



Mapping data ecosystems in the education sector - a foundation for green skills and jobs



This is licensed under a Creative Commons Attribution-ShareAlike 4.0 International license

Open Data Institute

Contents

Background	2
Approach	4
Terminology of data ecosystems	6
Data Ecosystem Map	8
Key findings	10
Barriers preventing access, use and sharing of data	10
Opportunities to improve access, use and sharing of data	11
Recommendations	13
Immediate priority	13
Short term recommendations (3 months)	14
Medium term recommendations (3 - 12 months)	16
Long term recommendations (12 months or longer)	17

About

This report has been researched and produced by the Open Data Institute, and published in November 2021. Its lead authors were Ben Snaith and James Maddison. If you want to share feedback by email or would like to get in touch, contact the project lead James Maddison at research@theodi.org.

Background

Developing and maintaining an inclusive and sustainable education and employment ecosystem plays an important role in the economic growth of local areas. Data is a core part of any effective modern ecosystem, for example, people seeking employment need to know what employers are looking for, and education providers need to understand how to align their offerings with market demands. In order to ensure that people have access to the data they need to make informed choices, the data that is part of those education and employment ecosystems need to be more open, while still maintaining the trust of stakeholders and those who have rights over that data.

West London Business is a business led non-profit whose members collaborate to raise West London's global economic competitiveness and catalyse action for people and planet. The West London Alliance is a public sector partnership between seven West London local authorities. Together these organisations have been working with stakeholders from across West London to understand how access to skills data could help to improve the responsiveness of skills provision within the region and ensure appropriate commissioning of education provision to support green skills and a transition to a zero carbon economy.

In 2020, the West London Alliance commissioned the Institute for Employment Studies and Rocket Science to deliver insights into demand for and supply of skills across seven West London boroughs, with a view to making better data available to enable the skills system to be more responsive. In parallel, West London Alliance and West London Business, with support from the Tech for Better programme, worked to develop a Minimum Viable Product (MVP) [website](#) that sought to collate data about green construction job profiles and information on relevant courses from Higher Education (HE) and Further Education (FE) providers in West London.

These initial efforts identified what issues currently affect how data is being accessed, used and shared across the West London education ecosystem, including:

- a lack of clear, openly available, timely and accurate data about the demand for skills, and at a sufficiently granular level of detail
- difficulty clearly identifying where specific gaps in skills provision lie, due to lack of available information
- a lack of data standards particularly around data about the supply side, for example general course profiles for FE courses, which makes it hard to compare and integrate similar datasets
- a lack of easy accessibility to supply side data, causing a larger manual working requirement for those wishing to use the data

As a first step towards addressing these issues and trying to identify opportunities to move forward in improving access to skills data for West London, the Open Data Institute (ODI) worked with West London Business, West London Alliance and partners to run a Data Ecosystem Mapping workshop. The study has been

commissioned by West London Business as part of Better Futures+, with funding from The Mayor of London's Green New Deal.

Approach

A data ecosystem map illustrates the value exchanges in a data ecosystem, which consists of the organisations, communities and people that create and benefit from the value created by the data flowing through our data infrastructure. Creating a map can help stakeholders understand and explain where and how the use of data creates value, and to identify the key users, the relationships between them, and the different roles they play.

On Friday 9 July 2021, the ODI ran a Data Ecosystem Mapping workshop for a group of 20 stakeholders from West London's education data ecosystem, with support from West London Business, West London Alliance and partners. The objectives of the workshop were as follows:

- Introduce the relevant ODI concepts of data ecosystem mapping, data infrastructure and the data spectrum.
- Teach participants how to map a data ecosystem, including data assets, value flows and stakeholders.
- Use the knowledge of participants in the room to create a data ecosystem map for the Education Sector.
- Use the data ecosystem map to identify what concrete steps can be taken to improve the data ecosystem or fulfill the use case needs.

To provide a balanced set of perspectives, stakeholders were assembled from varying organisation types and roles, including:

- Further education colleges
- Higher education institutions
- Independent training providers
- Local government skills and employment leads
- Greater London Authority representatives
- Department for Work and Pensions representatives

Participants worked together in two separate breakout sessions to map their understanding of the West London education data ecosystem. They used the following use cases to help them think about important stakeholders, data and other value flows, where the current gaps in the ecosystem exist and where there might be opportunities to improve access to data:

- Enabling the skills system to be more responsive through insights into demand for and supply of skills across seven West London Boroughs, with the overarching aim being to make available better data to enable the skills system to be more responsive, for example, by helping skills providers, funders and policymakers to better understand gaps in provision and how best to address them.
- Collating data about green construction job profiles and information on relevant courses from Higher Education (HE) and Further Education (FE)

providers in West London, with the aim of encouraging more people to take HE and FE courses in green construction, by making data about green construction jobs more easily accessible.

Terminology of data ecosystems

Actors - The people and organisations that are active within the ecosystem. They can include a wide variety of stakeholders such as data stewards, regulators, intermediaries, aggregators, creators, beneficiaries, researchers, policymakers and more.¹

Aggregator - A type of intermediary, an aggregator is a person or organisation that packages together datasets from many sources.

Algorithm - An algorithm is the list of instructions and rules that a computer needs to do to complete a task.² In the context of this work, algorithms might be used to match supply data with demand to identify gaps or oversupply.

Datasets - A dataset is a collection of data. Datasets can be raw data, collected for research or monitoring purposes, but a dataset could also be a combination of data collected by others, or the output of a model or analysis. Datasets are often used to enable the development of tools which can produce information, ultimately helping us to make more informed decisions.

Data assets - A data asset is what is considered to be a 'useful or valuable thing' relating to data. While this description is fairly vague, the term data asset most commonly refers to three types of asset: identifiers, registers and datasets.

Data governance - The exercise of authority, control and shared decision making (planning, monitoring and enforcement), over the management of data assets.³

Data identifiers - Identifiers are names that act as labels to help us uniquely identify physical and digital objects and services, and as pointers to information available online or stored in a variety of databases and systems.⁴

Data infrastructure - Consists of data assets, standards and technologies, guidance and policies, organisations and communities.⁵ We need strong data infrastructure in order to be able to collect, access, use and share data, which helps us to create value and minimises harmful impacts.

Data steward - A person or organisation responsible for collecting, managing and ensuring access to a dataset. A data steward may also provide other supporting infrastructure, such as data governance.

Data standards - Standards are documented, reusable agreements that solve a specific set of problems or meet clearly defined needs.⁶ Data standards can create a number of benefits if widely adopted, helping people to use the same vocabulary and language by using common models, attributes and definitions. Data standards can enable better exchange of data within and between organisations using common

1

<https://docs.google.com/document/d/1vSqoHOYT5u6vrCHlebCS0rze0gWwXOspeEowWzwake8/edit#heading=h.1jfx30agj1s6>

² <https://www.thinkautomation.com/eli5/what-is-an-algorithm-an-in-a-nutshell-explanation/>

3

<https://www.cio.com/article/3521011/what-is-data-governance-a-best-practices-framework-for-managing-data-assets.html>

⁴ <https://theodi.org/article/explainer-what-is-an-identifier/>

⁵ <https://open-data-institute.gitbook.io/data-landscape-playbook/who-is-this-playbook-for-1>

⁶ <https://standards.theodi.org/introduction/what-are-open-standards-for-data/>

formats and shared rules. Standards can also provide guidance and recommendations for sharing better quality data, and understanding processes and information flow.

File format - Refers to the structure of a file that tells a program how to display its contents.⁷ Common file formats include text, image, video and sound.

Intermediary - A person or organisation that provides value-added services that wrap, host or enrich a dataset.

Open API - An Application Programming Interface (API) is best described as a promise by one system to another on how the two can interact. The human API is our spoken and written language; it is how we interact.⁸ When APIs are made openly available, developers and other users can access and use data provided through the API to integrate with their own applications.

Registers - Registers are lists of data used to classify or categorise other data that help to improve the consistency and quality in how data is published and used. They help us to build confidence and trust in data by clarifying where different data stewards are referring to the same things in the same way.

Taxonomy - The process or system of describing the way in which different things are related by putting them in groups.⁹

⁷ <https://www.computerhope.com/jargon/f/file-format.htm>

⁸ <https://theodi.org/article/publishing-open-data-in-times-of-crisis/>

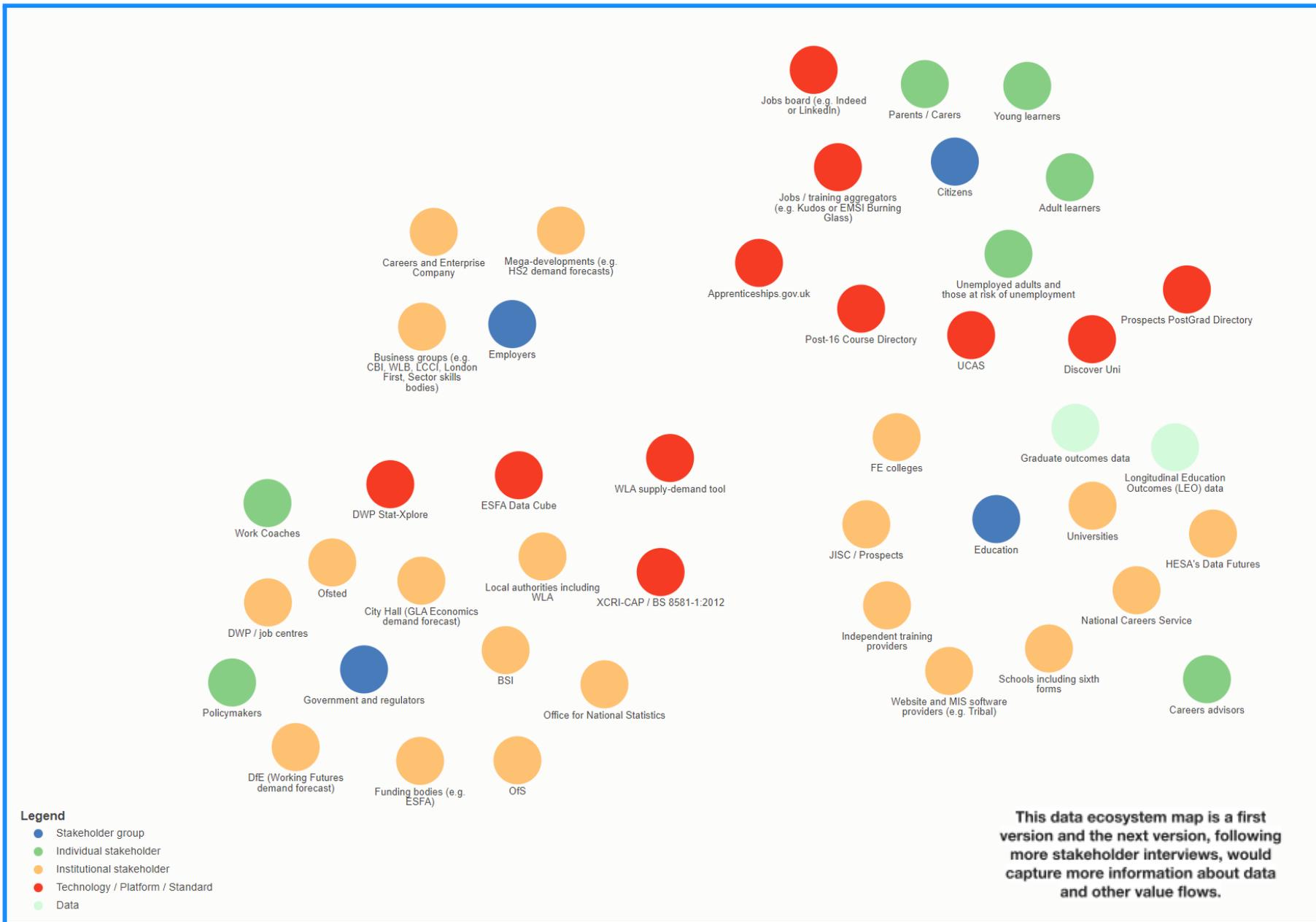
⁹ <https://www.merriam-webster.com/dictionary/taxonomy>

Data Ecosystem Map

Participants found the initial stages of the mapping exercise straightforward, as they were all familiar with the majority of stakeholders involved in West London's education data ecosystem. They could confidently identify data sources that many of those stakeholders collected and maintained, such as data about courses and their providers, or performance data, such as information on the success of schools and colleges in helping young people continue in education, apprenticeships or employment (known as destination measures). Participants could identify some of the data flows between organisations, for example, where an education provider is required to share data with a government department by law, or where an education provider shares data publicly via a particular platform. They could also clearly identify where the majority of the funding flows exist within the ecosystem.

Due to the size and complexity of the ecosystem at hand, participants found it difficult to map much more than the aforementioned information in the time available at the workshop. Key details about which organisations were sharing data with another, how they were approaching this and whether there were other value exchanges taking place were not able to be identified at this stage. As a result, a number of questions were raised, leading to a more detailed focus on the barriers and potential opportunities than is usual for a Data Ecosystem Mapping session.

Despite the challenges presented in mapping some of the value flows, as a result of the workshop, we were able to create the beginnings of a data ecosystem map for the West London education sector.



Key findings

Barriers preventing access, use and sharing of data

- Due to a lack of data, marrying up occupation outcomes of courses with labour market demand is difficult, particularly where courses have multiple occupational outcomes (HE), as opposed to having a specialist niche (more likely in FE). This could lead to inaccurate assumptions about which skills are required to satisfy labour market demand, as these assumptions are not based on evidence.
- There is a distinct lack of established standards, both around the taxonomies used to describe data and in the way that data is captured and identified. This is particularly prevalent in supply side data that is created by skills providers and makes it more difficult for organisations wishing to work with that data to compare different datasets. There are some existing frameworks, like the [XCRI-CAP](#), which could help to address this problem, but to date uptake in the UK has been limited to postgraduate provision from some universities. From the demand side, there is no standard way for employers to articulate their skills needs, which makes it difficult for skills providers to update their offerings to better meet these needs.
- It can be difficult to access national data that could be useful at a local level, as the data holders, such as the Department for Work and Pensions, require very specific and sometimes restrictive justifications as to why this data should be shared. For example, national data about skills and the labour market from the Department for Work and Pensions or elsewhere within central government, could help skills providers to provide more targeted and responsive services, such as aggregated data about the skills level and prior occupations of people who are on Universal Credit (or at risk of redundancy through the HR1 process) in a locality. Much of this national data is likely already aggregated and anonymised, or could be with relatively little effort, and therefore limiting access to the data unnecessarily limits potential innovation.
- There is a tendency for national government departments, like the Department for Education, to develop strategies without input from local stakeholders who are most affected by the resulting outcomes.
- There is little data about the people who enter and then leave the West London labour market (university students from other geographies for example), which means that it is difficult to know how much of the skills learnt in West London are being retained by the West London labour market.
- Some data sources which could support local education ecosystem decisions involve substantial cost to access making it less accessible for use by stakeholders (e.g. EMSI Burning Glass).
- An inability to send data quickly and efficiently can cause delays in data sharing between demand and supply, which affects the speed at which decisions can be made, particularly with regards to adjusting the supply side.

Opportunities to improve access, use and sharing of data

- Identifying which software providers are being used by the different skills providers and encouraging collaboration and communication between them in order to improve the way data is accessed and shared via their systems. West London Vice-Chancellors and Principals have set out opening up data as a shared objective of their new collaboration. This is strong advocacy for improving access to data in education, but it will be necessary to explore what incentives and leadership from their partners will be needed to accelerate open data sharing.
- Improving the way supply side data, such as course provision, connects to demand side data, such as skills needed by West London employers, is important in making sure that suppliers are able to respond dynamically to the needs of the market. There is an opportunity to build on the existing work done by Rocket Science using manual coding of courses with SOC/ SIC codes.
- Identifying new use cases for skills and education data could help to demonstrate the value of improving access to this data to key national data holders, such as the Department for Work and Pensions.
- Ensuring that residents of West London who are already in work have access to data about the provision of skills and how that relates to job opportunities, so that they can make decisions about adding to their skill set.
- Improving standardisation, starting with the taxonomies used to describe data benefits the whole ecosystem by giving everyone a shared language. Improving technical standards will improve the way that organisations are then able to access and share this information.
- Student barometer, provided by [Tribal](#), which tracks and compares the decision-making, expectations, perceptions and intentions of students from application to graduation. This information is not publicly available.
- There is an opportunity to engage with data initiatives and projects that were identified during the session, which may have useful datasets or data standards that could be made openly available, or may already be openly available. These initiatives and projects should be engaged as part of further work. Initiatives and projects include:
 - [Jisc](#), a digital solutions provider for UK education and research, which funds a number of projects around data, such as projects around [improving postgraduate course information](#) and [managing course-related information](#).
 - Jisc has also funded the creation of the [XCRI-CAP framework](#), which stands for eXchanging Course Related Information, Course Advertising Profile. XCRI-CAP is a standard that makes it easier to move data about courses around. It is used to improve the way staff do things within their own organisation and in partnership with other organisations. It gives organisations the capability to offer new and enhanced services to users of course related information. Its standard format and common definitions make it possible for information in different systems to be shared and updated without having to re-enter data. The XCRI-CAP framework could be

augmented by West London partners to include SOC/ SIC codes and course place information. It went on to underpin BS 8581-1:2012 Exchanging course related information - Course advertising profile and EN 15982 Metadata for Learning Opportunities (MLO).

- LinkedIn, which provides National and City level data that captures the evolution of skill requirements over time across the workforce based on updates to LinkedIn member profiles. Made available via [The World Bank](#).
- [Unistats dataset](#) from the HESA website, which provides comparable sets of information about full- and part-time undergraduate courses, as well as experimental statistics about graduate outcomes
- Given the size and complexity of the data ecosystem, a useful activity would be to identify key data holders from the map which were not engaged in the initial stages of the project, and work with them to map their internal data ecosystem. This would be done with a particular focus on the use cases around better skills provision in order to provide some parameters to the exercise.

Recommendations

Immediate priority

Create a priority list of use cases that wider data sharing should support, assess their viability and create a roadmap to execute them working with the relevant stakeholders

It is important for all work towards improving access to education and skills data be underpinned by clear use cases that outline how improving access to data can address a real-world problem. Without these examples, key stakeholders, such as skills providers and the software providers that they use, are unlikely to adapt their current approach to making data available.

Two clear use cases have already been established:

- Enabling the skills system to be more responsive through insights into demand for and supply of skills across seven West London Boroughs
- Collating data about green job profiles and information on relevant courses from Higher Education and Further Education providers in West London

In addition to the two existing use cases, two additional use cases were highlighted as worth exploring during the workshop:

- Understanding what better skills provision looks like for those on Universal Credit
- Understanding what better skills provision looks like for those at risk of redundancy, for example leveraging anonymised and aggregated insights from HR1 submissions broken down by local authority

These use cases should be presented to core project partners, and any additional key stakeholders that might need to be involved in delivery against said use cases, such as the Department for Work and Pensions, to get feedback from them as to whether addressing these use cases would prove useful to the wider ecosystem.

For each use case that is prioritised, a roadmap should be created to more specifically shape how the following recommendations are undertaken.

Short term recommendations (3 months)

Collate a structured list of relevant data sources that were identified during the data ecosystem mapping session and build on this list throughout the project

Using the prioritised use cases as a framing, collating a list of all relevant data sources, including those mentioned in the workshop, will help West London Business and West London Alliance answer the following questions:

- What does the current availability of data look like?
- Who is collecting data that hasn't been engaged yet?
- Are there any clear gaps in the available data?

This activity will help West London Business, West London Alliance and partners to understand who needs to be engaged in order to improve access to the necessary data and drive forward the chosen use cases.

Identify the main software providers in the ecosystem and establish their role in future project work

It is important that software providers are engaged with the project as early as possible, as their systems make it possible to improve access to the data currently held by skills providers. Engaging with software providers early will give West London Business, West London Alliance and partners the opportunity to identify how to incentivise each provider to make changes to their system and improve access to data. Adoption of XCRI-CAP by West London HE and FE partners may be a quick-win.

West London Business and West London Alliance should engage with key skills providers to identify what their software provision currently looks like and to request contact details for each provider.

Convene stakeholders who were identified as being a key part of the data ecosystem during the ecosystem mapping session, or as part of the data sources list. Work with these key players to understand their internal data ecosystem and what prevents them from sharing data more openly

During the ecosystem mapping session, a number of stakeholders were identified, such as skills providers and funding bodies, but participants were unable to go into detail about what data each organisation is currently collecting and how each organisation shares that information with other stakeholders, as they did not have the information to do so.

A clear next step would be to prioritise these key stakeholders and to reach out to those considered a high priority to establish if they would be willing to engage with West London Business, West London Alliance and partners around improving access to the data that they collect and maintain. Those organisations that are willing to engage would undergo a similar data ecosystem mapping session, using the parameters of the prioritised use cases as guidance. This session would focus more on internal data collection and sharing, as well as the blockers that prevent them from

improving access to data and the opportunities to progress towards this aim. For example, Jisc advised that data to create XCRI-CAP feeds was sometimes fragmented across multiple IT systems in a Higher Education Institution (HEI) which may then also demand collaboration across functions to bring the data together. Marketing departments are often strong champions of more open course data as it can ensure they reach as many audiences as possible.

Each session would need to involve individuals from relevant departments and roles that understand how data is collected and used by the organisation. In particular, having one or more senior stakeholders present, such as a Chief Data or Digital Officer, is crucial for securing buy-in for the future of the project. It is also recommended that, where possible, the organisation's software provider(s) also joins the session.

Build on existing Rocket Science research to provide a documented approach to connecting supply and demand data around skills

This recommendation applies specifically to the first use case around enabling the skills system to be more responsive.

Existing research from [Rocket Science](#), one of the core collaborators in the project so far, has found that there are a number of complicated nuances around connecting data about the supply of skills to data about the demand for skills. Examples include how to account for people receiving higher and further education outside of West London, who then bring those skills back to the region as they join the labour market, as well as how to differentiate between skills required for jobs that are fairly sector agnostic, such as a Data Analyst, compared to the skills required for jobs that are extremely sector specific, such as a Commercial Pilot.

In order to unlock value from supply and demand data, the Rocket Science research needs to be taken further, addressing nuances using activities like exploring possible standard taxonomies for supply data, or accessing and analysing more data about those entering and leaving the labour market. Existing standards such as XCRI-CAP should be examined to confirm that they can support the standardisation of supply side course data, and where and how they might be best deployed.

Important considerations for this research should be around how the findings are documented and managed. These considerations might underpin the logic in a future algorithm that matches supply and demand in real-time. It will need to be decided whether the approach to connecting supply and demand data is something managed centrally, by West London Business, West London Alliance and partners, a third party, or whether a decentralised approach should be taken, involving other stakeholders in the ecosystem, like the skills suppliers and their software providers. [Prospects](#) (part of Jisc that improves student and graduate career outcomes through information, guidance and opportunities) currently provide the third party intermediary role aggregating Postgraduate course information from HEIs using XCRI-CAP.

Medium term recommendations (3 - 12 months)

Work with key data providers and the software providers that they use to enable easier ways of sharing data between two systems or organisations

This recommendation builds directly on all four of the short term recommendations, by bringing together software providers, data holders, and project partners to establish how the systems that they currently use to collect and share data could be adapted to suit the needs of the use cases, and wider needs of improving access to supply and demand data.

By this point in the project, there should be a good understanding of the datasets available, so the stakeholders mentioned will need to come together to examine what adaptations need to be made in order to create a standardised approach to the following:

- The taxonomy used to describe supply and demand data, such as the way skills are described by each organisation
- The data standards that make data interoperable, such as file format and common identifiers
- The technology that enables sharing of data, such as open APIs that enable efficient and standardised sharing of data

There may also be other adaptations that emerge from the short term activities that should be incorporated into this activity.

Pivotal to the success of this recommendation is ensuring that key data providers and software providers understand the value of existing standards such as XCRI-CAP, and that education leaders champion the implementation of these standards to support greater cooperation and a better informed marketplace.

Stakeholders should first be brought together for a workshop to make sure they are on the same page. Some software providers' systems may already support XCRI-CAP. UCAS and other aggregators may already have an awareness of what data feeds are available. Some HEIs may have adopted XCRI-CAP and could help others to apply and adopt in a similar fashion. Once the state of play is established, through the initial workshop, this should be followed up by regular working sessions between data holders, their software providers and the relevant project partners. It is important to make sure that stewards of data standards, like Jisc, are involved at an early stage to ensure that all stakeholders can understand the value of adopting an existing data standard, stewards are better able to understand the needs of the stakeholders, and for both groups to be able identify where it may need to be augmented to meet current or future needs.

This activity should allow significant progress to be made towards the practical aspects of how supply and demand data is shared.

Work to build strategic partnerships with organisations that collect relevant data at a national level

This recommendation is particularly relevant to the new use cases around aggregated data about skill needs of people on Universal Credit and those at risk of unemployment, but could be beneficial to all of the suggested use cases.

One common barrier suggested in the workshop was the inability to access more granular data representing a national perspective around skills and the labour market from organisations such as the Department for Work and Pensions. However, it is worth noting that some early conversations have been taking place between project stakeholders and the organisations that steward these national datasets. The main reason cited for the lack of access was the difficulty in making a clear and strong justification for why access to this more granular data is needed, and the benefits and positive outcomes which could be achieved by better access. Making this recommendation a medium term priority will allow for time to build a really clear narrative as to why these organisations should be as open as possible with the data that they collect.

By building strategic partnerships with organisations that collect and maintain relevant national datasets, West London Business, West London Alliance and partners will be able to:

- Secure access to relevant national level datasets that will provide wider context and be useful for benchmarking against other regions
- Suggest where new datasets that do not already exist should be collected, for the benefit of the national labour market and education sector, for example, information about the current skills and occupational experience of those on Universal Credit or at risk of unemployment
- Ensure that software providers factor in national datasets as part of their efforts to adapt their systems to improve access to data
- Possibly secure additional funding for use cases that align with national priorities, or be able to draw on resource and capacity from teams within those organisations

Building these strategic partnerships will also put West London Business, West London Alliance and partners in a good position to create replicable solutions that benefit other regional labour markets across the UK.

Long term recommendations (12 months or longer)

Run small pilots addressing the priority use cases identified in the previous phase of work demonstrating the value of improving access to education and skills data

Once the data infrastructure has been adapted to better suit the needs of the use cases, a logical next step would be to run practical pilots that address the priority use cases and demonstrate the tangible value of improving access to data.

For the first use case around enabling the skills system to be more responsive, this might involve generating relevant insights for skills providers so that they can make informed decisions about how to adapt their course offerings to better suit the needs of employers.

For the second use case around collating data about green job profiles and information on relevant courses from Higher Education and Further Education providers in West London, this might involve building on the work done to create the Skills West London [platform](#), by integrating newly available data from HE and FE providers into the system.

The approach to the pilots for the other two use cases, if prioritised, is likely to be informed by interactions with national data holders, such as the Department for Work and Pensions and Office for National Statistics, and will need to align with their priorities.

Based on the outcome and learnings from the pilots, determine the need to formalise some elements of the data ecosystem in order to support wider scalability of use cases

Once the pilots are nearing completion, it is worth considering how the approach that has been taken to ensure better coverage of skills provision in West London could be made sustainable and scalable. It is important that the outcomes of this project have a long lasting, positive impact on the labour market.

In line with the outcomes and learnings from the pilots, this could involve the development of data standards that can support a wider variety of use cases, creating a formal governance structure such as a data institution to steward the data being shared and support wider engagement and collaboration between stakeholders, or supporting other regions to adopt a similar approach to their own use cases.