it broadly at module or degree programme level? Several respondents voiced support for the idea that it should be at module level, mostly because that's the level at which funding differences emerge, and because degree programmes can be derived from the modules that make them up, but not the other way around. On the other hand, all but one HEP currently code to JACS3 Principal Subject level only, and several made it clear they had no intention of coding modules with greater precision unless there was a significant new incentive. Whether by necessity or design, several sector bodies also indicated that they expected data to be classified at programme level, even if more detailed breakdowns could be interesting for a variety of reasons. Some also expressed the fear that requiring HEPs to provide more detail could result in increasingly poor and unreliable data.

4.4. How many top terms?

Development

This question was initially included in order to elicit opinions on how many top level terms in a hierarchical scheme would make it easier to use. The options originally listed were:

- A. Five or less;
- B. Between six and twelve;
- C. More than twelve terms.

In the event, because of the strong preference expressed for a flat list (solution 4.2.D.), the question was changed to "how many terms overall", and subsequently shelved, since that question is more appropriately posed and answered after further development of the NSCS prototypes.

Response

During the discussion of the question, the HEP representatives argued, sometimes vehemently, for both a total of "fifty codes or fewer" and "not one hundred and fifty; fewer than that". It was noted that if the NSCS contained a list of a hundred and fifty terms overall, each individual HEP would likely have to deal with no more than about fifty terms. There was consensus, however, about the need to "use evidence" for any list of terms the NSCS could contain, with current HESA return practice being one source for that.

Analysis

From the limited discussion of the question, it was possible to obtain a useful idea of the number of terms that HEPs would be happy with, and what would convince them of an appropriate number. In view of the factors that have a bearing on the total number of NSCS terms, such as desired level of granularity and the needs of multiple sector bodies, the optimal number will be best determined once more input has been received and more complete prototypes can be demonstrated.

4.5. How to deal with subjects not listed

Development

A fundamental question every vocabulary needs to deal with is how to treat exemplars for which no term currently exists. The solutions and examples suggested for this issue were:

- A. classify at the broader term - new subjects are classified at the nearest appropriate broader term;
- B. explicit “subject x not otherwise classified” code - as is currently used in JACS3;
- C. allow users to create their own codes - and use a to-be-determined mechanism to share new codes and/or have them added to NSCS.

Response

Solution 4.5.B was not popular, with some complaining that datasets with such classifications can not be effectively analysed. A representative of a stakeholder with a subject vocabulary of its own thought it a “really lazy cop-out”. Solution 4.5.A was acceptable to most, bearing in mind that a new term would have to be modelled as a narrower version of an existing term in the flat list solution of 4.2.D. There were two stakeholders who used the 4.5.A solution in similar vocabularies of their own. Most of the discussion, however, centred on solution 4.5.C.

Many HEP planners did not favour the idea of a NSCS that could rapidly change and expand, because of the concomitant loss of longitudinal comparability and a need to recode data. Political or current affairs driven requirements were not convincing to them: “[HEPs deal with] an ad hoc question by giving an ad hoc answer”. HEP representatives with a marketing background and one core sector body representative, however, did like the idea that the NSCS could rapidly absorb popular new subjects.

Analysis

Though the tension between stability and agility wasn't quite resolved, one solution that was discussed, and did get some approval, was to have a system whereby HEPs could propose new candidate terms that a representative body could consider for inclusion in a next version of the NSCS, to be released at regular and long intervals. A solution for dealing with re-coding historic datasets is key. In spite of the lack of enthusiasm for the option so far, the 4.5.B option of making explicit “not otherwise classified” categories does have some strengths in classic ‘closed world’ taxonomic practice, and may even be necessary within vocabularies defined in this way.

4.6. What should the NSCS codes look like?

Development

In theory, most stakeholders would only need to see the labels of the NSCS, never the code itself. However, experience with JACS and other coding schemes suggests that choices in the shape of the code have real consequences on their use. The suggested solutions were:

- A. a hierarchy reflected in the codes, much like JACS3;
- B. label-derived codes as a compromise between human readability and code flexibility;
- C. purely random codes;
- D. hierarchy in the codes, with space to add more by making the codes modular.

Response

Solution 4.6.B was not preferred for a variety of reasons, but mostly because meanings are unstable, and because label derived codes can be quite hard to unpack in the case of subjects, thereby negating any usability advantages. There was sympathy for 4.6.C but some feared that it would prove unpopular in practice, and some also thought it could hinder data quality assurance because the codes would be hard to memorise. Solution 4.6.A was used internally by one non-HEP stakeholder, but was not popular with HEP respondents because a non hierarchical list was preferred by them.

Analysis

In the discussion with HEP respondents the outline of a broadly supported version of 4.6.D emerged: one that had a memorable, but not label derived shape, preferably numerical, and that could be augmented with additional codes for specific purposes, such as more fine grained codes 'underneath' the common list terms. People with an IT system management responsibility thought that a fixed number of characters without leading zeros was crucial.

4.7. Initial design features

Taking the full set of responses on all of the initial design questions and examples, and setting aside constraints that are best treated as requirements (see section 5), the broad shape of a NSCS emerges. The features of such a new scheme could include:

- a single, common flat list of terms;
- a number of hierarchies that aggregate terms from the common list developed by stakeholders for specific purposes;
- a number of hierarchies that refine terms from the common list developed by stakeholders for specific purposes;
- common terms codifying between 40 and 150 subjects;
- avoiding 'not otherwise classified' terms;
- terms that have opaque but memorable identifiers of a fixed length.

These features describe the outline of the NSCS as it emerged in consultations so far, particularly the workshops. Though it was broadly supported by most stakeholders in this form, this outline emerged independent of all business requirements that were gathered at the same time, and prototypes of the NSCS may, therefore, not share all of these features.

5. Design goals, use cases and requirements

Though not all stakeholders were able to produce formal requirements or use cases, or had preferences that can be described as such, a number of requirements were implicit and explicit in the various sources of input we received. These sources include:

- the **Classifying subject of study** report (Ferrell, 2013) generally, and the recommendations in particular;
- the design principles developed by HEDIIP (see appendix 2);
- submissions to the comment website;
- the “likes and dislikes about JACS” exercise in the workshops;
- the user story exercise in the workshops;
- the initial questions and examples exercise in the workshops;
- the final discussion from the workshops;
- written submissions from sector bodies, both structured and unstructured by the interview protocol (see appendix 4);
- short phone interviews with sector bodies following the interview protocol;
- face to face meetings or long phone interviews with sector bodies following the interview protocol.

Because use cases were received in such a volume and variety of forms, they have been restated in terms of requirements in order to facilitate knowledge management.

Each of the requirements has been extracted, collated and classified for priority and design goals. The design goals are statements of desirable characteristics that define the design space for the NSCS, and were arrived at by clustering requirements that shared a common theme and/or had close dependencies on each other. In other words, they abstract across the details of a number of similar requirements. Because a requirement can have different facets in common with other different requirements, it can be associated with more than one design goal. Design goals have been used to constrain and guide the development of the prototypes as they emerged from the initial design phase.

The design goals are presented here by rank order of the number of mandatory requirements they cover. That is, the design goals that bundle the highest number of requirements that have been marked as ‘mandatory’ are presented first. The number of non-mandatory requirements is a secondary criterion. That means that the emerging NSCS prototypes will need to meet the design goals below in descending order of importance.

The earlier **Classifying subjects of study** (Ferrell, 2013) report listed a number of recommendations for the NSCS, and the work specification for the NSCS project listed a number of design principles as well (see appendix 2). As noted, these have been used as sources of requirements alongside others, but for ease of evaluation, they will be flagged and discussed with each design goal.

More detail about each of the requirements, such as formal descriptions and which stakeholder type it came from, is available in appendix 6.

5.1. Support policy implementation

The central goal of the NSCS has to be the need to support policy for a number of agencies in areas that include:

- Funding;
- Monitoring of specific subject areas (e.g. STEM, SIV);
- Widening participation;
- Progression;
- Accountability (including QA);
- Information to students;
- Benchmarking and performance indicators.

Clearly, these areas may have conflicting needs, and balancing them could be a challenge.

The specific requirements 5.1 summarises are:

	Mandatory requirements		Other requirements
R1	NSCS and JACS3	R10	Providing guidance on coding for specific purposes
R9	Governing and sector bodies	R12	Comparing degree programmes
R11	Providing training recommendations	R14	Supporting regulated professions
R13	Supporting operational and time series statistics	R16	Supporting multiple aggregation structures
R20	Enabling statutory returns	R21	Supporting student lifecycle comparisons
R28	Providing guidance on the NSCS and KIS	R23	Enabling workforce and capacity planning dataset comparisons
R29	Facilitating datasets that are fit for purpose	R24	Enabling disaggregation in health subjects
R31	Describing guidance purposes clearly	R25	Enabling aggregations for NHS planning
R32	Encouraging clarity in the description of data collection purposes	R33	Supporting aggregation of STEM and SIV course data
R45	Including external definitions of important subjects	R36	Classifying subjects or groups of subjects
R49	Structural position of Welsh as a subject in the scheme	R47	Supporting explicit aggregations of subjects
		R48	Supporting specific HEP functions
		R26	Providing more detail in medicine subjects

5.2. A unified approach to support a wide group of stakeholders

In order to function, the NSCS needs to be widely adopted, and in order to be adopted, it must meet the needs of the sector at large. Therefore, requirements must be confirmed with key stakeholders, and a high level of acceptance within the HE community achieved.

This need was listed as an essential design principle in the NSCS work plan.

The requirements that 5.2 summarises are:

	Mandatory requirements		Other requirements
R4	Replacing JACS	R8	Classifying subjects
R9	Governing and sector bodies	R12	Comparing degree programmes
R13	Supporting operational and time series statistics	R14	Supporting regulated professions
R18	Remaining static for an academic cycle	R15	Supporting service oriented approaches to publishing on the internet
R19	Facilitating annual reporting and review	R16	Supporting multiple aggregation structures
R20	Enabling statutory returns	R21	Supporting student lifecycle comparisons
R29	Facilitating datasets that are fit for purpose	R22	Enabling mapping to NHS occupation codes
R31	Describing guidance purposes clearly	R23	Enabling workforce and capacity planning dataset comparisons
R32	Encouraging clarity in the description of data collection purposes	R24	Enabling disaggregation in health subjects
R45	Including external definitions of important subjects	R25	Enabling aggregations for NHS planning
		R34	Supporting more fine grained classification of subjects
		R43	Providing support documents
		R47	Supporting explicit aggregations of subjects
		R48	Supporting specific HEP functions

5.3. A means of linking to data classified in other frameworks.

Ability to link to European and worldwide systems is desired on the grounds of creating interoperable systems. Also, furthering the open data agenda is a recognised common 'good' and the NSCS should adhere to recognised good practice in this area generally.

This goal encompasses the two essential and one desirable design principles from the NSCS workplan:

- a coding framework that is consistent with the open data agenda;
- a coding framework that corresponds to recognised good practice;
- a means of linking to data classified in other frameworks.

It also addresses several of the recommendations of the **Classifying subjects of study** (Ferrell, 2013) report:

- the new framework should provide a persistent URI (uniform resource identifier) for each of the entities in the classification;
- the authoritative URIs should be developed and maintained as a web service for the sector;
- the new framework should be explicitly assigned an open licence.

The requirements that 5.3 summarises are:

	Mandatory requirements		Other requirements
R2	Interoperating with other subject vocabularies	R14	Supporting regulated professions
R3	Persisting URIs	R15	Supporting service oriented approaches to publishing on the internet
R11	Providing training recommendations	R21	Supporting student lifecycle comparisons
R13	Supporting operational and time series statistics	R22	Enabling mapping to NHS occupation codes
R20	Enabling statutory returns	R23	Enabling workforce and capacity planning dataset comparisons
R28	Providing guidance on the NSCS and KIS	R24	Enabling disaggregation in health subjects
R29	Facilitating datasets that are fit for purpose	R25	Enabling aggregations for NHS planning
R40	Differentiating term definitions	R34	Supporting more fine grained classification of subjects
R45	Including external definitions of important subjects	R26	Providing more detail in medicine subjects

5.4. The NSCS should be simple to understand

In order to achieve wide usage and consistent application, the NSCS needs to be simple to understand. This may require a trade-off with a number of other design goals.

The requirements that 5.4 summarises are:

	Mandatory requirements		Other requirements
R1	NSCS and JACS3	R33	Supporting aggregation of STEM and SIV course data
R7	Having clear and concise definitions	R36	Classifying subjects or groups of subjects
R31	Describing guidance purposes clearly	R38	Excluding unstudied subjects
R32	Encouraging clarity in the description of data collection purposes	R39	Using common labels
R37	Supporting multiple, combined or interdisciplinary subjects	R35	Making codes memorable
R40	Differentiating term definitions		

5.5. Be robust

In order for the NSCS to be reliable over longer periods of time, it needs to be sufficiently well grounded in the HE sector such that usage can resist political change, particularly if maintenance resources are reduced. This requires a high level of acceptance within the HE community.

The requirements that 5.5 summarises are:

	Mandatory requirements		Other requirements
R11	Providing training recommendations	R43	Providing support documents
R13	Supporting operational and time series statistics		
R18	Remaining static for an academic cycle		
R19	Facilitating annual reporting and review		
R20	Enabling statutory returns		
R29	Facilitating datasets that are fit for purpose		

5.6. Be stable

It is clear that the NSCS should minimise change that would have cost implications for the sector. There is, however, a clear trade-off with flexibility and adaptability, and it also implies a schedule for maintenance, change and version control.

The requirements that 5.6 summarises are:

	Mandatory requirements		Other requirements
R3	Persisting URIs	R33	Supporting aggregation of STEM and SIV course data
R13	Supporting operational and time series statistics		
R18	Remaining static for an academic cycle		
R19	Facilitating annual reporting and review		
R20	Enabling statutory returns		
R40	Differentiating term definitions		

5.7. Backwards compatibility with legacy data.

An element of comparison with earlier data and ease of mapping from JACS3 to the NSCS was felt to be a prerequisite by many stakeholders. Clearly, this goal may put a brake on radical solutions to pressing other needs.

This goal addresses one essential and one desirable design principles listed in the NSCS work plan:

- [need to provide] an easy transition path from JACS to the proposed scheme;
- backwards compatibility with legacy data.

The requirements that 5.7 summarises are:

	Mandatory requirements		Other requirements
R1	NSCS and JACS3	R23	Enabling workforce and capacity planning dataset comparisons
R13	Supporting operational and time series statistics		
R19	Facilitating annual reporting and review		
R20	Enabling statutory returns		
R28	Providing guidance on the NSCS and KIS		
R29	Facilitating datasets that are fit for purpose		

5.8. An appropriate level of granularity (in relation to usability and meaningful data)

Feedback from stakeholders indicates that too much detail and complexity reduces usability, and thereby data quality. However, a NSCS that is too broad reduces the amount of meaningful data available to sector bodies. A careful balance will need to be struck for each application area.

This goal was listed as an essential design principle in the NSCS work plan. It also addresses the recommendation that the new framework should consist of three rather than four levels from the **Classifying subjects of study** (Ferrell, 2013) report.

The requirements that 5.8 summarises are:

	Mandatory requirements		Other requirements
R13	Supporting operational and time series statistics	R8	Classifying subjects
R20	Enabling statutory returns	R14	Supporting regulated professions
R28	Providing guidance on the NSCS and KIS	R16	Supporting multiple aggregation structures
R37	Supporting multiple, combined or interdisciplinary subjects	R23	Enabling workforce and capacity planning dataset comparisons
R45	Including external definitions of important subjects	R33	Supporting aggregation of STEM and SIV course data
		R34	Supporting more fine grained classification of subjects
		R36	Classifying subjects or groups of subjects
		R38	Excluding unstudied subjects
		R39	Using common labels
		R47	Supporting explicit aggregations of subjects
		R26	Providing more detail in medicine subjects

5.9. Provide comprehensive coverage of the range of subjects of study available in HE at an appropriate level of detail for its target users

This goal is mostly concerned with the ability to classify all courses and other entities in some way, as well as the ability of the NSCS to accommodate demand for more or less precision in specific areas.

The requirements that 5.9 summarises are:

	Mandatory requirements		Other requirements
R7	Having clear and concise definitions	R14	Supporting regulated professions
R13	Supporting operational and time series statistics	R17	Providing support for course searching
R20	Enabling statutory returns	R24	Enabling disaggregation in health subjects
R28	Providing guidance on the NSCS and KIS	R25	Enabling aggregations for NHS planning
R37	Supporting multiple, combined or interdisciplinary subjects	R34	Supporting more fine grained classification of subjects
		R36	Classifying subjects or groups of subjects
		R38	Excluding unstudied subjects
		R39	Using common labels

5.10. Increased consistency of application across institutions

Consistent application is crucial for ensuring data quality. This may have potential trade-off with other design goals, because different subject areas may have different needs with regard to granularity, for example.

This goal was listed as an essential design principle in the NSCS workplan.

The requirements that 5.10 summarises are:

	Mandatory requirements		Other requirements
R7	Having clear and concise definitions	R33	Supporting aggregation of STEM and SIV course data
R31	Describing guidance purposes clearly	R36	Classifying subjects or groups of subjects
R32	Encouraging clarity in the description of data collection purposes	R38	Excluding unstudied subjects
R37	Supporting multiple, combined or interdisciplinary subjects	R39	Using common labels
R40	Differentiating term definitions	R35	Making codes memorable
		R41	Avoiding leading zeros

5.11. Be clearly separate conceptually from the JACS system

Users will need to use the NSCS differently (better) than JACS3. Quality data relies on this. Yet at the same time, longitudinal comparability and ease of use may require strong ties with JACS.

The requirements that 5.11 summarises are:

	Mandatory requirements		Other requirements
R11	Providing training recommendations	R10	Providing guidance on coding for specific purposes
R40	Differentiating term definitions	R17	Providing support for course searching
R49	Structural position of Welsh as a subject in the scheme	R36	Classifying subjects or groups of subjects
		R38	Excluding unstudied subjects
		R39	Using common labels

5.12. A framework with scope for evolution

Various requirements from student recruitment facing functions in particular emphasise the need for a structure that enables regular maintenance, quick error corrections, and ongoing changes in subjects. Accommodating such agility has an obvious trade-off with stability, and implies that organisations will record and report current usage and requests for change.

This goal was listed as an essential design principle in the NSCS work plan. It also addresses the recommendation in the **Classifying subjects of study** (Ferrell, 2013) report that the new framework should be based on a six digit coding structure, to allow for growth in the number of terms.

The requirements that 5.125.12. summarises are:

	Mandatory requirements		Other requirements
R3	Persisting URIs	R10	Providing guidance on coding for specific purposes
R18	Remaining static for an academic cycle	R15	Supporting service oriented approaches to publishing on the internet
R19	Facilitating annual reporting and review	R33	Supporting aggregation of STEM and SIV course data

5.13. Have a broadly consistent level of detail

This goal supports goal 5.10, which is about keeping the NSCS simple and easy to understand. It suggests a framework approach, where the framework has a consistent level of detail and links to more specialist sub-classifications. The challenge will be to design an ecology of classifications such that it retains the simplicity of the framework, and doesn't drive a perception of complexity.

The requirements that 5.13 summarises are:

	Mandatory requirements		Other requirements
R13	Supporting operational and time series statistics	R38	Excluding unstudied subjects
R20	Enabling statutory returns	R39	Using common labels

5.14. The perceived need for a discipline-based system

JACS3 has been perceived to be a discipline-based system, even though it was designed to be a subject classification scheme. Specifically, this goal addresses two recommendations from the **Classifying subjects of study** (Ferrell, 2013) report that arose from this issue:

- the new framework should recognise the, currently implicit, assumption (at least by HE providers) that JACS is a discipline-based classification
- in developing the new framework the Higher Education Academy's discipline areas should be considered as a starting point.

During the consultation phase of stage 1, however, we found little support for a switch to a discipline based scheme, but many supported a requirement for a continued subject focus.

The requirement that 5.14 summarises is:

	Mandatory requirements		Other requirements
		R8	Classifying subjects

5.15. Have regard for the specific recommendations in the 'roadmap' about disciplines, levels and structure

The roadmap recommendations include the use of the HEA discipline areas as a starting point, having three rather than four levels, and a six-digit coding structure.

The requirement that 5.15 summarises is:

	Mandatory requirements		Other requirements
		R8	Classifying subjects

6. Areas of conflict, impact risks and their remediation

As the previous section makes clear, meeting one design goal can, but doesn't always have to, mean compromising another. In order to strike the right balance, assess impact risks on stakeholders and what can be done to reduce these risks, the main potential areas of conflict need further consideration.

6.1. Comprehensive detail for data analysts versus minimal coding effort for data providers

This conflict pits high priority design goals such as simplicity of understanding and application (5.4 and 5.7), against equally high priority ones such as the need to support policy implementation, and provide comprehensive coverage (5.4 and 5.9). It even splits some design goals: the need to support a wide group of stakeholders and the need to enable data creation and collection at the right level of granularity (5.2 and 5.8) require a balance of the two sides to succeed.

In terms of impact, there is a graded sets of risks in not getting the balance between ease of use and comprehensive detail right. The most extreme is the possibility that data providers will simply not be able or willing to adopt the NSCS, thereby endangering the flow of data in the sector generally. On the other hand, there is the slightly less extreme possibility that data that is too simple is too coarse to meet vital policy needs. Milder still, but still serious, is the possibility that the NSCS promises sufficient detail to meet the needs of providers, but is too difficult to code with in the more detailed areas, which leads to data that is too inconsistent and of poor quality to be useful.

Measures that can be adopted to mitigate these risks start at continuous and intensive contacts of the NSCS project with both data providers and consumers to ensure that proposals balance both sides' needs appropriately. The provision of early prototypes should help there too. Another measure that promises to reconcile conflicting needs is the idea to have the NSCS accommodate multiple hierarchies above and below a common core list in order to facilitate specific needs. Finally, thorough term definitions, training and tool implementation support and other support around the scheme could help people cope with a degree of complexity in the scheme itself.

6.2. Agility versus stability

Design goals that aim for flexibility and adaptability, that can accommodate new terms and links to other datasets, and that distance NSCS from JACS (5.12, 5.11, 5.9 and 5.2) clash directly with those that aim for stability, consistency, robustness and backward compatibility (5.6, 5.13, 5.10 and 5.5). Variants of this conflict are apparent in cases where the cost of implementation blunts a desire to adopt innovation that addresses other goals, or when the case for any change from the current status quo at all is not apparent to a key stakeholder.

The most dramatic impact of this conflict could be an impossibility to map NSCS data back to historic datasets. The need for time series is great enough that an absence of them could seriously threaten the value and therefore existence of the NSCS. A slightly less serious impact would be felt if the NSCS failed to accommodate new subjects appropriately over a period of time. Milder still is the impact of a delay in adoption of an otherwise attractive and innovative NSCS on cost grounds.

Measures to reconcile the need for agility with stability centre mostly around governance solutions. A deliberate and predictable process for changing the NSCS needs to be set up that allows relevant

stakeholders to come to consensus about the NSCS content as well as the speed of the adoption and change processes. These aspects will be addressed in the NSCS adoption plan. Other measures to manage the agility versus stability conflict involve the regular testing of mappings between JACS3 and the NSCS. Finally, the need for agility is particularly keenly felt in course discovery applications, but not so much in other uses of JACS3. Accommodating such tensions is the topic of the next section.

6.3. Supporting different purposes in a single scheme

The need to reconcile many purposes in a subject classification scheme is not a matter of striking a balance between different design goals. Instead, it is something inherent in goals such as the need to support a wide range of stakeholders and their policies, and the need to link to multiple existing other subject vocabularies.

The danger with the biggest impact in this area would be a scheme that, in an attempt to support all purposes equally well, ends up being able to serve none to any acceptable degree. This could threaten the adoption of the scheme, and thereby any remaining utility. A less serious outcome would be if the NSCS served only one purpose for one stakeholder well, but no more.

Measures that can help manage the tension inherent in having to support many different purposes include the possibility of accommodating multiple ways to aggregate subjects and add more fine grained vocabularies 'underneath' a common list of terms. Another measure is to look critically at which purposes can usefully be supported, and which have alternative solutions or higher priorities. It may well be, for example, that the relatively fast moving world of marketing courses to prospective students can only be supported indirectly when other purposes and goals are also taken into account.

7. The prototypes

Both prototype schemes came out of the development of the initial examples (see section 4), and were collaboratively and iteratively developed as the consultation with stakeholders progressed. Where the examples gave them shape, the requirements, and, later, the design goals, defined the constraints in which they had to remain. The design goals will be used here to describe and evaluate both prototypes.

The reason why there are two prototypes is to be able to gauge the response of stakeholders to two different solutions at either end of a spectrum: one close to current practice, and one a more radical departure. The expectation is that the two will converge in a compromise that suits most stakeholders.

7.1. Prototype 1

This prototype takes a moderately conservative view. It focuses on classifying 'subjects of study' as opposed to 'academic subjects' or 'disciplines' and stresses:

- ease of use for classifiers, particularly for navigation through the scheme to find the correct term;
- classification at programme level;
- continuity with the JACS3 approach.

It addresses key features outlined in section 4.7, but opts for continuity instead of radical design features. Modifying the features in 4.7, it offers:

- an extensive single list of terms at which classification would be carried out (the bottom level of a three level hierarchy);
- a default hierarchy aimed at ease of navigation and classification;
- while having its own level structure, it is envisaged that numerous aggregations would be derived from its main list by stakeholders for specific purposes;
- the extensive list of terms for classification would cover 'subjects of study' that would be of practical use for specific purposes without defining a limit – it is likely that this would be more than the number of JACS3 Principal Subjects but less than 3-character JACS3;
- it will use 'not otherwise classified' terms as a mechanism, coupled with 'make your own code' to manage new subject areas;
- the terms will have memorable identifiers of a fixed length that reflect a three level hierarchy.

Prototype 1 has been developed using the general area 'health and medicine', as an example of a domain that has been pointed up through consultation as a possible target for change. It has been populated using JACS3 supplemented by subjects suggested by NHS Occupation Codes, additional desk research, and influenced by requirements from Health Education England (represented by Health Education North West London) and other sector bodies. It has been designed using a mono-hierarchical thesaurus-based approach. Each term in the thesaurus has appropriate broader and narrower terms and a 6-digit code. This approach readily supports other forms of relationship, such as 'related terms' to link nodes across the sections of the hierarchy, synonyms, and recording of other classifications against each term for mapping purposes. It would include permanent URIs, rigorous definitions of each subject and group, as well as scope notes advising how to use the terms, and recording the development of each one over time.

The prototype proposes three levels of hierarchy with the following functions for each level:

- Top level: Very few broad areas (up to approximately 9), so that users can readily take a top-down approach to navigation. For example, a programme entitled 'Prosthetics and Orthotics' falls clearly into the 'health and medicine' area, and not any of the others.
- Second level: This level is used to group together cognate subjects, so that the classifier can navigate to an accurate term at the third level. A small number of very broad programmes may exceptionally be coded at this level. It also provides a 'default' aggregation mechanism. For example the 'Prosthetics and Orthotics' programme fits within the 'medical technology' group.
- Third level: Almost all classification would be at this level, which contains the vast majority of programme subjects. For example, our 'Prosthetics and Orthotics' programme sits directly in the 'prosthetics and orthotics' term.

The project's 'initial questions and examples' drew out views on five key issues: sharing subjects; discipline, subject or both; how many top terms; subjects not listed; and the code format. In relation to the structural issue of 'sharing subjects', this prototype takes a traditional 'unique subjects' approach. While there was a significant voice from the planning community in favour of a relatively small flat list approach, other stakeholders viewed a traditional unique node tree structure with favour, partly because it is readily understandable by the existing community of users, and partly because it can make reporting transparent and uncomplicated. However, as we recommend that all classification is handled at level three in this prototype, it could be considered a version of an 'extended flat list', albeit with a large number of terms.

The approach claims that each term is a 'subject of study' or 'cognate group of subjects of study', so it is in the ball-park of 'subjects' rather than 'disciplines'. These terms can be validated against the

real range of programmes on offer in UK Higher Education, as well as against authoritative lists such as HEA disciplines documents and QAA subject benchmark documents.

The prototype includes three specific measures designed to ease maintenance and classification, including handling of subjects not listed. For classification of broad programmes covering groups of subjects, a standard 'general' term is offered, for example a broad healthcare science programme might be classified at level three as 'healthcare science, general'. For subjects not listed in the scheme, an 'other' term is provided where it is needed, for example 'healthcare science, other' can be used for additional healthcare science programmes coming on stream after the launch of the scheme, prior to incorporation of the new subjects. However, this aspect of the prototype could be used in conjunction with the favoured 'make your own code' concept; users create their own new unofficial code constructed using the existing code format, map it to the generic 'other' term, then notify the central agency responsible for NSCS maintenance and governance. This agency would then publicise the origination of the new code for internal use prior to formal adoption.

Example:

healthcare science [second level; part only]

BT: **health and medicine** [top level]

NT: **audiology** [third level]

NT: **healthcare science, general** [third level]

NT: **healthcare science, other** [third level]

NT: **language pathology** [third level]

NT: **speech and language therapy** [third level]

[BT = Broader Term; NT = Narrower Term]

The top two levels provide the possibility of default aggregation categories for cognate groups of subjects of study. These groups are not simply convenient groups driven by limited space in the coding frame, but can be driven by stakeholder requirements.

While many coding frames could be used to represent this approach within information systems, it is suggested that consideration is given to a 6-digit code, with pairs of randomly-generated digits at each level. Using the example above, that might give the following examples: 'health and medicine, 550000'; medical technology, 552400'; 'prosthetics and orthotics, 552431'. This style of notation gives plenty of room for expansion (99 terms at each branch). The meaning of the code is very limited, but it contains the hierarchical location for ease of use.

Example (part only):

LEVEL 1

health and medicine

NSCS code: **550000**

NT: **alternative medicines and therapies**

NT: **complementary medicines and therapies**

NT: **health and medicine, other**

NT: **health sciences**

NT: **healthcare science**

NT: **medical sciences**

NT: **medical technology**

NT: medicine and dentistry

NT: nursing

LEVEL 2

medical technology

NSCS code: **552400**

BT: health and medicine

NT: biomechanics

NT: dental technology

NT: medical technology, general

NT: medical technology, other

NT: prosthetics and orthotics

NT: radiology

LEVEL 3

biomechanics

NSCS code: **552452**

BT: medical technology

dental technology

NSCS code: **552421**

BT: medical technology

medical technology, general

NSCS code: **552432**

BT: medical technology

medical technology, other

NSCS code: **552418**

BT: medical technology

prosthetics and orthotics

NSCS code: **552431**

BT: medical technology

radiology

NSCS code: **552447**

BT: medical technology

[BT = Broader Term; NT = Narrower Term]

The levels in this prototype are explicit, which means that data can be aggregated readily at each level: two digits for level 1, four digits for Level 2 and six digits for Level 3. In addition, the recommended level for classification is Level 3, so the scheme could support other hierarchies for different purposes, if desired. Level 3 terms should be sufficiently detailed to provide high definition at programme level and reasonable definition at module level, while this level could also

support a more detailed scheme articulating to it (for example for research or detailed module classification).

Examining this prototype against the design goals of section 5.1, we have the following broad evaluation.

Support policy implementation

Mapping to and from JACS3 and supporting time series statistics should be straightforward, as Levels 2 and 3 would contain roughly cognate terms to JACS3. While it has its own default hierarchy, aggregations can be built readily from Level 3 for additional purposes. In addition it has sufficient granularity at Level 3 to support currently unknown requests for subject-based data that might not be possible with a broader classification. It sacrifices some of the simplicity provided by a small list of terms, for the advantage of supporting a wider range of purposes, for example the easy identification of subjects of interest to PSRBs and policy-makers.

Facilitate linking of other data, including other classification systems

Prototype 1 supports usage of the W3C's Simple Knowledge Organization System (SKOS) specification, which is the most widely used vocabulary in the linked data world for the expression of schemes such as the NSCS. Each term will have a permanent URI. This approach would enable linking to other classifications, such as Research Council ones, or more detailed classification systems, such as LDCS.

The NSCS should be simple to understand

Prototype 1 retains some of the simplicity of a flat list structure, by using the top 2 levels for navigation and default aggregation, but not for coding. As classification is at Level 3 only, the approach should be readily understandable, subject to testing, especially as a given institution will use only a sub-set of the terms.

Be robust

Prototype 1 should prove to be more robust than the current JACS3, as it has a granularity geared to practical usage, not the amount of space in the coding frame.

Be stable

Prototype 1's relatively large Level 3 gives the opportunity for long-term stability, as it will be based on examination of course data to provide practical subject terminology.

Provide backwards compatibility with legacy data

Prototype 1 includes a map from all the subjects at JACS3 Principal Subject level and some at the next level down. This would support continuation of the JACS3 approach to general coding and analysis, as well as providing additional differentiation for improvement in data sets where there are acknowledged weaknesses in the detail.

Provide comprehensive coverage of the range of subjects of study available in HE at an appropriate level of detail for its target users

The extensive and well-defined list of subjects of study at Level 3 in Prototype 1 supports a wider range of users than existing JACS3. In addition it helps to resolve some of the quirks, inaccuracies, structural deficiencies and implementation difficulties identified in the consultation.

Be simple to apply consistently

With a simplified structure providing a single level for classification, compatibility with JACS3, continuity of approach and a redefinition of each term, this prototype should support consistent application. We envisage that part of Stage 2's adoption plan will be to wrap the NSCS with appropriate guidance and training, which can to some extent build on the existing JACS3 understanding.

Be clearly separate conceptually from the JACS system.

Prototype 1 has some clear concepts setting it apart from JACS, but maintaining important continuity of approach. The major breaks are the simpler concepts of classification at only one level, and a three level structure instead of four.

Be flexible and adaptable with scope for evolution.

While perhaps not as flexible in structure as a more radical solution, Prototype 1 supports evolution over time, as it has almost no practical constraint on the number of terms supported. It retains the ability to link to other schemes.

Have a broadly consistent level of detail.

Based on subjects of study at programme level, Prototype 1 has a broadly consistent level of detail at Level 3, as well as mechanisms to support the classification of more general programmes.

The perceived need for a discipline-based system.

Our consultations suggest that the primary requirement is for classification at programme level, rather than disciplines. This is directly supported by Prototype 1.

7.2. Prototype 2

This prototype is on the more radical, minimal edge of the NSCS design space, and emerged out of the consultation activities. The scheme therefore has all the features outlined in section 4.7:

- a single, common flat list of terms;
- a number of hierarchies that aggregate terms from the common list are developed by stakeholders for specific purposes;
- a number of hierarchies that refine terms from the common list are developed by stakeholders for specific purposes;
- the common terms codify between 40 and 150 subjects;
- it will not use 'not otherwise classified' terms;
- the terms will have opaque but memorable identifiers of a fixed length.

At this point, the prototype has one notable conservative aspect in that it has the exact same content as Level 2 of JACS3. The reason for this is twofold. One is the intent to develop the

structure of the NSCS framework before developing the content, and the other is that using JACS3 level 2 as a starting point ensures the prototype meets a number of design goals.

In order to best understand the nature of Prototype 2, and judge its fitness for purpose at the same time, it will be evaluated against the design goals of section 5.

Support policy implementation

By putting a limit on the number of terms in the common list, Prototype 2 makes it easier to attach an essentially unlimited number of external linked schemes for specific policy purposes.

HEFCE, for example, has a definition of STEM subjects that cuts across existing JACS3 subject aggregations. Prototype 2 addresses this problem as multiple definitions referring to the same subjects can exist side by side:

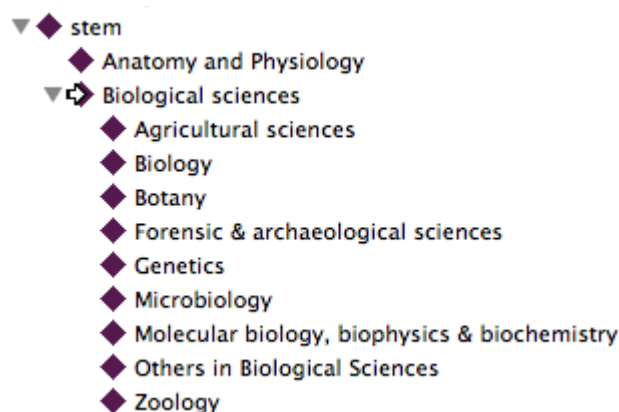


Figure 2 HEFCE aggregation of STEM subjects, biological sciences detail

Figure 2 shows the Biological Sciences fragment of the current HEFCE aggregation of STEM subjects. The first two levels are HEFCE's, but the subjects at the third level are Prototype 2's flat list. The JACS3 view of the same subjects is quite different:

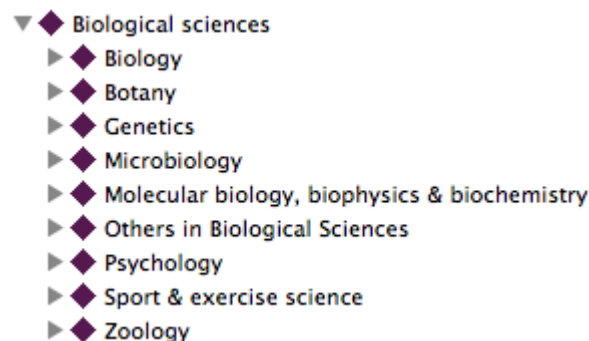


Figure 3 JACS3 aggregation of Biological Sciences

Even after the introduction of the NSCS, a JACS3 hierarchy could be maintained for a variety of reasons, and would be just another hierarchy on top of Prototype 2's common list of subjects.

Just as important as the range of aggregations that Prototype 2 allows, is the fact that each can develop as quickly and as radically as necessary, without affecting the others. If there is a strong

policy driver for HEFCE to rapidly and drastically alter its STEM definitions, for example, it can do that without triggering costly sector-wide system change for everyone who uses the NSCS.

Support a wide group of stakeholders

The need for the NSCS to support many stakeholders with many and varied drivers motivated Prototype 2's central characteristic: modularisation. By decoupling aggregations above and below the flat list, different sector bodies can fine tune their own structure to their own needs, while leaving a minimal core common to all. That way, fine-grained, but labour intensive detail can be defined where it is necessary for those who have an interest, but a simpler, coarser list can be used where that is adequate.

Facilitate linking of other data, including other classification systems

Prototype 2 has been developed as an application of the W3C's Simple Knowledge Organization System (SKOS) specification, which is the most widely used vocabulary in the linked data world for the expression of schemes such as the NSCS. Each term in Prototype 2 is a URI, and the schema as a whole qualifies as five star linked open data.

As a result, sample companion schemas for HEFCE's STEM definition and the nursing part of the NHS' occupation codes have been developed and formally linked to the Prototype 2 subject list within a few hours. Linking other classification systems should be just as easy, bearing in mind that voluminous schemes will take proportionally longer.

The SKOS nature of Prototype 2's list does not, however, oblige others to use the same technology, nor does it depend on it per se. Prototype 2 will use its own identifiers alongside URIs so that it can be deployed and used unambiguously in a variety of technologies, such as those used for JACS3 today.

The NSCS should be simple to understand

The original idea for a limited flat list of subjects came from those who would struggle most with a complex, difficult to understand scheme: HEP planners, who are responsible for accurate, timely and cost effective data returns. Though one hundred and fifty is a larger number of terms than was mentioned during the workshops, it is still a manageable number when taking into account that only a subset will be used in any given institution.

Be robust

Because level 2 of JACS3 is so widely used, it could be said that Prototype 2 merely formalises the part of JACS that has already proved to be most robust, whilst providing a flexible way forward. For example, if any scheme that links to Prototype 2 does not succeed for whatever reason, the common list will remain unaffected.

Be stable

Because modularisation means that change is compartmentalised, it is easier to provide stability for what the sector at large already appears to have settled on as the core: the one hundred and fifty or so principal subjects of JACS3 Level 2.

Provide backwards compatibility with legacy data

The crucial importance of time series data for a wide range of stakeholders is the main reason that the content of Prototype 2 is the same as JACS3 level 2 at this stage. Responses so far suggest that the majority of coding and data analysis is currently done at level 2. Most time series should therefore be able to continue with comparative little change, unless other requirements necessitate radical alteration to Prototype 2's list. Should this be the case, the ability of SKOS and other linked data vocabularies to express rigorous and precise relations can help ameliorate the loss of full backwards compatibility.

An appropriate level of granularity (in relation to usability and meaningful data)

One reason the core schema of Prototype 2 has adopted JACS 3 Level 2 as its departure point is that response from stakeholders suggest that it is close to the level where most data analysts can derive meaningful information, without imposing unduly on data providers.

Furthermore, the modular nature of Prototype 2 allows the level of granularity to be fine tuned within communities for specific purposes.

Provide comprehensive coverage of the range of subjects of study available in HE at an appropriate level of detail for its target users

While Prototype 2's list may not on its own be able to cover a comprehensive range of subjects of study for all target users and all purposes, its modular nature provides a mechanism to support as much additional detail as a particular community needs.

For example, responses from the consultation exercise indicate a significant interest in linking up medicine and medicine related degree statistics with NHS workforce planning. Prototype 2 facilitates this by enabling existing NHS occupational codes to be attached to NSCS subjects

- ▼ Nursing
 - ◆ Acute Elderly & General
 - ◆ Community Learning Disabilities
 - ◆ Community Psychiatry
 - ◆ Community Services
 - ◆ Education Staff
 - ◆ Maternity Services
 - ◆ Neonatal Nursing (including SCBUs)
 - ◆ Other Learning Disabilities
 - ◆ Other Psychiatry
 - ◆ School Nursing

unchanged:

Furthermore, a different audience looking at the same data can ignore the NHS occupational codes, or depending on the application they are using, conceal them, and treat nursing just like any other subject.

Be simple to apply consistently

This is an area where some adjustment may be required. While the absence of inconsistent lower levels and a disputed top level should be beneficial, there are still some areas of overlap and ambiguity that need addressing. However because the management of linked schemes will be in

the hands of the domain communities that use them, this should increase the chances of consistent application.

The role of system implementations is important here. Ensuring that Prototype 2 is easy to use, and supporting vendors who implement it, is likely to be key in the next phase of the NSCS project.

Be clearly separate conceptually from the JACS system

Though the content of Prototype 2 should look familiar, the absence of hierarchies, and the modular nature of the Prototype 2 framework should ensure a clean break with the past.

Be flexible and adaptable with scope for evolution

Modularisation by linking can be a flexible and adaptable solution to organising changeable information environments. As the NHS and HEFCE examples above make clear, having separate, but linked schemes allows different but related subject coding purposes to develop independently, while sharing a stable core.

Have a broadly consistent level of detail

Whether subject coding at low levels will be consistent across the board depends on whether there is a requirement for it: if there is a clear need, communities can ensure that linked schemes are consistent. If there is no perceived need, Prototype 2 will not enforce it. However, consistency of granularity – if not application – has already proved to be reasonably robust at the level of the flat list itself: that of principal subject.

The perceived need for a discipline-based system

Research so far suggests the main use for the NSCS will be the classification of degree programmes rather than subjects or disciplines per se. Since the contents of Prototype 2 have been used successfully for that purpose for a number of years, a good case for its efficacy can be made.

Have regard to the specific recommendations in the 'roadmap' about disciplines, levels and structure

Prototype 2 can accommodate HEA discipline areas as one among many aggregation structures on top of the common list. Multiple levels can be supported in the same way, should there be demand for it. 6-digit coding structures are not a priority, since codes for Prototype 2 do not encode levels.

8. Critical success factors for the new scheme

As the areas of conflict indicate, the critical success factors for the NSCS lie at the interface between the politics of data sharing and the features of the vocabulary itself. If the features of the NSCS are skewed too far in favour of data providers or in favour of one or all of the data consumers, the other parties may feel unable or unwilling to keep cooperating, and thus jeopardise the whole endeavour.

For that reason, the most critical success factor is to reach a compromise in scheme size, agility and function scope that is acceptable to all parties. Experience so far suggests that continuous engagement with concrete, if incomplete, prototypes is key in this regard, not just to build buy-in, but also to keep gauging the impact of the new scheme on daily practice. The same goes for the timing of the deployment of the NSCS.

Specific factors that arise from the development of the prototypes include:

- Checking whether the modularisation of the NSCS – whether in the auxiliary form of prototype 1, or the more fundamental form of prototype 2 – can bring the expected flexibility in supporting multiple purposes.
- Determining the right balance in number of terms and level of granularity to support data destination needs as well as data suppliers' resourcing.
- Developing a governance model that is supported by an appropriate selection of stakeholders, with sufficient authority and resources to manage the evolution of the NSCS.

9. Next steps

If the design goals, their constituent requirements and their prioritisation as well as the prototypes that emerged out of the initial design questions and examples are acceptable to the core HEDIIP stakeholders, then the next steps would be to continue to develop both prototypes with a view to converge them during the next stage of the project.

In stage 2, we envisage expanding consultation with groups that could add more to the content of the prototype scheme such as domain experts, the HE Academy and more of the PSRBs. Nonetheless, the main consultation work will remain with the core sector bodies and HEPs on the overall structure of the NSCS framework. As the development of the prototypes progresses, the consultation work is likely to be increasingly practical and hands-on.

Though the precise scope is yet to be determined, it will also be necessary to consider the development of an NSCS framework to guide the extension of the core NSCS scheme for specific purposes. For a scheme like Prototype 1 (see section 7.1), the framework is a relatively small addition, but for Prototype 2 (see section 7.2) the framework would be a centrally important means to guide the addition of the complementary schemes it assumes.

Another next step will be to use text mining tools and subject based analysis to see whether the developing contents of the NSCS can be validated with or augmented by the output of such tools and approaches.

Finally, work will start that takes the requirements on governing the future development of the NSCS, and develop a model that strikes the right balance between inclusiveness and authority, support and overhead, and decisiveness and agility.

10. References

Australian Bureau of Statistics. (n.d.). 1272.0 - Australian Standard Classification of Education (ASCED), 2001. Retrieved September 12, 2014, from <http://www.abs.gov.au/AUSSTATS/abs@.nsf/Latestproducts/56C477941726CFF0CA256AAF001FCA50>

Australian Research Council. (n.d.). Australian and New Zealand Standard Research Classification (ANZSRC) - Australian Research Council (ARC). Retrieved September 12, 2014, from <http://www.arc.gov.au/era/ANZSRC.htm>

Elsevier. (n.d.). Subjects | Elsevier. Retrieved September 12, 2014, from <http://www.elsevier.com/journals/subjects>

Ferrell, G. (2013). *Classifying subject of study; A roadmap to a new Joint Academic Coding System*. Regulatory Partnership Group. Retrieved from http://www.hediip.ac.uk/wp-content/uploads/JACS_Report_2013-07.pdf

Health and Social Care Information Centre. (n.d.). NHS Occupation Codes - Health & Social Care Information Centre. Retrieved September 12, 2014, from <http://www.hscic.gov.uk/article/2268/NHS-Occupation-Codes>

HESA - Higher Education Statistics Agency. (2012). Joint Academic Coding System (JACS) Version 3.0. Retrieved September 12, 2014, from <https://www.hesa.ac.uk/jacs3>

Learning and Skills Council. (n.d.). Providers - Data - Data Dictionary - Technical Definitions - Sector Subject Areas Data Definition 2008/2009. Retrieved September 12, 2014, from http://www.lsc.gov.uk/LSCGOVUK/Templates/Standard.aspx?NRMODE=Published&NRNODEGUID=%7BCCF3D60C-F6AD-4FF4-8A5A-2AB0D9015976%7D&NRORIGINALURL=%2Fproviders%2FData%2Fdatadictionary%2Ftechnicaldefinitions%2FA_SSA_T1%2BT2.htm&NRCACHEHINT=Guest

List of academic disciplines and sub-disciplines. (2014, September 6). In *Wikipedia, the free encyclopedia*. Retrieved from http://en.wikipedia.org/w/index.php?title=List_of_academic_disciplines_and_sub-disciplines&oldid=624393857

Organisation for Economic Co-operation and Development. (2007). *Committee for Scientific and Technological Policy: Working Party of National Experts on Science and Technology Indicators. Revised Field of Science (FOS) classification in the Frascati Manual*. (No. EAS/STP/NESTI(2006)19/FINAL). Retrieved from <http://www.oecd.org/science/inno/38235147.pdf>

QAA. (2014). Subject Benchmark Statements. Retrieved September 12, 2014, from <http://www.qaa.ac.uk/assuring-standards-and-quality/the-quality-code/subject-benchmark-statements>

Research Councils UK. (n.d.). Summary of Peer Review harmonisation activities - Research Councils UK. Retrieved September 12, 2014, from <http://www.rcuk.ac.uk/research/efficiency/researchadmin/harmonisation/>

The Higher Education Academy. (2014). Disciplines | The Higher Education Academy. Retrieved September 12, 2014, from <https://www.heacademy.ac.uk/workstreams-research/disciplines>

UNESCO. (1997). *International Standard Classification of Education-ISCED 1997*. UNESCO.

U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, & U.S. Department of Education. (2014). CIP user site. Retrieved September 12, 2014, from <http://nces.ed.gov/ipeds/cipcode/default.aspx?y=55>

U.S. National Library of Medicine. (n.d.). Medical Subject Headings - Home Page. Retrieved September 12, 2014, from <http://www.nlm.nih.gov/mesh/>

Appendices

1 Stakeholders contacted

Core sector bodies

These organisations were selected by HEDIIP because of their central role in the HE information landscape, or because subject-coded data is central to their operations. These bodies participated in the most intensive consultation activities; face to face meetings with a formal notes approval cycle. Some submitted written responses to the questionnaire as well.

HEFCE	Meeting notes approved and results incorporated
HESA	Meeting notes approved and results incorporated
Scottish Funding Council	All three met at the same time. Meeting report awaiting approval
Scottish Government	
Student Awards Agency for Scotland	
Research councils	Interviewed NERC, report awaiting approval
Student Loans Company	Meeting notes approved and results incorporated
UCAS	Meeting notes approved and results incorporated
DELNI	Meeting notes approved and results incorporated
Skills Funding Agency	Meeting report awaiting approval
Health Education England (via Health Education NW London)	Meeting notes approved and results incorporated; supplementary meeting notes to be approved
HEFCW	Meeting notes approved and results incorporated

Other sector bodies

These have been mostly contacted with the questionnaire by email, and invited to a one hour phone interview. Some have made written responses.

ARMA	Written submission
HESPA	Emailed, and met representatives face to face, and web response as well.
SROC	Written submission, phone interview, web response
UCISA	Written submission, phone interview
UHR	Written submission, met online, attended London workshop
Welsh Govt	Emailed - but see HEFCW
BIS	Written submission
AHUA	Emailed
AOC	Emailed
Guild HE	Emailed
NUS	Emailed
UUK	Emailed
ARC	Emailed
AUA	Emailed
Graduate Prospects	Phone interview and written submission
HE Academy	To be contacted in phase 2
HEDD	Run by Graduate Prospects, see above
Information Standards Board	informed, to be contacted further in phase 2
Jisc	Phone interview, notes approved and incorporated
LRS	Not believed to have specific role in relation to NSCS
OFFA	Not believed to have specific role in relation to NSCS

OIA	Not believed to have specific role in relation to NSCS
QAA	Interviewed
Software suppliers	To be contacted in phase 2
UKBA	To be contacted in phase 2
Universities Scotland	Emailed

Professional, Statutory and Regulatory Bodies (PSRBs)

These have been mostly contacted with the questionnaire by email, and invited to a one hour phone interview. Some have made written responses.

Engineering Council	Interviewed
General Medical Council	Interviewed
Architects Registration Board	Interviewed
General Pharmaceutical Council	Interviewed
Bar Standards Board	In email correspondence
Care Council for Wales	In email correspondence
Law Society of Scotland	Interviewed
Health and Care Professions Council	Interviewed
Royal Town Planning Institute	Interviewed
Association for Nutrition	Asked to be contacted in stage 2
Society of Sports Therapists	To be contacted in stage 2
CASE	To be contacted in stage 2
British Psychological Society	To be contacted in stage 2
Buttle Trust	To be contacted in stage 2
General Dental Council	To be contacted in stage 2
General Optical Council	To be contacted in stage 2
Nursing and Midwifery Council	To be contacted in stage 2
Royal College of Veterinary Surgeons	To be contacted in stage 2
Royal Institute of Architectural Technologists	To be contacted in stage 2
Joint Audio Media Education Services	To be contacted in stage 2
Scottish Social Services Council	To be contacted in stage 2

Higher Education providers

A broad selection of HEPs was invited to attend day long workshops. Some also responded on the comment site.

University of Sheffield	Attended the Manchester workshop
Manchester Metropolitan University	Attended the Manchester workshop
Bangor University	Attended the Manchester workshop
University of Salford, Manchester	Attended the Manchester workshop
University of Derby	Attended the Manchester workshop
University of the West of England, Bristol	Attended the Greenwich workshop
Royal Veterinary College	Attended the Greenwich workshop
Coventry University	Attended the Greenwich workshop
Birkbeck, University of London	Attended the Greenwich workshop

King's College London	Attended the Greenwich workshop
University of Greenwich	Attended the Greenwich workshop
University of Stirling	Attended the Edinburgh workshop
The University of Strathclyde	Attended the Edinburgh workshop
The University of Edinburgh	Attended the Edinburgh workshop
Imperial College, London	Responded online
Durham University	Responded online
London Southbank University	Responded online
Queen's University Belfast (Centre for Biomedical Sciences Education)	Responded online
University of Greenwich	Responded online
University of Northumbria	Responded online
Loughborough University	Responded online
The Open University	Interviewed

2 HEDIIP partnership NSCS design principles

As specified in the NSCS work specification.

The new subject coding scheme should be designed to meet the following design principles:

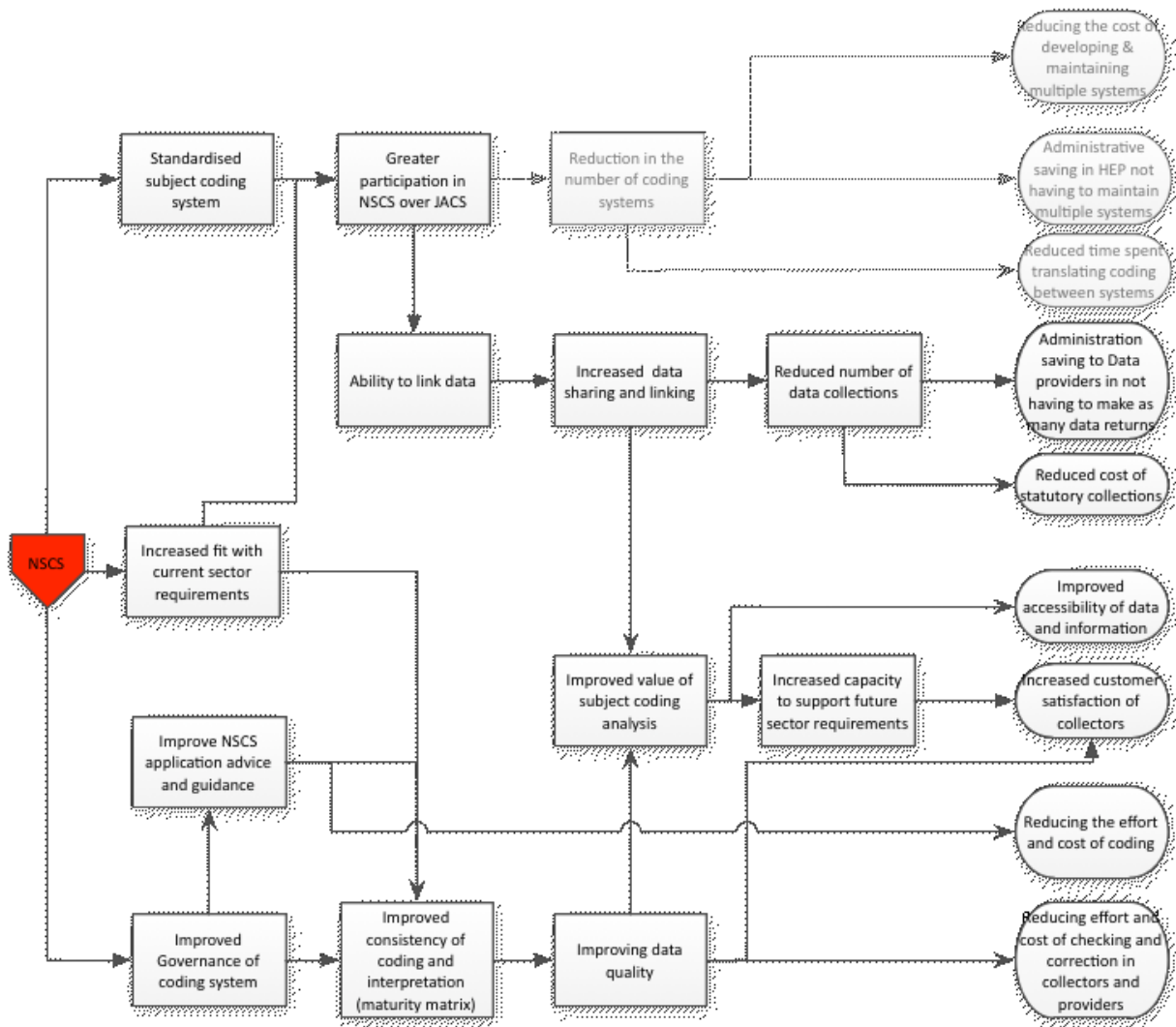
Essential:

- a coding framework that corresponds to recognised good practice;
- a framework with scope for evolution;
- a framework at an appropriate level of granularity (in relation to usability and meaningful data);
- an easy transition path from JACS to the proposed scheme;
- increased consistency of application across institutions;
- a coding framework that is consistent with the open data agenda;
- a unified approach to support a wide group of stakeholders.

Desirable:

- a means of linking to data classified in other frameworks;
- backwards compatibility with legacy data.

3 NSCS Benefits Map



The NSCS project found that the originally envisaged benefits of the scheme are still realistic, with two exceptions. One is the possibility that one of the prototypes may lead to a rationalisation and better coordination of coding schemes, but not a reduction in number. The other is that while there could be a significant reduction in the effort and cost of coding, this is unlikely to be true for the purpose of research management in relation to NSCS.

4 Interview Protocol

HEDIIP New Subject Coding Project: interview questions for sector organisations

Introduction

We would like to solicit your organisation's input into the redesign of the Higher Education information landscape, in particular in respect of a new subject coding system for the sector to replace JACS3 and cover a wider range of uses. The Higher Education Data & Information Improvement Programme (HEDIIP, www.hediip.ac.uk) has been established, following a number of studies sponsored by the Regulatory Partnership Group, 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. The work is funded by HEFCE, HEFCW, SFC and DEL.

Subject coding has been identified as a key building block that would yield benefits from sector-wide and consistent application. Following stakeholder consultations in 2013, HEDIIP has now contracted Cetus, in partnership with APS Ltd and Aspire Ltd, to develop a new subject coding scheme to enable sector-wide usage for a broad range of purposes, and to make recommendations for governance and adoption.

The first stage of this work runs to September 2014 and will consult widely to establish a robust set of requirements and impact analysis. The project team is now embarking on a set of activities to gather views of stakeholders across the sector, including Higher Education providers, sector-wide organisations, professional bodies, and others. This first stage will be followed by a second, longer period, during which the first version of the new subject coding scheme will be developed with close collaboration from its potential users.

Questions

We would like to collect information about your organisation's requirements for a new subject coding scheme for higher education, and its potential impact on you. We are interested in what you do now with subject-based information, how you do it, and why you do it, as well as your future requirements. We would like you to tell us about the potential impact of a new scheme on your organization's policies, planning, development, operations and information outputs. We would also like to know your views on how the new scheme should be controlled, maintained, and developed.

We've divided the questions into 4 groups:

- current usage
- future requirements
- impact
- governance.

Please treat these groups and the questions as a framework for discussion. If there are important points you'd like to raise that aren't specifically covered, please don't hesitate to make them, even if they might seem obvious or

you think they might be covered by others. We'll also ask you if there are any additional comments you'd like to make at the end.

One issue we are addressing is whether a new scheme should be subject or discipline based. Rather than constantly repeating 'subject, discipline, topic' in each question, we use the term 'subject' to encapsulate all these concepts.

Current usage of subject classification

- How does your organization currently use subject listings, coding systems or classification schemes, including the existing JACS3 scheme? A brief overview of your processes that use them would be helpful.
- What issues, if any, do you have at the present time with using these schemes?

Potential future requirements

Looking at your organisation's potential use of information about subjects in a more perfect future world:

- a. how might you wish to make use of subject information in the future?
- b. from your organisation's perspective what are the most important characteristics of a new subject coding scheme?
- c. how much detail would you need from a subject coding scheme?
- d. would you like the scheme to be based on academic disciplines, subjects of study, or something else? Please explain your reasons.

Impact

How would a new subject coding scheme affect your organisation's:

- a. policies and planning?
- b. systems development?
- c. operations?
- d. analysis and outputs?

Governance

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?

Additional

Any additional comments?

5 HEP Workshop Plan

New Subject Coding Scheme workshops

The purpose of the workshop is to gather initial input from the sector on the major design decisions for a new subject coding scheme (NSCS) for UK higher education. This event will also be an opportunity to discuss what the NSCS needs to be able to do for you.

Aims:

- to further develop the vision for a subject coding system that meets a wide variety of user needs
- to gain an HE provider view on some of the fundamental design decisions behind the creation of a new subject coding system
- to explore how basic design decisions might impact on the day-to-day work of staff in different functional areas
- to explore how basic design decisions might impact on HE providers' statistics, analytics, reporting and business intelligence
- to begin to explore systems implications of a new subject coding system
- to elicit HE provider views on the governance of a new subject coding system

Morning

10.00 Coffee and registration

Charlie to keep register

10.30 Welcome and introduction to HEDIIP and the NSCS

Gill Ferrell outlines the NSCS work and its place in the wider HEDIIP initiative

10.45 Discussion led by Gill

11.00 Black hat exercise led by Gill.

Objective: a list of JACS3 negatives

Requires smaller groups of about five each

Requires post it notes and black hat posters

11.15 User stories

Alan and Charlie Paull lead a small group activity that elicits what each stakeholder needs from the NSCS by constructing a fictive person and walking through what she needs

Objective: a described persona - individual who represents a typical user

Requires 3 small groups of about 4-5:

MCR: Planning, Recruitment & Admissions, Student Records.

LON: Planning, Student Records, Recruitment & Admissions

EDI: Planning, Student Records, Marketing

Gill and Wilbert to scribe

Requires magic whiteboard

12.30 Lunch

Afternoon

13.30 Initial questions

Wilbert Kraan leads a presentation and discussion of the most fundamental design questions of the NSCS, and how to answer them. Participants are encouraged to read through the questions beforehand and pick their favoured solutions.

Objective: one persona's choice out of 3/4 solutions to each of 4 questions, recorded with motivation

Requires the reconstitution of the user story groups

Requires initial question hand-outs

Alan, Charlie, Gill to scribe

13.35 Question 1: Sharing subjects or sub-disciplines and fields. Introduction

13.40 Persona group discussion of choice of solution and its motivation

13.50 Plenary round up of answers

13.55 Question 2. Discipline, subject, or both? Introduction

14.00 Persona group discussion of choice of solution and its motivation

14.10 Plenary round up of answers

14.15 Question 3. Making the NSCS easy to enter- how many top level terms? Introduction

14.20 Persona group discussion of choice of solution and its motivation

14.30 Plenary round up of answers

14.35 Question 4. How to deal with subjects not listed? Introduction

14.40 Persona group discussion of choice of solution and its motivation

14.50 Plenary round up of answers

14.55 Introduction of remaining questions online

15.00 Discussion

Gill Ferrell leads a general plenary discussion of some of the issues raised during the day and any other issues.

Topics to be covered include:

Governance

Backward compatibility

Linking to other vocabularies and datasets

15.45 Coffee and networking

16.00 Close

6 Full Requirements Matrix

The following list represents all requirements as they've been distilled from all responses that the NSCS team have received so far.

Note: the IDs of each requirement are not in sequence because some requirements have been merged or dropped in the process of requirements gathering.

ID	Title	Description	Link to design goal(s)	Stakeholder(s)	Rationale	Priority	Type	Grouping
R1	NSCS and JACS3	The NSCS system shall support, and provide guidance on, use of the NSCS with data classified with JACS3, for example for maintenance of time series.	1, 6, 14	analysts	Backwards compatibility for time series and easy transition.	Mandatory	constraint	Uses
R2	Interoperating with other subject vocabularies	Publishing mechanisms shall be designed so that the NSCS can be linked to and interoperate with other classification systems, including amongst others the Learndirect Classification System (LDCS).	11	SFA, HEFCE, HESA, Jisc	Support linking to other data. Widen usage across the sector. Enables development of services and applications with multiple data sets.	Mandatory	constraint	Uses
R3	Persisting URIs	A persistent URI for each of its entities shall be included, so that they can be addressed readily by systems using linked data.	2, 4, 5, 11	UCAS, HESA	Good practice. Supports open data.	Mandatory	functional	Codes and structure
R4	Replacing JACS	The NSCS shall replace all current uses of JACS3.	3	UCAS		Mandatory	constraint	Uses
R7	Having clear and concise definitions	The NSCS shall have clear and concise definitions for each of its terms. Where possible, these definitions should be drawn from an appropriate authority recognised as such by the HE sector.	1, 2, 10	classifiers, analysts, HEE	Aids understanding for analysis and classification	Mandatory	performance	Content

ID	Title	Description	Link to design goal(s)	Stakeholder(s)	Rationale	Priority	Type	Grouping
R8	Classifying subjects	The NSCS shall be used to classify HE data by subject of study, while supporting aggregation for usage via discipline, including courses throughout their lifecycle.	3, 7, 12, 15	academics	Provides mechanism for using disciplines within a subject framework	Desired	constraint	Uses
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.	3, 6	HEPs, sector bodies	Provides strong sector representation on development.	Mandatory	performance	Governance
R10	Providing guidance on coding for specific purposes	The NSCS shall include guidance on how codes are to be allocated with reference to specific purposes. Methods may be different for different purposes.	2, 3, 5, 6, 13	UCAS, classifiers, SFA	Recognises that the scheme sits within a 'service' implementation approach.	Desired	functional	Guidance and help
R11	Providing training recommendations	The NSCS documentation shall include recommendations for training in how to use the scheme.	2, 3, 6, 9, 11, 13	UCAS, classifiers	Recognises that the scheme sits within a 'service' implementation approach.	Mandatory	functional	Guidance and help
R12	Comparing degree programmes	The NSCS shall facilitate comparisons between programmes by applicants and advisers.	2, 3, 6	UCAS	Recognises importance of subject comparisons at programme level	Desired	functional	Uses

ID	Title	Description	Link to design goal(s)	Stakeholder(s)	Rationale	Priority	Type	Grouping
R13	Supporting operational and time series statistics	The NSCS shall enable production of useful operational and time series statistics by HESA, UCAS and others, that are compatible with JACS3-based statistics at JACS3 Principal Subject level (for example student progression rates, staff-student ratios, applications, acceptances, and so on).	3, 4, 6, 7, 8, 9, 10, 11, 14	UCAS, HESA, PSRBs, HEFCE, plus many others	Continuing requirement for planning and analysis of data by subject. Implies major thrust of requirement at roughly JACS3 Principal Subject level.	Mandatory	functional	Uses
R14	Supporting regulated professions	The NSCS shall ensure that specific subjects can be catered for, including subjects directly relevant to regulated professions, bearing in mind the requirements for codes and for support in correct usage: 'teaching' and teaching subjects, such that 'teaching' and individual subjects that are taught in teaching programmes can be analysed; pharmacy; healthcare sciences; planning;	3, 6, 7, 10, 11	UCAS, HEE, PSRBs, HEFCW	Recognises a particular problem area that could usefully be resolved.	Desired	functional	Content
R15	Supporting service oriented approaches to publishing on the internet	The NSCS shall be capable of deployment using a service oriented approach.	3, 5, 11	UCAS	Recognises that the scheme sits within a 'service' implementation approach.	Desired	performance	Codes and structure
R16	Supporting multiple aggregation structures	The NSCS shall support multiple aggregation methods, for example aggregation for league tables, for application statistics, for HEP planning purposes.	3, 6, 7	HEPs, UCAS, HEFCE	Supports more usages than JACS3.	Desired	functional	Codes and structure
R17	Providing support for course searching	The NSCS shall provide subject classification as the starting point for course search and marketing purposes.	3, 10, 13	UCAS	Supports more usages than JACS3.	Desired	functional	Uses

ID	Title	Description	Link to design goal(s)	Stakeholder(s)	Rationale	Priority	Type	Grouping
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.	3, 4, 5, 9	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.	3, 4, 5, 9, 14	UCAS, HESA, HEPs	Must be stable and robust, but also capable of change	Mandatory	constraint	Governance
R20	Enabling statutory returns	The NSCS shall form the basis of subject-based statutory and regulatory returns by HEPs to sector bodies.	3, 4, 6, 7, 8, 9, 10, 11, 14	HESA, UCAS, HEPs, sector bodies, PSRBs	Continuing requirement for planning and analysis of data by subject.	Mandatory	functional	Uses
R21	Supporting student lifecycle comparisons	The NSCS shall support comparisons of what students study and what they progress to doing later, e.g. occupation	3, 6, 11	HEE	HEE has a specific requirement to do this in respect of students who go on to NHS employment	Desired	functional	Uses
R22	Enabling mapping to NHS occupation codes	The NSCS shall have a mapping to NHS Occupation Codes, in order to support NHS workforce planning.	3, 11	HEE	HEE requirement	Desired	performance	Uses
R23	Enabling workforce and capacity planning dataset comparisons	The NSCS shall enable comparisons across UCAS, HESA, HEP and HEE data sets for capacity and workforce planning and for quality assessment.	3, 6, 7, 11, 14	HEE, SFA	HEE requirement	Desired	functional	Uses
R24	Enabling disaggregation in health subjects	The NSCS shall enable differentiation between critical health-based subjects.	3, 6, 10, 11	HEE	HEE requirement	Desired	constraint	Uses

ID	Title	Description	Link to design goal(s)	Stakeholder(s)	Rationale	Priority	Type	Grouping
R25	Enable aggregations for NHS planning	The NSCS shall permit aggregation for NHS planning purposes (workforce, capability and quality assessment).	3, 6, 10, 11	HEE	HEE requirement	Desired	constraint	Uses
R26	Providing more detail in medicine subjects	The NSCS should have more detail in healthcare science and medical specialisms than JACS3 to facilitate usage within the NHS and health professions.	6, 7, 11	HEE, PSRBs	HEE requirement	Best value	performance	Codes and structure
R28	Providing guidance on the NSCS and KIS	The NSCS should give clear guidance on how it should most usefully be included in the Key Information Set.	6, 7, 10, 11, 14		Covers existing requirement for use of JACS3	Mandatory	functional	Guidance and help
R29	Facilitating datasets that are fit for purpose	Governance shall facilitate the creation, maintenance and usage of authoritative data sets.	3, 6, 9, 11, 14	GPC, HEPs, HEE, other sector bodies	Provides for current and wider usage in analysis via HESA, HEFCE and others.	Mandatory	performance	Uses
R31	Describing guidance purposes clearly	NSCS guidance shall clearly describe the purposes for which it is designed to be used. It will also cover similar areas for which it is not designed to be used.	1, 2, 3, 6	HEPs, SFA	Competent usage for classification and analysis requires this.	Mandatory	performance	Guidance and help
R32	Encouraging clarity in the description of data collection purposes	NSCS guidance shall encourage data collectors to describe clearly the purposes for which the subject-classified data will be used.	1, 2, 3, 6	HEPs	Competent usage for classification and analysis requires this.	Mandatory	performance	Guidance and help
R33	Supporting aggregation of STEM and SIV course data	The NSCS shall support unambiguous aggregation of data for STEM and SIV subjects.	1, 2, 4, 5, 6, 7	HEPs	Important for critical policies using the data.	Desired	performance	Uses

ID	Title	Description	Link to design goal(s)	Stakeholder(s)	Rationale	Priority	Type	Grouping
R34	Supporting more fine grained classification of subjects	The NSCS shall support more fine-grained classification of subjects of study forming a separate coding frame, for example for modules or educational resources. This supports policy interventions.	3, 7, 10, 11	HEPs, HEFCW. HEFCE	Certain functions require module level classification, eg HEFCW funding model, reading list creation. HEFCE: nuclear technology, Islamic Studies, and specialisms at Masters.	Desired	functional	Codes and structure
R35	Making codes memorable	NSCS codes should be memorable, but should not encourage the use of the code as shorthand for the term itself.	1, 2	HEPs	Facilitates usage by classifiers	Best value	performance	Codes and structure
R36	Classifying subjects or groups of subjects	Each term in the NSCS shall be a subject of study or a cognate group of subjects of study.	1, 2, 6, 7, 10, 13	HEPs	Facilitates design and maintenance	Desired	functional	Content
R37	Supporting multiple, combined or interdisciplinary subjects	The NSCS shall include guidance on how to classify multiple, combined and interdisciplinary subjects.	1, 2, 7, 10	HEPs		Mandatory	functional	Guidance and help
R38	Excluding unstudied subjects	The NSCS shall not include subjects that are not subjects of study in programmes in the UK.	1, 2, 7, 8, 10, 13		Facilitates design and maintenance	Desired	constraint	Content
R39	Using common labels	NSCS terms shall use labels for subjects of study that are commonly used names within the subject area or discipline.	1, 2, 7, 8, 10, 13		Facilitates design and maintenance	Desired	constraint	Content

ID	Title	Description	Link to design goal(s)	Stakeholder(s)	Rationale	Priority	Type	Grouping
R40	Differentiating term definitions	Definitions of terms in the NSCS should not be confusingly similar.	1, 2, 4, 11, 13	HEPs	Facilitates usage by classifiers	Mandatory	constraint	Content
R41	Avoiding leading zeros	NSCS codes should not have leading zeros.		Project team	Facilities technical implementation	Best value	constraint	Codes and structure
R42	Using a consistent number of characters	NSCS codes shall have a consistent number of characters.	2	HEPs	Facilities technical implementation	Desired	constraint	Codes and structure
R43	Providing support documents	The NSCS shall have supported documents, such as guidance manuals, subject coding manual, context-sensitive help, scope notes within terms.	2, 3, 9	HEPs	Facilitates uptake of the schema.	Desired	functional	Guidance and help
R45	Including external definitions of important subjects	The NSCS shall include as terms those subjects of study included in SIV and STEM definitions, and other similarly recognised lists of important subjects. SIVS: chemistry, engineering, mathematics and physics; quantitative social science; and modern foreign languages and related area studies. STEM: anatomy and physiology; biosciences; chemistry; computer sciences; earth, marine and environmental sciences; engineering and technology; mathematics; pharmacy and pharmacology; physics	3, 6, 7, 11	HEPs, HESA, HEFCE, SLC	Important for critical policies using the data, including student finance policy	Mandatory	constraint	Content
R46	Supporting the classification of subjects not already included	NSCS shall recommend a mechanism to support the classification of subjects of study not currently covered by the schema.	5	HEPs	Facilitates maintenance	Desired	functional	Guidance and help

ID	Title	Description	Link to design goal(s)	Stakeholder(s)	Rationale	Priority	Type	Grouping
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.	3, 6, 7	HEPs	Supports use outside HEPs	Desired	constraint	Codes and structure
R48	Supporting specific HEP functions	NSCS shall support benchmarking and analytics for internal and external use of HEPs, as well as performance management, competitor/sector analysis and market intelligence.	3, 6	HEPs	Supports use inside HEPs and by HEPs	Desired	functional	Uses
R49	Structural position of Welsh as a subject in the scheme	Ensure Welsh is in a language category, not Celtic Studies.	6, 13	HEFCW	Improves on JACS2 and 3	Mandatory	constraint	Content