



Department
for Education

Developing and enhancing a labour market information database: LMI for All

Stakeholder Engagement and Usage, Data and Technical Developments (2018-2019)

Research report

January 2020

Sally-Anne Barnes, Jenny Bimrose, David Owen, Rob Wilson and Daria Luchinskaya – Institute for Employment Research, University of Warwick

Graham Attwell and Philipp Rustemeier – Pontydysgu

Contents

1.	Introduction	1
1.1.	The development of LMI for All: An overview	2
1.2.	Programme of work, 2018-2019	3
1.3.	Report structure	4
2.	Technical infrastructure and developments	5
2.1.	Objectives and outcomes	5
2.2.	LMI for All Infrastructure	5
2.3.	Server capacity and usage	7
2.4.	Review of the infrastructure	9
2.5.	Extract, Transform and Load (ETL) development	9
2.6.	Redevelopment of the LMI for All website	10
2.7.	Summary and next steps	11
3.	Data developments	13
3.1.	Objectives and outcomes	13
3.2.	Overview of the LMI for All database and organising framework	14
3.3.	Data refreshed	19
3.4.	Methodology and approach to providing data	20
3.5.	Potential future developments	21
3.5.1.	O*NET API	21
3.5.2.	Real time vacancies	22
3.5.3.	Career pathways data	23
3.5.4.	Linking education courses to occupations using LEO	28
3.6.	Summary and next steps	29
4.	Stakeholder engagement	31
4.1.	Objectives and outcomes	31
4.2.	Stakeholder plan	32
4.3.	Stakeholder activities and dissemination of the LMI for All service	33
4.3.1.	DfE Workshop	34
4.3.2.	CDI workshops and masterclasses	34
4.3.3.	Webinars	35

4.4. Online activity	35
4.5. Take-up of LMI for All and Careerometer	36
4.6. Summary and next steps	37
5. Future developments	39
References	41
Annex A: LMI for All reports	42
Annex B: Review of DLHE and Graduate Outcomes data for LMI for All	44
Annex C: Further information on LEO	45

List of figures

Figure 2.1 LMI for All infrastructure.....	6
Figure 2.2 Long-traffic to LMI for All, July-October 2019.....	8
Figure 3.1 Percentage of queries by dataset, July-October 2019	15
Figure 3.2 Overview of data refreshment cycle.....	20
Figure 3.3 Administrative data sources for LEO data	27

List of tables

Table 3.1 LMI for All Data overview, September 2019.....	17
Table A.1 Overview of LMI for All development.....	42
Table B.2 Review of DLHE and Graduate Outcomes data for LMI for All	44
Table B.1 LEO variables for UK-domiciled graduates.....	45

1. Introduction

The importance of high-quality career interventions throughout an individuals' lifespan has long been recognised. A pivotal feature of high-quality advice and guidance is reliable and up-to-date labour market information (LMI). One of the government's priorities set out in Careers Strategy (Department for Education, 2017) is to improve social mobility and well-being, offering opportunities to everyone. The Careers Strategy noted 'LMI for All' as a reliable source of LMI to support career transitions and enable access to information on opportunities (Department for Education, 2017). This follows on from a recommendation from Gatsby that 'LMI should be available to all schools through the 'LMI for All' service (The Gatsby Foundation, 2014, p.11). LMI for All continues to be recognised as a key resource throughout the careers stakeholder community in the UK (Alexander, McCabe, and De Backer, 2019; Barnes, 2015; The Careers & Enterprise Company and WorldSkills UK, 2016) and internationally (CEDEFOP, 2016; OECD, 2015).

The ultimate aim for the careers LMI database, known as LMI for All¹, has therefore been to provide a single access point for multiple, high quality sources of national and regional data, which is openly accessible. This development has enabled access to a number of publicly funded and open large-scale longitudinal databases, including the Annual Survey of Hours and Earnings (ASHE), the Labour Force Survey (LFS) and the Business Register and Employment Survey (BRES). By optimising access to these datasets, and others, through LMI for All, data are available in a way that allows them to be used in a number of career related interfaces. These are supporting individuals in making more informed decisions about learning and work. Data in the LMI for All database can help answer such questions as: where are the jobs of the future?; how much can I earn in this job?; what skills do I need to this job?; I have good technical skills, what jobs can I do?; and, what do people get paid in this job?

From 2012 to 2016, the LMI for All service was funded and managed by the UK Commission for Employment and Skills (UKCES). Since then the Department for Education (DfE) has managed and funded the service. A consortium led by the Institute for Employment Research at the University of Warwick has operated the service with Pontydysgu, an educational technology organisation². Details on the development of the service are presented in section 1.1 (Annex A provides an overview of the development phases of LMI for All and associated reports).

LMI for All has changed the way information on the labour market is made available and used by a range of beneficiaries, (e.g. government, careers organisations, schools and colleges), both across the UK and internationally, to support the delivery of careers

¹ LMI for All service website at: <https://www.lmiforall.org.uk/>

² The consortia is referred to as 'the team' throughout this report.

education, information and guidance services. As the database becomes increasingly widely used, it benefits both careers service providers and users of services, enabling access to up-to-date and robust data on the UK national and regional labour markets. In turn, beneficiaries are enabling careers professionals to provide high quality and impartial information to support career decisions.

1.1. The development of LMI for All: An overview

The LMI for All service was first piloted, then developed from 2012, with funding from the UK Commission for Employment and Skills (UKCES) to respond to a number of issues and concerns. First, despite the increasing emphasis on the importance of LMI for supporting individual labour market transitions and career choices, general access to a number of publicly funded and open large-scale databases was limited. In the absence of high quality and reliable data, a range of partial information and data with unknown provenance was increasingly becoming available online. For individuals searching for information on the labour market to help with their career choices, this resulted in conflicting and/or dated information being used. The publication of the 2012 Government open data white paper 'Unleashing the Potential' created new possibilities for the delivery of careers LMI and intelligence (HM Government, 2012). The white paper set out how the Government was opening up access to publicly funded datasets to unlock the potential of open data.

Second, UK public services were increasingly using the internet and technology to deliver services online and provide greater access to information. Alongside these changes, careers delivery services (in both the public and private sectors) were also using advances in technology to enhance the services offered to individuals. It was apparent that there was a need to strengthen and improve the quality of LMI for careers and employment practice, essential to inform the choices of individuals wishing to enter or re-enter the labour market, or to move between jobs. To respond to these changes and support the enhancement of careers service delivery, the feasibility of a careers database was piloted drawing upon access to open data and technological advances in data linking and processing. The careers database was piloted to evaluate whether such a service could provide easy access to improved data through a single portal and have the potential to enhance careers delivery services. The pilot confirmed the feasibility and viability of the foundation concepts and ideas to create a database with a test infrastructure created. The careers database became known as the 'LMI for All' service.

After successful completion of the pilot, from 2012-2015, the careers database was expanded with new and updated datasets, as well as undergoing further testing. The

technical infrastructure was improved and the application programming interface (API³), which enables access to the database through querying, was positively evaluated through two sets of hack and modding days. During this development period, a range of additional datasets were explored their potential to add value to the service. The service was also promoted through a number of stakeholder events to raise awareness of the resource and its potential to support those wishing to integrate LMI and intelligence effectively and creatively into a career's education and guidance context. A number of applications were developed, and third parties started to embed data from the LMI for All database into their own websites, integrating it as part of their careers offer. From 2015, the LMI for All service was given full project status by the Government and has continued to be maintained, updated and enhanced up to the present time. The overall aim of LMI for All continues to be:

- To increase and widen use of high-quality labour market information in order to support decisions about careers and learning.

It has been proven that the LMI for All service is able to meet the needs of a wider range of end users and their customers. A key element of continuing the successes of 'LMI for All' is to embed this significant and substantial resource not only through continued efforts in dissemination, but also in active engagement with stakeholders. This ensures that understanding continues to be developed of how the resource can be harnessed and integrated to enhance support to individuals and to careers organisations in a manner that increases the efficacy of individual labour market transitions.

1.2. Programme of work, 2018-2019

This report details work undertaken to support the delivery of the LMI for All service from October 2018 to September 2019, as well as the work to improve the service by extending it with new datasets and integrating new technologies. Work over this period has comprised three interrelated work streams: data, stakeholder and technical. Objectives for the programme of work were as follows:

- To maintain a comprehensive, high quality data offering that can inform career choices;
- To maintain the supply of data to current and future third-party users thereby meeting the requirement to provide open access to government data as set out in the eight Principles of Open Data⁴;

³ An Application Programming Interface (API) is a set of routines, communication protocols, and tools for building software applications. API queries produce formatted responses into a universal data format known as JSON (JavaScript Object Notation), before sent back to the developer/client.

⁴ The Eight Principles of Open Data, include: complete; primary; timely; accessible; machine processable; non-discriminatory; non-proprietary; and license-free (for more information, see: <https://opengovdata.org/>)

- To offer a technical infrastructure that provides a secure, engaging, accessible and reliable platform for LMI for All;
- To promote the widest possible take-up among third-party websites and applications as a means of opening-up the data to individual decision-makers; and
- To build wider awareness, understanding and support for LMI for All among key stakeholder groups, including policy, careers and technical communities.

LMI for All has continued to be developed in line with open data principles, with outputs from the three project components allowing community, voluntary, public and private sector organisations to access labour market data more effectively and efficiently. Consequently, these organisations are able to develop tools and interfaces that meet the needs of different customer groups in ways that augment existing careers education, guidance and information services being used directly by individuals to support the career decision process.

1.3. Report structure

Following this introduction, the report comprises three main sections following the three interrelated work streams, namely data, stakeholder and technical. Each section sets out the objectives and outcomes achieved in this reporting year, October 2018 to September 2019, as well as proposed next steps for each work stream.

Section two presents details on the LMI for All technical infrastructure, and includes a review and assessment of the infrastructure in terms of server capacity and usage. Over the last year, development work on an extract, transform and load (ETL) system has started and the LMI for All website has been redeveloped; these are also reported on in section four. Section three on data development provides information on the structure of the LMI for All database, datasets and indicators within the database, as well as an overview of the methodology used to provide data to the database. This section also includes a review of existing datasets that have changed and could now potentially be included in the LMI for All service. Section four details stakeholder engagement activities undertaken over the last year to disseminate the LMI for All service reporting on the various events and take-up of the service. Section five sets out future possible developments for the LMI for All service.

2. Technical infrastructure and developments

This section of the report provides an overview of the current IT infrastructure of LMI for All, a review of the current structure, a discussion on server capacity and usage, as well as review of the infrastructure. The latter part of the section details the development work that has been undertaken this reporting year, including the development of an extract, transform and load (ETL) system and the redevelopment of the LMI for All website. First, the objectives and outcomes for this phase of the project are detailed as these guided the work undertaken.

2.1. Objectives and outcomes

There are four key objectives guiding the provision and development of the technical infrastructure and support for the LMI for All service.

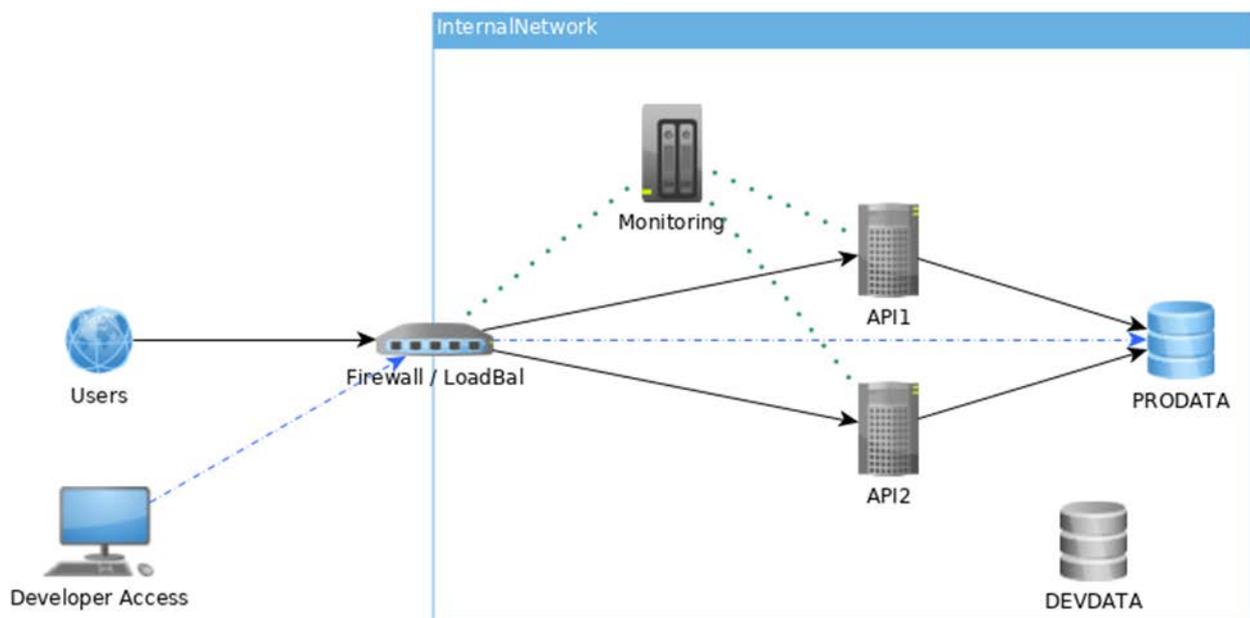
- *To maintain a secure and robust infrastructure for the data, including test and production servers and provision for backup services.* For this, the hardware and software environment has been maintained and monitored due to a robust and future-proof infrastructure. The infrastructure is also capable of being extended to respond to increased data provision and increasing numbers of queries.
- *To implement processes to ensure that the data made available to developers and end-users meets required quality standards.* Data are made available through the APIs, which have been programmed according to the rules of the data providers to suppress data that is either disclosive or unreliable due to sample sizes. Quality procedures are in place to check data before it is loaded to the server and before it is published.
- *To provide modern and flexible software tools to allow the querying of the database by external users.* Current tools in place to enable external users or developers to query the database have been reviewed and have been determined to provide the most appropriate technical solution at the current time.
- *To provide software tools and spaces for documenting the process and allowing public access to the LMI for All database.* The redevelopment of the website has addressed the need to support developers better in producing applications that utilise the database and API through targeted support materials, detailed and accessible information on the datasets in LMI for All, and new ways to contact the technical team.

2.2. LMI for All Infrastructure

Earlier reports (see annex A) detail the development of the LMI for All infrastructure and how it has changed to respond to new technologies and the increasing demands on the

service. The LMI for All infrastructure is currently comprised of UK based cloud based secure services; a data server (production environment); a development data server (test environment); a monitoring system; and two API servers (see figure 2.1, below). The two servers provide environments which: hold the actual data and process queries to serve data to the live API (PRODATA, production environment); and enable testing of new functionality and the integrity of new or updated data (DEVDATA, test environment). The cloud-based server and hosting system are continuously monitored to ensure it is operating as it should and running efficiently. Security updates are installed periodically, daily backups are made to ensure uninterrupted service and full backups to a remote backup server are undertaken incrementally.

Figure 2.1 LMI for All infrastructure



The API is used by developers to query the database and provides the only access for external users to the database. Rules (such as limitations of sample sizes) are in place to ensure that disclosive data are filtered for and restricted with API returns data based on a larger scale of disaggregation if data are disclosive. They rules are set in place by the data providers and are reviewed annual to ensure any changes are addressed. The system also provides warnings when data, that although non-disclosive, are based on small samples.

A load balancer is in place to optimise API response query time and active sessions by distributing queries among a set of identical API services. This multi-server redundant design addresses problems with a reliance on a single API server with no redundancy and a risk of downtime. It also provides an infrastructure that can be quickly and easily extended if API queries begin to rise above current capacity levels. The provision and deployment of new API servers has been automated.

An ETL system is in place to semi automate data uploads (namely ASHE Pay and ASHE Hours). This supports the refreshment of existing datasets, integration of new datasets and the maintenance of data quality.

To ensure data quality, internal checks are completed by the LMI for All data team before it is provided to the technical team (either through the ETL system or the University of Warwick secure data transfer service). The ETL system is able to identify and report any changes in the dataset, which has to be rectified before the dataset is accepted. Once in the database, checks of the original dataset against and the API data returns is undertaken to confirm that the dataset is consistent at all points before it is published.

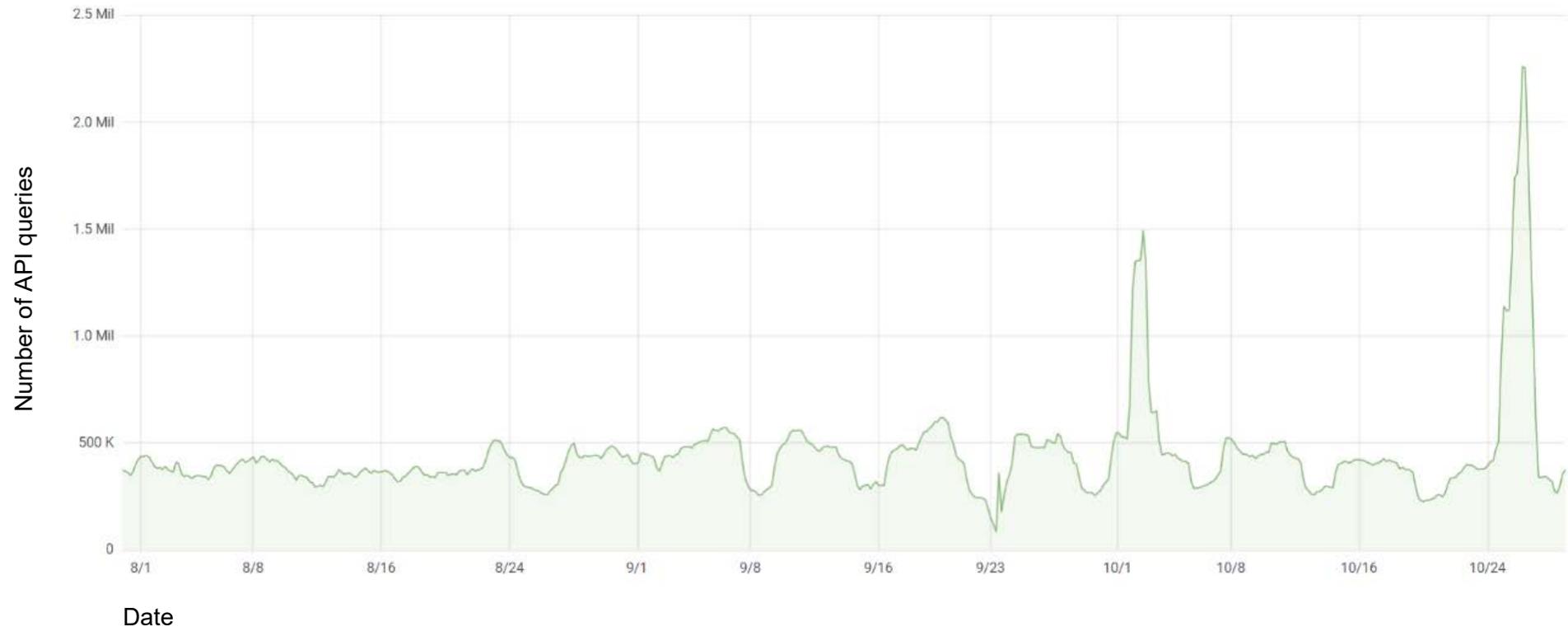
Query error (a message that is reported when there is a problem, request, or query to the database) and outage monitoring are also part of this infrastructure. The system reports on any errors, but production errors continue to be rare. The API monitoring dashboard provides details on the number and type of queries. A new reporting system has been added to the front page of the LMI for All website for users of LMI for All to report any issues; this is constantly monitored to ensure issues are quickly resolved. Technical support continues to be provided by the development team by email and telephone as part of the service along with a commitment to help third party users of the service to develop their own websites and applications.

2.3. Server capacity and usage

Present usage figures vary according to school hours and term-time, but there are on average 400,000-500,000 queries per day, which equates to 6-9 queries per second. This is managed by load balancing between two API servers and a somewhat overprovisioned database server, meaning the current infrastructure is more than sufficient for the current traffic and provides some capacity in case of unforeseen peaks. Figure 2.2, below, illustrates the number of queries to LMI for All over the long-term.

During October 2019, there were two such unforeseen peaks with 1.5 million requests and 2.2 million requests in one day each. Most of these requests were for SOC code lookups, which the API is specifically optimised to handle as quickly as possible and without querying the backend database, so this did not particularly tax the servers.

Figure 2.2 Long-traffic to LMI for All, July-October 2019



Source: LMI for All API Monitoring system, 28/07/19-28/10/19

2.4. Review of the infrastructure

The LMI for All infrastructure is well developed and it is able to manage current queries levels, small increases and short-term spikes in queries. The infrastructure is future-proof in that additional APIs can be implemented if required to handle an increase in the number of queries to the database. It is supported by various monitoring and error reporting services.

Whilst over the reporting year there has been a 100 percent uptime for the API itself, there were three instances of failure in the broader system; two of which affected some (but not all) API users, and one which was completely internal. The revised uptime is, therefore, estimated at about 99.7 percent. The two user-affecting failures were both related to SSL certificate expiration. During both of these times, which lasted about half a day, the API was unavailable to users connecting via SSL, and the monitoring service did not catch this (possibly because it only checks certificate presence, but not validity). Users connecting to the API over a regular, unencrypted connection were not affected and could use the API normally. Automated certificate expiration and renewal has been improved as part of scheduled maintenance and this problem should not occur again. The third failure was an unreported outage of the internal traffic monitoring system, which had no effect on API users, but led to the loss of about two weeks of daily traffic numbers for July 2019. In response, an upgraded monitoring system has been purchased and implemented, which provides immediate reporting of any failures to the technical team.

2.5. Extract, Transform and Load (ETL) development

As previously noted an ETL system was previously developed and tested for the ASHE dataset. As this development was a successful pilot, during this reporting year development work has been completed on an ETL system for additional datasets in LMI for All. The following details this development.

The ETL specification and development is based on an automated data updating process and the software to implement that process. This is based on fixed file formats and naming conventions for raw data files. The online data update interface allows administrators to:

- Upload raw data – Uploaded raw data files are semi-automatically picked up by structured query language (SQL⁵) Server Integration Services (SSIS) packages to be tested for integrity and start the import of data into the server Test environment.

⁵ SQL is a domain-specific language used in programming and designed for managing data held in a relational database management system.

- Review errors – If the raw data files are not in-line with the specifications or if they are truncated for any reason, the ETL process will reject them and report details of the rejection to the Administrator.
- Promote data to Production – After the data have been tested in the DEVDATA (test environment) environment to PRODATA (production environment) through the administration interface.

Part of the technical development process this reporting year has been to produce test suites for each dataset. The test suites are a python program, which can be run by the data team and replicates the ETL processes on the server. If the data passes the test suite then it will (in theory) be able to be automatically loaded into the test and main SQL servers and rapidly deployed through the API. Further desired options for the ETL process have been defined as the ability to:

- Transfer and/or move data between environments (i.e. test to production), backup in between, and enable full control by user;
- Monitor progress and produce review logs;
- Enable notifications (i.e. load completed, load error); and
- Allow data update and/or reload options with, again, end user control.

Most of this desired ETL functionality is now available. However, the system requires extension to include further datasets as well as the development of a front-end administration dashboard, which will allow administrators to upload raw data packages to start the automated ETL processes. The web interface will provide a live status report of the ETL process and prompts for user input when needed, as well as provide a tracking process for the administrator. This system will be tested during the next reporting year before it is fully deployed. It should be noted that it will not be possible to produce a fully automated system for all data uploads due to periodic changes in the structure of data provided by external agencies and organisations.

2.6. Redevelopment of the LMI for All website

Over the last reporting year (2018-2019), the content and structure of the LMI for All website has been reviewed and redesigned to meet the needs of users (those looking for LMI, those interested in LMI for All, and developers) and make it more user friendly. The website was developed during phase 2A (November 2012 – November 2013, see annex A) of the project, so the platform has now been updated together with a new design template. This new design template has been implemented across all dissemination materials and the LMI for All twitter account.

The LMI for All website is managed in a WordPress environment, which provides a full content management system. It is optimised for viewing on multiple platforms. Some significant developments undertaken this reporting year include:

- Implementation of new design and new Wordpress template;
- Redesign of the structure with distinct areas for different users of the website:
 - ‘Explore LMI data’ – those looking for LMI and resources;
 - ‘Designers’ – those looking to learn more about LMI for All and what is needed to develop their own website or application (usually careers organisations and service providers); and
 - ‘Developers’ – those (usually programmers and front-end developers) actively creating websites, applications or widgets using LMI for All needing further information on how to start or what data are available.
- Online user form for developers and end-users to complete in order to create a directory of those using LMI for All in their websites and applications. The aim is to create a better understanding of the range of users of LMI for All and potentially to co-author with end-users case studies that detail their development.
- Data pages which provide detail on each dataset (including data owner, provenance, potential use of data, currency and frequency of update, limitations of the dataset) and visualisations on how data have been used by others.

In the early stage of development is a range of online materials for careers practitioners, designed for self-learning focusing on understanding LMI and intelligence, key issues with regard to interpreting data, and descriptions of different types of labour market indicators. Materials specifically aimed at developers, such as how to mash LMI for All data with other available data, will also be produced. The online materials will be published in 2020 and reported in the next LMI for All annual report.

2.7. Summary and next steps

During this reporting period, the following tasks have been undertaken:

- A secure and robust database containing labour market data has been maintained and data have been refreshed and quality assured.
- A selection of modern and flexible software tools to allow querying of the database by external users, including an easily-accessible API has been monitored and maintained.
- A redeveloped project website has been published, which provides a range of new materials and new data pages to support those looking for LMI, designing their own sites and developers.
- The development of online materials focused on the LMI and its applications in career practice.
- A review and development of an ETL system to increase data upload efficiency and to automate some data quality checks.

Over the next reporting year, a second LMI for All widget will be developed. This is in response to the considerable take-up of the first widget, known as Careerometer, by those unable to develop their own interface. Feedback from users has been gathered through stakeholder events and through social media.

Technical work over the next year will also focus on using machine learning techniques to link education courses to occupations, and to pay (see section 3.5).

3. Data developments

This section provides details of the data and data refreshment undertaken for the LMI for All service during this reporting year. It first sets out the objectives and outcomes for this work stream which guided the work and data developments. An overview of the labour market data and indicators in LMI for All is presented first before more information on the datasets in LMI for All are detailed. The latter part of this section details some future potential developments for the service.

3.1. Objectives and outcomes

The data development work focusses on the management and development of the relevant datasets, detailed below, to provide the high quality data that underpins the database. As in previous phases of the project, the objectives of the data development strand have focused on the:

- *Refreshment of existing datasets held in LMI for All.* A timetable of data refreshment was developed in line with new datasets being published, which guided the work of the data team during the reporting year.
- *Management of documentation relating to the datasets.* All documentation relating to each dataset has been updated when data have been refreshed. This was used to develop new data pages on the LMI for All website.
- *Quality assurance and checking, including checks to ensure data rendered by the application programming interface (API⁶) are consistent with values contained in the dataset generated.* Quality procedures were implemented during the early stages of the project and are followed by both the data and technical teams. These procedures are reviewed periodically.

The data team have also provided advice on any issues and questions relating to the data as part of customer care support arrangements.

⁶ An Application Programming Interface (API) is a set of routines, communication protocols, and tools for building software applications. API queries produce formatted responses into a universal data format known as JSON (JavaScript Object Notation), before sent back to the developer/client.

3.2. Overview of the LMI for All database and organising framework

The organising framework for the LMI for All database is occupation classified at the unit group 4-digit UK SOC2010⁷ using the following dimensions and characteristics:

- 369 detailed occupational categories (SOC 2010 4-digit level);
- 75 detailed industries (roughly equivalent to SIC 2007 2 digit level);
- Employment status (full-time, and part-time employees and self-employment);
- Highest qualification held (9 levels of the National Qualification Framework (NQF)⁸);
- Countries and English regions within the UK; and
- Gender and age.

An index of around 28,000 job titles are mapped to SOC, which enables the end-user to search for and access data of interest and relevance to careers in an intuitive fashion.

The LMI for All service provides access to the following datasets:

- Actual and projected employment trends from 1990 to 2024 (from the *Working Futures* dataset);
- Projected number of workers needed by employers to replace those retiring between 2014 and 2024 ('replacement demand') (from the *Working Futures* dataset);
- Pay (for employees only) (from the Annual Survey of Hours and Earnings (ASHE) dataset);
- Hours worked (for employees only, not the self-employed) (from the ASHE dataset);
- Changes in pay by detailed 4-digit occupation (modelled on ASHE and Labour Force Survey (LFS) datasets);
- Unemployment rates (based on the LFS dataset);
- Vacancy data by occupation and industry (from the Employer Skills Survey (ESS) dataset);

⁷ For more information on the UK Standard Occupational Classification (SOC), see: <https://www.ons.gov.uk/methodology/classificationsandstandards/standardoccupationalclassificationsoc/soc2010>

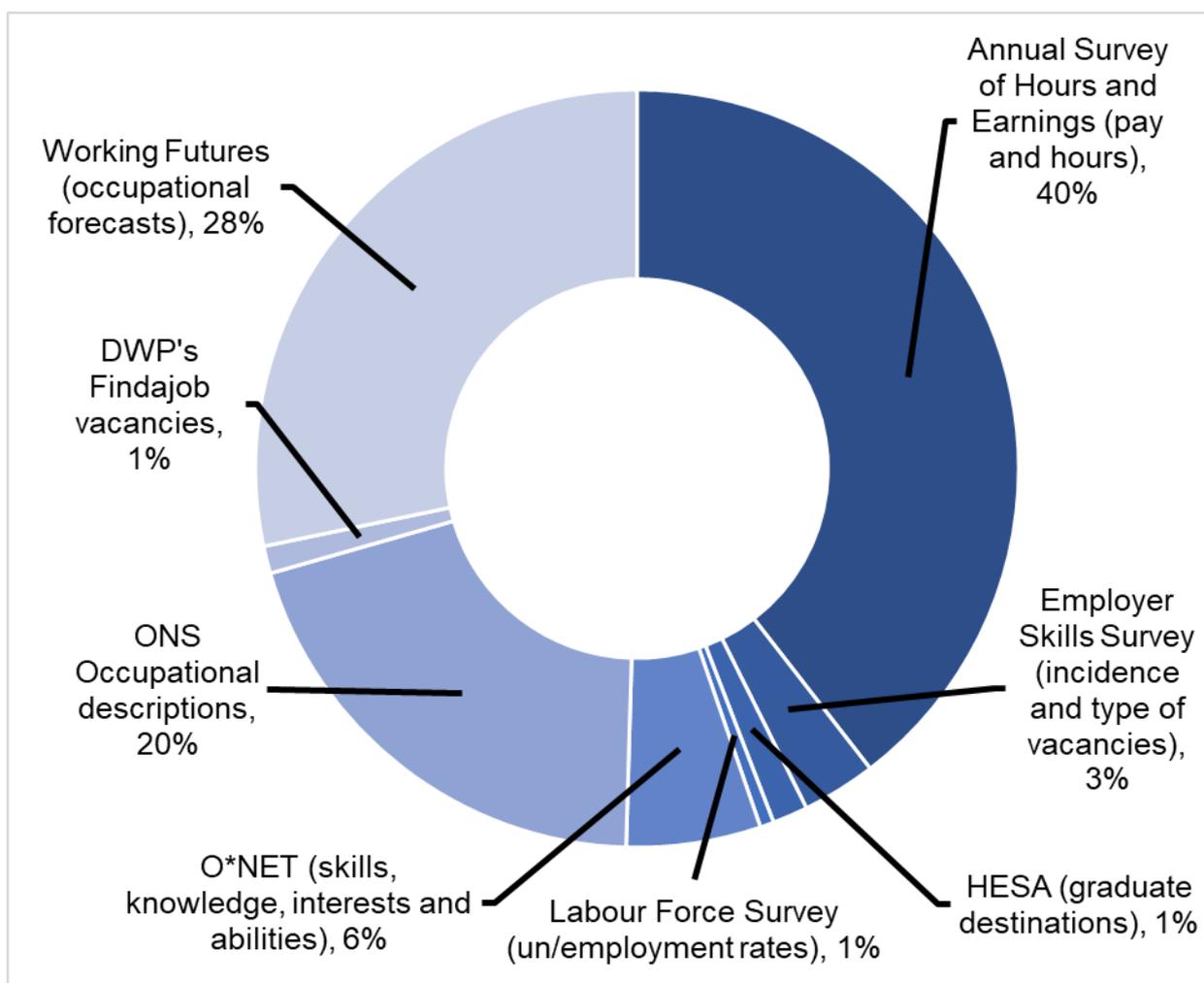
⁸ For more information on the NQF, see: <https://www.gov.uk/what-different-qualification-levels-mean/list-of-qualification-levels>

- Geographical patterns of employment and travel to work distances (from the 2011 Census of Population); and,
- Destinations of graduates (based on Higher Education Statistical Agency (HESA) data).

These data are enriched by linking to contextual information about occupations from two sources: first, occupational descriptions of 4-digit occupations (based on the Office for National Statistics (ONS) information); and second, the skills, abilities, knowledge and interests data mapped to 4-digit occupations based on US O*NET information and mapped to UK SOC.

Online vacancy information is provided by the Department for Work and Pension's Findajob API with query results returned from a fuzzy matching approach in the API.

Figure 3.1 Percentage of queries by dataset, July-October 2019



Source: LMI for All API Monitoring system, Based on 505,851 queries made between 28/07/19-28/10/19

Evidence on usage statistics from the LMI for All monitoring system (see section 2.2) have shown the ASHE dataset (providing data on pay and hours) and the *Working*

Futures dataset (providing employment data, past, current and projected) are the most accessed datasets. Figure 3.1, above, shows the percentage of queries for each dataset over a 3-month period.

Table 3.1, below, provides an overview of the data and indicators currently available in the database.

Table 3.1 LMI for All Data overview, September 2019

Dataset	Employment (Historical)	Employment (Projected)	Employment (Replacement Demand)	Pay and earnings	Hours	Unemployment Rates	Number of vacancies
Source	<i>Working Futures / Business Register and Employment Survey (BRES)/ Labour Force Survey (LFS)</i>	<i>Working Futures/ Business Register and Employment Survey (BRES)/ Labour Force Survey (LFS)</i>	<i>Working Futures/ Business Register and Employment Survey (BRES)/ Labour Force Survey (LFS)</i>	Annual Survey of Hours and Earnings (ASHE)/ Labour Force Survey (LFS)	Annual Survey of Hours and Earnings (ASHE)	Labour Force Survey (LFS)	Employer Skills Survey (ESS)
Indicator	Number of jobs (employee, self-employed)	Number of jobs (employee, self-employed)	Number of jobs openings between selected years (employee, self-employed)	Average full-time earnings; plus indicative estimates of medians and deciles	Average weekly hours	ILO Unemployment rate	Number of vacancies, Hard-to-fill vacancies, Skills shortage vacancies, Occupation
Dimensions*	Occupation, Industry, Qualification, Geography, Gender, Status	Occupation, Industry, Qualification, Geography, Gender, Status	Occupation, Industry, Qualification, Geography, Gender, Status	Occupation, Industry, Qualification, Geography, Gender, Status, Age	Occupation, Industry, Geography, Gender, Status	Occupation, Industry, Qualification, Geography, Gender, Status	Occupation, Industry, Geography
Period	2002-2018 #	2014-2024	2014-2024	2013, 2015, 2017, 2018 ##	2015, 2017, 2018 ###	2011-2018	2011, 2013, 2015, 2017
Available updates (if known)	every 2-3 years	every 2-3 years	every 2-3 years	annually	annually	annually	every 2-3 years

Dataset	Skills, Knowledge, Abilities, Interests	Occupational descriptions	Current vacancies	UK Census Population	Higher education destinations	Apprenticeships
Source	O*NET database (version 22.1)	ONS Standard Occupational Classifications	Findajob (DWP)	UK Census Population (England and Wales)	HESA	Apprenticeship Service
Indicator	Skills, Knowledge, Abilities, Interests	Structure and descriptions of occupations	(Available through fuzzy search, SOC mapping)	Information on geographical patterns of employment and travel to work distances	Destination of graduates immediately after graduation	Apprenticeship data and frameworks
Dimensions*	Occupation	Occupation	Occupation	Occupation (1,2,3 digit), Geography**	Occupation, Qualification, Qualification required for job, Subject of study	Occupation
Period	2018	2010 (new and revised job titles amended in 2019)	current	2011	2012/13-2016/17	(currently being tested)
Available updates (if known)	every 2-3 years	only required when SOC is updated	constant	every 10 years	annually	constant

Notes:

* Occupation (SOC2010 4-digit), Industry (SIC2007, 75 industries), Qualification (NQF 0-8), Geography (UK countries and English regions), Gender, Status (full-time or part-time employee and self-employed).

** Geography available for Output Areas, Lower and Middle Super Output Areas and the hierarchy of local government areas from wards to regions and nations

For 2000-2013 data are only available at the SOC2010 2-digit level

Estimates of Pay are also available for 2013 and 2014. However, comparisons of detailed estimates between years will not provide robust information at the 4 digit level. A more limited "change indicator" is available for those interested in changes over time.

3.3. Data refreshed

The following datasets have been refreshed during the reporting year:

- Pay and earnings (estimates based on ASHE and LFS);
- Hours (ASHE);
- Unemployment rates (LFS);
- Occupational descriptions (Office for National Statistics); and
- Skills, knowledge, abilities and interests (O*NET).

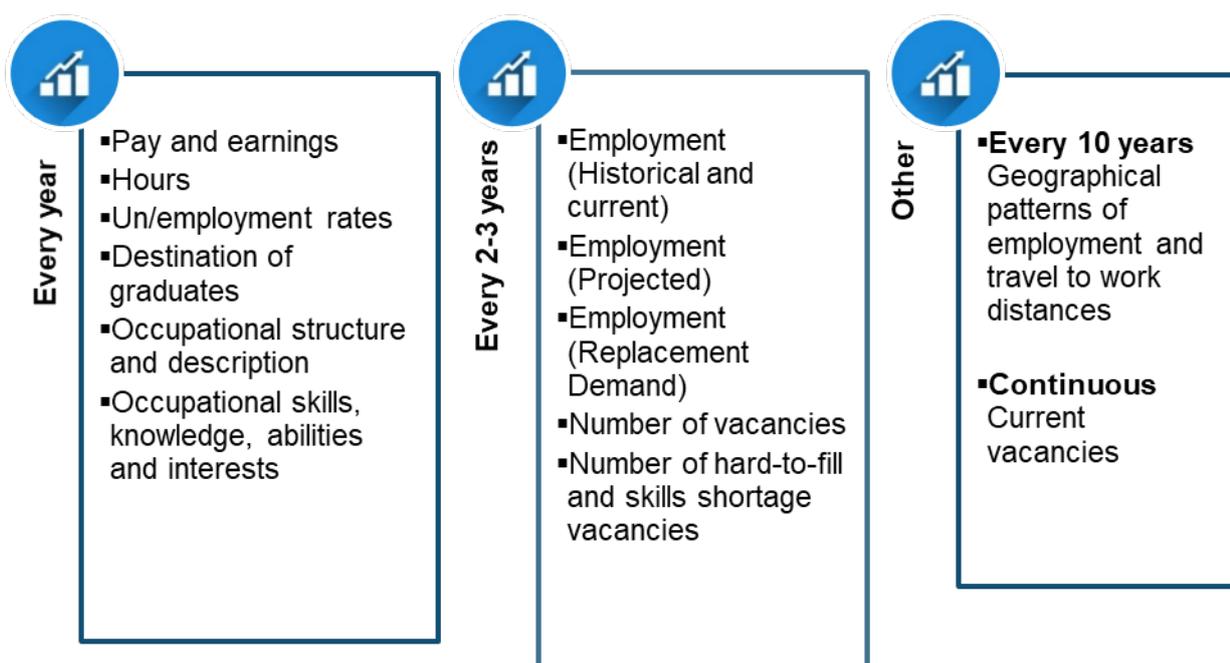
The following datasets have not been refreshed:

- **New employment and forecast data** (2017-2027) has been produced as part of the DfE *Working Futures* project. A dataset has been produced specifically for LMI for All, which has employment data at 4-digit level. This new employment data (historical, projections and replacement demand) will be available and integrated into the service in 2020 following publication of the new *Working Futures* reports.
- In previous years, data from the Higher Education Statistics Agency (HESA) on **higher education destinations** has been obtained, processed and added to the database. However, changes to the HESA survey means that new data will be integrated into the service in 2020. As assessment of the new HESA data that could potentially be available in future is detailed later in this section.
- The **number of vacancies, hard-to-fill and skills shortage vacancies** provided by the Employer Skills Survey has not been updated this year. New data will be available in 2020/2021 as expected.

Figure 3.2, below, details the refreshment cycle for all data in LMI for All.

Apprenticeship vacancy data were provided through the LMI for All service using the Apprenticeship service API. Previously the apprenticeship frameworks included information on sector and a number of job roles within one or more applicable levels including information on qualifications needed. A mapping to SOC was created in order for apprenticeship vacancies to be made available through LMI for All, but this was withdrawn due to the development of Apprenticeship Standards. These Standards have been gradually introduced since 2017 and frameworks withdrawn. During the last year, a new mapping has been produced for the Apprenticeship Standards. Each apprenticeship has been mapped to multiple SOC2010 codes where appropriate, as well as SOC2020 as this will be required in the future. This mapping is being updated to reflect very recent changes to the Standards before being implemented in LMI for All.

Figure 3.2 Overview of data refreshment cycle⁹



3.4. Methodology and approach to providing data

The LMI for All service has successfully achieved the provision of detailed data and is proven to be a useful tool in the development of applications and widgets aimed at those making career decisions. This includes the following aspects: occupation; sector; English regions and constituent countries of the UK; gender and age; employment status; and qualification level. The 369 SOC 4 digit occupational categories lie at the heart of the database and information at this level of detail is provided where possible. The core data provided comprises detailed information, as described above, for:

- Employment (time series of historical and projected levels, plus (for the future only) projected replacement needs (RDs));
- Pay (for a recent year (currently 2018, for employees only));
- Hours (for a recent year (currently 2018, for employees only)).

In addition, less detailed information is provided for a number of additional datasets.

Many of the official statistics are collected under the terms of legal instruments, which ensure confidentiality for those providing the data. These guarantee that these data are not published in a way which would disclose commercially sensitive or other confidential information about the companies or individuals concerned. The Office for National

⁹ Timescales based on previous update cycles.

Statistics (ONS), which is responsible for collecting and publishing the information, has guidelines in place to ensure that this is the case. This limits the level of detail that can be placed into the public domain. It should also be noted that whilst key data owners do not currently have APIs in place, data providers are considering the development of API to enable access to data.

The other important consideration is statistical reliability and the sample size on which the statistics are based. Many of the official sources are based on samples, which while large in statistical terms, are not large enough to provide robust information at a very detailed level. This applies to both the Business Register and Employment Survey (BRES), which is the main source of information on employment by industry in LMI for All, and the LFS, which is the main source of information on the structure of employment by occupation, qualification and employment status. Reliance on the raw survey data would, therefore, limit the level of detail that could be provided.

This issue has been addressed previously in the context of developing the *Working Futures* employment database. The solution adopted there was to combine the various official sources and to create estimates of employment at a more detailed level than it is possible to obtain from the official surveys alone. This has been combined with putting in place checks to ensure that the data generated are robust (in a general statistical sense) and that they do not breach confidentiality nor are disclosive. Following detailed discussions with ONS, it was concluded that: the aggregation of information on employment by industry to some 75 industries could avoid problems of disclosure¹⁰; and as long as sources such as the LFS and the ASHE were used to produce estimates for general groups rather than revealing information on individual cases, then this should not breach confidentiality. Data are, therefore, produced using these rules to ensure the integrity of data.

3.5. Potential future developments

New APIs, new datasets and revisions to existing datasets have been assessed during this reporting period. This sub-section includes conclusions of those assessments and details of potential future developments to enhance the LMI for All service.

3.5.1. O*NET API

O*NET data, mapped to UK SOC, is made available through the LMI for All database providing contextual data on skills, knowledge, abilities and interests for each occupation. This dataset is a powerful resource in LMI for All as enables an alternative pathway through the data that is not based on occupation. During the last year, it was noted that

¹⁰ Without the necessity for a Chancellor of the Exchequer's Notice (CEN).

O*NET is now providing real-time access to the raw information in the downloadable O*NET database through an API. This potentially allows the LMI for All service to use the O*NET Database Services as an alternative to downloading and regularly updating a local copy of the O*NET database. The O*NET services would allow LMI for All to:

- Programmatically discover files in the database, and what data each file contains;
- Retrieve all data from any part of the database; and
- Filter and sort results based on multiple criteria.

An assessment of the functionality of the new database services has been undertaken. Firstly, while usage restrictions apply a daily limit of 50,000 requests, it is possible to integrate these services into the LMI for All API since O*NET query numbers typically do not cross 10,000 requests per day. Secondly, the new O*NET services offer a great wealth of information and data - more than currently provided through LMI for All. However, some information that is in the original database files is omitted, most notably the numbers for weighting. As such, it is not recommended that the LMI for All replaces the existing O*NET API endpoints (which provide weighting queries) with the new O*NET service functionality. This service will be reviewed as it develops, as it may offer options for new queries around O*NET in the future.

3.5.2. Real time vacancies

With the increasing use of machine learning and web technologies, there is increased interest in the potential of scraping job vacancies advertised in a range of different agencies including official public employment services.

Whilst vacancies are available through LMI for All using the Department for Work and Pension's (DWP) Findajob API, these data do not show the complete picture of vacancies online, as there are a number of other (private) vacancy websites in operation across the UK. The vacancies provided through the Findajob service are not matched to UK 4 digit SOC, so an automated matching process is used by LMI for All resulting in some inconsistencies and gaps in the data.

An aspiration has also been to incorporate real-time vacancy data coded to SOC2010 4-digit occupation into LMI for All; it is data most sought by users. Burning Glass already provide vacancy data analysis to a number of UK public authorities. Burning Glass, like CEDEFOP (European Centre for the Development of Vocational Training) use an AI approach based on machine learning to aggregate and analyse the data. CEDEFOP have been developing a major European project based on this approach, called Skills-

OVATE (Skills Online Vacancy Analysis Tool for Europe¹¹), which offers insights into skills and jobs requested by employers and pre-defined dashboard visualisations for 18 countries, including the United Kingdom. Data are based on online job vacancies gathered from a defined list of online sources, including private job portals, public employment service portals, recruitment agencies, online newspapers, employers' portals etc. Data were gathered between 1 July 2018 and 31 March 2019. The tool provides considerable data about job vacancies and skills, classified by ISCO 2 level (International Standard Classification of Occupations¹²) and ESCO (European Skills, Competences, Qualifications and Occupations¹³), including analysis for more than 400 occupations and 1,500 skills. There is presently no API to the data. However, CEDEFOP's tool demonstrates the possibility of providing real time LMI on vacancies, which are classified by occupation, geographical location, qualification and skill. Over the next year, consideration will be given to whether data scraping technology¹⁴ has been sufficiently developed to be used within the LMI for All service.

3.5.3. Career pathways data

A long standing aim for LMI for All has been to provide data allowing end users to connect education courses to potential employment. This could be used both by students choosing school subjects and those selecting future higher and vocational education courses. Such data are sometimes referred to as career pathways. Career pathways could include a range of different activities and outcomes including linking careers and occupations to for example:

- School subjects – A levels, T Levels and vocational course;
- Apprenticeships and vocational courses;
- Degree and post initial degree studies;
- Work experience;
- Internships.

Obviously, there is no one-to-one accordance between school subjects and further education and training or university subjects; neither is there a link between university or apprenticeship programmes with future occupations. In most cases, it will be a many to many mapping. Ideally, three datasets would be needed to create career pathways data, including:

¹¹ For more information on Skills-OVATE, see: <https://www.cedefop.europa.eu/en/data-visualisations/skills-online-vacancies>

¹² For the ISCO classification, see: <https://www.ilo.org/public/english/bureau/stat/isco/isco08/>

¹³ For more information on ESCO, see: <https://ec.europa.eu/esco/portal/home>

¹⁴ Data scraping is a system where a technology extracts data from a particular codebase or program. The scraping provides results for a variety of uses and automates aspects of data aggregation.

- School based subjects linked to progression routes to apprenticeship or higher education subjects;
- Apprenticeships and university qualifications linked to occupations;
- Occupations linked to industrial sectors.

The major barrier to doing this is not only that data are from different sources, but different classification systems are used, for example:

- Occupations (which form the basis for LMI for All data) are based on SOC2010 (due to move to SOC2020 in the future);
- Apprenticeships are classified under the Apprenticeship Framework;
- University courses have been classified under JACS (Joint Academic Coding System¹⁵), but in 2019/20 academic year this is moving to the new HECoS (Higher Education Classification of Subjects¹⁶) system;
- Industrial sectors are classified under SIC2010.

Hence, there is a need for data that can link these different classification systems. It should be noted that for the LFS, which underpins some of the LMI for All datasets, the only data collected is on the level of highest qualification and does not specify the qualification.

LMI for All already provides data from the HESA Student Destinations survey providing jobs classified under SOC2010 against university subjects and courses although this will need to be updated to reflect the new HECoS classifications. This is based on aggregated data from the higher education sector. HESA are providing a look up table between JACS and HECoS. LMI for All has also developed a table classifying the Apprenticeship Framework against SOC2010 (and SOC2020) although this is yet to be implemented within the API (see section 3.4).

The development of career pathways data is discussed further in the next two sub-sections, which review and assess two new datasets for potential use in LMI for All.

3.5.3.1. Graduate destinations data

Until the last year, the LMI for All service has provided graduate destinations data from HESA. Data were collected in the HESA graduate destination survey, which contained SOC classification. This allowed mapping from courses studied to job destination. Whilst much of this information was only made available subject to a fee, following consultation and negotiation with the data owners (led then by the UKCES), detailed data were made

¹⁵ For information on JACS, see: <https://www.hesa.ac.uk/support/documentation/jacs>

¹⁶ For details of HECoS from HESA, see: <https://www.hesa.ac.uk/innovation/hecos>

available for use in LMI for All. This updating required a data request to be made to HESA via the DfE.

HESA has introduced a new graduate destinations survey, called Graduate Outcomes¹⁷, to replace the Destinations of Leavers from Higher Education (DLHE) survey; data from which was made available through the LMI for All service. Students who graduated in 2017 were the last DLHE cohort, whilst those who graduated in 2018 will be the first Graduate Outcomes cohort. Due to the change in the HESA destinations survey and the timing of collection, it is not possible to get the data this year.

The main indicators that LMI for All currently uses from the DLHE are listed together with information on their availability in the Graduate Outcomes in Annex B. The main data-related differences¹⁸ and innovations in Graduate Outcomes compared to DLHE, relevant to LMI for All, include:

- Shifting from asking graduates about destinations at six months after graduation to 15 months after graduation. This will be of benefit to LMI for All, because it will give graduates more of a chance to move to appropriate-level jobs (for instance, according to the DLHE, around 10 percent of graduates were working as retail, catering, waiting and bar staff at six months after graduation, and this proportion has been relatively steady over the last few years). It is expected that fewer graduates will be employed in such jobs 15 months after graduation.

However, the response rate to Graduate Outcomes may be lower than that to DLHE, owing to a longer gap between graduating and the survey, and the higher probability that graduates may have moved away or changed their contact details in that time.

- New questions about duration of employment with employer, number of jobs since graduating, responsibilities for supervising staff in their role (the inclusion of supervision questions suggest that more could be done on social mobility to derive the NS-SEC of the graduates' jobs¹⁹).
- Improved question wording for capturing information on graduates who are working freelance or in self-employment, which makes the survey more relevant in the face of the increasing importance of the gig economy, digital nomadism, etc.

¹⁷ For more information on the change to the Graduate Outcome Survey, see: <https://www.hesa.ac.uk/innovation/outcomes>; and for information on the new survey questions, see: <https://www.hesa.ac.uk/innovation/outcomes/survey>

¹⁸ For a summary of other differences from DLHE, see: <https://luminare.prospects.ac.uk/graduate-outcomes-the-changes-and-challenges>

¹⁹ For more discussion on this, see: <https://www.ons.gov.uk/methodology/classificationsandstandards/otherclassifications/thenationalstatisticsocioeconomicclassificationnssecbasedonsoc2010#the-data-you-need>

- Using linked data for more accuracy, consistency and parsimony and to reduce the burden on graduate respondents. Linked data will be used for salary information from Longitudinal Education Outcomes (LEO) data and for further study from student record data, and will be made available alongside survey responses in the future.

The first Graduate Outcomes survey took place at four different cohort stages over 2018-19. Data for the first cohort has already been collected and analysed, but based on current information this has not been released²⁰.

Current information, therefore, suggests that there is a good possibility of adding new indicators from the new survey to LMI for All to enhance the data. For instance, subject to discussion, new questions on supervisory responsibilities and the questions on freelance and self-employment work to enhance understanding of these areas can be added. It is envisaged that the LMI for All service will be able access the new Graduate Outcomes data in Spring 2020, with salary information from LEO available later.

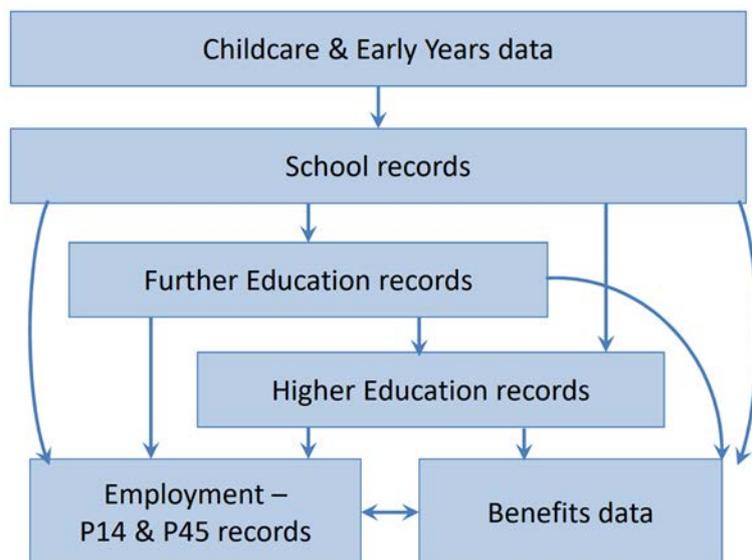
3.5.3.2. Longitudinal Education Outcomes (LEO) data

The Longitudinal Education Outcomes (LEO) study is a cross-government data linking project providing information about students' transitions from school, further education and higher education to employment. It contains data linked from different sources (see figure 3.2, below) showing UK students' levels of education and their levels of employment and earnings in later life. LEO also links together school, further education and higher education data to show how people progress through education, and how different courses can lead to different types of progression in education. The data would extend what is possible using LMI for All by enabling career pathways to be explored. For instance, the integration of LEO data could enable from education course to qualification to occupation to pay to be tracked.

LEO also currently contains information on earnings from employment and self-employment, benefits claimed, as well as detailed information about students, the courses they study, and the institutions they study at. This makes it a more comprehensive longitudinal source of information than national surveys. For example, the LFS does not contain self-employment earnings, and does not have a longitudinal component, although data on subject studied is available through a secure LFS data release. One of the main advantages of LEO is the near comprehensive coverage of those in employment.

²⁰ See: <https://www.hesa.ac.uk/innovation/outcomes/about/progress>

Figure 3.3 Administrative data sources for LEO data



Source: Judd (2018)

In summary, there has been increasingly more information added to LEO since it was first launched (e.g. adding earnings data for those in self-employment and the region of employment). Although LEO does not at present contain information about occupations, needed in LMI for All, it is possible that these data could become available in the future. At present, a mapping could be created from subject of study to SOC through using another source of data and applying the mapping to LEO. There is work underway to incorporate SIC into LEO, which will also be useful for LMI for All in the future.

The review of LEO data has highlighted the possible applications of LEO data in those careers websites and applications using LMI for All data. LEO lends itself well to being incorporated into careers websites and apps alongside other data. However, an important limitation of using LEO for informing prospective students' decisions is that the earnings data are retrospective, and does not distinguish between full and part-time earnings. It may be possible, however, to use LEO earnings data in combination with the *Working Futures* projections to estimate future earnings (and perhaps to compare these with LEO retrospective earnings), bearing in mind that prospective students have at least a four-year delay between applying to higher education and joining the labour market. This consideration is especially important for jobs in sectors at risk of automation that could see substantial change in the short to medium term. Ultimately, while the LEO dataset contains accurate information on graduate earnings, and is increasingly augmented to include more contextual data, it only captures one part of returns to higher education and experience of work in the labour market. Non-wage aspects, such as job quality, or skills required for a job, are not included. LEO should, therefore, be used in

combination with information from a variety of sources, but it is not possible to integrate into LMI for All at the current time.

3.5.4. Linking education courses to occupations using LEO

As part of the LMI for All data and technical development work, methods to link education courses to occupations and developing machine learning to link course data and SOC have been explored.

Interestingly, LEO contains information on education courses (see Annex C), but does not have information on SOC. The main benefit to using LEO for LMI for All would be to improve the information available about graduate destinations, in particular, more comprehensive earnings data over a longer time period. At present, such information about graduate earnings is not available in any other data source. An additional benefit is that, because the data are collected from administrative records, there is minimal scope for self-reporting error. Self-assessment data for those who are self-employed is self-reported, however.

As LEO does not contain SOC data at present, this limits the current use of LEO data for LMI for All. However, it is likely that, as LEO use becomes more widespread, information related to industry and/or occupation could be made available (this is unknown at present). If it were the case that occupational data at the 4-digit level were made available, this would potentially make LEO data a valuable addition to the LMI for All data sources.

For example, in the December 2017 release, LEO included information about earnings from self-employment, obtained through the self-assessment form. In the March 2018 release, LEO added information about graduates who got their degrees from a further education college, free school meal eligibility, and POLAR3 quintiles.²¹ A consultation document about the LEO before the first set of official statistics was released mentioned that the DfE were working with the DWP to look into making SIC and SOC data available.²² HESA is currently looking into whether it would be possible to include industry (SIC) data in the LEO data.²³

²¹ See for example:

https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/690859/SFR15_2018_Main_text.pdf

²² See for example:

https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/618628/LEO_survey_response.pdf

²³ See discussion at:

https://www.hesa.ac.uk/files/Graduate_Outcomes_London_Conference_complete_slidedeck_0.pdf

Alternative ways of linking education courses to occupations include using HESA Graduate Outcomes data that will link in earnings from LEO in mid-2020.²⁴ Therefore, while the main LEO dataset may not contain occupational information in the near future, it may be possible to make use of LEO's administrative information on earnings indirectly, via the HESA higher education outcomes data. This would improve the quality of information about earnings available up until now in the HESA data, as noted in section 3.6.3.

It may also be possible to approach HESA to look into developing a LEO to SOC mapping using Graduate Outcomes, and then applying that to LEO data as a proxy. At best, it would be a weak proxy, as HESA graduate outcomes collect data at 15 months since graduation, while LEO data are available for 1, 3, 5 and 10 years after graduation. This means that the SOC at 15 months since graduation available in HESA Graduate Outcomes may not be the same SOC at the time of LEO data collection.

Another possible way of calculating a proxy is using Secure LFS data to look at the proportion of people from particular degree subjects working in particular 4-digit occupational groups. The LFS data could allow one to proxy for years since graduation by using age or date left full-time higher education, which could allow better alignment with the corresponding years since graduation for LEO earnings. It should be noted that LEO currently includes HECoS subject classifications. While there is no link between HECoS and SOC, it may be possible to develop such a link using HESA Graduate Outcomes data, or, alternatively, to map HECoS to SOC indirectly via JACS (bi-directional HECoS-JACS mappings are currently available on the HESA website).²⁵

The resources required to generate a mapping of courses to SOC would therefore require collaboration with HESA, DfE or ONS. It may also be the case that some data (e.g. course studied, 4-digit occupation) could only be accessed through ONS Secure Data Service channels. In this case, a researcher would need to have or to obtain ONS Accreditation to access secure data. Some secure data could be made available remotely via Secure Pod access. Whilst this would support the use of the higher education data in LMI for All, this would not enable the use of school and further education data.

3.6. Summary and next steps

New datasets using and linking administrative data are being developed, so these will need to be reviewed as could provide valuable data for LMI for All and address gaps in terms of course data and education pathways.

²⁴ See: <https://www.hesa.ac.uk/innovation/outcomes/providers/responsibilities/analysis>

²⁵ For more information see: <https://www.hesa.ac.uk/innovation/hecos>

Furthermore, with constant developments in machine learning and data scraping techniques, the possibility of creating new datasets, particularly around vacancies, could be undertaken in the next few years. It is also worth noting that a number of organisations are becoming involved in open data and new taxonomies are being created²⁶, so there is a need for cooperation to share learning.

Over the next year, the LMI for All service will need to start preparing for changes from SOC2010 to SOC2020. Current information on SOC2020 suggests that major groups will remain unchanged, the minor groups and unit groups will change significantly. Changes are also noted in occupations that now require higher level of entry qualification, so the Professional and Associate profession groups have expanded. Unit groups have also been divided (or disaggregated) in recognition that occupations in the group have expanded as in, for examples, IT occupations. These changes will significantly impact on the LMI for All service with work required on the technical infrastructure of the database and the API in preparation for changes in the dataset. It will be a priority to maintain continuity of service and ensure any changes are communicated. This will be covered in the next LMI for All annual report.

²⁶ See for example the work of NESTA – the skills taxonomy <https://data-viz.nesta.org.uk/skills-taxonomy/index.html>; and open jobs project <https://www.nesta.org.uk/project/open-jobs/>.

4. Stakeholder engagement

The overall aim of the project continues to focus on the creation of a data service for developers to create products that support individuals in making better decisions about learning and work. Consequently, the stakeholder engagement strand of the project has continued to concentrate on awareness raising of the potential of the resource amongst a range of relevant stakeholder groups and audiences. The eight objectives for this phase of the project, therefore, mirrored and slightly expanded previous phases of the project, informing and shaping activities.

4.1. Objectives and outcomes

For this phase of work, there have been varied stakeholder activities guided by a stakeholder plan developed in consultation with DfE at the start of the project. Whilst many of the stakeholder activities may address several objectives, these and the outcomes are summarised next:

- *To facilitate and organise events and workshops to promote LMI for All to key stakeholders.* For this, the team has organised a number of targeted workshops, webinars, training events and presentations to inform and update stakeholders about LMI for All.
- *To participate in third party events and for a, which target relevant audiences.* The team have contributed to various events to reach a range of stakeholders, such as open data specialists, developers, policy makers, career practitioners and researchers. The aim has been to promote the LMI for All service, its current third party users, and Careerometer.
- *To pursue bi-lateral engagement with priority stakeholders to support them in the development of applications that use LMI for All.* Designers and developers of websites, applications and widgets using LMI for All have been engaged in informal conversations and provided support where requested.
- *To facilitate user fora and/or KIT meetings involving organisations that draw on LMI for All to gather feedback on support needs and potential improvements to the service and gain insights into who end-users are, how they use the data and the benefits in terms of careers decision-making.* This has been achieved through social media activities, blogs, webinars and meetings with developers.
- *To support DfE in understanding what more government can do to support the development and use of high quality careers apps and websites, including increasing awareness amongst schools and colleges of tools that exist.* The team supported a DfE run workshop, which brought together developers, stakeholders in the careers field and careers practitioners to discuss the LMI for All service.

Feedback from this workshop had informed the redevelopment of the LMI for All website and the second widget, which is currently being developed.

- *To undertake routine evaluation of activities in a manner appropriate to each particular activity and target group.* Where the team has organised events, evaluations have been undertaken and positive feedback received.

These objectives were achieved through integrated activities of work, which span this reporting year until the end of the project. Activities have also included the enhancement of LMI for All website (see section 2.6), the publication of blogs on the website and social media activities.

Like the previous phase, a consistent end-user focus was retained in these interrelated work clusters. The end-users comprising the third parties that have continued to be the data and service users, together with potential clients and customers of those end-users, for example. Perhaps unsurprisingly, the main geographical target for these stakeholder activities continued to be the UK, but the methodology underpinning the initiative has continued to attract international interest. This has been evident from attendance at conference presentations, together with invited workshops in countries like Finland and New Zealand, where there is a great deal of interest in the process of development, the technological challenges and their solutions.

4.2. Stakeholder plan

Shortly after inception, the first activity for this strand was the development of the stakeholder plan, in close consultation with DfE. Objectives for the stakeholder plan were agreed to be:

- To raise awareness and understanding of the LMI for All service;
- To ensure third parties develop applications or enhance existing applications in order to increase the reach of LMI for All among end-users;
- To ensure the implementation and use of Careerometer;
- To drive traffic to existing applications of LMI for All; and
- To foster positive and supportive attitudes to LMI for All service among target stakeholder groups.

In this plan was contained the broad forward schedule for activities and key stakeholder audiences. Target stakeholders were agreed to be: schools; career hubs; further and higher education providers; careers professionals and careers organisations; web and app developers; information and data service providers; third party users (defined as those using LMI for All in their online service); DfE, local authorities and other government departments; and parents and carers.

In addition, a protocol was agreed with DfE for vetting and approving specific activities that had been identified by the project team as potentially suitable. So, for example, whilst presentations at national and international conferences was agreed broadly as a suitable stakeholder activity, whenever a specific conference was identified, permission was sought from DfE, providing details of the conference, audience, costs, location, etc. Added value to the project was demonstrated by some events where invitations to present had meant costs were covered by the hosts.

For each event, audience/participant numbers were systematically recorded, subsequent to the activity, together with the exact nature of the activity carried out and any follow-up that had arisen. A short report was also presented by the relevant project team member at the monthly management team meeting with DfE that took place after the event.

4.3. Stakeholder activities and dissemination of the LMI for All service

During the reporting year, LMI for All was promoted at 17 dissemination events in the UK and internationally, ranging from national and international conference, workshops and presentations to local and regional events. In addition, targeted events and meetings were undertaken with over 30 individuals, developers and organisational representatives. These were follow-ups from activities as well as specific queries about the service. The number of stakeholders reached is an estimated 855. The five international events in which LMI for All was presented illustrated and confirmed the innovative nature of the service for careers provision. Additionally, the abstract for one of the international conference presentations was published in the Conference Proceedings, which was circulated to all members of the host professional association (estimated to be 19,000, worldwide).

It is worth noting that for the conference and workshop presentations, the approach adopted, which goes beyond presenting a factual account of LMI for All development, to examining the impact of this type of ICT development on the professional identity and practice of career and employment practitioners. This approach has been very well received by mixed audiences of practitioners, managers, trainers, policy makers and researchers.

Detailed evaluation of stakeholder activities is not always possible at all events, as some events are, organised by a host who undertake their own evaluation. Small-scale events are more open to evaluations. However, project team members have tended to assess the value of stakeholder activities based on interest expressed in terms of questions and requests for follow-up information during the activity.

In terms of what worked well for stakeholder engagement during the current project phase, ongoing dialogue is noteworthy. Details on specific activities follow.

4.3.1. DfE Workshop

In Spring 2019, the DfE organised and hosted an LMI workshop for sixteen experts in the field of careers and developers of digital careers tools²⁷. The aims of the workshop were to:

- Inform and update delegates about the LMI for All service;
- Provide a forum for careers experts and developers to share their views and experience;
- Encourage the development of user-friendly careers tools, which can be used in a tailored way with people who are less confident using digital tools; and
- Explore how the LMI for All service and DfE could better support the needs of current and prospective developers.

The workshop was supported by the team who presented on LMI for All focusing on the datasets and the types of careers questions LMI for All data could provide answers to. The workshop also included a presentation from a careers practitioner who spoke about how careers LMI is being accessed by young people and identified how services could better support their needs. Workshop discussions were focused on how LMI is currently used and how workshop participants could start, or continue, to use LMI for All.

Workshop participants provided useful feedback about how LMI for All and DfE can better support them to develop and use high-quality careers information tools. This feedback was used to create an action plan, which informed the redevelopment of the LMI for All website, the design of a new widget and future stakeholder engagement activities. The feedback given at the workshop has been used to inform the re-design of the new LMI for All website to make it as accessible and user-friendly as possible.

4.3.2. CDI workshops and masterclasses

An effective way to reach careers practitioners across a range of sectors has been collaborative events with the Career Development Institute (CDI). The CDI have taken the lead in organising events and the team have developed materials and delivered the events. Two workshops have been delivered by the team (the first in the previous phase of the project and the second in October 2018). These events were considered successful, so two further workshops have been agreed and scheduled to be located in London and as part of the CDI Annual Conference to be held in December 2019 in Gateshead.

²⁷ This has also been reported in a blog by Amy Hams on the LMI for All website, see: <https://www.lmiforall.org.uk/2019/07/department-for-education-lmi-workshop/>.

The workshop in October was attended by 30 careers practitioners, leaders and managers. At the start of the workshop, the majority of participants reported that one of the barriers to integrating LMI into practice was the difficulty in identifying trustworthy sources and that LMI is too complex to interpret. Of the 19 evaluations received from the participants of the CDI workshop; 100 percent rated the training event as 'good' or 'very good'; and 100 percent 'strongly agreed' or 'agreed' that, as a result of the training, they felt more confident in being able to use LMI more effectively in their practice. Positive comments about the training were the interactivity and the introduction to third party users of LMI for All. This suggests that these types of events are effective as supporting careers practitioners with their understanding and use of LMI in practice, and that introducing third party users of LMI for All give practitioners confidence in what sources to use.

4.3.3. Webinars

Three webinars²⁸ were organised during this period of reporting for the project. The webinars introduced LMI for All providing a brief overview of its development, overview of the data and how LMI for All could be accessed and used. The final part of the webinar was left open for attendees to ask questions. Webinars were organised for 30 minutes with a formal presentation and questions; they have run for about 45 minutes enabling further discussion.

The webinars have been targeted at key audiences such as Local Enterprise Partnerships, software developers, schools and colleges, careers professionals and other stakeholder audiences. Webinars have attracted both UK and international audiences. The webinars have aligned with other dissemination activities with the aim of continuing to raise the profile of the LMI for All service and driving traffic to third party applications. The LMI for All website, Twitter, the stakeholder list and our networks were used to disseminate these webinars. The first two webinars were hosted by the team and were attended by a total of 33 participants. A further webinar was delivered to the Career Industry Council of Australia (CICA) as part of the launch of new professional standards, with 31 participants. Looking forwards to the next year of the project, four further webinars will be organised; one per quarter starting Winter 2019.

4.4. Online activity

The team manages the LMI for All twitter account. Over the reporting year, there has been between 4-8 tweets per month. Tweets have focused on LMI for All developments, LMI for All news and blogs, and the promotion of dissemination activities and third party users' Twitter feeds. The Twitter feed has been embedded on the front page of the LMI

for All website. Examining the Twitter account analytics, the following summarises activity of the last year:

- 48,334 Tweet impressions (this is the number of times users on Twitter have view an LMI for All tweet in their timeline);
- 209 New followers (currently, LMI for All has 1,201 followers);
- 2,798 Profile visits (this is the number of time the LMI for All Twitter profile page has been visited).

On the LMI for All website, nine blogs²⁹ have been published covering a range of topics from careers activities, information on different LMI for All data, stakeholder activities, and examples of data in practice.

4.5. Take-up of LMI for All and Careerometer

It is often difficult to estimate the take-up of LMI for All, as there are no monitoring systems in place to track developers and new sites that come online. Even estimating those that use Careerometer widget could be an under-estimation as there is a reliance on search engines to identify keywords associated with Careerometer, plus where Careerometer is embedded on an intranet then it will not show up in search results.

Over the last year, the team are aware of a number of developments created by existing third party users of LMI for All, including expanding the use of the data (i.e. U-Explore's Start, KareerHub) and enhancing their service with guidance on how to interpret LMI (i.e. icould). The following is a list of new third party users of LMI for All:

- BBC Bitesize Careers, <https://www.bbc.com/bitesize/careers>;
- Career Pathways widget by Float, <https://www.career-pathways.co.uk/>;
- Eluceo, <https://eluceoeducation.org/>;
- Google's job information (still in beta version).

At all stakeholder activities, Careerometer has been promoted as one of the three ways in which to access the LMI for All data. Careerometer is a free widget; the code for which can be copied from the LMI for All website and pasted into another website.

Careerometer enables the user to type in an occupation and access top level data from LMI for All on that occupation. The user is then able to type another occupation into a second or third card and compare occupations. For schools and colleges, embedding

²⁹ The blogs on the LMI for All website can be viewed at: <https://www.lmiforall.org.uk/news-and-blog/>

Careerometer has been an easy way to provide access to LMI and helps to evidence their progress toward the Gatsby good guidance benchmarks³⁰.

There has been considerable take-up of Careerometer especially by those unable to develop their own interface. In 2017, it was estimated that over 60 institutions had implemented Careerometer in their website which has enabled their users to access data from the LMI for All service. It is now estimated that 145 organisations embed Careerometer in their website.

In response to user requests, coding for Careerometer has now been provided that ensures it is mobile friendly and the user is able to embed it with colours that match their website. Guidance on implementing Careerometer was also enhanced with some suggested text to include with Careerometer and links to examples of it in practice.

Careerometer has generated good publicity and awareness around LMI for All, so a second widget will be developed during the next year of the project.

4.6. Summary and next steps

There continues to be a substantial and sustained uptake of the LMI for All resource. The ambition is to expand and secure this sustained uptake. The integrated stakeholder activities provide a clear demonstration of a consistent focus on end-user need and represent a transparent engagement strategy with those third parties that continue to be the primary data/service users. The overall aim of these activities is to continue stimulating demand and promoting the data tool.

The stakeholder activities are as diverse as in previous phases of the project, having reached a similarly diverse audience. They have all, without exception, been aligned with the stakeholder engagement objectives and plan. It is evident, however, that there are additional stakeholders that have not yet been effectively engaged which could be the focus of further effort, specifically parents/carers (some successful work has been completed); further education colleges and career teachers/leaders. Over the next year, stakeholder activity will focus on:

- Increasing activities relating to targeting and awareness raising in schools and further education colleges, targeting careers leaders as well as parents and carers;
- Continuing work with those stakeholders who have started designing and developing LMI for All interfaces and applications.

³⁰ For more information on the Gatsby Good Guidance benchmarks, see: <https://www.gatsby.org.uk/education/focus-areas/good-career-guidance>

Reflecting on developments for the next stage of the project, effectively supporting practitioners in the use of LMI to stimulate greater engagement with LMI for All is a focus going forwards. Delivering workshops and presentations is one way that has been successfully implemented. Additional thinking will be done on how best to support the confidence and competence of career and employment practitioners operating in compulsory and tertiary education.

Over the next year, a range of stakeholder activities will continue to be organised as well further development and refinement of case studies that demonstrate how third party developers have used LMI for All and the benefits accrued. The team will also explore how to promote the service through various publications, including its own newsletter. Currently, approximately 300 people have registered an interest in receiving further information on the service.

5. Future developments

The data, technical and stakeholder engagement developments outlined in this section will be undertaken over 2019-2020 and progress reported in the next annual LMI for All report.

Data developments

1. To monitor the development of new datasets using and linking administrative data with the aim of extending LMI for All with course and education pathways data.
2. To review and assess developments and techniques in machine learning and data scraping techniques, with the aim of creating new datasets, particularly for vacancies.
3. To connect with organisations that are becoming involved in open data with the aim of sharing learning.
4. To prepare for changes to the UK standard occupational classification system, which is being updated from SOC2010 to SOC2020.

Stakeholder engagement activities

5. To continue to stimulate demand and promote the LMI for All service to a range of stakeholders, with a focus on careers hubs, further education institutions, local governmental organisations and open data specialists.
6. To maintain support and dialogue with existing third-party users of LM for All and continue working with those stakeholders who have started designing and developing LMI for All interfaces and applications.
7. To increase activities relating to the targeting and awareness raising in schools, targeting careers leaders as well as parents and carers.
8. To continue collaborations with organisations that support the professional development of careers practitioners to ensure practitioners are confident in their use of LMI and confident in choosing between sources of LMI.

Technical developments

9. To continue to maintain a secure and robust database with the aim of ensuring the infrastructure that uses the most appropriate technologies, and review new data APIs that may be relevant to LMI for All.
10. To pilot and deploy the ETL system to support a more automated data upload process.
11. To implement new mappings to enable access to apprenticeship standard vacancies and higher education courses using HeSOC.

12. To continue to develop the LMI for All website with the aim of optimising visitor experience of the website.
13. To design and develop a new widget that will be made feely available and aimed at those who do not have the resources to develop their own interface.

References

- Alexander, R., McCabe, G. and De Backer, M. (2019). *Careers and labour market information: An international review of the evidence*. Reading: Education Development Trust. Available online: <https://www.educationdevelopmenttrust.com/EducationDevelopmentTrust/files/3a/3a1edb39-3ce4-46fb-958d-6e4f2fe3bc78.pdf>
- Barnes, A. (2015). *Ten Best Resources for LMI*. Cegnet, http://www.cegnet.co.uk/uploads/resources/The_ten_best_LMI_resources_final.pdf
- CEDEFOP (2016). *Labour market information and guidance* (CEDEFOP research paper; No 55). Luxembourg: Publications Office of the European Union.
- Department for Education (2017). *Career strategy: making the most of everyone's skills and talents*. London: Department for Education. Available online: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/664319/Careers_strategy.pdf
- Gatsby (2014). *Good Career Guidance*. London: The Gatsby Foundation. Available online: <https://www.gatsby.org.uk/uploads/education/reports/pdf/gatsby-sir-john-holman-good-career-guidance-2014.pdf>
- HM Government (2012). *Open data white paper: Unleashing the Potential* (Cm8353). Norwich: The Stationery Office. Available online: <https://www.gov.uk/government/publications/open-data-white-paper-unleashing-the-potential>
- Judd, A. (2018). Improving graduate earnings measures using LEO data. Paper presented at the *Graduate Outcomes measuring success and providing opportunity Conference*, London. Available online: https://www.hesa.ac.uk/files/Graduate_Outcomes_London_Conference_complete_sliddeck_0.pdf
- OECD (2015). *OECD Reviews on Local Job Creation Employment and Skills Strategies in England, United Kingdom*. Paris: OECD Publishing.
- The Careers & Enterprise Company and WorldSkills UK (2016). *How to use labour market information*, <https://www.worldskillsuk.org/media/4461/how-to-use-lmi-booklet-a5.pdf>

Annex A: LMI for All reports

The following is a list of LMI for All reports which document the development of the service from piloting, to full project status, to its continued development.

Table A.1 Overview of LMI for All development

Phase	Date	Activity
1	2012	<p>Tested the feasibility and viability of the foundation concepts and ideas</p> <p>Bimrose, J., Wilson, R., Elias, P., Barnes, S-A., Millar, P., Attwell, G., Elferink, R., Rustemeier, P., Beaven, R., Hay, G. and Dickerson, A. (2012). <i>LMI for All Career Database Project - Processes Adapted and Lesson Learned</i>. London: UKCES. Summary available to download at: http://www2.warwick.ac.uk/fac/soc/ier/research/prototypecareerslmi/bimrose_et_al_2012_lmi4all_summary.pdf</p>
2A	2012-2013	<p>Expanded and updated the database, tested and evaluated the API, improved the technical infrastructure, explored data sets that could add value to services, raised awareness of the service through a number of stakeholder events</p> <p>Bimrose, J., Wilson, R., Barnes, S-A., Owen, D., Green, A., Li, Y., Millar, P., Bosworth, L., Holden, A., Attwell, G., Rustemeier, P., Elferink, R., McGreggor, A., Wicks, D., Long, A. and Dickerson, A. (2013). <i>LMI for All Developing a Careers LMI Database Phase 2A report</i>. London: UKCES. Available online: http://www2.warwick.ac.uk/fac/soc/ier/publications/2012/bimrose_et_al_2012_lmi4all_summary.pdf</p>
2B	2013-2015	<p>Continual development of the LMI for All service, implemented recommendations from Phase 2A</p> <p>Bimrose, J., Wilson, R., Barnes, S-A., Owen, D., Li, Y., Green, A., Bosworth, L., Millar, P., Holden, A., Attwell, G., Rustemeier, P., Elferink, R. and Higginbottom, J. (2015). <i>LMI for All: Developing a Careers Database (Final report)</i>. London: UKCES. Available online: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/446332/LMI_for_All_Final_report_v020715_FW.pdf</p>

Table A.1 continued

Phase	Date	Activity
3	2015-2016	<p>Continual development of the LMI for All service, updated database and technical infrastructure, continued to raise awareness of the service through stakeholder events, worked with stakeholders to develop LMI for All applications</p> <p>Bimrose, J., Sofroniou, N., Barnes, S-A., and Attwell, G. with Rustemeier, P., Stieglitz, D., Elferink, R., Hughes, D., Li, Y., Owen, D., Wilson, R., Bosworth, L. and Millar, P. (2016). <i>LMI for All: Technical, Stakeholder engagement and data development services</i>. Wath-upon-Deerne: UKCES. Available online: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/545324/LMI_For_All_for_web.pdf</p>
4	2016-2017	<p>Continued development of the LMI for All service through a range of stakeholder engagement services and refreshment of the data, as well as the implementation of a new IT infrastructure to manage increasing queries to the database</p> <p>Bimrose, J., Barnes, S-A., Owen, D., Hughes, D., Wilson, R., Attwell, G. and Rustemeier, P. (2018). <i>Labour market information (LMI) for all: Stakeholder Engagement and Usage, Data and Technical Developments</i>. London: Department for Education. Available online: https://www.gov.uk/government/publications/labour-market-information-lmi-2016-to-2017-data</p>
5	2017-2018	<p>Continued maintenance and development of the LMI for All service</p> <p>Barnes, S-A., Bimrose, J., Owen, D., Wilson, R., Attwell, G. and Rustemeier, P. (2019). <i>Labour market information (LMI) for All 2017 to 2018</i>. London: Department for Education. Available online: https://www.gov.uk/government/publications/labour-market-information-lmi-for-all-2017-to-2018</p>
6	2018-2019	<p>This report describes the activities undertaken during this phase of the project</p>

Annex B: Review of DLHE and Graduate Outcomes data for LMI for All

Table B.2 Review of DLHE and Graduate Outcomes data for LMI for All

Topic	DLHE	Graduate Outcomes questionnaire	Graduate Outcomes Question
SOC	Standard occupational classification (4 digit)		Presume this will be made available from administrative data
Qualification obtained	Level of qualification obtained (Doctorate/ Masters/ Other Postgraduate/ First degree/ Other undergraduate)		Presume this will be made available from administrative data
Qualification required for job [TO GET THE JOB]	Qualification required for job	Available	ASK IF IN WORK OR DUE TO START WORK (A1=01 OR 05 OR A3=01 EXCEPT IF A4=02) B16 Did you need the qualification that you completed 15 months ago to get the job? (QUALREQ)
Subject	Subject of study (2 digit JACS)	Not in questionnaire	Presume this will be made available from administrative data
SIC	Standard Industrial Classification (4-digit)	Not in questionnaire	Presume this will be made available from administrative data
Sex	Sex	Not in questionnaire	Presume this will be made available from administrative data
Domicile	Domicile (UK/ EU)	Available	C10 What was the town, city or area in which your business or the main business you were working for was based during the [census week]? (BUSEMPCITY) WRITE IN.

Annex C: Further information on LEO

The variables included in the LEO for UK-domiciled graduates (correct as of 16/07/2019) are shown in B.1.

Table B.1 LEO variables for UK-domiciled graduates

Variable	Description
tax_year	Tax year in which outcomes and earnings relate to.
academic_year	Academic year in which the group completed their first degree.
YAG	Number of years after graduation.
sex	Sex of graduate.
Subject	Subject studied.
Ethnicity	Ethnic group of graduate.
InstType	Type of institution attended during study (HEI, FEC or AP).
StudyMode	Mode of study of graduate.
ageBand	Age of graduate at start of course.
POLAR3	POLAR quintile of graduate prior to study (young (under 21 at start of course) graduates only).
PriorAtt	Academic attainment of graduate prior to study (young (under 21 at start of course) graduates only).
FSM	Free school meal eligibility of graduate between the academic years 6 and 11 (young (under 21 at start of course) graduates only).
HomeRegion	Home region of graduate prior to study (young (under 21 at start of course) graduates only).
Residence	Residence of graduate in their final year of study (young (under 21 at start of course) graduates only).
grads	Number of graduates included in calculations.
unmatched_percent	Percentage of graduates that have been classed as unmatched.
matched	Number of graduates that have been classed as matched.

Variable	Description
activity_not_captured	Percentage of matched graduates whose activity could not be captured.
no_sust_dest	Percentage of matched graduates with an unsustained destination.
sust_emp_only	Percentage of graduates with a record or sustained employment only.
sust_emp_with_or_with_out_fs	Percentage of graduates with a record or sustained employment (these graduates may or may not have a further study record in addition to a sustained employment record).
sust_emp_fs_or_both	Percentage of graduates with a record or sustained employment, a record of further study, or both.
earnings_include	Number of matched graduates included in the earnings calculations (graduates in PAYE employment or self-employment or both).
earnings_LQ	Earnings lower quartile (PAYE earnings annualised, self-employment non annualised).
earnings_median	Median earnings (PAYE earnings annualised, self-employment non annualised).
earnings_UQ	Earnings upper quartile (PAYE earnings annualised, self-employment non annualised).

Source:

https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/789666/UKdomiciled_metadata.txt



Department
for Education

© Department for Education 2020

Reference: DfE-00024-2020 or RR 971

ISBN: 978-1-83870-090-4

The views expressed in this report are the authors' and do not necessarily reflect those of the Department for Education.

Any enquiries regarding this publication should be sent to us at: LMIDFE
LMI.DFE@education.gov.uk or www.education.gov.uk/contactus

This document is available for download at www.gov.uk/government/publications