



Funded by
UK Government

London LSIP Open Data Project Report

May 2023

Contents

Contents	1
About	2
Executive summary	3
Background	4
Project background	4
Creating sustainable data access	5
Project team	6
Approach	7
Scene-setting: survey results	10
Current systems and data collection	11
Benefits of skills demand data	12
Building an open data standard	12
Findings and Recommendations	14
Summary of interviews and workshop	14
College–student: course planning and registration	14
College–employer: understanding skills demand	16
Course capacity and fulfilment	17
Linking roles to courses	18
Student–employer: destination tracking	19
Summary of Recommendations	20
1) Develop an open dataset that maps SOC to qualification reference identifiers.	21
2) Establish a working group to define and test the qualification to four-digit SOC data mapping	22
3) Extend the course data that DfE aggregate from colleges and make available in real-time	23
4) Identify the governance or stewarding process	23
5) Test the aggregation of data from multiple sources using the new course data standard	24
6) Investigate how to improve efficiency and impact of learner destinations tracking	25
Conclusion	26
Appendix 1: Rocket Science report summary	28
Appendix 2: Cetus report summary	30

Appendix 3: Survey	32
London LSIP Open Data Project Survey	32
Demographics	33
Current systems and data collection	33
Data usage	34
Future expectations	34
Ongoing participation	35

About

This report was researched and produced by the [Open Data Institute](#) (ODI), and published in May 2023. Its lead authors were Ed Evans and Ian Cheng. If you want to share feedback by email or would like to get in touch, contact the LSIP Open Data project lead Ed Evans at ed.evans@theodi.org

To share feedback in the comments, highlight the relevant piece of text and click the 'Add a comment' icon on the right-hand side of the page.



**Funded by
UK Government**

The project is funded by UKG. The 'Funded by UK Government' logo has been designed to help the citizen quickly identify government backed work.

Executive summary

This document summarises the activities, findings and recommendations from the stakeholder engagement work in phase 1 of the Local Skills Improvement Plan (LSIP) open data project, as set out in the [project terms of reference](#). Alongside this, project partners [Rocket Science](#) and [Cetis](#) LLP worked on the proposed data-mapping process and proposed open data standard for linking further education courses to skills; their reports are summarised in the appendices.

The main aim of phase 1 of the study was to work with selected 'pioneer colleges', key software suppliers and sector experts to examine the feasibility of implementing a standard for publishing available or planned FE course information and then linking courses to the likely occupation outcomes of learners. This could then be matched to labour market demand forecasts by policymakers to assist with the identification of gaps in training provision. We also set out to build an understanding of the likely costs and effort involved.

The phase 1 project has fulfilled its aims and has six key recommendations:

1. Develop an open dataset that maps SOC-to-qualification reference identifiers, enabling colleges, government and the private sector to link existing data
2. Establish a working group to define and test the qualification to four-digit SOC data mapping
3. Extend the course data that the Department for Education (DfE) aggregates from colleges and make it available in real-time
4. Identify the governance or stewarding process
5. Test the aggregation of data from multiple sources using the new course data standard
6. Investigate how to improve the efficiency and impact of learner destinations tracking

Before this investigation, it was felt that the implementation of any change in this area may involve significant software change, either to the various management information systems (MIS) used within the colleges, or to the typically bespoke websites where colleges advertise courses and recruit learners. However, we discovered that stakeholders feel that these recommendations will *not require* significant software changes, nor indeed any large-scale data transformation.

Background

Project background

In the white paper on [Skills for Jobs: Lifelong Learning for Opportunity and Growth](#), the UK government set out its plans to bring together employers, further education (FE) colleges and other local stakeholders, to determine the measures needed to ensure technical skills training is responsive to employers' skills needs. Improving skills boosts economic growth and gathering data about skills provision plays a vital role in this.

Increasing access to course provision data and linking this to occupational outcomes can help improve the responsiveness of provision by FE colleges in the UK and ensure appropriate commissioning of education provision to support the development of the skills required in the economy.

As part of the London Local Skills Improvement Plan (LSIP), the [Open Data Institute](#) (ODI) is working with [Rocket Science](#) and [Cetis](#) LLP to support [West London Business](#) (WLB) and [BusinessLDN](#) in piloting course directories in an open data standard with FE Colleges.

The project is divided into two phases and this report is the summary of phase 1. Our objective for phase 1 is to help four FE colleges understand the likely benefits of using open data standards, and to find out if any upgrades would be required to their systems in order to pilot the implementation of an agreed open data standard for course information. Test implementation will be in phase 2.

In this phase 1 discovery stage, the ODI project team built on previous work around [mapping data ecosystems in the education sector](#). We used a range of research methods to gather information about what data and systems are already being collected or used by FE colleges in managing and publishing course information; understanding labour market skills' demand for course planning; and tracking learners' career destinations. In the process, we also engaged software developers to assess the appetite for, and challenges of, implementing the proposed open data standard. There is more detail about the approach in the next section.

Subject to further funding, the project will move into a second

implementation phase, where the project team will convene a stakeholder group to co-create a methodology for mapping qualifications' learning aims against occupation outcomes (SOC level 4), work on the data aggregation process and test these with one or more pilot implementations. We will also fine-tune the open data standard developed in phase 1 and develop guidance for wider adoption.

Creating sustainable data access

The ODI is a not-for-profit organisation focused on encouraging more open and shared data to create social and economic value. It is recognised that in order to create long-term change and long-term value, sustainable data access supported by data infrastructure is required.

Sustainable data access requires strong data infrastructure: by this, we mean standards, processes and governance to ensure continued collection and availability of a high-quality shared dataset. The data ecosystem supported by the data infrastructure needs to be stewarded responsibly by an entity, and that entity must have the financial and operational resources to run the system for the benefit of the stakeholders.

For this project, the ODI is considering these wider elements and is making recommendations designed to generate a sustainable data ecosystem.

For more information, see:

- [Sustainable data access](#)
- [Data infrastructure for common challenges](#)
- [Responsible data stewardship](#)
- [Data Landscape Playbook \(beta\)](#)

Project team

The project is sponsored by the **Department for Education** (DfE), which in addition to funding, provides the national policy perspective and guidance on project approach.

West London Business (WLB), on behalf of **BusinessLDN**, managed the project and was responsible for the overall success of the project.

Rocket Science documented its past [work for the West London Alliance](#) on defining data requirements and how to allocate likely outcomes to courses. Rocket Science used Cascot, an algorithm designed by the University of Warwick, to map text information course descriptions to the [Standard Industrial Classification \(SIC\)](#) and [Standard Occupational Classification \(SOC\)](#) codes. The executive summary of Rocket Science's report is in [Appendix 1](#).

The **ODI** led the strategic and engagement work with FE colleges and software providers to understand the data ecosystem surrounding course and skills data. The ODI will also provide recommendations on implementation of the proposed open data standard.

The open data standard was developed by the technical team at **Cetis LLP**, which translated data requirements into the appropriate use by FE colleges and other stakeholders. The summary of the report that documents their process of developing the open data standard is in [Appendix 2](#).

Throughout the project, the project team provided regular progress reports to BusinessLDN and WLB for ongoing monitoring and evaluation.

Approach

In phase 1, we took a multi-pronged approach to help us understand the

problem and develop solutions for it.

Firstly, we designed a survey to gather basic information around course information collection and usage, and to confirm the suitability of any FE college for the pilot. The survey asked for information about the systems used to store basic course information and manage course capacity, as well as for links to SIC and SOC codes relevant to their course offer.

In addition, we captured opinions on the balance between benefits and effort when capturing, analysing and reporting additional data in regard to business demand and student employment. Toward the end of the survey, we invited colleges to indicate whether they wanted to be a pilot college and to trial the open data standard. The results of the survey will be discussed in detail in later sections.

After this initial industry assessment, we carried out a series of focussed interviews with the data, technology or marketing teams of the pilot FE colleges. These interviews formed the basis of our research into the mechanics of sourcing, aggregating and publishing course and skills data using the open data standard. The following organisations were interviewed:

FE colleges

- Harrow, Richmond & Uxbridge Colleges (HRUC)
- Morley College
- Waltham Forest College
- West London College

Software providers

- One Advanced, provider of ProSolutions
- RCU, provider of Vector
- Parent Pay, provider of UNIT-e

Authorities

- Education Skills and Funding Agency (ESFA, part of DfE)

The interviews took the form of one hour discussions and helped us understand each individual's role in supplying or using course and skills data.

Combining information from both stages, allowed us to visualise the data ecosystem using a data ecosystem map which gave a generalised model of how data flows within the FE sector.

The ODI also facilitated a workshop with stakeholders in mid-April to set out the project team's findings so far, and to collaborate on next steps and recommendations for phase 2. More details on the ecosystem mapping, and our aggregate findings from the interviews and workshop appear later in this report.

Data ecosystem mapping

A data ecosystem is the people, communities, and organisations that are stewarding data; creating things from it; deciding what to do based on it; influencing any of those activities; or are affected by any of those activities. A data ecosystem map illustrates the different actors in a data ecosystem, and shows how value is exchanged across it. It aims to provide a high-level understanding about how data creates value.

Based on the project team's conversations with FE colleges, software providers and the ESFA, we created a data ecosystem map to identify stewards and users of different data in the FE sector, and show the relationships between them.

The green bubbles represent the FE ecosystem and how it relates to skills and course provision. This could be understood as a three-way relationship among learners, employers and colleges. This will be explored in detail later.

The orange bubbles denote the different datasets that are exchanged across the ecosystem, involving authorities such as government departments or public bodies (brown bubbles), as well as other contributors like software providers (blue bubbles). The arrows between bubbles show what type of data is being exchanged between actors.

[Link to [ecosystem mapping](#) tool here]

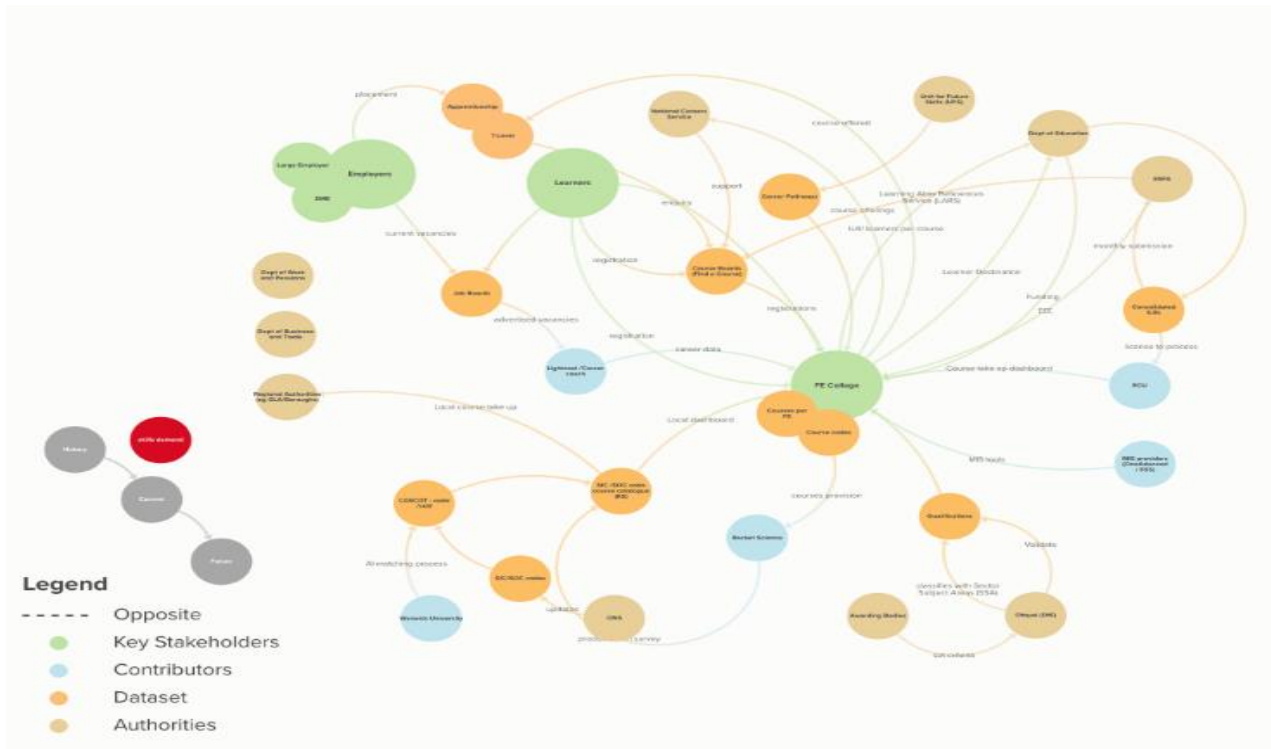


Fig 1: Data ecosystem map employers, learners and FEs

Scene-setting: survey results

The ODI project team developed an online survey targeted at data or technical leads in FE colleges. Its purpose was to understand the course information and website content management systems used by FE colleges, and also to gather views about the changes required to adopt a future open data standard for capturing and reporting course data matched to required skills. The full questionnaire is found in [Appendix 3](#).

The survey was open for two weeks between March 6 and March 17 2023. It was disseminated by BusinessLDN and the Association of Colleges to their respective member colleges. We received responses from 13 FE colleges¹ across all four sub-regions of London, which gives a response rate of at least 26%.

We received responses from technical leads, business leads and careers leads engaged in Management Information Systems (MIS), course marketing and data management.

¹ These colleges include Barnet and Southgate College, Big Creative Academy, Capital City College Group, Harrow, Richmond & Uxbridge Colleges (HRUC), London South East Colleges, Morley College London, NCG, New City College, Newham College of Further Education, South Thames Colleges Group, United College Group, Waltham Forest College, West London College and West Thames College.

Current systems and data collection

The majority of respondents use the following MIS systems:

- ebs by Tribal
- ProSolution by One Advanced and
- UNIT-e by Parent Pay

These systems help colleges manage student records for applications, enrollment, attendance, progress monitoring and exam results. They also support them in their Individualised Learner Record (ILR) return.

According to the colleges, course information is the basis of their business, from planning to running courses. Since it is an essential business need, there should be robust processes to support it, and centralised and integrated systems for curriculum planning and management.

While colleges generally do not find it difficult to collect the information, it was clear that a large amount of work goes into the creation of course data every year, and the situation is further complicated by the high volume of courses. In particular, the cadence of how a school year runs affects how frequently data needs to be collected – one college said that although course planning is completed by April, enrollment data would only come in August, with part-time courses starting even later.

All colleges reported that they publish course information on their own website. However, course information is also published on the DfE's [National Careers Service](#) website. In addition, other arms-length bodies such as the [ESFA](#) and [UCAS](#) have access to the course information.

Benefits of skills demand data

Most colleges believe having additional information about the occupational links for courses will help inform curriculum planning and allow them to match course supply to industry demands, thereby strengthening the offer for students.

In the long run, this helps learners leave the colleges with the relevant skills that allow them to progress to employment and with the skills required by businesses. It would also help colleges if they could look at emerging sectors to develop new provision.

Colleges are also able to make comparisons between their courses and those offered by rival colleges in other ways. One example is Vector, a software package that aggregates ILR data under licence to produce dashboards showing how the colleges compare to their competitors.

Some respondents said a centralised, open data source of labour market information that also reflects local demands, or student volumes by secondary school in each borough, would be desirable.

Building an open data standard

The biggest concern for colleges about adopting an open data standard around course information is about the interoperability of data with their existing system provisions. Some said that their current data is tailored for submission to ILR or requirements of funding bodies, and the new standard must cater for this.

If the new standard requires data that is not easily collected, then the vast number of courses, coupled with limited technical skills, might result in an undue administrative burden on those implementing the open data standard.

Colleges said they need to be given sufficient time to adopt the standard and comply, depending on the level of detail the data standard might require.

In terms of support, many felt they needed to get software suppliers or web developers on board, as colleges rely on their MIS or web content management systems to generate the required output at ease. They would also want to see clear and thorough specifications for data collection, especially if the new standard requires data that is not currently held.

There should be reasonable flexibility regarding requirements such as timelines for updating, checking or publishing new information, mandatory fields, and data validation processes. Respondents said colleges should also be given clear guidance, with an easy way to allocate SIC or SOC codes if they are mandatory fields in the data standard. One suggested incorporating the codes into the National Learning Aims database to ensure consistency.

Many felt a responsive and supportive helpdesk to handle any problems from colleges was an essential support. Training for staff is also needed.

Findings and recommendations

Summary of interviews and workshop

To investigate the questions framed by the project in more detail, a number of key stakeholders were interviewed (full list in the [Approach](#) section above). Following this series of interviews, the same key stakeholders were invited to a workshop where we were able to share findings and generate consensus on the main recommendations, as reported here.

The information gathered focused around course planning and registration, understanding skills demand, and links between FE colleges and employers. Additional information was shared about approaches to tracking learner destinations, as this is seen as an important indicator for prospective learners and key information for policymakers.

Each college has its own blend of course types. We focused on those leading to qualifications from level 3 to level 5 so that we could assume that the courses discussed linked to a qualification recognised by the Office of Qualifications and Examinations Regulation (Ofqual) and came from an awarding body. Respondents also concluded that courses at these levels would generally have clear intended occupational outcomes.

College–student: course planning and registration

- Annual changes to course provision is a relatively small percentage of overall course provision
- Increasingly colleges recognise the value of describing a course with regard to developing careers and career skills
- Learner registration is reported to the DfE via ILRs
- Funding is based on the ILR

Each college has an established process for course planning, which is

critical to its continuing success. Not all courses follow the same pattern, but the course planning cycle typically begins in October and completes during the following September.

During this window, colleges must assess which types of courses and what number of places to offer. The key people in this process are the curriculum team.

Provision is largely based on history and a track record of success, with new course offers, or changes in course offering, making up a relatively small percentage each year.

Colleges describe the courses to attract learners, noting the activities, formats and skills developed leading to a particular qualification. Qualifications are set by awarding bodies and accepted by Ofqual.

Offered courses are posted on the college website and also on course boards – like the NCS’ [‘Find a Course’](#) and [‘IDPconnect’](#), and registration is linked through to the college website.

Increasingly, colleges recognise the value of describing a course in terms of developing careers and career skills, with some colleges presenting ‘career options’ using a range of offered courses at different levels with indications of potential future earnings for someone with that qualification. Some colleges use tools like [LMI For All](#) to do this.

When a learner is registered for a course, an ILR is created or updated. These records are submitted to the DfE, and the college receives funding on this basis. The DfE gathers the number of learners per course (and qualifications) per college. Learner information comes with additional data, such as contact address, indicating location and likely travel. However, these returns do not capture the total learner capacity on a course, which is needed for LSIPs and strategic decision-making.

Consolidated ILRs are available from the DfE under licence. Data aggregators, like RCU, provide dashboards based on ILR data for colleges within the Vector system (that provide information about learners and courses within a local or national context). By using these systems, colleges can make more informed judgments about the decisions learners are making, and better plan their course provision.

College–employer: understanding

skills demand

- Matching provision to skills demand is essential to making courses more attractive to prospective learners
- While many sources of data are used, estimation of demand is a matter of judgement
- School-leaver data is very accurate, but data around employer demand is not accurate
- Links with employers take a number of forms and are strengthened through apprenticeships and T-levels

Each college supported the view that more information and more accurate information about the demand-side of the equation for course provision would be welcome for future course planning. While changes to course provision and capacity each year may be a relatively small percentage, offering attractive and up-to-date insight is recognised as essential to ensure that a college is responsive and can add to the evidence base that underpins future [Accountability agreements](#). Reliable historical data is most important.

The colleges reported a number of different information sources, while recognising that estimating demand remains a matter of judgement. There is an expectation on each faculty team to make this judgement call. There are no formal ways of producing expected demand data, beyond basic cohort numbers. The colleges expect to juggle courses and availability throughout the registration process.

For FE colleges, demand coming from local schools is a key part of the picture. Reliable information about the number of students in a particular cohort, the schools attended, and 'home' location are all available. RCU's Vector tool provides this aggregated information along with information, based on the ILR about recent take up of FE college courses.

Colleges use services provided by other data aggregators. For example, [Lightcast](#) is an organisation that uses AI to process information from job boards and other open job postings to provide useful information about demand, roles and potential salaries.

It is recognised that links with employers are very important, though there is no formal process for collecting local demand from them. Colleges reach local employers through a number of different channels (for example, faculty teams arrange work experience and job fairs). It is recognised that employers have a key role in setting demand but it is difficult to make estimates at the local level, or for individual employers. National, regional

and sector-level demand can be better estimated. It is recognised, too, that larger employers and new large local projects may provide specific information around local demand, but that most employers are Micro, Small and Medium-sized Enterprises (MSMEs) that are less able to forecast demand.

Links with employers are strong in colleges where there are courses with elements of work experience. Where T-levels are being planned, links with employers will be strengthened, as the course demands a significantly increased amount of work experience hours.

Course capacity and fulfilment

- Actual course take-up numbers are very accurate
- Colleges approach course capacity with a degree of flexibility, so numbers related to the take-up of capacity are estimates

During the investigation, a question arose about the ability to measure course capacity and fulfilment. The actual numbers of learners attending each course is available via the aggregated ILRs. The data is available to trusted data aggregators under licence. The calculation of how this relates to the take-up of available or potential capacity can only be roughly estimated.

Any individual college will be able to provide details of course utilisation from their perspective, in retrospect. It is noted that the capacity numbers for courses can be very flexible until the course actually begins. For example, where there is a low take-up, courses can be cancelled and resources reassigned. On the other hand, additional capacity beyond that originally planned may be found where demand is high. It is noted that the level of flexibility of capacity is very different for courses requiring specific equipment, such as hairdressing, compared to courses where the teaching is online or not limited by classroom space.

Linking roles to courses

- FE colleges were generally not so familiar with SIC/SOC
- Some MIS systems provide a SOC recording function
- Rocket Science developed and tested a manual process for linking SOC codes to courses
- We recommend building and testing a process for linking SOC codes to qualifications/ learning aims, and therefore automatically to courses

Courses can be linked to skills, and therefore likely occupations/job roles, through the use of SIC and SOC codes. FE colleges were not particularly familiar with these codes and did not use them. At least one MIS provider has available data fields for listing SIC and SOC codes to be completed when courses are defined. In this example, the code/ MIS system capability was not being used in practice and it was left to users to populate the data field with no detailed guidance.

The Rocket Science report in [Appendix 1](#) provides a clear illustration of a process to assign SIC and SOC codes. Tested with courses for six London FE Colleges, the reported process shows that it is a relatively straightforward but time-consuming manual task. If this were to be the preferred linking method, then the process could become more automated and scaled to a national level with an identified and agreed service provider.

While limitations with SIC and SOC were noted through the focused discussions and the workshop, participants felt that the SOC code should be used as it is:

- the best hierarchy available
- recognised internationally, and
- will continue to evolve.

At the workshop, the group felt that SOC codes should be linked at the qualification level to reduce overall effort and eliminate an additional burden on FE colleges. A SOC code or codes linked to a qualification would be 'inherited' automatically by any courses assigned to that qualification. This report recommends that the process to link SOC codes to qualifications is defined by a group of convened experts.

Student–employer: destination tracking

- A lot of effort goes into finding out the destinations of learners, as it is seen as a very important dataset for future planning
- For most colleges (with the exception of Waltham Forest), the data collected is patchy and unreliable
- From the FE perspective, an investigation into how to improve destination data and destination-data collection would be very welcome

Learner destination information came up as a key topic in discussions with the FE colleges. With one exception, it was felt that collecting this information required a lot of effort for little reward, typically generating a patchy and unreliable dataset. One college acknowledged a very high response from leaving learners, and that it collected a wide range of additional information through a process of continual engagement wrapped up within career counselling.

The destination of learners following completion of a course, whether continuing in education or entering employment, is information that is very interesting to learners considering a course. Therefore it is important information to FE colleges too, and contributes to making course offerings attractive. Destination information is also useful for FE colleges to help them plan future course provision.

Collecting destination information is a process that is part of the ILR submission. FE colleges are expected to complete destination submissions per learner within four months of course completion. The information in this process is limited to high-level descriptions, for example 'went on to employment'. This leaves out useful information such as the employer and the role.

With the one exception noted above, collecting destination information is limited to attempted phone conversations with former learners. This process is expensive and yields low response rates. It was recognised that better destination information that is much easier to collect would be very useful to FE colleges. More granularity is required to show roles and skills, making it easier to establish a link to potential careers and earnings.

Summary of recommendations

The main purpose of phase 1 of the study was to examine the feasibility of implementing a developed standard for linking skills to courses, and to explore the cost and effort involved in doing that. During this work, we have been able to review the ‘manual’ course and occupations mapping process developed by Rocket Science and for Cetus LLP to develop the basis of a data standard.

We have been able to create a group of stakeholders, including four volunteer ‘pioneer’ FE colleges, software suppliers and education experts who are prepared to continue working on the recommendations summarised below in a phase 2.

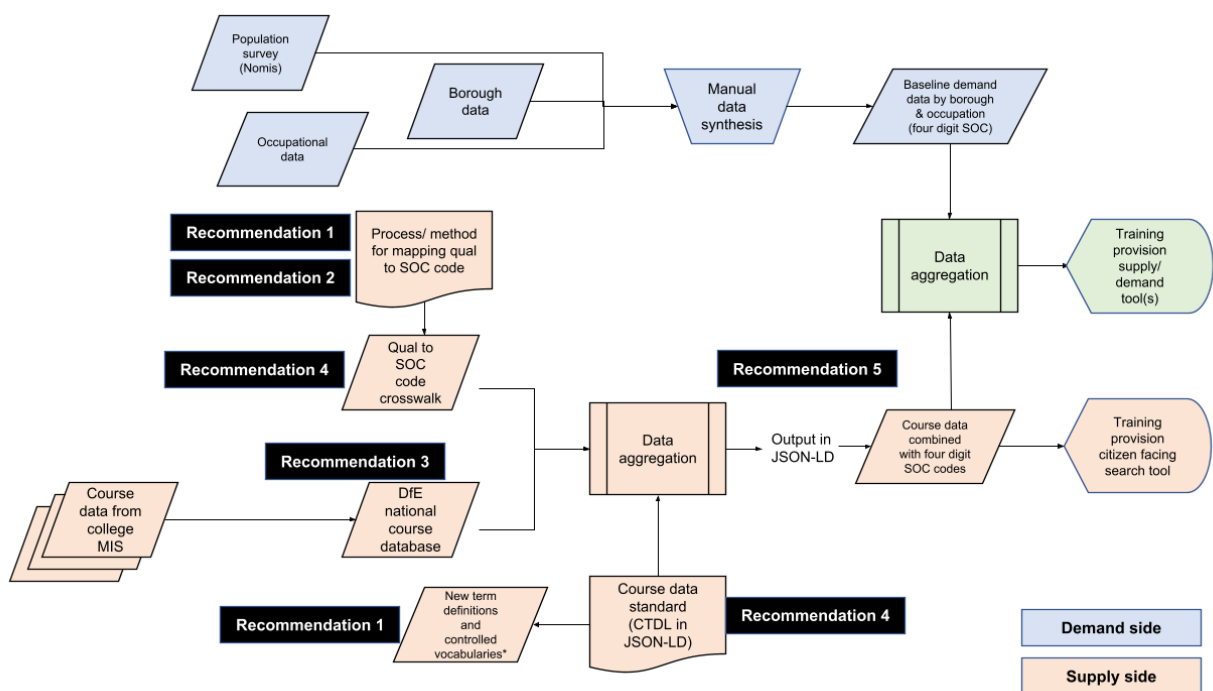


Fig 2: Summary of process and recommendations, for 1-5

1) Develop an open dataset that maps SOC to qualification reference identifiers

Work with stakeholders to develop an open dataset that maps SOC to qualification reference identifiers, enabling colleges, government and the private sector to link existing data.

We support the Cetus LLP preferred option with the data mapped to a specialised data standard for qualifications, Credential Transparency Description Language ([CTDL](#)), in [JSON-LD](#). Although not as widely used as Schema.org, CTDL is designed specifically to describe the types of resources that this project will deal with, and its use in similar projects is gaining some traction. It also uses the linked-data paradigm to enhance reuse of data. This option would require the creation, maintenance and publication of some term definitions and controlled vocabularies in RDF, and a JSON-LD context file, as well as a JSON-Schema and/or SHACL files to allow validation of any data provided.

We do not expect any significant changes will be required to MIS systems or the course registration process, although a formal note from DfE to the major software vendors may be useful at the right time. We do not believe that the process will require additional data effort from the FE colleges and we do not expect to need any large-scale data migration or data transformation.

We are able to recommend a number of additional steps, building on phase 1 activity, to advance the open data ecosystem.

2) Establish a working group to define and test the qualification to four-digit SOC data mapping

The aim is to develop the mapping of SOC code(s) linked to awarding bodies' qualifications. These will then be 'inherited' automatically by any courses assigned to that qualification.

- Focus on level 3-5 FE courses for the mapping (perhaps level 2 for some quals)
- Resource and convene a working group to define the method and do the data mapping: Cetus, Future Skills, RCU, Rocket Science, ESFA
- Steps:
 - Create the data mapping (team)
 - Define a repeatable process for mapping the data (team)
 - Apply as test/refine/plan to automate (with selected courses)
 - Working with pioneer colleges and particular qualifications
 - Define as an open standard (Cetus LLP)
 - Specify host platform and format for storing/accessing Mapping (Crosswalk) table
 - Process owner identified/ownership accepted

3) Extend the course data that DfE aggregates from colleges and make it available in real-time

Work with the ESFA/National Careers Service Course Finder team as an established data aggregator to determine:

- How and when learner volume (actual/ potential capacity) could start to be captured as part of colleges' data returns
- How course directory data currently published regularly in CSV format under an Open Government License² could become a live feed to support citizen-facing websites and services for decision-makers. This external facing feed could adopt the new course open

² Education and Skills Funding Agency (2023), '[ESFA: course directory](#)'

data standard to maximise its socio-economic impact.

4) Identify the governance or stewarding process

The aim is to find a suitable organisation to take 'ownership' for the qual-to-SOC mapping process, its data table and the course data standard moving forward.

Investigate and report on options for stewardship for the process and the standard with stakeholders.

Consider:

- An existing organisation — ESFA/DfE/Ofqual
- An independent data steward
- Establishing a stewarding body
- Potential for additional standards in destination tracking

5) Test the aggregation of data from multiple sources using the new course data standard

The aim would be to:

- Publish the course data standard — creation, maintenance and publication of some term definitions and controlled vocabularies in RDF, and a JSON-LD context file, as well as a JSON-Schema and/or SHACL files to allow any data provided to be validated.
- Use a new version of Rocket Science's supply/demand tool to test the draft course open data standard against the following changes to the data ecosystem:
 - MIS access to course/qual-to-SOC code data, data mapping access; and
 - Live feeds from ESFA/National Careers Service Course

finder.

- Match with demand-side data produced by Rocket Science at Local Authority level (for London) by 4-digit SOC codes and with a sector subject area consistently linked to each 4-digit SOC code; it is assumed producing this dataset will remain a manual process for the moment.
- Agree who might be the long-term owner of the tool (for example, the [DfE Unit for Future Skills](#))

6) Investigate how to improve efficiency and impact of learner destinations tracking

Though out of scope for this project, it is clear that learner 'destinations' data has a key role to play within the landscape of matching job requirements to course provision. Most of the required data is within the UK Government's data ecosystem and could provide a very accurate picture. At present, the FE colleges are asked to check on learner destinations and most find it an expensive and time-consuming way of collecting poor quality data. Waltham Forest College is the exception to this.

An investigation into developing destination data could identify a solution with many benefits in this space.

Conclusion

Phase 1 has delivered the outcomes identified at the start of the project within a compressed timescale and within budget. The team has made key recommendations that will move the project forward through significant next steps.

Specific objectives for phase 1 are reported below with a note to confirm what happened:

1. Agree four FE colleges ready to become pilot projects for this initiative - the 'pioneers group'
Measure: A list of 4 FE colleges
Actual: HRUC, WLC, Morley College and Waltham Forest College volunteered to become 'pioneer colleges' and are ready to continue participation in phase 2
2. Cetus LLP to specify the data standard and articulate the benefits and barriers to adoption (based on data requirements from RocketScience)
Measure: Input for the required reports
Actual: See Cetus LLP and Rocket Science reports in the appendix
3. Capture a summary of the pioneer colleges' high-level ecosystem for course management, course outcomes reporting and marketing (through website CMS) and note the software suppliers
M: Summaries of the systems of each of the 4 colleges
Actual: We produced notes from each college interview and generated a useful, high-level ecosystem map. These notes were shared within the project team as they were collected, and are the basis for the findings and recommendations
4. Work with their current software providers to specify what upgrades are required to their course management systems, and any changes required to websites to publish course directories and supplementary information through open APIs to the agreed open data standard.
Measure: Each supplier/software developers to develop a cost/effort plan for the upgrade
Actual: The major software providers were approached and three of them were interviewed. Each was given detailed information about the aims of phase 1 and the wider project. Based on findings, it is not expected that significant software changes will be required
5. Survey other London FE colleges' IT ecosystems to understand likely pan-London costs of implementing standard
Measure: ODI produce likely cost/effort plan for pan-London implementation

Actual: A survey of all London FEs showed strong support for the work and indicated that the software/data set-up as seen in the pioneer colleges could be considered typical. Based on findings, it is not expected that significant software changes will be required

6. Share plans (final phase 1 report and roundtable) for implementation with DfE, pioneer colleges and other interested parties

Measure: Shared plans with key stakeholders (possible roundtable meeting)

Actual: We held a roundtable discussion with a range of stakeholders to review the findings and to make the recommendations in this report. We invited the pioneer colleges, the software providers, the DfE, GLA and the phase 1 project team

The recommendations in this report are supported by the project team and the pioneer colleges. The group of stakeholders are interested in continuing to provide input for the next phase of the project. A phase 2 project should be planned to deliver the recommendations from this report:

- Steps to define and test the data mapping for SOCs linked to qualifications
- Identifying options for the governance, stewarding process

Consideration should be given to commissioning a project to look at improving the collection and quality of learner destination data.

Appendix 1: Rocket Science report summary

This working paper sets out the process developed by Rocket Science in 2020 to create the supply data from West London colleges to feed into the supply and demand tool launched by the West London Alliance in 2021. This is to inform a data standard project being led by West London Business for BusinessLDN and delivered by the Open Data Institute (ODI). The project will ultimately investigate whether this manual process could be automated in some way.

The paper also sets out Rocket Science's understanding of XCRI-CAP, although it is not clear whether the data standard developed for Post-Graduate courses could be extended to accommodate Further Education data and tackle the issue of understanding both what courses are available and the places available on those courses. This will be addressed by Cetus LLP. A data standard or process is needed to bring greater transparency to understand the skills pipeline, whether existing provision meets job demands, and what courses are needed to plug the gap.

Since 2021, there have been developments on these issues such as the National Careers Website, which features relevant courses and providers, and the development of Vector. This is software used by colleges to align courses with labour market data. However, this is proprietary software and, like most economic demand studies, is not freely accessible.

As part of this project, we will be conducting an updated review of data sources and opportunities as well as gaining a better understanding of the Working Futures initiative and developments around London's Local Skills Improvement Plan (LSIP) applications, all of which have commissioned some form of skills demand forecasts that may be helpful in the next phase of this work.

This report should be used in conjunction with the data capture form and the existing data on 1000 courses that have already been allocated Standard Occupational Classifications (SOC) and Standard Industrial Classifications (SIC) codes already provided to the project team by Rocket Science.

The [full report is available here](#) with detailed recommendations on pages 12-15 that have been largely integrated with the ODI's headline recommendations.

Appendix 2: Cetis report summary

From discussions with stakeholders so far, there is a need to join up data on the supply of training and education with demand in the labour market in the London area (and probably nationally).

From the education and skills training side, we can adopt a standard that can capture the supply of courses and qualifications and provide an articulation to the demand side in the form of related occupational outcomes.

Data from the demand side is out of scope for this work, but we are assuming that at some point we can connect this data to a source of demand data (for example, based on existing data sources such as Working Futures, the Vacancies Survey, and related government and devolved authority datasets) and [draw on the experience of Rocket Science](#).

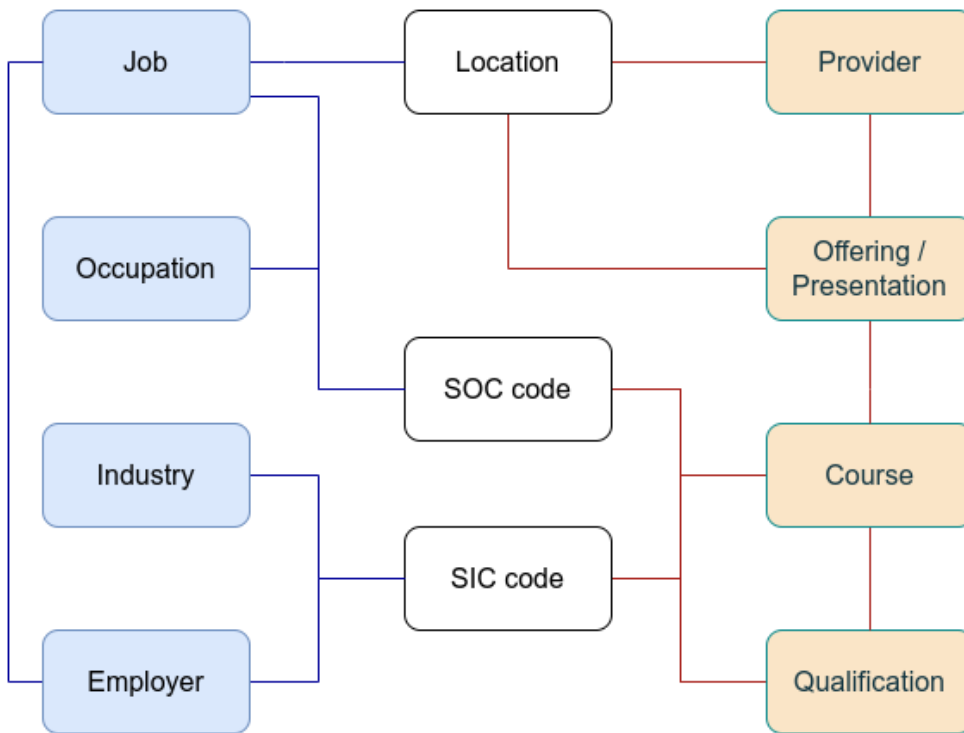
As set out in 'The case for change' in the project [Terms of Reference](#), the intended use of data conforming to the standard includes, but isn't limited to:

(a) dashboards and analytics to inform decision making, for example at a college, borough or regional level that link course 'supply' with forecast labour market 'demand', and

(b) services, such as course search portals, that aim to connect people to education and training opportunities that are connected with local labour market demand.

Cetis has created a logical data model that supports the London Local Skills Improvement Plan (LSIP) process by allowing the provision of relevant data about courses and the qualifications they lead to, and linking this data to relevant occupations and industries.

High-level view of the entities and relationships relevant to the supply of training and education with demand in the labour market of any locality



The full report is available [here](#) with Cetus LLP’s recommendations on page 41. Their preferred ‘option 4’ is adopted in the ODI’s headline recommendations.

Appendix 3: Survey

London LSIP open data project survey

As part of the Local Skills Improvement Plan (LSIP), which BusinessLDN is coordinating across London, the Department for Education (DfE) has funded a workstream looking at how to apply open data standards to course directories, also augmenting this information with learner volumes and occupational outcomes.

This will ultimately help colleges to better plan provision against market demand and citizens to find courses more easily and understand their likely occupational outcomes.

This project is in the investigation stage, with coordination from colleagues at West London Business. The Open Data Institute, Rocket Science and Cetus LLP are working together to look at the data and IT systems implications of introducing a common data standard.

As part of the project, we are looking for **data or technical leads** within your college to fill out this short survey, which seeks basic information about your current course sign-up or fulfilment tracking systems/ website Content Management Systems and your initial views about changes required to adopt the future open data standard.

Participation is voluntary. The responses to this survey are confidential (results will be shared in an aggregated format). There is an option at the end for a small number of colleges to volunteer to join the **project pilot**; ideally we would like one '**pioneer college**' from each of the capital's four sub-regions, but need all colleges to respond to this initial survey exercise.

This survey should take approximately 10 minutes, depending on the level of detail in your responses.

Demographics

1. What is the name of your college?

2. What is your job title?

Current systems and data collection

Please provide information about the systems you are using to store basic course information (title and description), manage course capacity and places taken-up, as well as any links to Standard Occupational Classification (SOC) and Standard Industrial Classification (SIC) codes relevant to your course offer. We are also interested in any systems you may use to track the employment destination of students.

Typical course information includes course name, faculty or course grouping, qualification, capacity, actual take-up, student employment, etc.

3. Please list the IT systems you are using to collect course information? (If you can, please provide system name, supplier, main purpose, data store, local or remotely hosted for each).

4. Please list any other IT systems where the data sets are stored (e.g. City Hall, EFSA, DfE systems)?

5. On a scale of 1 to 5, how difficult is it for you to collect information about courses, capacity and take-up?

Very easy

Very difficult

1

2

3

4

5

6. What makes it easy or difficult for you?

Data usage

7. From your perspective, how valuable is information about courses linked to business needs information?

Not valuable

Very valuable

1

2

3

4

5

8. Where do you publish your course information? Please include information on your website Content Management System here and where/how data is sourced internally for your website.

9. What tools are used to analyse and report course information?

10. What other outside parties have access to your course information?

Future expectations

As changes are being considered, this section asks about your views on the balance between benefits and effort when it comes to capturing, analysing and reporting additional data in regard to industry needs and student employment.

11. How do you think your college could benefit from additional information about business needs?

12. If a new open data standard on course information were to be adopted

by your college, what would be challenging and why?

13. If a new open data standard on course information were to be adopted by your college, what support would you need?

Ongoing participation

We are looking for a pioneer college from each of the capital's four sub-regions to help us further with this study. Initially, this will involve a 2-hour meeting to capture more details about your IT systems, introducing our project team to your software suppliers to discuss potential upgrades, and a later invitation to a roundtable event to discuss results and recommendations. Pioneer colleges will get the first option to participate in piloting any agreed implementation plans.

14. Would you like to be considered as a 'pioneer college' to help further with this study?

- Yes
- No

15. If so, please provide your contact details. Leave this blank if you do not wish to.

16. Are there any other comments you would like to make?