



Classifying subject of study

A roadmap to a new Joint Academic Coding System

July 2013

About this report

This report was commissioned by the Regulatory Partnership Group as strand 4a of the broader project **Redesigning the data and information landscape**; it was delivered by Gill Ferrell in March 2013. Strand 1 of this project led to the creation of HEDIIP and the subject classification work is now being taken forward by HEDIIP. This report was published on the HEDIIP web site in July 2013.

Contents

1	Executive summary	4
1.1	Current position	4
1.2	Vision	4
1.3	Recommendations	4
1.4	Acknowledgements	5
2	Project aims, scope and approach	6
2.1	Aims and Scope	6
2.2	Approach	6
3	JACS background and revision history	7
4	Uses of subject classification	10
4.1	Supporting policy implementation	10
4.1.1	Funding	10
4.1.2	STEM and SIV Monitoring	10
4.1.3	Widening participation	11
4.1.4	Supporting progression in HE	11
4.2	Regulation and accountability	12
4.2.1	Funding eligibility/audit	12
4.2.2	Quality Assurance	13
4.2.3	Academic Technology Approval Scheme	13
4.3	Providing information to students	13
4.3.1	Marketing and recruitment	13
4.3.2	HEAR	14
4.4	Benchmarking/performance indicators	14
4.5	Learning and teaching/resource discovery	15
4.6	Supporting research	17
4.7	Public information	17
4.8	Open/Linked Data	18
5	Designing a new subject classification	19
5.1	Characteristics of the existing coding framework	19
5.2	Comparison with other related classification schemes	20
5.3	Level of granularity	21
5.4	The relationship between subjects and other attributes	23
5.4.1	Subject and discipline	23
5.4.2	Subject and cost centre	27
5.4.3	JACS and UCAS	29
5.5	Linked/Open Data and Semantic Approaches	29
6	Next steps	32
6.1	Implementation	32
6.2	Ongoing Maintenance	32
7	References	33
8	List of appendices	35

1. Executive Summary

1.1 Current position

Subject coding is one of the key building blocks of student data that ought to have sector-wide and consistent applicability. The current method of classifying subjects is based on a scheme known as the Joint Academic Coding System (JACS). JACS has been in use for a number of years and requires a major review because:

- 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.

1.2 Vision

The vision is for a revised subject classification that can form the basis of a common language across many datasets and which is sufficiently flexible to meet multiple stakeholder requirements. At its base level it should be sufficiently simple to be applied consistently and sufficiently high level to maintain stability over a long period of time. It should however take full advantage of developments in technology that will provide the ability to relate other keywords to the controlled vocabulary and facilitate linking of data from various sources and between classifications that are optimised for different purposes.

This report proposes what appears to be an optimal solution in value for money terms that will deliver the following benefits:

- 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;
- increased consistency of application across institutions;
- a means of furthering the open data agenda;
- a unified approach to supporting a much wider group of stakeholders;
- a means of linking to data classified in other frameworks;
- backwards compatibility with legacy data.

1.3 Recommendations

Approaches to classifying and sharing data have evolved considerably since the JACS classification was first conceived. Many of the newer approaches are aimed at finding ways of discovering, sharing and making sense of, the vast body of unstructured information that exists on the World Wide Web. The Semantic Web (sometimes called Web 3.0) is a term used to describe the ways in which technology can be used to make more intelligent connections between data elements that exist in different places. One concept of relevance is that of Linked Data (there is no generally accepted definition but the term refers to publishing and connecting structured data on the Web).

The concepts are of relevance to the work in hand because they may offer a pragmatic means of meeting the needs of a variety of different stakeholders. Rather than design an all-encompassing classification system that meets the needs for different levels of detail and handles some quite specialised vocabularies, it may be possible to solve the problems one at a time. It could, for example, allow us to make a, relatively straightforward, JACS replacement work effectively with specialist vocabularies in use in research and the world of medicine as well as relating it to the more general keywords that prospective students might use when undertaking a Google search for information.

This report therefore makes a number of recommendations some of which are intended to address issues with the current classification system and some of which are intended to facilitate data sharing and linking in the future.

The specific recommendations in relation to the development of a new framework are as follows:

- the new framework should recognise the, currently implicit, assumption (at least by HEIs) 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;
- the new framework should consist of three rather than four levels;
- the new framework should be based on a six digit coding structure;
- the new framework should provide a persistent URI¹ (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;
- a project to implement the new coding framework will require a clear and targeted strand of communications activity.

Other points.

- Consideration should be given as to whether the levels of the new classification should be explicitly defined and named e.g.
Level 1 - Discipline
Level 2 - Broad Field of Study (or sub-discipline)
Level 3 - Narrow Subject (or Topic) of Study.
- Consideration should be given as to how the levels of a new classification should be articulated i.e. whether level 3 needs to have a 1:1 relationship with the upper tiers or not.
- Consideration should be given as to whether the new coding framework should be renamed in order to make a clear distinction between the new approach and any previous iterations of JACS.
- Although outside the scope of this particular piece of work, it would be helpful if future iterations of the sector level data model, lexicon and thesaurus (strand two of this programme of work) could seek to ensure that the attribute 'subject' is defined to avoid ambiguity and confusion with other terms such as discipline and cost centre.

1.4 Acknowledgements

The project team would like to thank the many individuals and organisations who have contributed to the research behind this report including the staff of HEFCE, HEFCW, HESA, NHS, NUS, QAA, RCUK, SFC, SLC, TA, UCAS and the University of Oxford who made themselves available for interview; staff of the Bodleian libraries and discipline leads from 13 of the HE Academy's discipline areas who contributed views and the staff of over 80 separate HEIs who participated in the survey and/or in informal discussion sessions on the topic as part of other events.

Particular thanks go to CETIS for their contribution to the appraisal of the technical options and to staff from HESA, HEFCE and UCAS who reviewed earlier drafts of this report.

This report remains a starting point: the main findings reflect broad consensus amongst the contributors and we have tried to identify areas where further work is needed to evaluate the merits of alternative possible approaches.

¹ A URI is a globally unique and unambiguous identifier that can be used in any system without fear of conflicting with other identifiers. This is explained further in section 5.5. The term is often confused with the other common, but informal, term URL (uniform resource locator).

2. Project Aims, Scope and Approach

2.1 Aims and Scope

Subject coding is one of the key building blocks of student data that ought to have sector-wide and consistent applicability. Project B however identified that the JACS coding framework, currently used in the core HESA and UCAS systems, is in need of revision (having reached the limit of the current 4-character structure) and is not meeting the requirements of some data collectors – most notably the NHS and Research Councils.

Building on the experience of using JACS versions 1, 2 and 3, the aim of the project was to establish requirements for a sector-wide subject coding system and develop a framework that facilitates future development of the system and use by a broader set of stakeholders. The project initiation document identified the following as the key deliverables.

- Analysis of requirements for a sector-level subject coding system
- Recommendations for a coding structure design
- Recommendations of next steps

2.2 Approach

The initial focus of stakeholder engagement in this project was with Regulatory Partnership Group member organisations in order to determine their perspectives on the issues and the practicalities and timescales of implementing change. We have however also consulted with CETIS, the HE Academy and over 80 HE providers (79 HEIs responded to a survey and others were involved in a mini workshop and other informal discussions) as well as looking at a range of international comparators.

Early discussions indicated general agreement amongst key stakeholders that it is time for a root and branch review of the JACS system (not least because it is widely recognised that the group of key stakeholders is now far wider than the originators of the system). The project team has taken the view that such a review should not be constrained by approaches that have been taken in the past but nor should it stray too far from the pragmatism that has characterised the previous evolution. The idea is to meet the needs of a wider group of stakeholders and to future proof the system whilst ensuring that the cost of any change is commensurate with the benefits. With this in mind the following principles have guided the proposals in this report:

- explore a wide set of current and potential uses for JACS whilst accepting that some use cases must take greater priority than others
- make every effort to future proof the system whilst having due regard for the value of existing time series
- endeavour to take account of current and likely trends in technology whilst attempting to minimise the cost of transition

On first reading this may sound as if the project is hamstrung by starting with a series of compromises. In fact there were no such constraints imposed at the outset and indeed the principles stem from looking at a wide range of options and realising the risk that seeking to deliver the most all-encompassing and, technically elegant, solution could be expensive and technically complex, perhaps even tantamount to attempting to 'boil the ocean', and of unproven value.

The project is therefore proposing what is believed to be an optimal solution that will deliver the following benefits:

- 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
- increased consistency of application across institutions
- a means of furthering the open data agenda
- a unified approach to supporting a much wider group of stakeholders
- a means of linking to data classified in other frameworks

3. JACS Background and Revision History

The Joint Academic Coding System (JACS) is used UK wide to describe subjects of study at the course/programme and module level. The coding bears no relationship to level of study so can be used at undergraduate, postgraduate, research and continuing education levels as well as for FE delivered in HE.

JACS was developed through a collaborative effort between HESA and UCAS who previously operated two similar, although different, subject coding systems: HESACode and Standard Classification of Academic Subjects (SCAS). The development was explicitly framed as a 'merger' between HESACode and SCAS. This point is important in understanding the characteristics of JACS: it never started with a 'clean sheet of paper' and many of the constraints of the current system relate to its inherited characteristics. The first operational version of JACS (v1.7) was published in 1999 and became operational for entry to HE in 2002 and HESA reporting for 2002/03.

JACS codes are formed from a letter and three digits: the letter indicates the subject area; the numbers indicate the topics within the main subject area. The coding framework appears hierarchical thus:

V Historical and philosophical studies			
	V100 History by period		
	V200 History by area		
	V300 History by topic		
	V400 Archaeology		
		V470 Classical art & archaeology	
			V471 Roman art & archaeology

Courses are therefore likely to be coded at levels 2 and 3 of the classification whereas individual modules may be coded at the finest level of detail. Key stakeholder however differ in the extent to which they view the classification as a rigid hierarchy or one which is merely 'implied'.

A course may have between one and three JACS codes depending on whether it is classed as a single subject, a balance of two equally weighted subjects, a major/minor combination, or a triple combination of three equally weighted subjects. A JACS code can only ever be in the 'letter plus 3 numbers' format.

The evolutionary nature of subject disciplines means that any classification will require regular review. JACS was reviewed in 2005 and JACS version 2.0 was introduced for 2007 entry and the reporting year 2007/08.

A further review began in 2008 resulting in version JACS 3.0 which was introduced for 2012 entry and the 2012/13 HESA reporting year. The review was deliberately limited in scope and changes made only in those few areas where it was anticipated that they would significantly improve the fitness for purpose of the classification to an extent which was judged to justify the cost of change.

A couple of example of changes made during reviews may serve to highlight some of the issues with the current coding framework and potential issues to consider when implementing any revised system.

Group G (Mathematical and Computer Sciences) was identified as an area requiring review for JACS 3.0 because Mathematics and Computer Sciences are both large academic areas and, as Computer Sciences in particular are continually evolving, there was little room to expand. The decision was taken to split them into two different subject areas and to allocate the previously unused letter 'I' to the new Computer Sciences classification. This permitted some expansion in Mathematics (without any reuse or renaming of codes which is poor data management practice and, hence, explicitly prohibited in JACS) whilst supporting the larger of the growth areas. This has solved an immediate problem but the capacity for this type of amendment is extremely limited and there are other subject areas which will ultimately run out of codes.

The creation of Group X (Education) in JACS 1.7 represented a fundamental departure from the earlier HESAcodes system which assumed that any course coded to X1 represented initial teacher training (ITT). JACS represented an improvement in data terms, as it did not seek to embed a qualification aim within the subject code, the change however did mean that automated mapping to ease the transition was not possible. In implementing any new system we would want to limit the amount of manual adjustment that needed to be made but, as this example illustrates, this is not always possible where a shift is significant.

4. Uses of subject classification

This section looks at the existing uses of JACS but also covers new areas where stakeholders have indicated they feel a revised JACS classification system could be helpful to them or where stakeholders are currently applying alternative approaches to meet perceived shortcomings in the existing system. It is not intended to be a comprehensive analysis of all of the current and potential uses and, in line with the project brief, it focuses in on some aspects, particularly the potential uses of JACS to the NHS and in supporting medical education and classifying research activity.

Analysis of the main requirements reveals that they fall into distinct groupings which demand different characteristics of a classification scheme:

- the main areas of support for funding, regulation and accountability demand a system that is sufficiently simple to be applied consistently and sufficiently high level to maintain stability over a long period of time;
- support for learning and teaching practice and retrieval of information by students and potential students, demands a greater level of detail and the ability to relate other keywords to the controlled vocabulary;
- support for coding research activities and for using HE information in other domains relating to the broader economy and society (as well as some in-sector use by institutions) demands a system that is sufficiently open (in the technical sense) to facilitate linking of data from various sources and between classifications that are optimised for different purposes.

It is not impossible to develop a classification that meets all of these criteria provided certain fundamental definitions can be agreed. This is discussed further in the following section.

4.1 Supporting policy implementation

4.1.1. Funding

In the past most teaching funding throughout the UK was allocated on a subject basis. Alongside the shift from block grant to tuition fees there have been other changes in funding approaches. Currently only HEFCW still operates a funding policy based on academic subject categories (ASCs). HEFCE has moved to a more cost-based model with a concentration on Science, Technology, Engineering and Mathematics (STEM) and Strategically Important and Vulnerable (SIV) subjects and from 2012/13 the SFC model is also based on price groups. Although JACS does not directly underpin teaching funding models to the same extent that it did previously, there is still a relationship of which institutions are very conscious when approaching classification. This is discussed further in the section on subjects and cost centres.

Administration of the student loans scheme does not currently rely on JACS at all and, indeed, the Student Loans Company does not currently collect JACS information. However, the SLC is increasingly finding that government information demands, mainly in the form of parliamentary questions, cannot be met without recourse to some form of subject classification. What this means in practice is that the SLC has been obliged to generate a set of algorithms in its own systems to 'recreate' the JACS code. This is clearly an unsatisfactory state of affairs and the SLC is keen to ensure that the use of JACS, or its successor, is built into its future plans.

4.1.2 STEM and SIV Monitoring

Monitoring of STEM and SIV subjects is a key use of JACS for the HE funding bodies and there is likely to be an ongoing requirement for the promotion and protection of certain subjects in spite of the move to a more marketised HE system. Whilst STEM subjects are an established group, there may be differences in emphasis across the four nations of the UK e.g. the list of 'subjects of broader importance to Wales' is narrower than HEFCE's STEM list. DELNI does not identify any supply-side problems with STEM at the present time (DELNI 2011) but is concerned about low employment returns on STEM education investment.

With regard to SIV subjects however, current events can change the emphasis relatively quickly e.g. bringing subjects such as Islamic studies or climate change to the fore. This does not necessarily occur in any kind of predictable fashion but once an entity is defined as a SIV subject there is likely to be an interest in time series as well as one-off analysis. There are again national aspects to this, e.g. Gaelic is important to the SFC and the lack of a top level code for Welsh causes problems for HEFCW, but the broad issues are the same. SIV monitoring may differ from other types of use in the broad category of supporting policy in that it is an example of where greater detail and the ability to link across data sources may be beneficial. HE providers have indeed expressed some scepticism about the use of JACS for SIVs analysis and it has been suggested a number of times that semantic approaches (more on this in section 5.5) may be a better means of meeting this requirement in future.

4.1.3 Widening participation

Currently JACS does not appear to feature prominently in widening participation monitoring by the regulatory bodies other than its use by QAA to compile statistics on the Access to HE Diploma.² HE providers are however conscious of a relationship between an institution's subject profile and its likely student profile hence this is a feature of their own benchmarking activities as well as of HESA's performance indicators (more on this below). Action on Access³ (2009) reports that universities have targets to address gender imbalances in certain subjects i.e. males under-represented in some areas and females under-represented in STEM although it goes on to say **'Although a wide range of retention interventions is identified, very few subject-specific references are made, with a couple of notable exceptions. Nursing and related courses are identified by a few institutions as subject areas where specific retention interventions are being developed.'**

4.1.4 Supporting progression in HE

Although the English Lifelong Learning Networks⁴ (LLNs) are no longer funded by HEFCE there remains a need to articulate progression routes into higher education from further education and other vocational backgrounds. Some LLNs continue to operate funded by other means and other initiatives exist in other parts of the UK such as the series of Scottish Articulation Hubs⁵ funded by the SFC. All of these bodies appear to present course information, on a subject basis, in a way that does not correspond to the JACS classification. The section on designing a new subject classification looks at approaches that could help link higher education information to other data sets in ways that could be strategically useful for monitoring progression as well as for looking at other aspects of impact on broader economy or society.

An example of bottom up development in this area is work being undertaken by Plymouth College of Art and ULCC to support student progression from undergraduate to postgraduate study. The 'Moodle block⁶' project is creating an open source application that will be available to any institution using Moodle (the most widely used virtual learning environment (VLE) in the sector). As students are required to engage with the VLE as part of their programme of study this provides the opportunity to target information about further study opportunities at the appropriate point in their course. The application relies on having course information structured in a particular format (known as XCRI-CAP) and JACS is the ideal controlled vocabulary to manage the subject element of this.

² <http://www.accesstohe.ac.uk/>

³ <http://actiononaccess.org/>

⁴ LLN <http://www.hefce.ac.uk/pubs/rereports/year/2010/llnsummevaln/#d.en.64131>

⁵ <http://www.ggap.org.uk/aboutus/hubs/>

⁶ Moodle Block <http://www.jisc.ac.uk/whatwedo/programmes/elearning/coursedata/demonstrators/moodle-block.aspx>

4.2 Regulation and accountability

4.2.1 Funding eligibility/audit

Although not a core component of the teaching funding model, JACS is used by HEFCE as a means of cross referencing and auditing data provided by institutions. This is discussed further in the section on subjects and cost centres. In our survey 10% of respondents said JACS played a part in internal resource distribution mechanisms within HEIs. An interesting observation has also been made in relation to the role JACS may play in attracting overseas students.

'The current JACS codes are proving to be a bargaining chip, particularly at research student level and particularly with overseas students whose sponsors want them to study in a certain sub discipline area.' (HE Academy Discipline Lead)

The examples below highlight some potential additional uses of JACS.

The NHS does not currently use any standard coding framework for the subject of study: its main interest is in the final qualification which renders the student eligible to register with the relevant professional body (who validated the course) and hence to work in the profession and adopt one of the standard professional titles (regulated by the professional bodies).

The NHS has however indicated that a standard subject nomenclature could be extremely useful in discharging its responsibility for the management of bursary funding and in interrogating standard student information collected by HESA. The NHS (in England only) operates a bursary scheme that covers pre-registration nursing, midwifery, medicine, dentistry and a number of the allied health professions. The student loan element is funded by the NHS but administered by the SLC. The NHS is responsible for indicating which courses are eligible for bursaries and there is currently an issue in that the NHS collects data from the learning providers in the form of free text course titles and then has to interpret those titles. If the learning provider was obliged to supply some form of agreed subject classification, it has been suggested that this would leave the NHS bursary unit in no doubt as to whether the programme was eligible for a bursary as titles such as 'Foundations of professional practice' can currently cause confusion. Further work is needed to define whether subject of course is itself sufficient to determine eligibility or whether this would need to be viewed alongside professional body accreditation status.

The introduction of a standard nomenclature would require agreement from the Department of Health, Health Education England, Local Education and Training Boards (LETBs) and the regulatory bodies and is therefore not a trivial matter although it does not appear as if the addition of a subject classification code would have major resource implications in relation to the work needed to amend legacy information systems. The introduction of a standard framework would simplify administration for learning providers and could also offer direct benefits to prospective students e.g. through adoption by NHS Careers⁷ and making a direct connection between this information and the Key Information Set (KIS). Such a development would require careful consideration of the attributes involved so that the subject classification is not compromised by trying to accommodate other attributes e.g. the registerable qualification. The various recognised classifications widely used in the world of health form the logical starting point for the development of an appropriate vocabulary and a suggested way to progress this is discussed further in the section on comparison with other related classification schemes.

The Teaching Agency (TA) does not make any particular use of JACS as its main interest is in qualifications on entry rather than any breakdown of specialisms once people achieve qualified teacher status. The TA does however have an issue, not dissimilar to that described above in relation to the NHS, in establishing whether some learners with postgraduate and non-traditional qualifications are eligible for a targeted bursary. In this case, the problem is rather more difficult to resolve as it would involve subject coding of the entry qualifications at the point of application. Whilst the requirement can be noted, and cross referenced against strand 4b of this work stream (which looks at potential uses of the Personal Learning Record), it is unlikely to be met in the medium term. Lest this requirement be viewed as an 'outlier' it should however be noted that the HESA Performance

⁷ <http://www.nhs Careers.nhs.uk/>

Indicators Steering Group has recently investigated whether its benchmarks take sufficient account of admissions policy/qualifications on entry⁸. Although this analysis concluded that inclusion of A level subject information within the performance indicator benchmarks was not justified⁹ it is further evidence of interest in this area. The TA has also indicated that it would like to see some further work in relating entry qualifications to JACS codes in order to be able to say how many people achieving qualified teacher status are qualified to teach a particular subject (for example have a prior qualification in physics). This requirement can readily be met from the existing HESA student record via a subject coding system used by the awarding bodies for level 3 qualifications but there would be advantages in full integration with JACS.

4.2.2 Quality Assurance

The move to a risk based approach to quality assurance means that JACS is of less immediate relevance to QAA than under the previous regime when the top level of JACS was used as the starting point for development of subject benchmark statements although, as noted above, it is still an important component of QAA's monitoring of the Access to HE Diploma.

4.2.3 Academic Technology Approval Scheme

JACS is used by the Foreign and Commonwealth Office (FCO) in the administration of the Academic Technology Approval Scheme (ATAS)¹⁰. ATAS is deemed to be an essential part of the UK's commitment to counter proliferation. Its aim is to help stop the spread of knowledge and skills that could be used in the proliferation of weapons of mass destruction (WMD) and their means of delivery. The scheme is designed to ensure that those applying for postgraduate study in certain sensitive subjects do not acquire knowledge that could potentially be used in WMD programmes. Where the student requires a visa to study in the UK and has been accepted onto a course with a JACS code on the ATAS list, they will have to apply to the FCO for ATAS clearance.

4.3 Providing information to students

4.3.1 Marketing and recruitment

JACS appears currently to be little used in directly student-facing contexts particularly with regard to marketing and recruitment other than in the Unistats website discussed in section 5.4.1. Despite UCAS being one of the main 'owners' of JACS, it has grown its own systems in other ways to better meet student needs. The UCAS site currently offers a course search facility for potential students. One of the options available is a 'keyword' search by subject. UCAS notes that: *'Recent analysis of unique IP address use of the UCAS Course Search indicates that the subject search is by far the most popular of the 3 search options currently available.'* The keyword search is currently based on a controlled vocabulary of 4500 keywords and does not correspond precisely to the JACS classification. UCAS is currently undertaking considerable development work to its systems and building on lessons learned from successes such as this which provide valuable pointers for the work of this project.

Similarly, other major course information aggregators in the sector e.g. Graduate Prospects¹¹ (the commercial arm of Higher Education Careers Services Unit (HECSU) owned by Universities UK and GuildHE), for postgraduate study and the National Careers Service¹² base their searches on subject groupings that do not correspond to JACS. In the survey carried out as part of this project, a number of institutions (24% of respondents) indicated that JACS played a role in their marketing. It is however assumed that this is as a behind-the-scenes classification because examination of all of the websites of those institutions who responded positively in this context did not reveal a single example where any form of course search facility matched exactly to JACS.

⁸ HESA PISG 2011 http://www.hesa.ac.uk/dox/performanceIndicators/PISG/PITG_2011_11/PITG_11_06.pdf

⁹ HESA PISG 2012 http://www.hesa.ac.uk/dox/performanceIndicators/PISG/PISG_2012_01/PISG_12_01.pdf

¹⁰ ATAS <http://www.fco.gov.uk/en/about-us/what-we-do/services-we-deliver/atas/>

¹¹ <http://www.prospects.ac.uk/>

¹² <https://nationalcareersservice.direct.gov.uk/Pages/Home.aspx>

It is unsurprising that the search terms used by prospective students, particularly those who may not have previous experience of higher education, may differ considerably from the 'expert' terms in a controlled vocabulary. It is unlikely, for example, that a prospective student interested in psychology would choose to search under biological sciences. A feature that might be considered for future iterations of JACS is to combine the controlled vocabulary with a thesaurus containing synonyms and related but non-preferred terms. This would add to the implementation effort and the maintenance overhead but, given the extent to which this effort is already being undertaken across the sector (as well as by those offering commercial services), may be a worthwhile investment. This would be a separate, but complementary, development that could be undertaken as part of the initial implementation or subsequent to it.

There are a number of starting points for this type of development including the work already undertaken by UCAS. The approach has also been successfully demonstrated in a Jisc funded project, known as AX-S Widget¹³, which has linked courses classified with JACS 3.0 codes to a specially constructed thesaurus. When a user performs a search, each retrieved record is ranked in the search results in accordance with how close its JACS subject is to the user's search term within the thesaurus which also matches broader and narrower linked concepts. Work by the University of Oxford on automated generation of keywords from text¹⁴ further demonstrates the potential for technology to ease this task.

Use of JACS within the Key Information Set (KIS) is discussed in the section on benchmarking/performance indicators.

4.3.2 HEAR

There appears to be equally little use made of JACS when providing student-facing administrative information. The most obvious example where one might expect to see JACS is applied is probably in filling in the 'Field of study' information in the Higher Education Achievement Report (HEAR¹⁵) which also fulfils the requirement to produce a European Diploma Supplement for each student. The glossary in the outline structure for the European diploma supplement provides this definition: FIELD OF STUDY - the main disciplines or subject areas of a qualification¹⁶ and the explanatory notes state: 2.2 Show only the major field(s) of study (disciplines) that define the main subject area(s) for the qualification e.g. Politics and History, Human Resource Management, Business Administration, Molecular Biology etc. In the survey conducted as part of this project only 11 out of 110 responding institutions replied that they use JACS for this purpose.

'... especially from the student point of view, it will be increasingly important for the JACS codes to speak to European equivalents in any HEAR document.' (HE Academy Discipline Lead)

4.4 Benchmarking/performance indicators

HESA's published performance indicators¹⁷ for the HE sector recognise the importance of subject mix in developing its benchmarks: '... a medical school and a college that mainly concentrates on engineering subjects are not comparable, as medical students have much lower non-continuation rates than engineering students.'

¹³ AX-S Widget <http://alanepaull.wordpress.com/category/xcri/demonstrators/ax-s-widget/>

¹⁴ <http://www.jisc.ac.uk/whatwedo/programmes/ukoer3/rapidinnovation/spindle.aspx>

¹⁵ HEAR <http://www.hear.ac.uk/home>

¹⁶ http://ec.europa.eu/education/lifelong-learning-policy/ds_en.htm

¹⁷ http://www.hesa.ac.uk/index.php?option=com_content&task=view&id=2072&Itemid=141

The importance placed on this by HESA may be one of the reasons why HE providers identified benchmarking and performance indicators as one of the main uses of JACS in our survey. Exactly 50% of respondents said that JACS played a part in their external benchmarking and c.36% said also used JACS in relation to internal benchmarking and performance indicators. Some 33% also said that JACS played a part in helping them identify market trends although others were more sceptical about its use for this purpose:

'We tried to use JACS in market trends, but it is not reliable, so we tend to go by programme titles instead.' (Survey response)

The use of JACS codes in compiling statistics from the National Student Survey (NSS¹⁸) and in various forms of league tables is included here. As discussed further below, JACS has taken on increased significance because of its application within the Key Information Set (KIS) and this has thrown up some interesting issues relating both to the classification itself and the way it is applied by different institutions.

In many cases there are transformations necessary before publicly produced statistics based on JACS can be meaningful in the individual institutional context.

'For benchmarking purposes, people want to see the data that relate to their own departments. So it is necessary to map JACS, NSS, league table subjects, cost centres, UoAs to the academic departments they are most relevant to.' (Survey response)

A brief search of some university planning documents available on the Internet suggests that impact on league tables is a consideration when providing guidance on how to apply JACS codes that are open to interpretation.

4.5 Learning and teaching/resource discovery

An authoritative and consistent subject classification ought to have a multitude of uses in relation to the development and sharing of resources to support learning and teaching activity. The educational benefits and cost efficiencies associated with the use of Open Educational Resources (OERs) have been widely acknowledged (e.g. White and Manton 2011) and many millions of pounds worth of public funding have been poured into development projects e.g. c.£14 million in the 3 phases of the Jisc/Higher Education Academy OER programme¹⁹ (2009-2012). JACS however appears to have played a very minor role in supporting these significant initiatives in data sharing to date.

In the survey that supported this research only 3 HEIs indicated that they used JACS codes to support the sharing of learning resources internally, only 3 used JACS in the management of a formal repository and only 1 indicated JACS use in sharing learning resources externally. This picture may be a slight under-representation as the target audience for the survey was biased towards those responsible for administration rather than direct learning and teaching support but it backs up anecdotal evidence from the learning and teaching community.

¹⁸ NSS <http://www.thestudentsurvey.com/>

¹⁹ OER Programme <http://www.jisc.ac.uk/whatwedo/programmes/elearning/oer.aspx>

One response from the management of some significant publication and research repositories indicated that they are considering implementing JACS within their repositories and that the ability to link to keywords and to map to other terminologies e.g. FAST²⁰ (Faceted Application of Subject Terminology) that apply open standards (more on this in section 5.5) would be a significant benefit enabling them to integrate academic search and administrative reporting (both internal and external) functions. A current barrier to JACS use is the apparent mismatch with current terminology in some areas (more on this below):

'I have never run across the term "Cell zoology" in 40 years of teaching cell biology, and even I, as a biochemistry graduate, would be hard pushed to distinguish between "Biochemistry", "Biological chemistry" and "Biomolecular sciences".' (Academic, University of Oxford)

Jorum²¹ is a Jisc funded service, based at the University of Manchester, which provides a free online repository of learning resources for further and higher education. Jorum does use the first level of the JACS hierarchy as the basis for its top level subject search but any more detail searching relies on a set of keywords that are unrelated to JACS. Jorum relies heavily on the metadata provided by institutions thus JACS would need to become the *de facto* standard to impact on a major repository such as this.

The University of Oxford is one of a small number of universities with an open data policy and it has a very clear commitment to using recognised data standards. It was therefore very keen to use JACS in relation to its own work on OERs. The University has a Continuing Education department that runs over 400 courses a year and it has created a repository of learning objects²² (covering 150 separate JACS codes) to support these courses. The use of JACS however proved problematic:

'At an early stage we decided to use JACS subject codes, which we hoped would make data about our OER more transferable; however this has been a frustrating experience on a practical level for end users, with subjects we needed not on the list. As a result we are not sure we would use this list again.' (Manton 2012)

The problem in this case was largely down to the fact that entities that seemed to be 'self evident' search terms for the average user e.g. 'History' do not exist as a JACS code (history is broken down by period, topic and area). This is similar to the issue in the use of JACS for marketing to prospective students although in this case, experienced academics had difficulty in matching their chosen search terms to JACS codes.

This is only one example of a widespread commitment to openness: the Oxford Open Spires²³ project presents a challenge to proprietary elements of the widely used iTunes U by making quality podcasts openly available for reuse. The reach of some of the open resources being produced by HEIs is extraordinary and has been shown to have an impact on both recruitment and widening participation (Jisc infoNet 2008). A JACS successor that could adequately support the sharing of open content would therefore have significant economic impact.

'It is still harder than it should be to find OER. Directing resources to simplifying the process of discovering and identifying OER, to make it as simple as using Google, is still needed.' (Manton 2012)

²⁰ FAST <http://www.oclc.org/research/news/2012/02-15.html>

²¹ Jorum <http://www.jorum.ac.uk/>

²² <http://open.conted.ox.ac.uk/>

²³ Open Spires <http://openspires.oucs.ox.ac.uk/faqs/index.html>

4.6 Supporting research

The UK has seven Research Councils who work together in a strategic partnership branded as Research Councils UK (RCUK). RCUK was represented on the Steering Group for this review and RCUK is used as a shorthand term for the collective unless a specific research Council is mentioned. RCUK was consulted as part of the review that led to the development of JACS 3.0. That work took place at the same time as the research councils were attempting to harmonise the classification scheme in their own areas and the creation of a single classification framework that covered both education and research was felt to be a step too far at that point. RCUK currently therefore operates its own subject classification framework which is entirely separate from JACS. The RCUK scheme²⁴ is used by six of the seven research councils: the Medical Research Council uses a classification called HRCS (which is discussed further in section 5.2).

The RCUK classification combines a polythetic hierarchy of 2 levels with a third level of keywords. The top level of the RC classification aligns in broad terms with level 2 of JACS with the main differences being a greater breakdown of physical sciences in RC and a greater breakdown of languages in JACS. The RC scheme also features some genuinely interdisciplinary fields that tend not to exist as courses at undergraduate or taught postgraduate level e.g. omic sciences and technologies. There is more on a comparison between the two classifications in appendix 7. The research councils are currently questioning the value of the third, detailed level of the hierarchy and are experimenting with the use of more flexible thesauri.

The tendency for the research councils seems to be to classify at a broader level of detail than they did in the past because the possibilities afforded by semantic approaches provide a better means of answering more sophisticated questions. The overall approach is a very pragmatic one and there appears to be less of an emphasis on time series than in the rest of HE. Alignment with JACS has been a stated goal of the research councils since they began harmonisation activities and discussions so far have identified no real 'showstoppers' in this regard. Indeed the reduced demand for both detail and time series, as well as the ongoing commitment to harmonisation with JACS, seems to point to an increased likelihood that it may be possible to develop a joint classification system covering the whole of higher education and research. At the very least it ought to be possible to produce a means of mapping and linking so that two classifications, optimised for slightly different purposes, could work well together.

4.7 Public information

Although there is overlap with many of the uses already discussed, it is worth highlighting the important part JACS plays in statistical reporting and the production of national statistics which are an integral part of the wider public information agenda. Whilst statistical reporting is not an end in itself, the availability of authoritative and consistent statistics is sufficiently important to warrant mention and enhancement of the classification system in order to improve the reliability and consistency of those statistics will inevitably improve decision-making by a range of public and other bodies as well as by individuals. The ways in which the core national dataset may support other types of business intelligence are discussed further in the section below.

²⁴ <http://www.rcuk.ac.uk/research/Efficiency/Pages/harmonisation.aspx>

4.8 Open/Linked Data

The sector is currently awash with material on 'Big Data' and learning analytics but, setting aside the hype, universities are becoming increasingly serious about making the most of their data assets and are beginning to develop some sophisticated analytical capabilities. It is incumbent on the custodians of the sector's authoritative data sources to ensure that they do everything possible to facilitate further development of these capabilities hence the emphasis in the 'Redesigning the higher education data and information landscape project' on the application of open standards.

It is impossible to identify all the potential uses of subject related information but the section on open/linked data and semantic approaches outlines the kind of things that are going on under the broad heading 'linked data'. The University of Oxford's open data policy and work with OERs has already been mentioned.

The University of Southampton is a leader in the field of open data (Professor Sir Tim Berners-Lee is based there). The University has an open data policy and is providing much of its administrative data for others to use via its Open Data Service²⁵. The University has created its own set of URIs (essential for others to apply similar approaches to the data) for JACS codes²⁶ as no authoritative source currently exists.

The University of Nottingham has also been at the forefront of developments in linked data²⁷ working to understand the course and labour market information needs of a range of stakeholders (including learners, potential learners, Sector Skills Councils, IAG organisations, schools, careers professionals, development agencies and local authorities) and using linked data techniques to collate and use labour market information. The proposals for a new classification system take note of these developments and try to ensure that the new framework is made available in such a way as to maximise its potential use.

²⁵ <http://data.southampton.ac.uk/>

²⁶ <http://data.southampton.ac.uk/dataset/jacs.html>

²⁷ <http://www.jiscinfonet.ac.uk/infokits/optimisation/experiences-of-linked-data/>

5. Designing a new subject classification

This section takes a broad look at the approach to designing a subject classification bearing in mind:

- issues raised with the current classification system,
- comparison with other related classifications,
- the concepts that underpin the classification and
- technical developments that may have a bearing on any future system.

5.1 Characteristics of the existing coding framework

Two principal factors characterise the existing framework:

- JACS is a controlled vocabulary (and thus, as with any vocabulary of this type it represents a particular world view) but it does not apply consistent criteria to form the classification although it has an 'implied' hierarchy;
- the four digit structure of the coding framework itself imposes limitations on the classification.

The two factors taken together mean that the evolution of JACS has necessitated a series of compromises. On the plus side a pragmatic approach has made transition possible and delivered a workable solution for a number of key purposes. Some of the compromises do however become increasingly unsatisfactory as the range of uses of JACS expands.

'Currently use of JACS has fragmented to the point where there isn't a common understanding.'
(Stakeholder interview)

Taking the first point it can indeed be argued that the existing JACS framework is not a classification system in the strictest sense in that it does not apply a consistent set of criteria in ordering entities within a hierarchy e.g. history is broken down by period, topic and area, hence a module on the European economy from 1700 to 1750 could easily appear in any one of the three branches. The 'hierarchy' in JACS is described as 'implied' rather than rigid (and has indeed been adapted in some instances e.g. relating to KIS).

Furthermore the existence of 128 codes of the 'other' or 'not elsewhere classified' variety may hint at issues with the extent to which it provides comprehensive coverage of the current curriculum as few of these codes are unused despite the fact that selection of a more general code may often offer a simple solution to classification issues.

'The structure is inherently flawed: it is not a detailed branching classification system and we are often fudging to classify something ... It is an obstacle to transmission and understanding of data.'
(Stakeholder interview)

Comparison with a range of other subject classifications (discussed further below) reveals the extent to which JACS displays its history. It may be seen that many of the most significant differences between JACS and other classifications probably relate either to government priorities particular to parts of the UK or to a shortage of codes in certain areas.

5.2 Comparison with other related classification schemes

Mappings were performed against the following classifications:

- Australian Standard Classification of Education (ASCED)
- Australian and New Zealand Standard Research Classification (ANZSRC)
- Organisation for Economic Co-operation and Development (OECD) Field of Science (FOS) Classification
- RCUK Research Classification
- UNESCO International Standard Classification of Education (ISCED)

As well as the Higher Education Academy's discipline areas (appendix 9).

As a general observation JACS has about the same number of top level headings as most other related classifications although it breaks language related studies down further than most and groups physical sciences more broadly (possibly to compensate for the limited number of top level codes available). At this level it is, unexpectedly, slightly more detailed than classifications that cover a range of educational levels and slightly less detailed than research focused classifications although the observations about languages and physical sciences apply across the board.

The most pertinent of the comparisons is with work undertaken in Australia and New Zealand and this is discussed further in some detail in appendix 5. Further details on all of the mappings are included in appendices 5-9. The mappings were undertaken for illustrative purposes only and should not be taken as definitive.

A particular aim of the project is to explore the possibility that a revised subject classification could be of use to the NHS and provide the basis for better alignment between data collections specifically for NHS purposes and other HE data collections as well as better supporting medical education generally. There are a number of controlled vocabularies in use in the world of health. A system named the Health Research Classification System (HRCS)²⁸ has been developed by the UK Clinical Research Collaboration (UKCRC)²⁹ partners and has been widely adopted by UK research funders (including the Medical Research Council³⁰) as well as some international users. UKCRC partners include all of the key stakeholders pertinent to the delivery of higher education in this field. UKCRC provides the main steer for the development of health education policy and has developed career and training pathways for doctors, dentists, nurses, midwives and allied health professionals. It thus appears that for statistics from the higher education sector to have maximum strategic value, they ought to relate to the requirements of this body.

A brief examination of UKCRC publications reveals that data relating to HE level activity is considered to be a problem area. As an example a report on developing research professionals (UKCRC 2007) noted:

'This lack of data about nurse researchers and educators represents an over-riding barrier to effective integration of research into nursing career pathways, with all the implications for education and training, employment structure and workforce capacity that such an obstruction brings.' and went on to recommend: 'We recommend that a single data source be developed to provide information on labour market intelligence that relates to nurses engaged in training to be researchers and educators.'

²⁸ HRCS <http://www.hrcsonline.net/>

²⁹ UKCRC <http://www.ukcrc.org/>

³⁰ MRC <http://www.mrc.ac.uk/index.htm>

It would be glib however to assume that that HRCS provides a classification that can simply be adopted as it stands in HE. HRCS is itself based on the World Health Organisation (WHO) International Classification of Diseases (ICD)³¹ and, for example, has required some adjustment as the ICD codes only describe diseases and ill-health and do not therefore fit research into the normal development and functioning of parts of the body. N.B. The HRCS is a two dimensional framework that encompasses both 'health categories' and 'research activity codes'. HRCS is however gaining traction on an international scale particularly via the European Medical Research Councils (EMRC). A report on current approaches and future directions (European Science Foundation 2011³²) provides a good overview of the benefits of common approaches and the issues to be overcome in achieving this. The fact that the recommendations of the report are supported by representative bodies from 19 nations (counting the UK as one) suggests that achieving workable solutions across agency and other boundaries is achievable when the benefits are sufficient to make this worthwhile. The approach suggested is a pragmatic one that recognises the value of semantic approaches and is based on the desire to integrate different systems that are optimised for different purposes. Relationships between data on higher education, research and the workforce and labour market amount to exactly the kind of joining up that linked data approaches can facilitate (more on this in section 5.5).

'A standard international classification system for health research does not necessitate individual agencies forfeiting their own particular classification systems: it may be implemented in parallel or there may be methods to map existing classification information to it or "translate" it to a more widely used categorisation system. Using automated approaches for classifying research has the potential to speed up work in this area, lower the cost of implementing classification processes and support the use of a common categorisation approach.' (European Science Foundation 2011³³)

The conclusion from this analysis is that as the approach suggested in this report is not dissimilar to that taken by HRCS, and as the main RPG and NHS stakeholders are all UKCRC partners, it ought to be possible to work through this body (housed with the Medical Research Council) to develop the means to make the JACS successor useful across the wider health arena.

Although the above discussion relates specifically to the means of relating JACS to established classifications used in medical research, there is no reason why similar approaches should not facilitate data exchange between higher education and other fields of research.

5.3 Level of granularity

The appropriate level of granularity for any new coding framework has been discussed with a range of stakeholders. It appears that, whilst institutions are strongly urged to use four digit coding, for many purposes the funders and regulators are content to look at statistics and trends at the top level of the JACS classification system (20 broad headings). Monitoring of SIVs and responses to Parliamentary questions were cited as the most frequent reasons requiring analysis at a more detailed level.

'We see-saw between levels: 20 subjects is enough for most purposes then occasionally you want to know about nuclear waste recycling.' (Stakeholder interview)

³¹ WHO ICD <http://www.who.int/classifications/icd/en/>

³² : <http://www.esf.org/research-areas/medical-sciences/activities/science-policy/health-research-classification-systems-current-approaches-and-future-recommendations.html>

³³ : <http://www.esf.org/research-areas/medical-sciences/activities/science-policy/health-research-classification-systems-current-approaches-and-future-recommendations.html>

A survey of HEIs revealed mixed views on the level of granularity in the existing classification. Around 7% of respondents felt it was too general, 15.5% thought it too detailed and 27% found it 'about right'. The bulk of respondents (47%) however noted the variability in the level of granularity between different subject areas and there were many comments on this topic.

'Some subjects go into excruciating details ... but there are huge gaps in other areas.' (Survey response)

'Rather than increase the detail of the existing JACS codes, or push for institutions to use more detailed JACS codes than they are currently doing, I would rather see existing gaps in the classification plugged ...' (Survey response)

This is an important observation which has a bearing on the value of any statistics produced as comparisons across subjects which are broken down in detail and others which are greatly aggregated are unlikely to give meaningful results. A goal of any new framework ought to be to achieve a broadly consistent level of detail across each of the subject areas. N.B. Whilst this is relatively simple to articulate as a principle, we should not underestimate the likely difficulties in achieving a shared understanding of what does constitute consistency across different subject areas.

'Small numbers at the most detailed level undermine the value of trends/ comparisons.' (Survey response)

The survey had a question on the extent to which HEIs routinely try to code to the finest level of detail. The figures here are slightly surprising when compared to HESA's own analysis of returns and suggest (see appendix 4) that staff in HEIs have a perception that they are generally classifying to a greater level of detail than is actually the case. Even more surprising is that HESA figures suggests that modules are rarely classified to a greater level of detail than courses (in 2010 62% were classified to 1 digit; 30% to 2 digits; 8% to 3 digits and none to 4 digits).

When asked why (despite requests by bodies such as HEFCE for 4 digit coding) they did not always classify to the lowest level of detail possible many respondents cited the variability in JACS itself across different subjects as well as the difficulty in establishing consistent application across the whole organisation.

'Those who understand the course/module contents don't understand the JACS coding structure, and vice versa.' (Survey response)

'Part of the problem is that it's hard to work out what is required. When in doubt, people tend to be vague ...' (Survey response)

The most significant portion however felt that the benefit in this was simply not worth the extra effort and, in particular, was not justified in terms of the information returned to HEIs by the data collectors. The fact that most published statistics institutions are using as comparators are at a high level (and the fact that the inconsistency in the detailed level undermines the value of trends/comparisons) was noted.

'The value of detail in JACS codes is limited by the detail of the data we get back. In HESA published data, league tables, NSS results, the data we get back with very few, if any, exceptions are provided at 2-digit level.' (Survey response)

For many of the regulatory and administrative purposes it therefore appears that a very broad classification is generally sufficient. More worryingly it also appears that lack of consistency, both in the classification itself and in its application, mean that the fine-grained data used for other purposes may not be so robust as is believed.

The question of granularity however takes on a different aspect when we look at a broader set of uses. JACS has many potential applications in learning and teaching, most notably in facilitating the sharing and re-use of learning resources. In this set of use cases a finer level of granularity is needed (although JACS is only one part of the metadata jigsaw).

The conclusions of this analysis in relation to the development of a new framework are as follows:

- the existence of a fourth level of detail in some parts of the classification and not in others is unhelpful and the evidence suggests that this is a source of inconsistency in the data rather than an enabler of better analysis;
- there is probably a need to 'fill gaps' rather than differentiate more finely in some areas: a coding framework which permitted the creation of some new codes at the higher levels would solve this problem without further compromising existing time series;
- in order to facilitate the wider use of JACS there needs to be a mechanism for linking to more detailed subject specific vocabularies: this does not mean that those vocabularies need to be hardcoded into JACS itself.

5.4 The relationship between subjects and other attributes

As with many aspects of the Redesigning the HE Data and Information Landscape work (e.g. What is a Course?³⁴) terms that are widely used and generally understood in broad terms become problematic when one tries to define them more precisely.

JACS is a means of classifying subjects of study. In the HESA world view a subject is an objective field (or topic) of study regardless of the programme of study (or course), the broad academic discipline in which it is taught or the precise curriculum of which it forms a part.

A particular module of study will:

- have a main subject of study (identified by one or more JACS codes);
- belong to a cost centre in the HESA record (although this attribute derives from the teacher see below);
- be taught by one or more members of staff who are assigned to a cost centre and belong to a (or possibly more than one) 'home' department (or school);
- be 'owned' by a particular department (or school or other organisational unit which may itself have multiple cost centres or which may indeed be part of a larger cost centre);
- belong to one or more courses or programmes of study (also identified by one or more JACS codes) which is also 'owned' by a particular department (or school) which may be different to the teaching department.

5.4.1 Subject and discipline

The above description is far from being the basis of an accurate entity/attribute relationship diagram but, even at this vague level of description, there is already a concept that is fundamental to the academic world view but hard to fit into this administrative picture: that of academic discipline. There is no authoritative definition of the term but it is taken to refer to a cognate grouping of fields of knowledge that are developed, researched and passed on through a shared set of data sources, research methods and professional practices i.e. subdivisions of what is often referred to as 'scholarly activity'.

'Academics identify with a discipline as being more than a collection of subjects being taught. It encapsulates a way of thinking, a framework for deciding on which problems are worth studying and a way of approaching these problems.' (Phil Barker, CETIS)

³⁴ http://www.hesa.ac.uk/dox/publications/The_Course_Report.pdf

An HEI's organisational structure will often reflect its particular mix of academic disciplines although there may not be a direct mapping between organisational units and disciplines.

Why does any of this matter in the context of this report? It matters because, as already noted, the JACS classification is a particular way of representing the academic world. If it were as simple as following precise HESA guidance to produce a particular dataset that is a consistent interpretation of this world view, it would not matter. It is not however that simple: the use of JACS is open to interpretation and there is considerable evidence of stakeholders viewing JACS through different lenses and thus bringing into question the integrity of the dataset.

The assertion that a subject is a subject regardless of who is teaching, and to whom it is taught, is itself questionable. A good example of this is statistics, which is taught across many fields from mathematics through to social sciences but there are many other examples: languages taught in mediaeval studies and business translation courses, elements of computing taught in electronic engineering and computer science and so on. To ignore the discipline dimension, to say the subject is the same regardless of the different disciplinary slants taken, that is to say statistics taught to mathematicians is the same as statistics taught to physicists is the same as statistics taught to social sciences is problematic. This may be true at a very superficial level, but the relevance of theoretical versus practical elements will vary between those disciplines, as will the nature of the data collected (typically a physicist will design an experiment to control each variable independently so as not to deal with multivariate data, this is not often possible in social sciences and so multivariate analysis is far more important). The distinctions are more subtle than can be dealt with by simply looking at other attributes e.g. level of study.

This is not just a theoretical problem. Early discussions with staff from HEIs revealed differences in approach that appeared to warrant further investigation. The survey conducted as part of this work asked HEIs how they would approach classifying an example module 'Statistics for Archaeologists' in both the context of the module being taught by a Department of Mathematics and in the context of teaching within a Department of Archaeology. HESA guidance suggests that this particular module amounts to 'Applied Statistics' (JACS code G310) in any teaching context. Only 10% of HEI respondents gave that answer (11 out of 110).

The most common response (14 responses) was to view the subject as 'Applied Statistics' (JACS code G310) if it was delivered in the Department of Mathematics and 'Archaeological Techniques' (JACS code V460) if it was delivered in the Department of Archaeology. The variability in actual responses was considerable (see appendix 4) but, overall, 31% (34 respondents) gave an answer that applied a G code to delivery in the Department of Mathematics and a V code if it was delivered in the Department of Archaeology.

Although this is not a comprehensive survey of all staff responsible for data input, it represents a significant and well-informed group and suggests that, despite HESA guidance to the contrary, HEIs view the relationship between subject and discipline as integral to the definition of 'subject'. Furthermore the wide variation in responses across the board is sufficient to suggest that current interpretation of JACS codes is inconsistent across the sector and that any data beyond the broadest level should be viewed with caution.

In determining what statistics are of use to support its policy-making, HEFCE has taken decisions based on its concept of subject. For example, in relation to student retention rates HEFCE has historically viewed the subject of study, as exemplified by the JACS code, as the stable and consistent measurable factor in determining the rate of retention. The analysis above suggests that in practice the figures may relate as much to the support infrastructure provided by the home department.

Issues arising from the perceived subject/discipline relationship appear likely to increase as a result of increased use and visibility of the data. This is highlighted most strongly by the application of JACS in relation to the Key Information Set (KIS). Whilst active debate about the presentation of data on the Unistats website (particularly relating to the aggregation of small cohorts) is taking place in other contexts. It warrants inclusion here as there are significant issues relating to how JACS is designed and applied.

KIS may indeed reveal many issues with institutions' internal data management (one institution in the Jisc course data programme³⁵ described how KIS led to the realisation that JACS codes were being applied '*fairly randomly*' in some areas) but other characteristics of the approach give rise to further questions about the relationship between subject and disciplines in the hierarchy. The combination of Classics with Linguistics in JACS is a case in point and an example of where JACS diverges from the, recently reviewed, set of discipline areas recognised by the Higher Education Academy.

'... any perception that the topic they [academics] teach is being placed under a someone else's discipline will be resisted as misrepresenting what is actually being taught, indeed as a threat to the integrity of the discipline.' (Phil Barker, CETIS)

A respondent to the review survey cited one particularly anomalous example from the Unistats website where their BA Fashion Marketing course returns satisfaction statistics from Materials & Minerals Technology and work and/or further study statistics from Engineering & Technology. Without going into the detail of whether this is down to the institution's own coding or the KIS algorithms, it is clear that such data is of little use (and indeed misleading) to potential students and that the current classification system is not lending itself to enabling clarity. Other comments from HEIs include:

'Unhelpful aggregation of subjects into groups which those outside HE and even those inside HE but not familiar with JACS do not understand.' (Survey response)

'The way that higher level results are used when lower level results are not available means that KIS results, ostensibly for a particular subject, often relate to something quite different. This is a particular problem for an institution with a diverse range of subject such as ours, where JACS codes in the same group may be used by very different departments.' (Survey response)

'Some areas of our provision do not match with how they are mapped within JACS. ... An example might be Economics. This sits within our Business School but in the JACS framework it appears in Social Studies. This means it gets grouped with Social Work (Health) and Human Geography (Life Sciences). ... If provision is aggregated to JACS Level 1, then DLHE responses in particular could be skewed (the employability outcomes for Social Work, for example, may be considerably different to those for Economics and Geography) and also aggregates very different and separate student experiences for the NSS.' (Survey response)

The ideal coding scheme would be able to represent both the subject and the discipline context in which it was taught. There are various ways in classification terms to deal with the fact that similar subjects (or topics) can be addressed in many different disciplines: one is to create a polythetic hierarchy where detailed terms can relate to more than one of the higher levels and another is to accept some proliferation of terms.

The former would allow discipline to be selected independently of Field or Subject/Topic e.g. Discipline = Archaeology and Topic = Statistics. This would help with the description of multidisciplinary activities but it represents a significant conceptual departure from the existing system and has technical and training implications that are not inconsiderable (and due, to the complexity of application, it might also result in even less consistent data). Nor would such an approach resolve the issue that a course or module needs to sit in one place for statistical reporting purposes.

³⁵ <http://www.jisc.ac.uk/whatwedo/programmes/elearning/coursedata/stage2.aspx>

The latter is more akin to the genesis of JACS to date as it has been developed through a discipline-focused lens rather than any cross cutting examination of the various branches of the classification; this has the added benefit of encapsulating any significant differences in vocabulary between disciplines and it is unsurprising that it largely mirrors how JACS appears to be operated in practice. N.B. The fact that people tend to classify within 'their own' subject area appears to be borne out by the limited use of more than one JACS code: over 97% of modules have a single JACS code indicating that the use of multiple JACS codes per module to highlight interdisciplinarity is probably seen as a burdensome level of administration at this level of detail. The down side to this approach is that it there will be more overlap in the granular information and as one stakeholder put it: 'by enforcing the 1:1 relationship this offers the very real risk of becoming an academic football game where a consensus cannot be reached.'

These are not the only ways of dealing with the issue. The Australian Standard Classification of Education (ASCED) requires the understanding of a number of different concepts that define the criteria for each level of its hierarchy including: theoretical content; purpose of learning; objects of interest; methods and techniques; and tools and equipment. This would represent a significant degree of complexity in the administration of modules for most HEIs and would not appear to further the goal of increasing accuracy and consistency across institutions. This conclusion is borne out by experiences in the learning and teaching sphere of using the IEEE³⁶ standard classification for learning objects IEEE LOMS (IEEE Learning Object Metadata Standard) which identified both topic and discipline for different purposes. The standard was found to be poorly understood and cumbersome to apply and is little used in the UK these days.

One example of a classification that specifically builds discipline into the classification is the DFG (Deutsche Forschungsgemeinschaft) German research classification.³⁷ DFG statistics include profile analyses showing the particular disciplinary mix of each German University.

Of the bodies supporting UK HE, the HE Academy is the one that aligns itself and its work most closely with the academic concepts of subject and discipline. The Academy has recently undergone a restructure in which it moved from a network of 24 subject centres to an approach of identifying and targeting support for 36 different disciplines grouped under the broad headings of: Arts and Humanities, Health and Social Care, STEM, and Social Sciences.

The disciplines recognised by the Academy bear a closer relationship to HESA cost centres and research units of assessment than to the current top level of JACS (see appendix 9 for a mapping to JACS). The definition of these areas has however also been subject to pragmatic considerations relating to the number of students and staff engaged in each of the areas. For example Philosophical and Religious Studies has been combined as an umbrella term despite the fact that *'philosophy and theology have a history which includes the perception that there have been, and may continue to be, active efforts by the practitioners of the one to suppress the other.'* (HE Academy Discipline Lead). The two are identified separately as cost centres and as research units of assessment.

The practical difficulties of any classification scheme keeping up with the emergence of new sub disciplines are as well understood in the academic world as by the data specialists. Criminology has been cited as a subject/discipline which has grown significantly over the last decade (it was designated as a QAA Subject Benchmark in 2007 and a UCAS course search on the keyword currently returns 682 courses) yet is not currently named within the HEA framework. Another 'outlier' in the academy scheme is the separate identification of 'Islamic Studies' in order to support a specific initiative (now ended) by the four UK HE funding bodies.

³⁶ IEEE the Institute of Electrical and Electronics Engineers is the world's largest technical professional association but the formal name is used only on legal documents

³⁷ DFG http://www.dfg.de/download/pdf/dfg_im_profil/gremien/fachkollegien/dfg_fachsystematik_en_08_11.pdf

Whilst it would therefore be overly simplistic to say that the HEA academy's discipline areas could immediately replace the top level of JACS, they provide a useful and well-informed starting point for consideration.

The conclusions of this analysis in relation to the development of a new framework are listed below.

- The new framework should recognise the, currently implicit, assumption (at least by HEIs) 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.
- Consideration should be given as to whether the levels of a new classification should be explicitly named as in the Australian and New Zealand system e.g. in the UK we may adopt something along the lines of:
Level 1 - Discipline
Level 2 - Broad Field of Study (or sub-discipline)
Level 3 - Narrow Subject (or Topic) of Study.
- Consideration should be given as to how the levels of a new classification should be articulated i.e. whether level 3 needs to have a 1:1 relationship with the upper tiers or not.
- Guidance should be provided to data users as to circumstances in which it may, or may not, be appropriate to aggregate data from different branches of the classification.

Although outside the scope of this particular piece of work, it would be helpful if future iterations of the sector level data model, lexicon and thesaurus (strand 2 of this programme of work) could seek to ensure that the term 'subject' and its main relationships are defined.

5.4.2 Subject and cost centre

Ostensibly subjects and cost centres are quite different entities and, in data terms, there is no formal relationship between the two. In practice however the relationship is less clear-cut. HEFCE has moved away from a teaching funding model based on subject to one based on cost centres (other funders are less concerned with the concept of cost centre). In determining the cost centre for a particular module, institutions need to make reference to the individual member of staff mainly responsible for teaching the activity.

'Student FTE on both HESes and HESA returns should therefore be returned based upon the cost centre of the member of staff most directly associated with delivering the activity.'³⁸

This view of the relationship between the module and the locus of teaching may appear to be somewhat at odds with both HESA and HEFCE's assertion that a subject is a subject regardless of who is teaching and/or being taught. If the two are not directly contradictory (subjects and cost centres are not the same thing after all), it is nonetheless not difficult to see how the different definitions can be a source of confusion. Furthermore, in auditing institutional data returns, HEFCE bases judgements on 'cost centre norms' that are themselves derived from the JACS codes used in each cost centre.

³⁸ <http://www.hefce.ac.uk/data/famd/fundingdatareconciliations/heses10re-creationbasedoncostcentresectornorms/mappingcalculationforthehesesre-creationbasedoncostcentresectornorms/>

For each institution and JACS code, the cost centre with the largest FTE is assumed to be the institution's 'cost centre preference'. On a sector-wide basis the cost centre with the greatest number of preferences is taken to be the norm for that JACS code. HEFCE's mapping³⁹ of JACS codes and cost centre norms for 2010/11 indicates the considerable degree of variability across the sector with the number of institutions that teach a particular subject and correspond to the cost centre norm being as low as 10% in the case of one particular JACS code (this is not an isolated example).

Our survey backed up anecdotal evidence that the relationship with cost centres is a further challenge to the objectivity of JACS as a subject classification. A limited number of institutions (15%), often those with a strongly centralised planning function, went so far as to suggest that they only accept a limited range of JACS codes as being 'valid' in particular cost centre (although taking into account free text comments, it appears that the figure who generally assume a close alignment is more like 20%). Others, whilst being less strongly directive, make reference to having an eye both to norms and to league tables.

'Ensuring that the JACS code reflects the academic content of the module for the appropriate cost centre is difficult when JACS subject descriptions appear in several different cost centre possibilities.' (Survey response)

'There is no direct relationship operationally within the University and no mapping is presented to departments however the planning team maintain oversight of this and make appropriate changes if issues arise with cost centre norms etc.' (Survey response)

It is unsurprising that any classification that has implications for funding drives behaviour. Given the points already made about the extent to which JACS is open to interpretation, it is equally unsurprising that, in this case, the behaviour is neither consistent nor predictable.

The conclusions of this analysis in relation to the development of a new framework are as follows.

- The evidence presented in this report has come as a surprise to a number of stakeholders in key sector bodies who believed the definitions of subject and cost centre to be well established and understood and consistently applied and some stakeholders who previously believed this to be a 'non-issue' have gone so far as to question whether we do indeed need both subject and cost centre.
- Consideration of whether both subject and cost centre are needed hinges on whether a model that makes sense conceptually is delivering value if it is not applied consistently and this debate is outwith the scope of this project.
- Very clear guidance and a targeted communications plan is needed to underpin implementation of a revised subject classification (and may indeed be needed as an interim measure to improve consistency if implementation of a revised JACS is some years away).
- There is already relatively close alignment between HESA cost centres, the HE Academy disciplines and, indeed, REF Units of Assessment: the suggested revisions to the top level of JACS to give it a clearer disciplinary focus will bring it more closely into alignment with existing cost centres and make the exceptions easier to highlight and explain.

³⁹ <http://www.hefce.ac.uk/data/famd/thefundingandmonitoringdataexercise/hesadata/>

5.4.3 JACS and UCAS

One final relationship, or indeed non-relationship, requires mention. Each full-time undergraduate course that recruits through UCAS will also have a single UCAS code. A UCAS course code is made up of four characters and can be any combination of four numbers, four letters, or a combination of letters and numbers in any order. The UCAS course codes have no meaning other than '*this course is offered by this institution for this application cycle*'. In the past, however, UCAS has created course codes based on the JACS codes of the course subject. The similarity in appearance and meaning between the two sets of codes has therefore long been a source of confusion.

UCAS has attempted to deal with this by changing the format of its course code although UCAS' preferred schedule for full implementation had to be delayed due to HE providers wanting more time to adjust to the change. Recent interactions with HE providers suggest that, despite best efforts by UCAS to clarify the matter, misconceptions about an ongoing relationship between UCAS codes and JACS codes are still common. The point required clarification by UCAS at a January 2013 meeting of 63 institutions involved in the Jisc Course Data programme and a brief search of some university planning documents available on the Internet also revealed some evidence confusion with at least one example where the terms UCAS code and JACS code were being used interchangeably. There is some evidence that this perceived relationship may be perpetuated in institutions' internal systems: one respondent to our survey suggested that the institution uses the JACS code to generate its own course codes for postgraduate courses.

The conclusions of this analysis in relation to the development of a new framework are as follows:

- any project to implement a new coding framework requires a clear and targeted strand of communications activity;
- it is worth considering whether any new coding framework should be renamed in order to make a clear distinction between the new approach and any previous iterations of JACS.

5.5 Linked/Open Data and Semantic Approaches

Approaches to classifying and sharing data have evolved considerably since the JACS classification was first conceived. Many of the newer approaches are aimed at finding ways of discovering, sharing and making sense of, the vast body of unstructured information that exists on the World Wide Web. Those used to traditional methods of handling data may be forgiven for thinking that dealing with the vast body of information on the Internet would require a huge, and hence unachievable, top-down exercise to classify and standardise the data. Many communities are using flexible, non-hierarchical means of classifying information, often termed 'folksonomies' based on individuals and communities tagging data using vocabularies relevant to them. Whilst this might sound like a 'free-for-all' unsuited to meeting the needs of the higher education sector, such approaches need not be viewed entirely separately from the more formal systems of classification and controlled vocabularies.

The Semantic Web⁴⁰ (sometimes called Web 3.0) is a term used to describe the ways in which technology can be used to make more intelligent connections between data elements that exist in different places. One data sharing standard that can serve to bridge different means of organising and classifying information is the Simple Knowledge Organization System (SKOS). Another concept of relevance is that of 'linked data': There is no generally accepted definition of linked data but the term refers to publishing and connecting structured data on the Web. The Semantic Web is made up of linked data: the Semantic Web is the whole, while linked data is the parts. Rather than envisage a situation where everybody needs to agree on all terms in order to make sense of the data, linked data works in a bottom up way with small datasets being loosely joined and vocabularies that are flexible and evolving.

⁴⁰ The potential of the Semantic Web is outlined very clearly in this five-minute video by Professor Sir Tim Berners-Lee. He describes how the release of open data by the UK Government is enabling individuals to find information of relevance to them e.g. linking accident statistics to geographical data to establish whether their route to work is a cycling blackspot and how the relief effort after the Haiti earthquake in 2010 was aided by the open source community providing accurate maps of the affected areas. <http://www.ted.com/talks/view/lang/en/id/788>

The concepts are of relevance to the work in hand because they may offer a pragmatic means of meeting the needs of a variety of different stakeholders. Rather than design an all-encompassing classification system that meets the needs for different levels of detail and handles some quite specialised vocabularies, it may be possible to solve the problems one at a time. It could, for example, allow us to make a, relatively straightforward, JACS replacement work effectively with specialist vocabularies in use in research and the NHS as well as relating it to the more general keywords that prospective students might use when undertaking a Google search for information.

The most important thing at this stage is to ensure that the approach taken does not preclude the use of semantic approaches in the future. Some of the concepts behind developments such as SKOS can be quite sophisticated and require a technical interface layer before the outcomes can be human readable. At the most basic level there are however a simple set of good practice rules that those looking to facilitate data sharing and linked open data should follow. These are outlined in a five-star system⁴¹ (Berners-Lee 2006) which defines a number of attributes of data to make it usable in an open data context.

- ★ make data available on the Web (whatever format) under an open licence
- ★★ make it available as structured data (e.g. Excel instead of image scan of a table)
- ★★★ use non-proprietary formats (e.g., CSV instead of Excel)
- ★★★★ use open standards (e.g. URIs, RDF⁴², SPARQL⁴³) to identify things, so that people can point at your data
- ★★★★★ link your data to other data to provide context

Currently JACS fails to meet the requirement to gain its first star as it does not have an explicit open licence (or indeed any specific licensing arrangement). It appears to be generally assumed that JACS is open to all and that its adoption for a range of purposes is encouraged but it has not been explicitly assigned an open licence. Whilst organisations within the HE sector would probably make the assumption that JACS was available for them to use for other purposes, it cannot be assumed that other potential users of the data would think likewise (particularly as many library classification systems are proprietary and their use restricted). Questions both of 'ownership' and licensing of JACS relate to strand 1 of this work programme which is looking at a broader issues around the governance of the data and information landscape and therefore should not be tackled in isolation.

Assuming the licence issue is a technicality, because the intention behind JACS is that it is freely available for others to use, then currently JACS operates at about three-star level. It would be a very large, yet relatively simple, step forward for the new version of JACS to conform to four-star data and facilitate others generating five-star data.

This could be accomplished by ensuring that each of the entries in the classification scheme has a uniform resource identifier (URI). A URI is a globally unique and unambiguous identifier that can be used in any system without fear of conflicting with other identifiers. The term is often confused with the other common, but informal, term URL (uniform resource locator) and with URN (uniform resource name) both of which relate to early ideas about how a URI should be constructed i.e. partitioned into a set of discrete attributes including URN, URL and others such as URC (uniform resource citation where the reference is to meta data rather than the entity itself).

⁴¹ <http://www.w3.org/DesignIssues/LinkedData.html>

⁴² RDF (Resource Description Framework) is one of the key ingredients of Linked Data: it provides a generic graph-based data model for describing things, including their relationships with other things.

⁴³ SPARQL is a query language for linked data, a SPARQL endpoint is the interface to linked data.

The JACS URI would need to be persistent and named in such a way as to make it clear that this was the authoritative source of the information.⁴⁴ The ideal is for the URIs to be published via a web service (JACS is currently available as a .CSV file from the HESA website). The outcomes of strand 1 of this wider programme of work ought to indicate the most appropriate means of providing this web service on an ongoing basis. The JACS subjects could then be referenced globally by humans and by computer systems in/from data feeds and web applications.

Shared URIs are integral to the links in linked data thus: envisaging two data sets both relating in some way to subjects in HE, if they both use the same URIs the data sets can be linked using web-based technologies (linked data) and users can know what those links mean (semantics).

The subjects could thus be used in semantic web approaches to representing ontologies which allow relationships with other concepts to be represented independently of the coding scheme used. SKOS, mentioned above, is an example of a semantic web application of RDF which provides a model for expressing the basic structure and content of concept schemes such as thesauri, classification schemes, subject heading lists, taxonomies, folksonomies, and other similar types of controlled vocabulary. It allows for the description of a concept and the expression of the relationship between pairs of concepts. In short these approaches free JACS from the limitations of a rigid hierarchy and allow relationships to be defined between JACS subjects and things in other schemas.

A practical example of how this might be used in practice is that Bodleian Libraries have indicated they would like to be able to use a range of subject specific vocabularies (in SKOS/RDF formats) in their publication and research repositories and map them to JACS for administrative reporting purposes.

The above is only one example and implementing such an approach would be of considerable benefit to a number of universities who are already adopting semantic approaches yet having to deliver workarounds themselves to compensate for the fact that the existing classification system does not facilitate this type of data exchange. The benefits would thus be achievable regardless of whether the or not the major agencies such as HESA, UCAS etc adopt semantic approaches themselves.

The conclusions of this analysis in relation to the development of a new framework are as follows:

- the new framework should provide a persistent URI 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 to encourage its adoption and use by the widest possible group of stakeholders.

⁴⁴ See Berruta, D. and Phipps, J. (2008) for good practice in connection with this <http://www.w3.org/DesignIssues/LinkedData.html>

6. Next Steps

The primary audience for this report is as yet unformed because it is assumed that the report will inform the work of whatever future entity emerges from strand 1 of the wider work programme which is putting in place effective mechanisms for the governance of the HE data and information landscape. The Steering Group is asked to consider the report and make some recommendations in principle as to whether the proposals should be accepted.

6.1 Implementation

The following next steps are suggested on the basis that the new entity will instigate a project to implement the successor to JACS 3.0.

- Consider and agree the overall timescale for implementation with key stakeholders: in view of the issues with the current system it is proposed that the implementation should be as rapid as is practical to achieve.
- In parallel with the above, determine the scope/phasing of the implementation e.g.
 - develop a series of use cases and agree which stakeholder requirements can be met
 - determine whether the development of associated keywords is an integral part or a later phase.
- Develop a communications plan to run throughout the project (including an early decision on the name for the classification).
- Determine whether any interim action is necessary to deal with inconsistencies in the current system.
- Identify the authority that will provide and maintain the URIs as a web service (N.B. there may be merit in the short term in creating URIs for JACS 3.0 and this may aid migration)
- Agree the coding framework for the core and provide early information to system suppliers.
- Work with subject matter experts to collapse the classification from four levels to 3 seeking to align as far as possible with HE Academy and Research Council classifications.
- Develop keyword mapping in tandem with the above if this is agreed to be part of the initial implementation.
- Provide a mapping of JACS 3.0 to new codes for planning purposes.
- Create URIs.
- Create guidance materials.

These activities should take place in tandem with work on the other strands of this programme e.g. developing robust definitions for the attributes as part of the ongoing work of strand 2 (data model, lexicon and thesaurus) and looking to identify where the collection and use of subject related data can be streamlined or add value as part of the ongoing work of strand 3 (inventory of data collections).

6.2 Ongoing Maintenance

Ongoing maintenance will also be informed by the work on the governance of the HE data and information landscape (strand 1).

- The logical locus of responsibility for maintenance and the home for the web service (not necessarily the same body) should derive from the work on the governance strand.
- There will be a need to balance stability (the core classification) with flexibility and adaptability i.e. more regular updating of associated keywords hence a regular review cycle should be agreed.
- There will be a need for change control processes e.g. does a change to the definition associated with an entity (including correction of an error) demand a new code/URI (technically this ought to be the case).
- There will be a need to maintain an overview of the classification as a whole and not allow it to become skewed by over emphasis on particular branches.

7. References

- Action on Access (2009) Review of Widening Participation Strategic Assessments. Available at: http://actiononaccess.org/wp-content/files_mf/wpsareview.pdf [Accessed 13 Feb 2013]
- Berners-Lee, T. (2006) Design Issues: Linked Data. Available at: <http://www.w3.org/DesignIssues/LinkedData.html> [Accessed 10 Feb 2013]
- Berruta, D. and Phipps, J. (2008) Best Practice Recipes for Publishing RDF Vocabularies. W3C Working Group Note. Available at: <http://www.w3.org/TR/swbp-vocab-pub/> [Accessed 10 Feb 2013]
- Currier, S., Campbell, L. and Beetham, H. (2005) Pedagogical Vocabularies Review. Report 1 of Jisc Pedagogical Vocabularies Project. Available at: <http://www.jisc.ac.uk/whatwedo/programmes/elearningpedagogy/vocabularies.aspx> [Accessed 14 February 2013].
- Currier, S., Campbell, L. (2005) Vocabulary Management Technologies Review. Report 2 of Jisc Pedagogical Vocabularies Project. Available at: <http://www.jisc.ac.uk/whatwedo/programmes/elearningpedagogy/vocabularies.aspx> [Accessed 14 February 2013].
- Campbell, L and Currier, S. (2005) Jisc Pedagogical Vocabularies Project: Priorities, Issues and Recommendations to JISC. Report 3 of Jisc Pedagogical Vocabularies Project. Available at: <http://www.jisc.ac.uk/whatwedo/programmes/elearningpedagogy/vocabularies.aspx> [Accessed 14 February 2013].
- Department for Business, Innovation and Skills (2011) Higher Education: Students at the Heart of the System Available at <http://discuss.bis.gov.uk/hereform/white-paper/> [Accessed 6 June 2012]
- Department for Employment and Learning, Northern Ireland (2011) Success through STEM: STEM Strategy. Available at: http://www.delni.gov.uk/es/2857p_stem_booklet_v5.pdf [Accessed 6 February 2013]
- European Science Foundation (2011) Health Research Classification Systems - current approaches and future recommendations. Science Policy Briefing 43. Available at: <http://www.esf.org/research-areas/medical-sciences/activities/science-policy/health-research-classification-systems-current-approaches-and-future-recommendations.html> [Accessed 10 Feb 2013]
- Higher Education Statistics Agency (2011) What is a Course? Report produced for Jisc. Available at: http://www.hesa.ac.uk/dox/publications/The_Course_Report.pdf [Accessed 20 May 2012].
- Jisc infoNet (2008) Exploring Tangible Benefits of e-Learning: does investment yield interest? Available at: <http://www.jiscinfonet.ac.uk/publications/exploring-tangible-benefits/> [Accessed 13 Feb 2013]
- Interim Regulatory Partnership Group (2012a) Mapping the higher education funding and regulatory system in England: summary report. Report produced by Deloitte. Available at: <http://www.hefce.ac.uk/about/intro/wip/interimregulatorypartnershipgroupirpg/> [Accessed 20 May 2012].
- Interim Regulatory Partnership Group (2012b) Redesigning the HE data and information landscape: a pathway to reform. Available at: <http://www.hefce.ac.uk/about/intro/wip/rpg/redesigningthedatalandscape/> [Accessed 29 November 2012].
- Manton, M. (2012) Sesame Project Final Report. JISC/HE Academy funded OER Phase 3 programme. University of Oxford. Available at: <http://www.tall.ox.ac.uk/research/current/sesame.php> [Accessed 5 February 2013]
- Miles, A. and Bechhofer, S. (2009) SKOS Simple Knowledge Organization System Reference. W3C Recommendation. Available at: <http://www.w3.org/TR/2009/REC-skos-reference-20090818/> Accessed 10 Feb 2013]

Organisation for Economic Co-operation and Development (2003) The Measurement of Scientific and Technological Activities: Proposed Standard Practice for Surveys on Research and Experimental Development "Frascati Manual" Sixth Revision 2002 OECD Publishing ISBN 978-92-64-19903-9. Available at:

<http://www.oecd.org/innovation/innovationinsciencetechnologyandindustry/frascatimanualproposedstandardpracticeforsurveysonresearchandexperimentaldevelopment6thedition.htm> [Accessed 5 February 2013]

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. Available at:

<http://www.oecd.org/innovation/innovationinsciencetechnologyandindustry/38235147.pdf> [Accessed 5 February 2013]

Pink, B. and Bascand, G. (2008) Australian and New Zealand Standard Research Classification (ANZSRC) 2008. Australian Bureau of Statistics and Statistics New Zealand. ISBN 9780642483584. Available at:

[http://www.ausstats.abs.gov.au/ausstats/subscriber.nsf/0/2A3A6DB3F4180D03CA25741A000E25F3/\\$File/12970_2008.pdf](http://www.ausstats.abs.gov.au/ausstats/subscriber.nsf/0/2A3A6DB3F4180D03CA25741A000E25F3/$File/12970_2008.pdf) [Accessed 5 February 2013]

UK Clinical Research Collaboration in collaboration with Modernising Nursing Careers (2007) Developing the best research professionals. Qualified graduate nurses: recommendations for preparing and supporting clinical academic nurses of the future. Report of the UKCRC Subcommittee for Nurses in Clinical Research (Workforce). Available at:

<http://www.ukcrc.org/publications/reports/> [Accessed 11 February 2013]

UNESCO (2012) International Standard Classification of Education ISCED 2011. Published by UNESCO Institute for Statistics Montreal, Canada ISBN 978-92-9189-123-8. Available at:

<http://www.uis.unesco.org/Education/Documents/isced-2011-en.pdf> [Accessed 5 February 2013]

White, D. and Manton, M. (2011) Open Educational Resources: The value of reuse in higher education. JISC-funded OER Impact Study, University of Oxford. Available at:

<http://www.jisc.ac.uk/media/documents/programmes/elearning/oer/OERTheValueOfReuseInHigherEducation.pdf> [Accessed 5 February 2013]

W3C (2001) URIs, URLs, and URNs: Clarifications and Recommendations 1.0. Report from the joint W3C/IETF URI Planning Interest Group. W3C Note 21. Available at: <http://www.w3.org/TR/uri-clarification/> [Accessed 22 February 2013]

8. List of Appendices

The appendices to this report are presented in a separate document.

- Appendix 1** Steering Group Members
- Appendix 2** Glossary of acronyms and terms
- Appendix 3** Institutions participating in Survey
- Appendix 4** Survey Responses
- Appendix 5** Comparison with approaches in Australia and New Zealand
- Appendix 6** Comparison with OECD Field of Science Classification
- Appendix 7** Comparison with Research Councils UK Classification
- Appendix 8** Comparison with International Standard Classification of Education
- Appendix 9** Comparison with HE Academy Discipline areas