UK Tech Innovation Index 2 and The Data City

January 2018

Thomas Forth, Paul Connell, Peter Laflin, The Data City

UK Tech Innovation Index 2 is produced by The Data City with support from the Open Data Institute (ODI). The project is part of the ODI's innovation programme, a three-year, £6m programme to support and build upon the UK's strengths in data and data analytics, funded by Innovate UK.

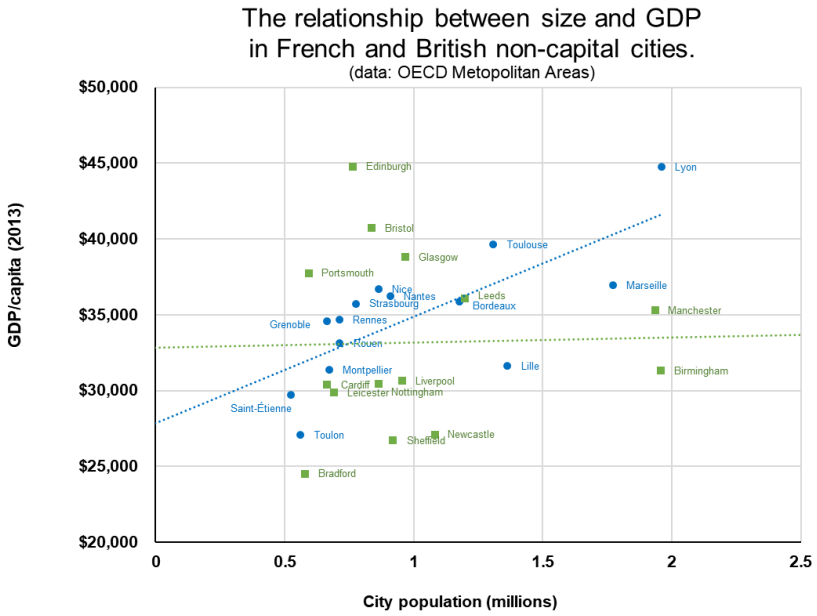
Throughout this project we have blogged at thedatacity.com/blog and will continue to do so as The Data City grows. At the time of writing there are seven blog posts which we refer to in this report and which cover much of the same content. There will soon be more. You may prefer to start there.

The final product is at thedatacity.com/products/uk-tech-innovation-index-2/

Contents

Motivation: Why study Industrial clusters?

The increasingly urgent desire to study industrial clusters in the UK can best be explained in a single graph. In most countries (below I have chosen France, but the pattern holds across Europe), the larger a city is, the more prosperous it is likely to be. In the UK this pattern does not hold and Manchester, Birmingham, Liverpool, and Newcastle significantly underperform international equivalents.



UK mid-size cities underperform EU equivalents — a cluster-focused industrial strategy aims to close the gap.

The most widely accepted reason why larger cities are more prosperous is that the gathering together in one place — the clustering — of industry, events, skilled people, and institutions leads to faster innovation and more rapid gains in productivity. The challenge for the UK is to understand why it almost uniquely has failed to harness these productive powers.

The UK Government's 2017 Industrial Strategy White Paper seeks to make the UK better at forming and growing clusters. Three quotes summarise the goals well,

- "my belief in a strong and strategic state that intervenes decisively wherever it can make a difference".
- "we must promote growth through fostering clusters and connectivity across cities, towns, and surrounding areas".
- "we must earn wide support for our Industrial Strategy through the quality of our decisions and by sharing the evidence on which they're based".

The aim of this work is to support the last of these quotes.

We are sharing a first output of our ongoing work to provide an evidence base for better-informed decisions within the UK government. We are sharing many of our methods so that others can understand and trust them. We are documenting the datasets that we use and releasing many of the datasets that we create so that others can use them. We are showing the limitations of existing data and showing what would be possible if more data were available to us.

The Data City: how we are different.

Focused on our diverse customers and users[1]

The Data City is a private company. We are completely focused on our customers because their investment is the only way that we will be able to continue our work. In the past three months we and our project partners have spoken to BEIS (The Department for Business, Energy, and Industrial Strategy), Nesta, Bradford City Council, Leeds City Council, KPMG, Lille City, Lille Metropole, The Royal Society, Innovate UK, Transport for the North, The Northern Powerhouse Partnership, The Cabinet Office, Raspberry Pi Foundation, Oxford Insights, and Regeneris about our work.

This diverse group are asking us to help them inform,

- Private investors looking to invest in companies.
- Existing businesses looking to relocate or expand.
- National government departments looking to assign investment nationally.
- Local and regional governments looking to assign investment locally.
- Local and regional governments looking to make the case to national governments and private investors.

This diversity is our strength because every new dataset that we collect and every new question that we ask improves The Data City for all of our customers. Everyone benefits from each other's curiosity and ambition.

Working in the open

Throughout this project we have blogged at thedatacity.com/blog. We are committed to working in the open. We do this partly because it helps other people, but mostly because it helps our team communicate, it helps us hear about new ideas, and it helps us to improve the quality of our work.

Our open approach has already created a lot of extra value.

- Publicising our early work with the French companies open dataset, Sirène, led to us being featured on the French government's state innovation lab blog, étalab.
- We are now partnering with the Métropole Européenne de Lille to deliver The Data City in France.
- Our open data from the first UK Tech Innovation Index was used in the UK Government's Artificial Intelligence Strategy and referred to in the UK Government's Industrial Strategy.

There is always a risk that working in the open lets others copy your innovations without sharing back. This has happened to us in the past, but overall we think that it is worth the risk.

[1] Full blog at -- http://thedatacity.com/blog/#Listening

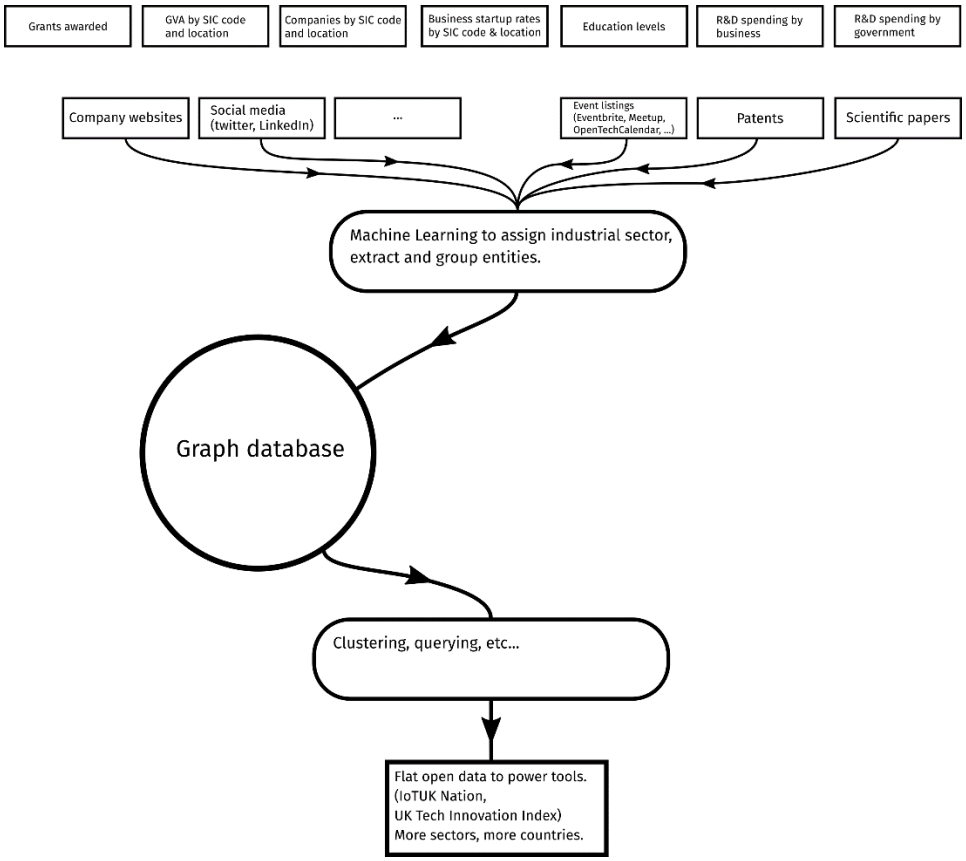
UK Tech Innovation Index 2.

UK Tech Innovation Index 2 is powered by The Data City and for this purpose it is best to understand The Data City as the combination and development of two complimentary products.

The first product is the IoT UK Nation database. This uses natural language processing (NLP) and machine-learning to find over 600 UK firms working with The Internet of Things (IoT).

The second product is the UK Tech Innovation Index. This ranks UK cities by their innovation performance and potential in niches of technology. To do this it uses innovative measurement techniques such as Eventbrite and Meetup events, and scientific publication records.

UK Tech Innovation Index 2 combines these two approaches into a single process that takes in more data, updates more frequently, and covers many more businesses.



Data flow through The Data City makes continually-updated data from over a dozen sources accessible via a single querying interface.

Improvements have been made in five key areas,

1. The classification of businesses into industrial subsections using **machine-learning**.
2. The inclusion of many more scientific papers and events.
3. Moving from pre-defined cities to **functional dynamic clusters**.
4. The adoption of categories of technology from The UK Government's Industrial Strategy.
5. More frequent and efficient data collection and a shift from a flat data structure to a **graph database**.

A more detailed summary, including planned improvements for the future can be found in the Tech Innovation Index development progress and roadmap (past, current, future).section at the end of this report.

Further explanation of these five changes, and more, are included on our blog. In this report, we will focus on explaining the two most important methods that power UK Tech Innovation Index 2; industrial sector classification using machine learning, and clustering.

Classification using machine learning[2]

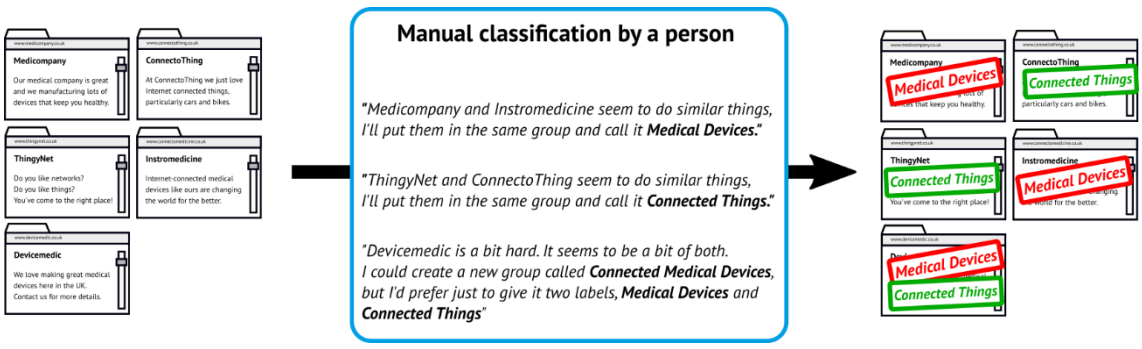
One of the most important features of The Data City workflow is the classification of businesses. This is important because of the limitations of SIC codes. SIC codes define what activities every business in the UK performs, but they are poorly-suited to large companies that span many industrial sectors and technology companies whose small niches of operation change frequently.

To better classify businesses in the UK we use machine-learning.

At the start of this report we showed where machine-learning fits into our data processing pipeline. Here we'll explain more about how that works. It's simpler than it sounds.

We start with the list of UK companies available as open data from Companies House. Searching for each company name on the internet usually finds a website for the business and we can collect every website for analysis. The difficult part is deciding what a company does.

We could try and classify each business manually; it is usually quite easy to tell what a business does just from its website. We call this manual classification. It's not perfect and no two people will agree in how they classify business, but it's pretty good.



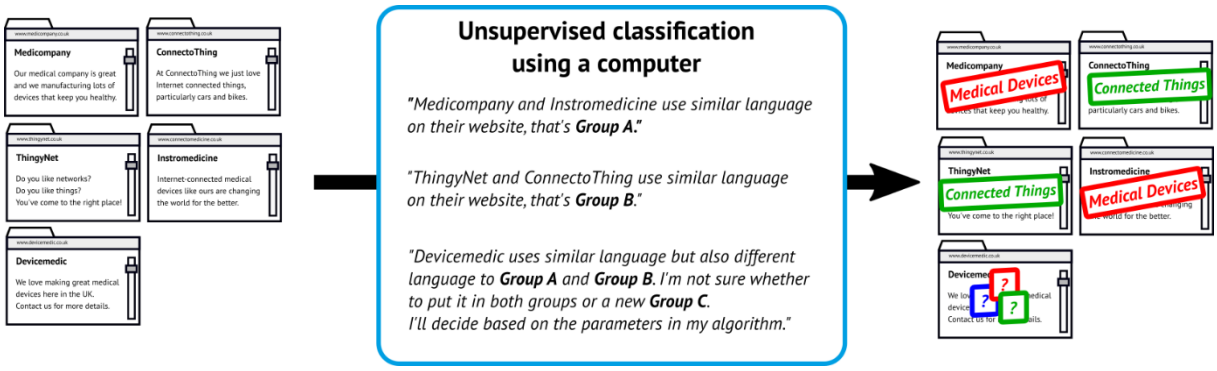
It is usually quite easy for a human expert to tell what a business does by looking at its website.

The obvious problem with manual classification is that with over a million UK businesses, it takes far too long. By the time industries are classified, new industrial sectors have sprung up.

The solution is to get a computer to classify industrial sectors.

[2] Full blog at -- http://thedatacity.com/blog/#Classification. A longer blog is at -- http://odileeds.org/blog/2017-06-13-mapping-economic-structures

Advances in machine-learning in the past decade mean that this is much easier than it once was. Unsupervised machine-learning algorithms can cluster websites into groups quite quickly.

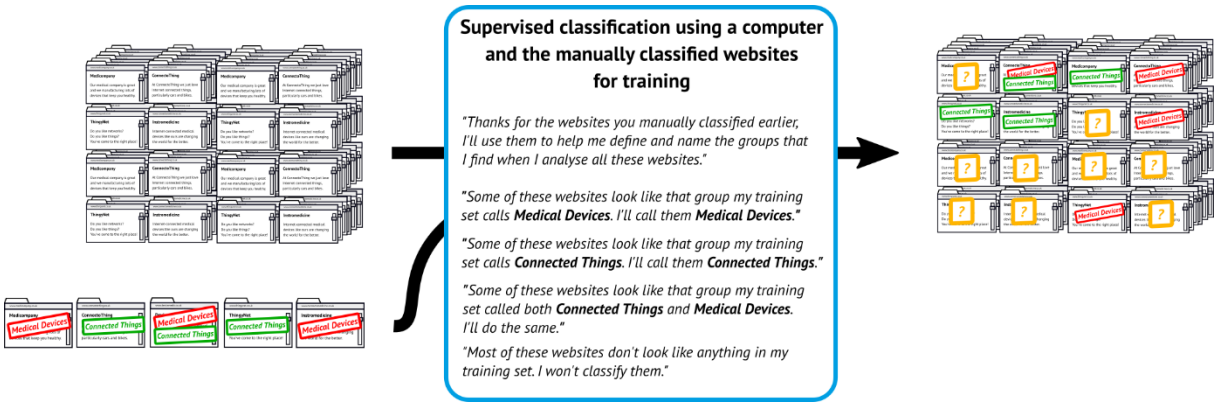


An unsupervised machine-learning algorithm can identify similarities between websites and group them. It cannot usually name the groups, and will require guidance as to how many groupings to create.

The problem with unsupervised classification is that the results lack context about what is being grouped. Techniques might group companies by things we don't care about, like how optimistic the language on their webpage is, or whether they use Wordpress or Squarespace for hosting. Unless we make it clear that the classification we're interested in is by industrial sector this information can easily be lost. And even once we've got this right, creating clusters without giving them human-understandable names isn't much use.

The solution to this problem is to manually classify a small number of websites and businesses, and then use this to train the machine-learning algorithm so that it can spot similar businesses.

Since the groupings are named manually, they are meaningful to experts. Since we use a computer to do the classification, we can analyse millions of websites.



A supervised machine-learning algorithm uses a collection of websites that have been manually classified to learn how to classify a much larger set of websites. "Learning" can be as simple as setting parameters in an otherwise unsupervised algorithm. It can be much more complicated and iterative in cases where deep learning techniques are used.

The power of this classification technique is increased enormously because of our approach to open data. When we released the first IoT UK Nation dataset as open data we were told that some

companies were missing. We received other feedback that some companies on the list were not in fact involved in IoT.

This was not a surprise to us — our machine-learning based approach isn't perfect and never will be. We incorporated this feedback into our classification model by adding those companies that were missing to the training set. When we reclassified our businesses using this new model many similar companies that were previously excluded from the list of IoT businesses were added. In a similar way, this new classification model removed both those companies that were wrongly classified as involved in IoT and some similar companies that no-one had explicitly alerted us to.

In this way, by sharing our outputs openly and by continually checking our classifications, we improve the classification model and thus the quality of our classifications over time.

It's a bit more complicated than that.

The diagrams and the explanation we've given are simple, but they manage to cover the important parts about how our industrial classification system works. Additional complexities are added to support more languages, more types of industry, and to try and keep the industrial classifications reasonably stable over time.

The classification system is continually learning as more websites are scraped, and more manual classifications are added. It also learns because we use more data sources than just websites. We use company tweets, links from the company website, and links to the company website from elsewhere on the web. These inbound links can include LinkedIn, open lists of companies from specialised industry groups, and public grant winners where we know the theme of the grant.

These are all pieces of information that a human expert might miss or be too busy to consider. But a computer can include them when deciding how to classify a business.

The final additional complexity for this blog post is how we use our classification model to classify more than just businesses.

We use the same system as we've just described to classify events from services like Open Tech Calendar, Meetup, and Eventbrite, and from patents and scientific paper abstracts. In this way we have a single learned ontology of industrial classification that we can use to combine data from many data sources, each with errors and uncertainties, to understand the evolution of industrial clusters.

Our classifications are currently pretty good, but we know that they can get much better. The most exciting part of The Data City is that the more data we collect, the more questions we answer, and the more manual corrections we make, the better our classifications get.

Current and potential clustering methods and why graph databases help[3]

Wards, local authorities, combined authorities, primary urban areas, local enterprise partnerships, statistical regions, nations. They're just some of the geographies that we deal with when using official statistics within the UK. The history and emotion captured in each of their definitions makes comparisons difficult. Some people even argue that we shouldn't compare at all — though we've never heard a good alternative for how limited resources should be allocated.

Things gets even more complicated when we talk with governments and communities. The definition of a place changes enormously depending on who we speak with and the context of the discussion. It's great that people are as passionate about places as we are, but it makes things complicated.

We deal with this complexity in two ways; simplification and starting from the beginning.

Simplification: The French way.

Geography is simpler than in the UK in many countries. For example, France has at various points in recent centuries thrown away old geographical definitions and created new simpler ones. The organisation of the country into communes, agglomerations, metropoles, départements, and régions makes most data analysis, and the discussions that arise from it, simpler than in the UK. Where similar simplified and functional geographies exist, we use them.

Starting from the beginning: clustering

In the absence of similarly good geographies in the UK we've developed an alternative: ignore existing geographies completely and use clustering algorithms to create better geographies.

With this approach, we don't care which city an event happens in, which local enterprise partnership a businesses is in, or which statistical region a university publishes papers in. We throw away all of that data and for each industrial category that we're investigating we just consider the precise location of each event, business, or university, and the links between them.

The strength of a link between two entities can be as simple as how close they are on a map. It might be the proportion of common attendees across two events, a shared supply chain for businesses, an innovation grant that multiple companies won, or staff who've worked at multiple places.

These links are why there is a graph database at the core of The Data City and the importance of the links is clear in the fact that the graph database has many times more links than entities.

The biggest problem with our clustering approach is naming. Cities have names, but clusters almost never do. In the first UK Tech Innovation Index we used a city definition called the primary urban area, which often merges together neighbouring cities and towns, then gives the whole area the name of the largest city.

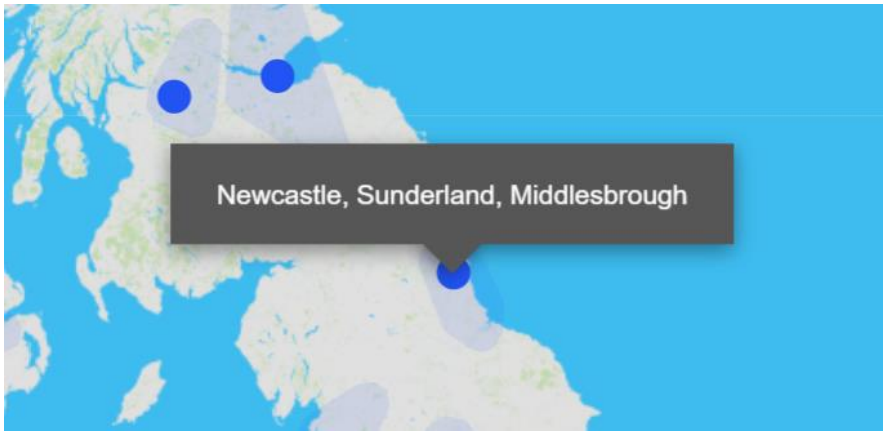
We heard from a lot of people that they didn't like that, and they kept on asking us where their city was in our ranking.

---

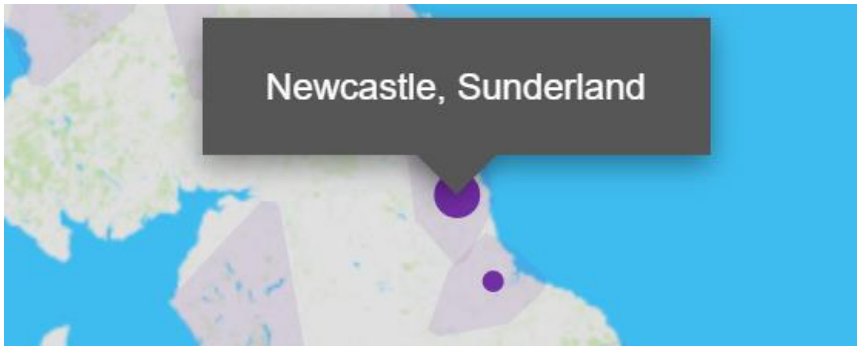
[3] A full blog, with video, on clustering is at thedatacity.com/blog/#Clusters.
Another blog, more fully explaining the graph database is at odileeds.org/blog/2017-06-13-mapping-economic-structures

So for UK Tech Innovation Index 2 we're trying something new. We create our clusters and then look up what significant towns and cities are within them. Then instead of naming the cluster after the largest place, we name it after them all. And we make the extent of the cluster clear on our map.

We know that this isn't perfect, but we hope that it will be less controversial and more useful.



Should we call it Newcastle? Newcastle, Sunderland, Middlesbrough? Or split the cluster and push each place down our rankings?



Because clustering is specific to each industrial category it is not consistent across all industrial categories.

If this new approach to clustering proves popular then it will be extremely useful to us as we continue to develop The Data City.

In Great Britain we understand most of the politics and local tensions around combining places into clusters and then naming them. But we don't understand these concerns anywhere else, and we already working with global data and thousands of clusters. It would be extremely difficult to name these clusters by hand.

For us, UK Tech Innovation Index 2 is a chance to see how well-received out new cluster-definition and cluster-naming system is. Let's see.

What data is being used and how much is open?

The Data City contains data from many sources, most with complicated and often incompatible licenses. This is a big challenge.

We try hard to operate within all of the restrictions on the data that we have and we think that by extracting summaries and then deleting raw data we achieve this. But it is impossible to be sure.

The following are the main data sources that we use for UK Tech Innovation Index 2, with our notes on the quality of the data, the licensing and conditions of use, and how we comply with these conditions.

- Eventbrite API. Eventbrite has an excellent API that gives access to all events from the search date into the future. It has a non-open custom license with some important restrictions. We have interpreted these restrictions as not allowing us to store full events, and so we instead store a summary. In accordance with the terms of the license we will not share raw Eventbrite data, but will share that an event took place at a location in a year.
- Meetup API. Meetup has a very good API with access to all events from the search date into the future. The event search function is not as powerful as Eventbrite but this can be worked around. It has a non-open custom license with some important restrictions. We have interpreted these restrictions as not allowing us to store full events, and so we instead store a summary. We treat this data similarly to how we treat Eventbrite data.
- Open Tech Calendar. Events are listed from the download date into the future as a CSV file and under an open license. Historic events are available on request.
- GetInvited to is an event service that is widely used in Northern Ireland. We had hoped to add its events to The Data City but since it has no API and no data policy we are unable to.
- UK Patent Office (UKPO) releases patent data under The Open Government License (OGL) which permits re-use including in commercial settings. Patents in this data are classified according to the International Patent Classification (IPC), but these definitions do not seem to be completely open. For our purposes UKPO is poor because it typically identifies the address of the patent holder and not the addresses of the primary inventors. In an example where a deodorant technician in Leeds invents a new product and the patent is filed at Unilever's headquarters in London or Rotterdam this misidentifies the true location of innovation.
- The International Patent Classification list is available via Eurostat and appears to be free to use, with copyright remaining the property of The World Intellectual Property Organization.
- European Patent Office full-text data requires payment to access in bulk form and is made available under a 6-page custom license. There is a good online full-text search tool but the results are limited to 1000 results per query via the custom UI. With no API and no free method for bulk data download we have fallen back to using UK Patent Office data for this project.
- Companies House data is free and available to use without condition. When it was published on data.gov.uk it was marked as 'Licensed under "Supplied under section 47 and 50 of the Copyright, Designs and Patents Act 1988 and Schedule 1 of the Database Regulations (SI 1997/3032)' although exactly what this means is unclear. The main Companies House website states that "All content is available under the Open Government Licence v3.0, except where otherwise stated" but it is unclear whether this extends to The Companies House data product. But clarification from Companies House suggests that there is no license, and no license is required, because the data is public information and free from any restriction on use.
- Microsoft Academic Knowledge API data is not free and not open. Licensing conditions are not stated anywhere and no restrictions are flagged for data that you've paid for. Following discussions with Microsoft Research staff we have assumed that it is okay to use but not share in a raw form. We treat this data similarly to how we treat Eventbrite data.

- We also scrape data from a wide variety of websites and access data using APIs to social media feeds. We discuss this more, and whether such data is personal and/or sensitive on our blog[4]. In general, data scraped from public websites are not restricted by licensing as much as by data protection law and copyright. Where we access data via twitter's APIs the licensing is quite restrictive.

What open data is being created?

We share open data both on portals and GitHub. Examples from previous projects by The Data City are,

- Leeds Digital Businesses on Data Mill North.
- IoTUK Nation Business List on Data Mill North.
- UK Tech Innovation Index 1 on GitHub.

We are currently sharing the following outputs of UK Tech Innovation Index 2 on GitHub.

- GeoJSON files of cluster boundaries, for each industrial category.
- Relative scores for each cluster, for each category, with a breakdown by contribution (business, events, papers, patents etc…).
- A list of every event, business, and scientific paper used to calculate the cluster boundaries and cluster scores, including their precise location.

In the future we intend to publish the following open data,

- Classified businesses, similar to IoT UK Nation Database, but for the categories defined in UK Tech Innovation Index v2.
- Classified events, where those events came from open events services such as Open Tech Calendar that allow us to republic events. This will let us show that machine-learning can classify events and get feedback that might improve our method. Republishing of events from Eventbrite and Meetup is not permitted under their license so we will not include their events.

What data would make it better?

As we have mentioned frequently in this report, more data and more questions means that The Data City can answer more questions and answer them better. The following improvements would have made UK Tech Innovation Index 2 even better.

Events data

- Historic data for Eventbrite, Meetup, and Open Tech Calendar. The more events we have, the better we can understand where innovation is happening. Historic data is currently only available from Open Tech Calendar and it requires a manual request process.
- Event attendance data from Eventbrite and Meetup. This would help us understand the strength of links between places. We caution that such information would need to be dealt with carefully to respect users' privacy. As part of this project we wrote more of our thoughts

[4] thedatacity.com/blog/#PersonalData

on the privacy and personal data consideration for The Data City and such approaches could work here.

One idea for sharing such data safely is provided in how Eventbrite and Meetup suggest events currently. The smart radius and auto-radius features in their APIs suggest nearby events when searching in a particular place. We believe this is based on event co-attendance and shows that such data can be shared without compromising privacy.

Patents data

- More data from the UK Patent Office (UKPO) (Intellectual Property Office). The UK makes a lot of historic patent data available under an open license. It is quite detailed but it could be even better. By including machine-readable abstracts or full text of the patents we could use machine-learning to classify the industrial sector of the patents better. Currently UKPO data includes the filed address of a patent. Often this is a large company's headquarters, often abroad, and frequently in somewhere like Luxembourg. The invention was unlikely to have occurred in this place. Data on the inventor's address is much better for understanding where an invention took place, and what cluster that knowledge might have fed off and contributed to.
- We should have got access to the European Patent Office during this project but we struggled to understand the licensing and access options. There are a lot of options on the EPO's raw data page and we got confused. Other people who we asked for advice also got confused. Our current understanding is that EPO full text data is available under a closed but reasonable license. You can't create a copy of their download or patent search site, but pretty much anything else is okay. The fee for access is €800 for a one-off backfile (1.9TB, shipped on a hard drive) and €150/year for updates. These costs seem reasonable to cover the cost of providing the data and if our understanding is correct, we will be accessing this data soon.

Scientific papers data

Anne-Wil Harzing is one of the world's leading experts on scientific paper analysis. Her excellent software Public or Perish is widely used by individual researchers and universities to track output and impact. She has written a number of blogs[5] [6] [7] explaining how the extremely expensive old systems like Web of Science and Scopus have been nearly matched by cheap or free alternatives like Microsoft Academic Knowledge and Google Scholar.

Of these two options we use Microsoft Academic Knowledge because it allows bulk querying and analysis in a way that is disallowed by Google Scholar's terms and conditions. In both cases the data is licensed under a restrictive license, although the terms are unclear.

Open data would be fantastic and efforts such as The Initiative for Open Citations[8] are trying to achieve that. We think it is unlikely that such efforts will come close to the quality and coverage of Microsoft Academic Knowledge or Google Scholar any time soon. Given that, a clearer license from both of these companies would help us and others.

[5] https://harzing.com/blog/2017/02/publish-or-perish-realising-google-scholars-potential-to-democratise-citation-analysis
[6] https://harzing.com/blog/2017/02/google-scholar-is-a-serious-alternative-to-web-of-science
[7] https://harzing.com/blog/2017/06/microsoft-academic-is-one-year-old-the-phoenix-is-ready-to-leave-the-nest
[8] https://i4oc.org/

The future: UK Tech Innovation Index and the Data City.

The Data City is constantly improving as we add more data, make more manual corrections, and answer more questions. Although we have taken a cut of the data so that we can release UK Tech Innovation Index 2 this doesn't mean that development has stopped.

On the next page we summarise the improvements in UK Tech Innovation Index 2 over UK Tech Innovation Index 1 and chart the improvements that we expect to see in the coming months. Since The Data City is always on and always improving we are calling this potential new product Tech Innovation Index Live. Instead of a static cut of data, it will be powered by live data. Instead of being restricted to the UK, it will work for the whole world.

THE DATA CITY

Tech Innovation Index development progress and roadmap (past, current, future).

	UK Tech Innovation Index 1.	UK Tech Innovation Index 2.	Tech Innovation Index Live.
Businesses	Density of all "tech" businesses as defined by SIC Code.	Density of businesses in subsectors of tech, classified by machine learning. We only have IoT business data for now and in other fields we use SIC code data.	Density of businesses in subsectors of tech, classified by machine learning, for both pre-defined sectors and custom sectors.
Scientific papers	Search on Microsoft Academic Knowledge API for papers in a selected range of topics from UK research institutions only.	We use a full extract of all research papers and geolocate every institution. Papers are classified by keywords related to the categories.	Full extract of all research papers with geolocated institutions and information on citation count and inter-institution co-publishing.
Events	Eventbrite, Meetup, Open Tech Calendar. UK tech only. Classified by keywords. Searched by city.	Eventbrite, Meetup, Open Tech Calendar – technology events only. More events captured. Classification by improved keyword searching. UK only.	Eventbrite, Meetup, Open Tech Calendar – all event types. Classified by machine learning. All events in the world over a longer period.
Patents	No patent data.	UK patent data, geolocated using the owner's registered address and classified using the IPC code. (removed from public website build to meet deadline)	Full European patent data, geolocated using primary inventor's address and classified using a combination of machine-learning on abstract, full text and IPC code.
Skills	Percentage of adults with a degree, by PUA.	No skills data.	Skills data from TechNorth on demand for skills by subsector of industry/technology and supply of skills by sub-subsector of industry/technology.
Geography /clustering	We use existing geographies (PUAs, NUTS, Local authorities) and assign businesses, scientific papers, and events to those geographies inconsistently.	We use no pre-existing geographies (definitions of cities, city regions etc…). Clustering is by K-means clustering using latitude/longitude and weak extra data on linkages. Number of clusters is set manually (with the assistance of entropy measures). Clusters are category-specific.	We use no pre-existing geographies, instead we use clustering (HDBScan or K-means or X-means) with latitude/longitude and other connectivity data (co-authorship of patents and papers, co-attendence at events, etc…). This is important because the number of clusters and the shape of each cluster will be different for each category.
Categories	Defined by Innovate UK and The Digital Catapult. Overly broad in areas (Health, Creative) and redundant in others (AI and Data).	Defined in collaboration with users. Based on The Industrial Strategy White Paper, 2017.	Defined in collaboration with users. Based on The Industrial Strategy White Paper, 2017, but with the ability to visualise data for any user-defined category within minutes.
Data Collection	One-off collection, manually assembled in Excel. We publish open data.	Scripted data collection, saved to multiple databases, accessed by queries, and assembled in Excel. We publish open data.	Continual data collection saved to a single database and accessed by queries that produce files that directly power our tools. We publish open data.

The Data City Lille

The Data City doesn't rely heavily on formal national statistics so we are able to expand to new countries easily. Language and subtle differences in economic structure are the only large barriers.

We know this because we've tested it. We already have large parts of The Data City working in Ireland and Scotland where national statistics are different to those in England & Wales. Our method is mostly unaffected. We can today provide comparable assessments of industrial strengths and potential for innovation in small niches of technology in Dublin, Belfast, Glasgow, Cardiff, and Leeds.

This January we are expanding to France. We'll be starting in Leeds' twin city of Lille, producing a version of The Data City for the Lille City Region. At the same time we'll be updating our version of The Data City for Leeds so that both tools can be presented to both cities at the same time at the beginning of March.

We've chosen France for three big reasons,

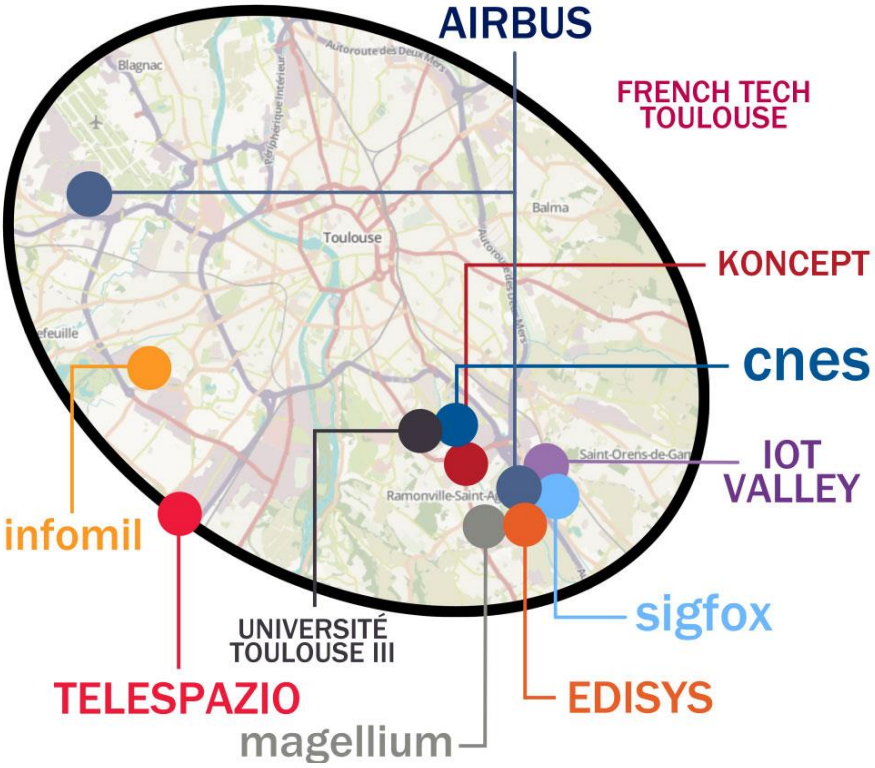
- France's cities and national government have embraced open data, so we can easily access everything that we need to expand our tool. We can publish our results easily too, on great data portals run by our friends at OpenDataSoft.
- French Cities have strong regional governments and business groups, both of which have the power and money to invest in innovation.
- France is one of world's leading countries for artificial intelligence[9]. Places like Station F in Paris and Euratechnologies in Lille host both start-ups and big companies like Microsoft.

We've already done a lot. Our early work has been featured by étalab, the French government's digital services team. Our workflow for analysing scientific papers works without change, since most papers today are published in English. And we've already shared a lot of the additional methods we've developed to use French datasets and compare them with English & Welsh ones.

Our work on IoT UK Nation provides a fantastic basis on which to compare the UK and France. Famously, France is strong in The of Internet of Things, with nearly a third of exhibitors at 2017 CES (and even more in 2018) in Las Vegas coming from France. Our initial work suggests that this excellence is widely spread and often deeply linked with local industries.

In Rennes, IoT businesses are linked with Orange and Télécom Bretagne, a leading university and research institute. In Toulouse, IoT businesses are linked to Airbus and Ariane, world-leading aviation and airspace companies. What we see in France is similar to what we've found in the UK. One example is The West Midlands, where the automotive industry plays host to world-leading companies in IoT that slip beneath the radar of many policy experts and investors.

[9] https://www.wavestone.com/en/insight/deep-tech-global-investor-survey-2017/



IoT businesses in Toulouse. We are building the same, but better, for Lille.

We still have lots of work to do in France. For a start all of our machine-learning needs translating. Internet of Things is probably Objets Connectés but we need to teach a machine that, and we'll need French tech people to help us.

Our next update will be in March.