

HOW TO CREATE AN IMPACT DATA LAB



Tracey Gyateng

November 2017

NPC's (New Philanthropy Capital) mission is to support not-for-profit organisations to achieve the greatest impact. We think it is essential that charities understand the outcomes of their services to establish and improve their impact.

Since 2011, NPC has researched and supported the development of Impact Data Labs (which we refer to as 'Data Labs') within the UK. This model aims to unlock government administrative data to enable organisations to understand the impact of their services on the people who use them.

This short guide sets out the key things we have learned while setting up the Ministry of Justice [Justice Data Lab](#), and working to develop Data Labs within education, employment, health and substance misuse.

Our work to develop Data Labs has been focused in the UK. This paper presents the core factors that have enabled us to advocate for Data Labs here. However, since there is international interest in the Data Labs model, we hope these insights will prove useful for anyone seeking to use this model.

What is a Data Lab?

Charities have a need to evaluate their programmes to assess if their services are operating as expected and to provide evidence to service users, donors, and funders of their work. Good evaluation requires robust data, but many charities struggle to access or acquire such data. Few also have the skills needed to analyse these complex datasets even if they were to be made available to them.

Meanwhile, governments routinely collect administrative data from services that are offered to the public. This administrative data is increasingly recognised as being of value¹, especially if it is opened up for other organisations to make use of. It is in government's (and society's) interest as key purchasers and deliverers of social interventions to support evaluations of these services. Understanding which programmes are effective in addressing social issues would support effective decision-making for allocating scarce resources. However, administrative data about individuals is personal, and highly sensitive, and must be kept secure.

That is where Data Labs come in. There are a variety of models for a 'Data Lab'—from services that enable researchers to access de-identified data within a secure setting, to services where bespoke research is conducted. This paper focuses on the Data Lab model NPC has used, which is a standardised evaluation service provided by government departments or government approved agencies. The model enables providers of services to measure their impact on known social outcomes, such as reoffending, through use of government administrative data sets.

¹ Cabinet Office (2012) [Open data: Unleashing the potential](#). Cabinet Office.

An 'impact' Data Lab

NPC has been advocating for a Data Lab model that aims to unlock government administrative data safely and securely, enabling organisations to understand their impact of their services on the people who use them.

The Data Labs NPC has worked on have the following characteristics:

- Not-for-profit organisations can access government-held data concerning a group of service users either directly or through a government or third-party service.
- A comparison group can be established either through statistical matching, or by drawing on a previous process of random assignment.
- The impact of a not-for-profit's service can be presented as a comparison of treated versus non-treated groups at an aggregate level—as a group, rather than as individuals.
- Results of impact measurement is shared across the sector to build a body of evidence for what works to achieve particular outcomes.

What value does a Data Lab bring?

A Data Lab addresses the need for impact evaluation by providing a high quality, statistically robust and routine evaluation service. It brings value by:

- removing the barriers for all service providers—not just charities—to evaluate their services
- preventing duplication of effort by giving service providers access to data that already exists rather than collecting their own
- expanding the evidence base of effective interventions, supporting better commissioning, and better service provision

Table 1 summarises the benefits of a Data Lab for each potential stake holder:

Table 1: Benefits by audience

Benefits for organisations using the Data Lab	Benefits for government
<ul style="list-style-type: none"> • clear route to analysis of government-held data • accessible, tailored and useful reports • a more robust way to measure impact • improved understanding of the outcomes achieved for beneficiaries • insight into what works and what doesn't—informing service development and delivery • ability to demonstrate impact and respond to commissioners' requirements 	<ul style="list-style-type: none"> • provide accessible and tailored analysis through a secure mechanism • increased evidence base of effective interventions—can focus on what works and save money overall • opportunity to develop staff skills to produce complex statistical analysis • route to engage and collaborate with the charity sector and learn from their work and expertise • cost-effective—offering a cheaper way to conduct routine analysis than commissioning separate evaluations

Benefits for policymakers, commissioners and researchers	Benefits for beneficiaries
<ul style="list-style-type: none"> • more evidence to inform commissioning and policy developments • routine access to high-quality, low-cost evaluations of social policies and interventions • increased body of evidence—can be used for meta-analyses to better understand what works • enables shift from quantitative data collection to data analysis, synthesis, interpretation and qualitative research 	<ul style="list-style-type: none"> • improved outcomes as interventions are more effective (informed also by other sources of research and impact measurement eg, qualitative research)

What is needed to set up a Data Lab?

In our time advocating for Data Labs to be set up in various government departments, we have identified the most important things required to aid this process. These are:

- **Environment:** A broader culture that supports impact measurement.
- **A leading advocate:** A Data Lab project team, and strong networks.
- **Data:** Individual person-level administrative data.
- **Customers:** Identifiable organisations that would use the Data Lab.
- **Data owners:** Engaged government civil servants.
- **Legal and ethical governance:** A legal framework for sharing data.
- **Funding:** Resource and support for driving the set-up of the Data Lab.

We talk through each one of these in the following pages.

Environment

NPC has been influential in the UK in developing an environment where charities view evaluation as the norm. Fifteen years ago, few charities would have considered measuring their impact. Yet funding pressures—in the light of tightening public sector expenditure—has been an important driver for increasing demand.

Evaluation take-up has also increased as the ease and availability of data due to digital technology has increased. This has been driven by the private sector, with the largest companies in the world being digital- and data- led, who have driven down the cost of IT, and quickly embraced by governments. This has created an expectation that all organisations should be making better use of data.

The UK government has been a world leader in providing ‘open data’—data that is non-personal and can be accessed and used by anyone. This has created an environment in which many stakeholders, especially charities and governments see value in using data to build and improve services.

Charities need to evaluate services + government & funders need to allocate scarce resources effectively + data available + IT infrastructure costs low =

The right environment for a Data Lab

A leading advocate

Whilst a conducive environment is needed, Data Labs do not then just manifest by themselves. It takes an organisation to: lead the development of the service; work with potential users and data owners to define the service; and support both parties to make use of and advertise the service. (This organisation will also need funding—see section on funding below.) This is the role NPC has played.

A Data Labs is a rigorous evaluation service. It requires expertise in the social sector, evaluation, statistical methodology, and community relationship building, which is difficult to find in one organisation. Partnering with social sector umbrella groups, and expert advisors is strategically important. NPC partnered with [Clinks](#), the charity infrastructure body for charity and voluntary organisations that work with offenders for criminal justice work to support NPC's work on the Justice Data Lab. In developing an employment Data Lab, NPC has partnered with [ERSA](#), the employment infrastructure that provides employment support body for all organisations (private, public and charitable).

Working with experts

Throughout the Justice Data Lab, NPC has convened an expert panel with representation from the Ministry of Justice, charities, funders and academia.

NPC's overall Data Lab programme has been overseen by an expert [advisory panel](#) comprising of sector leaders in the areas of homelessness, employment, young people, economics, health, charity data analysis.

Data

A key requirement for a Data Lab is access to large-scale administrative datasets that contain personal identifiable data (eg, name, date of birth, and gender) that enables it to be linked to the data that a charity or other service-provider may hold. Unique identifiers are often collected as part of the administrative process, such as a Police National Computer (PNC) number. This strengthens linking between service users and their administrative data. The data should also include the following features:

1. The data should record an **outcome or event**. For example:
 - whether people have been convicted of a crime
 - abstinence from drugs or completion of drug treatment programmes
 - health, or usage of health services
 - attendance and/or attainment in education
 - employment, income, benefits receipt
 - other variables that reflect specific policy areas (eg, levels of pension saving, council tax payment)
2. The data should contain, or can be linked to data that contains demographic and contextual variables that can be used to measure the likelihood of receiving the intervention. These enable statisticians to derive a 'counterfactual group' which mimics randomisation (also known as quasi-experimental design). This is used to understand what would have happened to the group if they *had not* received the intervention. The counterfactual will be a group of people who have been statistically matched, and share similar

characteristics to people who did receive the service, with the important difference being that they did not receive the service.²

3. The data source is comprehensive enough to ensure that all—or nearly all—relevant individuals can be identified, eg, in measuring re-offending, the dataset should record all re-offending offences. In some countries, offences are recorded across local and national government. Unless these datasets can be combined, neither database alone would be sufficient in tracing an individual re-offending outcome.
4. The data source should be reliable enough to produce official national statistics that can be checked by a relevant statistical authority.

Setting the Data Lab strategy

Following implementation of the Justice Data Lab, NPC investigated other core datasets relevant to homelessness, which was the focus of the project funder, the Oak Foundation. We made an that the other datasets that would be pursued for Data Labs would be: education, employment, health and substance misuse—areas relevant to those experiencing homelessness. We outline this process in [Creating a Data Lab](#)³.

Customers

There is a current movement for governments to be open with their data, and Data Labs capitalise on this. But providing a service without a customer base is a waste of resources. There needs to be a market: organisations willing to measure their impact using the Data Labs model.

In the UK there has been a long and continual drive to build the market for evaluation practice to become an expected norm across the sector. Charities have increasingly become aware of the need to evaluate their services to inform service delivery and provide evidence of impact. Additional stimulus has been provided by the increase in [outcomes based commissioning](#)⁴ and a small, but growing market for [social impact bonds](#)⁵—which link funding to achieving impact. This new acceptance for outcomes measurement has meant a gradual receptiveness by charities and funders to solutions which provide easy access to analysis of impact. This is also relevant to public and private sector organisations, many of which work to measurement/performance frameworks or KPIs. For example, local authorities are required to consider how delivered public health services meet the Department of Health's [Public Health Outcomes Framework](#)⁶.

² In the absence of contextual variables, a less statistically robust form of Data Lab might be possible, such as comparing before and after results over a set period.

³ Gyateng, T., Pritchard, D., and de Las Casas, L. (2013) *Creating a Data Lab*. New Philanthropy Capital

⁴ See <http://golab.bsg.ox.ac.uk/about>

⁵ See https://data.gov.uk/sib_knowledge_box/home

⁶ Department of Health (2016) [Public Health Outcomes Framework 2016 to 2019](#).

Identifying a customer base

Charities: NPC investigated how charities accessed reoffending data, finding that charities used a range of sources, with none providing stable success. After considering different models, NPC advocated for a Data Lab that would support charities in evaluating re-offending, and surveyed the charity and voluntary criminal justice community (see [Unlocking offending data](#)⁷). The survey indicated that there was appetite from the sector for a Justice Data Lab.

Private and public-sector organisations: Private and public sector organisations have made use of the Justice Data Lab. This was initially surprising as it was assumed that public sector organisations could already access this data and conduct the analysis, and that private sector organisations would have the funds to commission independent evaluations. However, access to statisticians, and to data can be just as difficult for public sector organisations as it is to charities. Meanwhile some private sector organisations are run for social purposes and are committed to developing an evidence base of effective interventions. This reason is also shared by charities and public sector organisations.

In setting up a Data Lab, organisations that can champion the service need to be identified. They will need to agree to pilot the Data Lab, engage in developing the service for the general market, and advocate for the service with peers. These organisations need to have personal data on the individuals they have worked with (the Justice Data Lab requires data on a minimum of 60 people) so that data can be linked to the Data Lab. This data needs to be reliable, accurate and in an electronic format. Not all charities will, or should use a Data Lab. Data Labs are best used for medium to large organisations who have provided a consistent programme. If the programme is new, and still being embedded, it is likely that the results of a Data Lab, which is a retrospective evaluation, will not reflect current or future practices, nor will it reflect the true potential of the programme.

Finally, organisations need to be open to testing their work. This is essential as in this model there is a compulsory requirement for all evaluation results to be published publicly. Openly publishing reports is important for building sector knowledge of effective programmes. Yet this acts as a disincentive to many organisations that are concerned that receiving an insignificant or negative result may impact on their funding. Organisations that have used the Justice Data Lab have been a mixture of beneficiary-focused organisations unafraid to publicly test their programmes, or they have been a funder submitting the data—for example the National Offender Management Service (NOMS) evaluated an [employment programme](#) through the Justice Data Lab.

Data owners

In most cases, a government administrative body will be the data controller of the dataset of interest. To develop a Data Lab, NPC has found that the data controller either needs to be a visionary: someone who is interested in making use of data for the benefit of society (or can be persuaded as to the merits). We have found that engaging, and getting 'buy in' from a variety of data controllers at differing levels of seniority has been essential. Mapping out these key players will be an essential exercise. Broadly, it is important to have engaged:

- **Senior-level officials.** Senior level support is needed to a) ensure resources are allocated to setting up a Data Lab; and b) drive forward a strategy for implementation. Two types of officials are crucial, an official that is the head of policy and programme management, and an official that is head of research and statistics.

⁷ Rickey, B., and Pritchard, D. (2012) *Unlocking offending data: Charities' views on how access to offending data could help them improve outcomes for offenders*. New Philanthropy Capital.

- **Senior analysts.** Analysts need to be enthusiastic, both in lobbying internally and proactively working to resolve technical data sharing and analytical issues.
- **Political support.** Politicians can be instrumental in pushing for Data Labs. At the very least they need to sanction it. Gaining their approval is helpful in supporting analysts and senior level officials to set up a Data Lab. Careful consideration needs to be placed as to when to involve a Minister so that senior level officials and analysts do not feel undermined by a Minister telling them what to do. However, NPC has found a directive from a Minister has been useful at times where there is little, or slow engagement from senior level officials and analysts.

Having identified the staff to be targeted, it is important that the Data Lab development organisation(s) can provide a strong and compelling case for a Data Lab. Incentives will vary across contexts, but NPC has developed the following arguments that are applicable in most cases:

- **Transparency/open data:** Many public authorities have made commitments to transparency. Given that Data Lab reports are made public, Data Labs are a convenient way to deliver this objective.
- **Maintaining privacy and confidentiality of individuals:** A Data Lab makes use of personal and sensitive data, whilst protecting the privacy and confidentiality of service users. This is a core consideration for information governance teams.
- **Public good:** This data will help organisations to understand whether their services are making a difference and how to improve them, ultimately benefitting society.
- **Good PR for government departments:** The Data Lab is therefore a friendly mechanism for government departments to engage with the social sector, providing good PR. The Justice Data Lab has won several awards over the years of service, for example, Excellence in Official Statistics from the Royal Statistical Society in 2014. The Ministry of Justice is viewed as a digital and data leader across government departments.
- **Efficiency:** Data Labs help organisations to cut down on duplicative and difficult to collect longitudinal data. It also enables a data processor (often the government department) to specialise and speed up the process for evaluating key outcomes—a skill that can be utilised across a government department. A related argument is that it can be used to pilot or test programmes in advance of a more expensive randomised control trial.
- **Commissioning scarce resources:** Where public or charitable money is funding social programmes, there is a strong argument that authorities should allow them to be tested through a Data Lab to ensure they are working as intended. When collated with costs for running the service, it can aid in making decisions on value for money.
- **Political expediency:** The UK Justice Data Lab was partly set up due to the minister at the time wanting the department to be engaging with the voluntary sector in a period of major reforms. In addition, the minister wanted to support government's transparency work of opening up data. Data Labs provided a convenient solution.

Legal and ethical data sharing

There needs to be a legal framework that permits the transfer of data from customer organisations to the administrative data processors. Ideally, this framework would not require the consent of individual service users as data is analysed at an aggregate level. Legal assurances will be needed from the data processor to ensure personal data submitted by organisations is not used for any other purpose than to provide statistics.

If the data processor requires consent from service users to share data, then customer organisations will need to collect this. This may require amending current consent forms, meaning that an organisation will not be ready to use a Data Lab for several years until a database of consenting service users has been acquired. This is not desirable, as it prevents organisations from using a Data Lab for existing data. However, in line with data protection legislation, service users should be informed of what will happen to their data. Organisations should

communicate that data collected is used for research, and name likely data processor on the privacy policies they provide their clients.

NPC considers that submitting data to Data Labs is ethical because the data is aggregated and therefore cannot be—and is not designed to be—used to single out or target individual service users. The point of it is to use it to assess services, not people.

Privacy Impact Assessment

The Ministry of Justice published its [Privacy Impact Assessment](#)⁸ for the Justice Data Lab. The report explains how data can be legally shared from organisations to the Ministry of Justice.

Funding

The Data Lab model is not particularly expensive. Part of its appeal is that it provides access to longitudinal and counterfactual data at a small fraction of what it would otherwise cost. The standardisation of the service creates efficiency savings—it is cheaper for organisations to use a Data Lab than to commission impact evaluations from various sources. It can be argued that identification of services that are proven to have statistically significant impact, if replicated, would produce savings for society that would justify the cost of a Data Lab. The UK Justice Data Lab began operations as a pilot with three staff, a senior statistician and two junior statisticians. No additional infrastructure was required as the Lab utilised resources that were already available within the Ministry of Justice and accessed free statistical software (R). Costs for a Data Lab would differ according to sector and current state of infrastructure, but staff costs will be the significant factor.

Customers could be asked to pay or contribute to costs. This is likely to be successful if customer organisations are large enough, or if their own funders are willing to invest in this (possibly instead of other evaluation activities).

Funding is also required for the lead Data Lab team to advocate and support the development of Data Labs.

Funding for infrastructure type projects can be difficult to obtain. NPC began facilitation and campaigning for Data Labs without project funding. It would not have been possible to develop Data Labs without receipt of a four-year grant by [the Oak Foundation](#). The Oak Foundation shared NPC's vision to make impact measurement more accessible to charities, particularly those in the housing and homelessness sector.

NPC continues to advocate for Data Labs even though core funding has ended; the need for the service still exists and piloting of Data Labs in education and employment is ongoing. NPC has received grant funding from [Ingeus](#) to continue efforts for an employment Data Lab, and in partnership with [The GovLab](#) a [Digital Impact Grant](#)⁹ from Stanford Centre on Philanthropy and Civil Society to explore the potential for Data Labs internationally. This will enable NPC to continue working on Data Labs until early 2018. NPC continues to seek funding to keep campaigning for Data Labs.

⁸ Ministry of Justice. (2013) [Justice Data Lab privacy impact assessment report](#). Ministry of Justice

⁹ See <https://digitalimpact.org/grants/>

What are the common barriers to the process?

Developing Data Labs across sectors has brought along unique and varied barriers to overcome. However, three common barriers have been identified across customers and data owners/processors:

Questions about how people's rights to privacy can be upheld

There is a concern that the use of data will compromise individuals' rights to privacy.

It is essential that legal assurances are given by data processors that personal data is only used for producing statistics, which alleviates concerns that data might be used to target individual's and perhaps adversely affect them. Outputs from the Data Lab are aggregated and are anonymous, which reassures those who initially raise concerns. The UK privacy lobby have been sanguine about the Data Lab model.

Methodological arguments about the use of quasi-experimental designs

Second, there is a methodological argument against identifying a comparison group by using statistical methods and the potential for bias. This is an ongoing debate between evaluators. There are some variables which are difficult to control for within a comparison group, such as an individual's motivation to engage with and benefit from a service. However, the Justice Data Lab has continued to add new variables to its models without much change in the findings. The model can always be refined and improved, and perceived flaws should not detract from attempting the analysis. You just need to be clear about these risks and limitations when results are interpreted.

Concerns that Data Labs will be perceived as a complete evaluation of a service

A third argument that mainly arises from potential customers is that it is not the answer to everything. We agree. A Data Lab provides an important piece of an evaluation process that can be expensive and too complex to attempt to measure alone. The results are powerful to inform judgement on how effective a service is, but additional data is needed to enable a service to form a full view on how to develop the service. For example, to answer questions about efficiency, finance data is needed; to understand the practical running of the service and its perceived quality, qualitative research is essential. Data Labs is an important piece of the jigsaw, and organisations should certainly continue with other evaluation activities.

Summary

This paper has outlined the key factors that NPC has discovered across six years of working design and advocate for Data Labs. In that time, NPC has engaged with hundreds of key stakeholders across charities, government and organisations interested in impact evaluations. It takes time to build up relationships, develop the model and the market, support implementation by data owners and encourage usage. The Justice Data Lab pilot began within two years from initial scoping work. Piloting for education and employment has taken three years, whilst NPC efforts to develop a [Health Data Lab](#) continues. Yet if the factors identified within this paper had been present in education, employment, and health, it is likely that piloting could have taken place sooner.

NPC remains committed to Data Labs primarily because charities, public and private sector organisations continue to highlight a common need to evaluate standard outcomes that they work towards, such as increased employment, and see the benefit of using existing administrative data. When this need for support stops, this is when we will end our campaign for Data Labs.

If you are thinking about developing a Data Lab do visit our website www.NPCdatalabs.org and get in touch with Tracey.Gyateng@thinknpc.org or Tris.Lumely@thinknpc.org.

TRANSFORMING THE CHARITY SECTOR

NPC is a charity think tank and consultancy. Over the past 15 years we have worked with charities, funders, philanthropists and others, supporting them to deliver the greatest possible impact for the causes and beneficiaries they exist to serve.

NPC occupies a unique position at the nexus between charities and funders. We are driven by the values and mission of the charity sector, to which we bring the rigour, clarity and analysis needed to better achieve the outcomes we all seek. We also share the motivations and passion of funders, to which we bring our expertise, experience and track record of success.

Increasing the impact of charities: NPC exists to make charities and social enterprises more successful in achieving their missions. Through rigorous analysis, practical advice and innovative thinking, we make charities' money and energy go further, and help them to achieve the greatest impact.

Increasing the impact of funders: NPC's role is to make funders more successful too. We share the passion funders have for helping charities and changing people's lives. We understand their motivations and their objectives, and we know that giving is more rewarding if it achieves the greatest impact it can.

Strengthening the partnership between charities and funders: NPC's mission is also to bring the two sides of the funding equation together, improving understanding and enhancing their combined impact. We can help funders and those they fund to connect and transform the way they work together to achieve their vision.

New Philanthropy Capital
185 Park Street, London SE1 9BL
020 7620 4850
info@thinkNPC.org
[@NPCthinks](https://www.thinkNPC.org)

Registered charity No 1091450
A company limited by guarantee
Registered in England and Wales No 4244715

www.thinkNPC.org