Data Labs: A new approach to impact evaluation

An update from NPC

Rosario Piazza, Dan Corry, James Noble, Sally Bagwell

August 2019
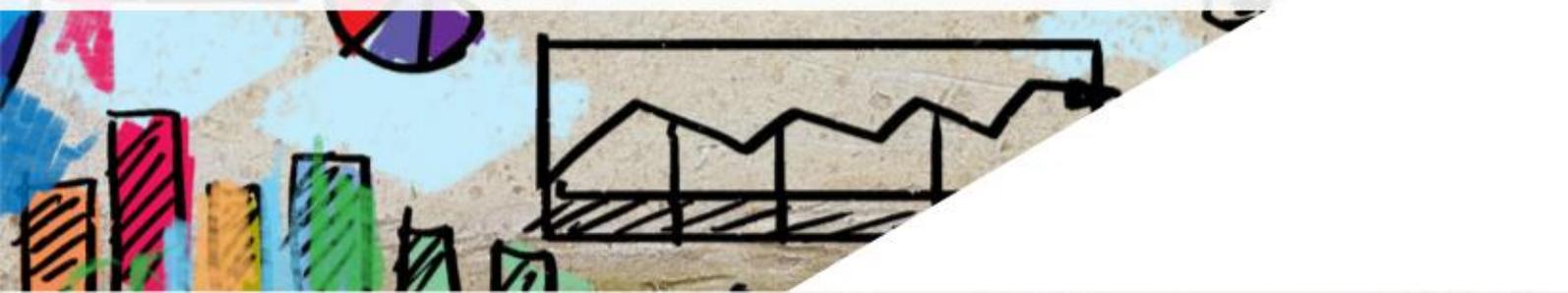
Contents

1. Why we got going on Data Labs

At NPC we want to make the social sector as impactful as it can be so it can help more people and do it better. A big part of this is to help and encourage charities, funders and other social organisations to allocate their scarce resources to fulfil their mission based on a credible assessment of what works and what does not.

We call this impact measurement. There are many barriers to doing this of course, such as the capacity and skills within the sector, and a mindset that can often put heart way above head.

More practically however, one of the key barriers to better impact measurement across the sector is the genuine difficulty of collecting longitudinal data to track what happens to those you've worked with, and with sample sizes big enough to obtain statistical significance. This is compounded by the further trouble of finding a reasonably robust control group so you can make some real assessment as to whether your actions were responsible for any of the changes that occurred.

What on earth could we do to help with this?

It was then we had a brainwave. Could we not attempt to address this problem by harnessing all the administrative data the government already collects for other reasons? This led us to developing the idea of a Data Lab, the basic idea of which is explained in the box on page 3 (using health as an example)[1]. This paper explains our journey and progress so far and what we need to do next.

For Data Labs to be of real use to the social sector, and not just a plaything for academics or major organisations with big data and evaluation teams, we knew they would have to be:

1. **Useful and truthful,** with adequate identifiers for those people charities had worked with so as to allow them to link up with government administrative data.

2. **Straightforward, simple and easy to use,** without requiring charity staff to have detailed knowledge of statistical techniques.

---

[1] You may find reference to other similar approaches elsewhere, with common names including Data Labs, Impact Labs, Policy Labs or Big Data Impact Labs. Although similar in principle, they may differ on the detail from our model.

3. **Legal and not flouting any data privacy laws,** it was always our intention that the data would be aggregate and anonymous.

4. **A substantial improvement on where we were**, even though it would never be perfect – it's not a randomised control trial (RCT).

We were confident that if one could construct such a thing then it could be transformational. Data Labs would be a cost effective and simple system that would allow us to begin to understand which interventions run by which organisations worked best.

How a Data Lab works

The basic idea behind a Data Lab and the way it will work is quite straightforward and is illustrated below in the case of health:

IMPACT CYCLE

1. Charity submits individual-level data on the beneficiary and intervention they want to measure

2. Beneficiary data is matched to NHS data about their use of services

3. A matched control group is created

4. Outcomes for the groups are compared

5. Analysis report is published & charity learns about impact of intervention
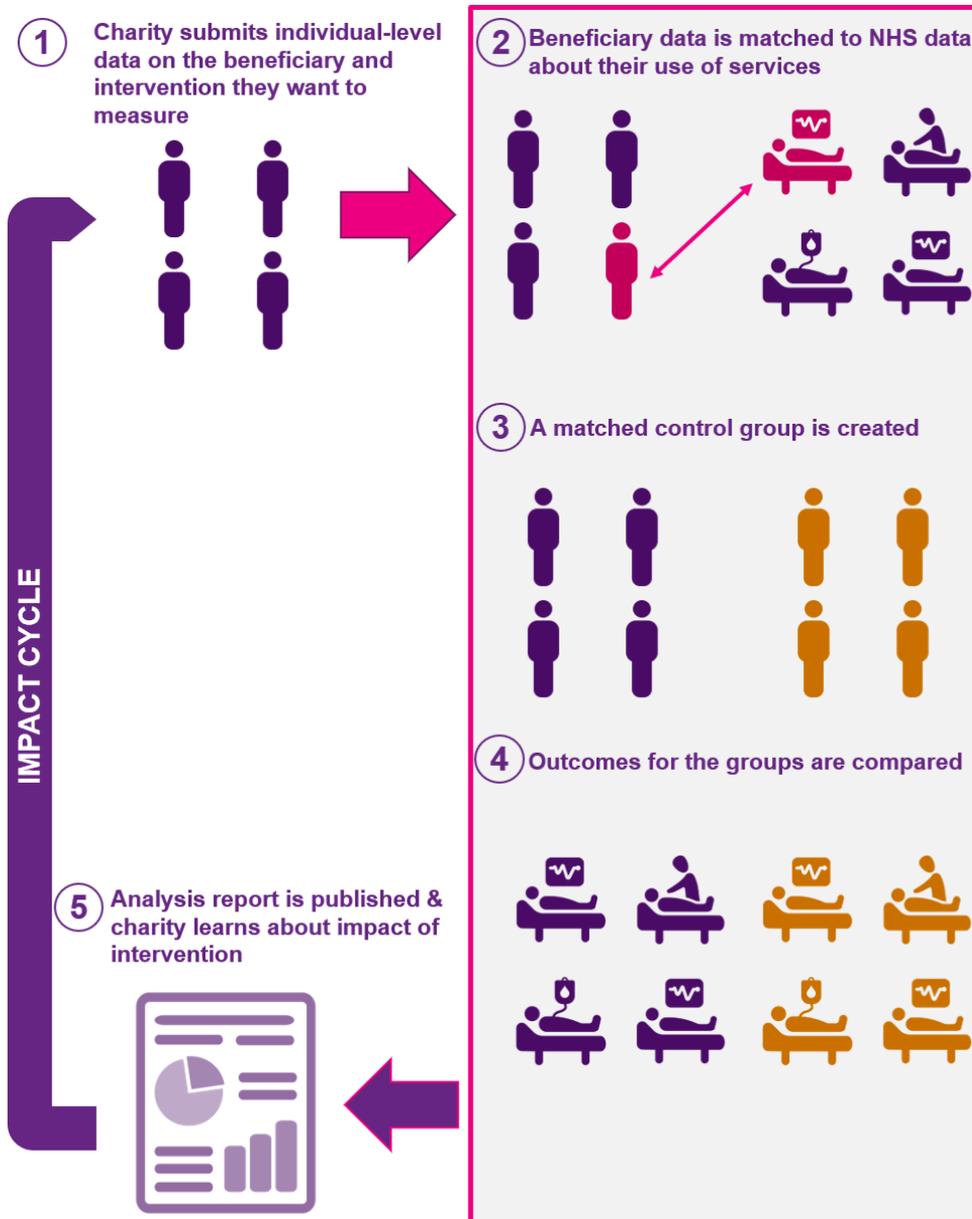
For this work to get going, we were very grateful for funding from the Oak Foundation through their homelessness work, which fitted well with our own emphasis on looking at reoffending as the first place for a Data Lab, given the strong overlaps between the two.

We were always aware that a Data Lab would not be able to solve all problems:

- It cannot tell you definitively about **causation**, just correlation.
- It is not comparing the intervention group against a randomly selected control group. **Motivation can play a role** (such as the individual's choice on whether to take part in an intervention or not), as can other factors, so the results cannot be definitive.
- It tells us in a summative way whether the intervention worked or not, **but not why it worked**, or whether it would have worked just as well with half the 'dosage' (e.g. number of mentoring sessions).

That said, we believe the progress we have made proves we're onto something, so we're keen to expand the model further.

2. Progress on the idea

The key to Data Labs was to get access to data already collected by government for administrative reasons, and then to use it to create a longitudinal series and so help measure impact by doing 'quasi RCTs'. Clearly a prerequisite goal was to get government on board. After all, they hold the data. So, we have spent a lot of time talking to people in various departments across Whitehall.[i]

The tightening of the public purse from the ongoing fiscal climate has spurred a demand for evidence of what works. Various What Works Centres have been set up in recent years, so we believed Data Labs should have been a slam dunk idea for the civil service, at least in principle. Additionally, the academic community were starting to get to grips with the idea of administrative data, with the Economic and Social Research Council setting up several strands of funding[ii].

In practice it has generally not been so easy; enthusiasm has certainly been uneven across departments. We sometimes describe the progress in different areas as like tortoises and hares, although who is what at any time varies. There were of course serious questions to be answered by both ourselves and government if we were to take this forward, including:

- What **privacy and consent rules** allowed with respect to using the data of individuals (albeit in an anonymous and aggregated way).
- **Technical issues** like how best to do create a control group using the statistical method of propensity score matching and how close to a true control group that is.
- Other **conceptual issues**, for example, what if everyone is on a Work Programme? Who and what is the control group if everyone must be on it?
- **Capacity** within the civil service analytical community to actually do these things; the resources needed to set up and run a Data Lab; and the question of who would pay for it and run it, government or some outside body?

It's fair to say however that the conversation and rhetoric have advanced in the nine years[iii] we've been pushing this. The government has now got the idea that it needs to use its own data better, with the narrative now very much pro-data. We're pleased to see more government departments looking at the Data Lab type of approach amongst others to support our wider knowledge as well as their own aims, but a greater sense of urgency would be welcome.

Our goal now has been to ensure that whatever Data Labs emerge are designed to work for charities as well as for the public and private sectors. As the Justice Data Lab shows, while it was pushed by us for the charity sector, it's now being used by and for the benefit of all sectors, public and private too.

Another challenge has been convincing charities and their funders that once up and running they should use the Data Labs to measure and evaluate their work and to further develop their strategies.

3. Justice Data Lab – The first of its kind

The first area we pressed on for Data Labs was to do with prisons and re-offending. Many charities work with prisoners and they think what they do reduces re-offending. **But does it?**

We worked closely with Clinks, the membership body for these charities, as well as with a steering group of academics and sector representatives to develop the initial concept. We surveyed over 230 charities back in 2012 to see how many thought they would use a Data Lab if it existed, and how much capacity whoever ran the Justice Data Lab (JDL) would need.

Testing the water like this mattered because at this early stage we were not sure if the government would run and fund a Data Lab or whether they would simply sanction the use of the data in some way but require administration to be done a by a third body that would need funding. There were many debates on all sides, and we made progress largely due to the amazing support of a handful of senior officials on the analytical and policy side within the Ministry of Justice.

To start with, we assumed that there may be a charge to use the Justice Data Lab, and the need for a business model as to how to run it, so attempts to measure potential usage and demand were very important. In 2012 we published a paper where we tried to asses demand for such a product.

At the outset there was naturally a lot of discussion about which outcomes should be measured. The easy thing was simply to look at whether prisoners reoffended or not after a set time period. But charities rightly pointed out that offenders rarely go from offending to never reoffending; the journey was more likely to be through less frequent reoffending. So, what charities really wanted to know more about was how frequently someone reoffended, not just a simple yes/no. They were frustrated that the Justice Data Lab did not kick off like that, but to make progress it was judged that the project had to start simple before more nuanced outcomes could be added.

Equally, the control group had to be as alike to the intervention group as possible. At first it was fairly crude, although still better than almost anything to have come before. Over time more data has been added, such as mental health problems and substance misuse, to try to make the control group better (via propensity score matching).

A lesson we have learnt is that it is better to get going first to test the model, show the approach works and produces sensible and useful results and then to make it more sophisticated as we go along, rather than to wait until we have perfection before moving forward. By taking it step by step, we made progress. Now things can be even more complex. For example, the Justice Data Lab now has bespoke analysis on different levels or types of intervention within a charity and the effect they have on employment outcomes as well as justice outcomes.

So now we had a Data Lab that worked for individual charities and which was beginning to allow them to know if what they were doing worked or not. Furthermore, it was bringing in common or shared metrics, a rare thing in our sector, which enabled comparisons. Charities were now able to learn from others doing similar things. In recognition of all of this, the Justice Data Lab was given the 2014 Excellence Award in Official Statistics by the Royal Statistical Society.

Once the Justice Data Lab was established and it was decided it would be free to use, a key goal was to explain to the charity sector what it was and to persuade them to use it. Our 2014 paper with Clinks made the case when it had become clearer how it would work.

By July 2019, 233 analyses had been conducted by the Justice Data Lab, around 40 per year. Of these, 58 had shown a statistically significant positive effect on reducing the headline reoffending rate, 14 had shown a negative effect (i.e. more people reoffended) and 161 had been statistically inconclusive (mainly due to sample size). The data lab has been used 98 times by the voluntary sector, 73 times by the public sector and 62 times by other sectors. However, because some of these analyses are repeats by the same organisation, the total number of organisations using the Justice Data Lab is lower. Alongside public and private organisations, around 40 charities have used it since its launch, lower than we might have expected but still a major step forward.

We think the main reasons for low take up are that the justice sector has many micro charities who haven't had enough service users to use it, or that many charities are poor on the identifiers that they have on the people they work with. There is also the fact that results are published by the Ministry of Justice on their website, which may have put off some charities who might be too nervous about getting a disappointing result to really want to know if their intervention works or not. We think the underlying reason for some people not using the data lab is down to misunderstanding or misinterpretation of how it works, which we recently tried to address with a set of FAQs.[iv]

Aside from their usefulness to individual organisations and charities, the results in aggregate are already starting to show important things. On one hand they show that the average impact (when positive) is much lower than 'usual' impact studies. This is helping to reduce expectations that had become unrealistic due to the use of crude Social Return on Investment (SROI) calculations and other methodologies with data that was not strong enough and often ended up making exaggerated suggestions, like weekly art classes reducing reoffending by 60% or so.

The results so far show certain classes of interventions work better than others. For instance, education-based initiatives seem overall to give stronger results. As you'd expect, they suggest the best results come when you work with people who are less likely to reoffend. Conversely, those projects whose users have a likely reoffending rate above 50% do appear to struggle; there even seems to be a higher chance of a negative result.

We have also learned a lot that we can apply to further work on Data Labs:

- Many results are inconclusive due to small sample sizes (we did a guide as to what that means[v]).
- Ultimately as it is not an RCT we cannot exclude motivational differences between those who take up the intervention and other people, which may at times influence the results.

We have also hosted an annual meeting of our Justice Data Lab expert advisory group, including Ministry of Justice analysts, academics, charities and funders to consider progress, and where to go next. It's certainly a rather unique gathering.

Our next steps

The Justice Data Lab is alive and active, and it is a great credit to the Ministry of Justice that they have kept it running even as they have faced major budget pressures. Our next steps are:

1. To continue to add to what the Justice Data Lab can do, such as adding in detail about employment outcomes (already happening).
2. To get more organisations using it, especially charities[vi].
3. To see more meta-analysis of the results conducted and published. We are pleased that the first of these, by Dr Sandra von Paris of Middlesex University, should be published later this year.

4. Employment Data Lab – Moving to the pilot phase

We think an obvious place to have a Data Lab is in employment. Here the case for longitudinal follow up is clear, and there are good longitudinal data sets which academics have used for years. There are also many organisations working in this area without definitely knowing whether and to what extent they are increasing the employment rate of those they aim to help beyond what might have happened anyway.

We've seen some keenness from the Department for Work and Pensions, and over the years several secretaries of state and ministers have expressed support, including Ian Duncan Smith. It has been a long road as we all tried to overcome legal issues (the legal basis for using the data is different for different departments, so the Ministry of Justice approach could not just be transferred over unchanged); technical issues to do with propensity score matching; and the capacity of the department to assign staff to the project given other priorities like Universal Credit.

We have worked very closely with the team at the Employment Related Services Association on this portfolio and have benefited from a great steering/advisory group. We recently made substantial and exciting progress as ministerial decisions led to the DWP allocating some resources for this and appointing some dedicated analytical capacity. Pilots are now being planned and capacity put in place.

One of the big issues will be around what outcomes to measure, especially in the first wave. An easy first step is to assess whether a person is in employment or not after a fixed period. But we will also want to know how long the job was kept for, what type of job it is, what the pay was, whether there was any progression and promotion within it and so on. We will also want to see what other data sets can be linked to it, not least to get the control group better (through propensity score matching).

There was an issue as to what effect the comparator group was when everybody with certain characteristics had to be on the Work Programme. It was alright in this period to look at the effectiveness of different schemes within the programme, but it was hard to compare with somebody not on it at all.

Currently there are no major national employment programmes, and we expect that the need now is to understand the effectiveness of a whole range of programmes, some commissioned and run by a combined authority or mayoral area, some by councils, some simply run by a charity with some grant funding.

Our next steps:

We want to get this Employment Data Lab up and running as soon as possible. This means supporting the DWP to move on this; reviewing the pilots and ironing out any issues; checking in particular that what is established works well for charities; agree outcomes to measure; get it working and build demand.

Learning from our experience with the Justice Data Lab, we are committed to supporting the development of the Employment Data Lab by bringing together charitable organisations who will be involved in the pilot, and bridging gaps between statutory and voluntary organisations. We will also create an advisory group of people from the sector and a range of professional backgrounds and expertise to assess, discuss and review the development of the Employment Data Lab.

5. Education – Slower progress but things happening

Another place the Data Labs model can have leverage is with school's data. Here the case for longitudinal follow up is clear. There is a longitudinal data set (the National Pupil Database) and there are many organisations who say they help improve outcomes, especially for attainment, but do not really know if they do or not.

In the What Works space, there is already the Education Endowment Foundation who mainly try to use a very high-end evaluation technique known as randomised control trials (RCT). The Education Endowment Foundation is well funded by government and philanthropists and carries out or funds many RCTs. However, it is mainly geared up for big educational programmes rather than charities. Its methods are a long way from the easy to use model that we are after, where a charity just gives in its identifiers of those it works with and is handed back an easy to understand assessment of its impact. In principle the National Pupil Database is available to any approved researcher (indeed we have used it ourselves to help some schools assess their impact[vii]) but in practice we've found it tricky to access and use, which means few charities are equipped to take advantage.

Ideally then we want an Education Data Lab akin to the Justice Data Lab. It needs to operate at the more advanced 'pupil level' in which individual participants in programmes are matched to and studied against a cohort of similar pupils. The alternative 'school level' approach, which simply compares schools in which interventions have been or not been delivered, is much easier but also much less useful. It has limited use except for the very largest charities that work across many schools.

Both government and the Education Endowment Fund have perhaps been more attracted by the fact that they would like a relatively easy way of assessing whether an intervention is showing promise, before putting it through an expensive and time-consuming RCT. **It's clear then that we both want something like an Education Data Lab.**

The Education Endowment Fund have been plugging away at this in the face of some significant obstacles, including sensitivity about the use of the National Pupil Database, and recent changes in how that data set can be accessed[ix], as well as the sheer size of the dataset and the expertise needed to use it. They did put out a tender[x] to develop a data lab along these lines which was won by the Fisher Family Trust.[2] But there appears to have been limited progress in the two years since then, other than a study which shows that at the school level the broad approach delivers largely consistent results to the more expensive randomised control trials.[xi]

We see our role here as keeping in close touch with the Education Endowment Fund to ensure progress is made and that whatever emerges is useful for charities working with school-aged children.

Our next steps

NPC have occasional contact with the Education Endowment Fund and have been keen to help wherever we can. Our aim is to ensure that the 'Analytical Service' can meet the needs of the wider children and young people's sector, beyond those organisations interested in or applicable for EEF trials.

---

[2] There is already an "Education Data Lab" in the UK, run by the Fischer Family Trust. However, this is different to NPC model. It focusses on secondary analysis of existing education data sets and can't be used for service evaluations.

6. Health – The biggest prize of all?

Health is often seen as being predominantly about medical interventions. Many charities involved in the medical research world engage with this and mimic the way researchers go about their evaluations. The reality though is that much charity work that can improve health outcomes and the effectiveness of health spending are actually about approaches which are less medical. These include prevention, support to navigate systems, and support to stay at home rather than be admitted to hospital. Such interventions improve quality of life whilst also saving money for the health service.

However, these outcomes are hard to measure. For example, does a charity's work mean someone attends their GP less? Do people charities work with spend less time in a hospital bed due to fewer admissions and quicker exits? Without answers, charities cannot understand what they are achieving. They can't learn or adapt, and they are unable to pitch to the NHS to spend more of its money on these kinds of approaches. Overall it may mean that the NHS underinvests in these sort of approaches that could bring real benefits, whilst overinvesting in well-meaning initiatives that don't necessarily secure results.

A Health Data Lab looks like it could be a very valuable tool for the charity sector, for the NHS and for the government itself. This argument was made in a letter we put together signed by 25 health charity CEOs[xii] as we tried to get the idea up the health agenda. This turns out to be difficult as there are many players in the health world. Who could ask for this and who could deliver it? Was it NHS England, the Department of Health, or Clinical Commissioning Groups and other local bits of the system? This led to us having many discussions with NHS England and especially with NHS Digital and its predecessors. We even had one of our staff work in the Leeds HQ of NHS Digital for several months to produce a draft business plan.

In health we have confronted the usual issues addressed in other Data Labs. But there are extra ones too. Consent and privacy are even more important here, so the system is nervous not least following the experience it had with Care Data[xiii]. We also found it confusing to know who has the data and who has brought it together.

Progress has therefore been slow but there has been some. Back in 2018, the then Secretary of State for Health, Jeremy Hunt, gave a directive establishing a legal basis for a Health Care Analytics Service. Meanwhile the Health Foundation has been doing good work using similar methods to help assess the progress of some of the vanguards coming out of the Five Year Forward View, although this is primarily about organisations within the NHS already (so less an issue of data going 'outside' the sector) and a more explicit focus on system outcomes rather than exclusively individual outcomes.

These analyses are bespoke and tailored to answering the questions most relevant to a given area, so it's not the standardised methodology we aspire to in a Data Lab. Nonetheless, the basic idea of drawing together data to assess the effectiveness of a service is accepted.

Within the charity sector and beyond, there is a major issue of what outcomes data is available and what outcomes data is actually wanted. Is the main aim to see impact in terms of reduced GP and hospital appointments? Not every charity supports 'reduced use of health services' as an outcome they aspire to, as it could be an indication that people are not accessing the service they need, instead of an indication that people don't need the service.

Some other charities want to focus much more on quality and therefore on wellbeing issues, but there is always the limitation of what data is actually collected or available at sufficient scale and in standardised form to provide an effective basis for understanding the effectiveness of any given intervention.

The other problems are what data is held and who has it. The best available national data set is the Hospital Episode Statistics. This is very powerful but only relevant for a subset of charities, namely those who work with cohorts of people who have had hospital-based interventions, for example a 'home from hospital' support service.

The sort of data contained in the Hospital Episode Statistics, according to NHS Digital, includes:

- Clinical information about diagnoses and operations.
- Patient information, such as age group, gender and ethnicity.
- Administrative information, such as dates and methods of admission and discharge.
- Geographical information such as where patients are treated and the area where they live.

Use of primary care data (trends in visits to GP) is likely to be better for most charities. The problem is this data is held only at a local level at present and there is currently no mechanism or proposal for it to be aggregated at a national scale, which would give an opportunity to make this part of a national data lab.

We have looked at whether one can do more by working at a local level. Indeed there have been local attempts at things that are a bit like what the Data Lab does (not the easy to use approach but the basic methodology). We have been investigating this approach and it could offer a valuable service, but we feel it is a second-best option which will run into challenges where the geographic boundaries of local authorities, primary care networks and the hospital's area do not neatly overlap. Local Data Labs would also lack the efficiency and standardisation of a national approach, which would make it difficult for charities and other organisations who work on a regional or national scale.

Our next steps:

We want to push the idea of a Data Lab up the agenda for NHS Digital and NHS England. If successful we hope to work with partners in the charity sector to define outcomes and pilot the service. In the meantime, we continue to learn about interesting examples of organisations using health data in powerful ways, and we are exploring what we can do at a local or regional level that might be able to include primary care data.

---

6. International – A growing movement

The potential for government administrative data to be harnessed by the social sector is not just being explored in the UK. Others around the world have picked up on the initial success and promise of the Data Labs model, and NPC is engaged in a number of conversations and partnerships aiming towards replicating the model across the globe:

- **In Canada,** the Transforming the Sector conference convened by Powered by Data identified the use of government administrative data as a high priority for supporting the non-profit sector in its efforts to measure and understand impact. This is now being explored through a multi-stakeholder process, working towards practical implementation in the coming years.

- **In the USA,** NPC partnered with GovLab in a project funded by the Stanford Data for Good Grants programme, to explore the applicability of the Data Labs model in a US context. Despite challenges around data being held at federal, state and county levels, there is great potential in the US Policy Lab model which has been developed in multiple places. Policy Labs currently focus on the use of administrative data to improve policy and public services. We think they could be harnessed to serve a non-profit sector customer base by taking on the Data Lab model.

- **In Israel,** NPC is working in partnership with sister organisation Midot to explore the replication of the Data Labs model.

NPC is in ongoing conversations with others about replication of the model. We are open to further discussions, particularly with governments and those with existing responsibility for administrative data and policy oversight. **If you want to be part of this conversation, get in touch with our Data Lead: Rosario Piazza.**

7. Conclusion – A call to action

Data Labs were always a great idea. Now, through the Justice Data Lab, we have proved it is more than a great idea, it is an idea that works and can unlock new knowledge and learning without anyone having to collect a single extra piece of data.

Surely given the potential everyone should be trying to build one and we salute those who are getting on with it like UCAS even if they are not yet perfect. Our experience suggests though that, despite the compelling logic, these things will rarely exist without pressure on government to do it, and will rarely succeed without the desire of charities and their funders to use the evidence they produce.

Going forwards then, we need:

Charities to:

1. Get involved.
2. Use Data Labs (where they exist); and don't just quote the results selectively!
3. Exert pressure for more Data Labs; join a campaign where they don't.

Funders to:

1. Support and fund more work around Data Labs, including the work NPC is carrying out to pursue this agenda.
2. Get your grantees to be aware of all this and use Data labs where they exist.

Government to:

1. Walk the talk.

2. Instruct every Whitehall department to start setting up a data lab or explain publicly why it is a bad idea for them[xiv].

NPCs involvement with Data Labs and our need for funding

NPC approached the issue of Data Labs with the aim of helping charities understand and act upon the impact they are creating. As it has turned out, Data Labs are very useful for a whole host of organisations only some of which are in the social sector.

It has been a transformational programme. The costs are very small given the massive benefits of getting better allocation of public, philanthropic and private resource.

We are very proud of what we have achieved. We need to keep campaigning, keep helping the charities understand what is happening, keep promoting the idea so that when we have data labs they are used.

But we are not funded to do this work anymore so if you are interested in supporting this work, please do get in touch.

---

[i] Gyaten, T. (2017), *How to create an Impact Data Lab*. See https://www.thinknpc.org/resource-hub/how-to-create-an-impact-data-lab/

[ii] Administrative Data for Public Good. See https://adruk.org

[iii]Rickey, B. (2012). Unlocking Offending Data: How Access to Offending Data Could Help Charities Improve Outcomes for Offenders. See https://www.thinknpc.org/resource-hub/unlocking-offending-data/ and *Unlock Prison Data* (2010) https://www.thinknpc.org/blog/unlock-prison-data/

[iv] Noble, J. (2019). *Justice Data Lab: FAQ Update*. See https://www.thinknpc.org/resource-hub/justice-data-lab-faq-update/

[v] Ibis

[vi] Noble, J. (2019). *A Rare Opportunity to Get Data on the Impact of Interventions of Reoffending*. See https://www.clinks.org/community/blog-posts/rare-opportunity-get-data-impact-interventions-reoffending

[vii] Bertram, C., Meierkord, A., Day, L. (2018). *The Cadet Experience: Understanding Cadet Outcomes*. See https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/689698/The_cadet_experience-understanding_cadet_outcomes.pdf

[ix] Humpherson, E. (2018). *National Pupil Database Update*. See https://www.statisticsauthority.gov.uk/correspondence/national-pupil-database-2/

[x] Education Endowment Foundation (EEF) Education Data Service Pilot. See https://educationendowmentfoundation.org.uk/public/files/Education_Data_Service_pilot_-_Invitation_to_Tender.pdf

[xi] Weidman, B. (2017) *Lurking Inferential Monsters: Exploring Selection Bias in School-Based Interventions*. See https://slideplayer.com/slide/12930428/ /

[xii] Boswell, K. (2017). *Health Data Lab: An Initiative Whose Time Has Come*. See https://www.thinknpc.org/blog/health-data-lab-an-initiative-whose-time-has-come/

[xiii] Triggle, N. (2014). *Care.data: How Did It Go so Wrong?* See https://www.bbc.co.uk/news/health-26259101

[xiv] NPC's Manifesto: A Vision for Change. See https://www.thinknpc.org/resource-hub/npcs-manifesto-2015-a-vision-for-change/