Public Sector Data Analytics

A Nesta Guide

Eddie Copeland
Tom Symons
Hilary Simpson
Nevena Dragicevic

nesta

nesta

About this guide

This guide is for public sectors organisations who are interested in using data analytics to make better decisions and improve public services.

The methods and advice are primarily based on Nesta's Offices of Data Analytics (ODA) programme. During pilots in London, the North East of England and Essex, we've explored how cities and regions can establish ODAs to join up, analyse and act upon data sourced from multiple public sector bodies to improve public services.

For more information on our ODA programme, visit: https://www.nesta.org.uk/project/offices-data-analytics

nesta

Learning modules

This guide contains the following modules:

Module 1: How data analytics can help the public sector
Module 2: Barriers to using public sector data
Module 3: The ODA method
Module 4: The 8 phases of an ODA project
Module 5: Using data legally and ethically
Module 6: Designing and running an ODA pilot
Module 7: How to make it easier next time round
Module 8: Setting up an Office of Data Analytics

Module 1:
How data analytics can help the public sector

nesta

Data analytics: from insight to action

Data analytics is the discovery, interpretation, and communication of meaningful patterns in data.

It can be used by individual teams and organisations to better inform their own decisions and activities.

It can also be used to help multiple teams or organisations collaborate more effectively.

At Nesta, we believe data analytics has the most value when it leads to better *actions*. This guide therefore focuses on achieving actionable insights from data.

Old approach; new technologies

Using data to deliver actionable insights is nothing new. In 1854, John Snow famously plotted the location of deaths in London's Soho to show that a cholera outbreak was caused by contamination of a local water pump. The map he created led to the pump handle being removed, saving many lives.

A key difference today is that computers enable us to analyse greater quantities of data in more sophisticated ways.

nesta

What can data analytics do?

Public sector organisations are working towards many different goals. So where does data analytics fit in?

Data analytics can be particularly helpful for:*

- Identifying specific cases in a wider group
- Prioritising cases based on risk or need
- Creating early warning tools
- Making better, quicker decisions
- Optimising resource allocation

Let's look at a few examples...

*Categories derived from NOLAlytics: https://datadriven.nola.gov/nolalytics/

nesta

Targeting HMO building inspections

12 London boroughs and the Greater London Authority analysed data on the known risk factors associated with unlicensed Houses in Multiple Occupation (HMOs) to help building inspectors find other properties that were likely to be unlicensed HMOs. Its aim was to increase licence revenues and protect vulnerable tenants.

Read the full casestudy

Image: AD_Imgaes | Pixabay CC0 Creative Commons

nesta

Tackling modern slavery

Essex Police and Essex County Council are exploring how data sharing and analytics could help them develop a better understanding of local business inspections.

The aim is to enable improved collaboration between the many different public sector organisations involved in assessing businesses' safety and compliance.

Image: Brian A Jackson / Shutterstock.com

nesta

Optimising ambulance standby locations

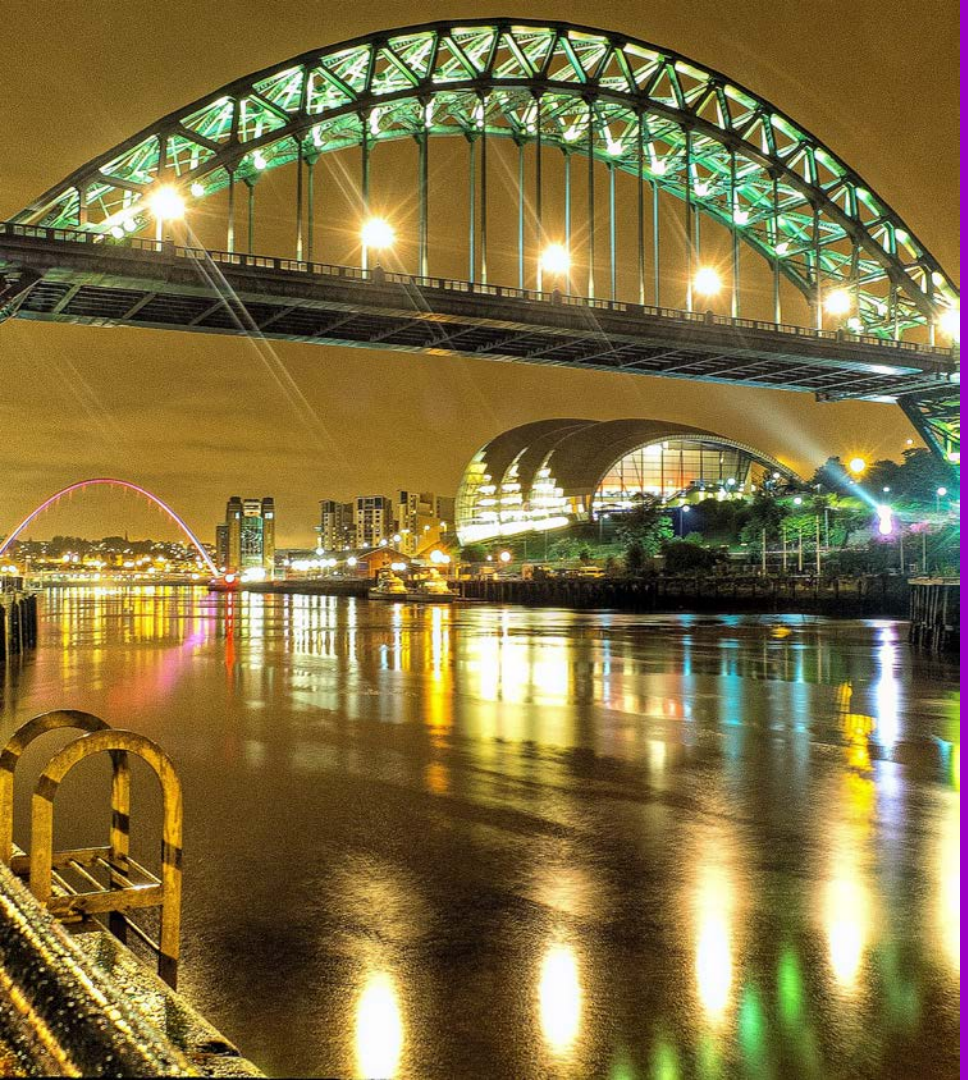
In New Orleans, data on the nature, location and timing of past emergencies was analysed in order to predict where and when future emergencies could happen.

This analysis helped identify the optimum places to park ambulances on standby to reduce response times.

Read the full casestudy

Image: Emergency Medical Transport, Inc.

nesta



Newcastle NEET analysis

Newcastle City Council's analysis of NEET individuals helps the local authority identify children most at risk of not being in employment, education or training.

Read the full casestudy (page 26)

Understanding motoring accidents

The Behavioural Insights Team analysed data from East Sussex on KSIs – road accidents leading to catastrophic injury or death. The analysis helped debunk widely held assumptions about the causes of accidents, helping the local authority see where they could design interventions with the most impact.

Read the full casestudy (page 21)

Image: SteelFish | Pixabay CC0 Creative Commons

In short, data analytics is useful because…

It enables many of the tried and tested ways of working better

But it’s not always straightforward…

Module 2:

Barriers to using public sector data

Public sector organisations face a number of barriers to using their data.

The first is a basic issue around data quality. Common quality issues include:

- Records are only recorded on paper
- Records are digitised, but in hard-to-analyse formats like PDF
- Data is recorded inconsistently, such “Smith Street” and “Smith Str”.
- Records about the same person or thing lack a common unique identifier
- Records are unknowingly duplicated

nesta

A second issue concerns discoverability. Public sector organisations tend to hold thousands of records that have accumulated over time, and find them hard to search. As a result, individuals may have little knowledge about what useful data is held by other teams.

In other cases, the existence of the data is known, but is thought too hard to use as records are in the form of free-text fields, old emails, meeting minutes, etc. Where services or IT are outsourced, a public sector body may even find that it cannot access the data relating to its own service, or must pay an additional fee! (Our advice: you should explicitly prohibit this bad practice in future contracts with external suppliers.)

Nesta outlines some further potential solutions in our report:

Can Government Stop Losing its Mind?

The jigsaw problem

An additional challenge is that many data analytics projects require sourcing, analysing and acting upon data sourced from different teams and organisations.

This is made hard by the jigsaw problem: every team has their own piece of the data puzzle, but rarely can anyone put all the pieces together to see what the big picture shows.

nesta

Challenges caused by the jigsaw problem

The jigsaw problem hinders public sector organisations from using data to enable some tried and tested ways of working more effectively. For example:

1) Shared Services: it's hard for organisations to see where they could share resources with their neighbours if they don't have data on the scale and location of the problems, demand and opportunities beyond their boundaries.

2) Target areas of greatest need: it's hard to target resources effectively if organisations don't have access to data that shows where the people and places of greatest need are located.

3) Intelligent coordination of teams: it's hard for teams to efficiently coordinate their activities on complex areas like adult social care (where up to 30 organisations may need to collaborate to support one individual or family) if they don't have data on what each other is doing.

4) Prediction and Prevention: it's hard to intervene in problems early if organisations cannot bring together and analyse the datasets that could collectively point to cases of highest future risk.

The jigsaw problem has several well known causes. These include:

Technical: Different organisations and teams use different IT systems, some of which don't easily talk to each other or make data accessible.

Data: Records may be stored in different formats and according to different conventions, making the matching and analysis of data about the same person or place hard.

Legal: There are some things the law does not allow. There are many more things that can be done with data that never happen because most public sector staff are not confident in what data protection laws permit and prohibit.

Organisational: Every public sector body was set up to serve a certain community in a certain way. It can take significant organisational and cultural changes to start systematically collaborating.

Why data sharing is hard (and how to make it easier)

Why it's hard

Technology

- Bespoke, siloed IT systems
- Legacy IT that makes data hard to extract
- Outsourced IT providers charging for data access
- Lack of common platform for data sharing

Data

- Data in hard-to-use formats like PDF
- Data inconsistently entered
- Use of different standards
- Lack of common identifiers
- Lack of open data

Legal

- Risk averse leadership
- Staff unsure about data protection rules
- Most senior data professional in org is Data Protection Officer
- Lack of template data sharing agreements

Organisation

- Teams created to focus on their siloed remit
- Lack of dedicated time and resources for data collaboration
- Leaders lack understanding of role and means of using data

How to make it easier

Technology

- Use tech conforming to common standards for interoperability
- Insist all IT has open APIs
- Ensure contracts give full access to data
- Invest in common platform for data sharing

Data

- Record all data in machine-readable format
- Enforce consistent data entry
- Use common standards
- Use unique IDs, e.g. UPRNs
- Release non-personal data openly by default

Legal

- Train all staff in Privacy Impact Assessments
- Appoint Chief Data Officer tasked with responsible data sharing
- Make use of template Information Sharing Protocols

Organisation

- Establish Offices of Data Analytics
- Free up time of in-house analysts to work on data science projects rather than KPI reporting
- Leaders insist on using data to inform decisions

nesta

The ODA method

To help overcome some of these challenges, Nesta has been experimenting with and refining a methodology for running public sector data analytics projects. We call this the Office of Data Analytics (ODA) method.

This work was originally inspired by the activities of New York City's Mayor's Office of Data Analytics (MODA), established during the tenure of Mayor Michael Bloomberg.

We'll now explore this method and show how it can help you determine whether a given challenge can be tackled with data analytics.

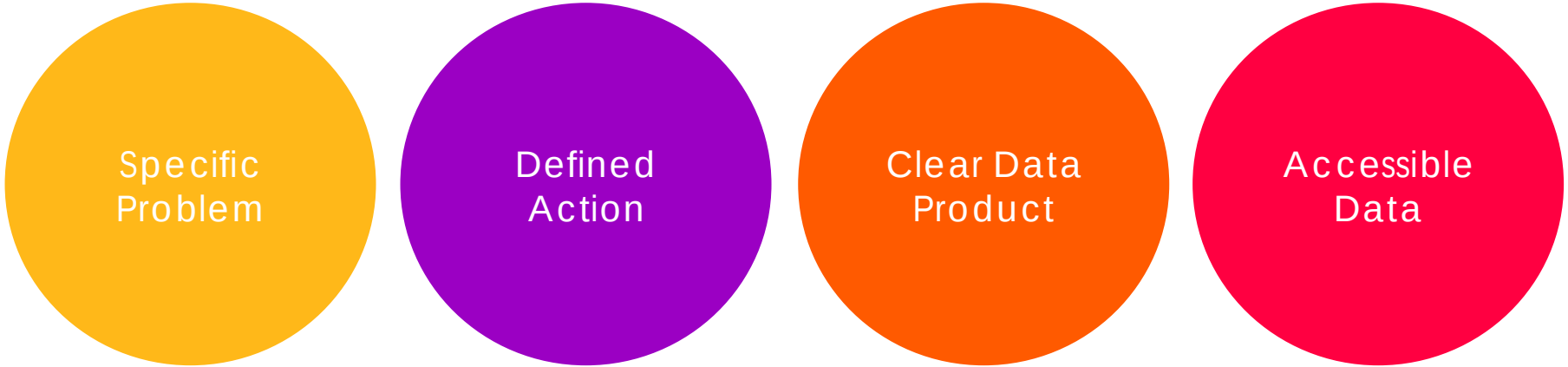
nesta

Module 3:
The Office of Data Analytics (ODA) method

How to decide if a given challenge can be tackled with data analytics

Public sector organisations face many different challenges. How can you tell which of them might be tackled with data analytics?

We believe that successful data analytics projects consist of four core elements:



Let's explore each in turn.

KEY POINT: It's vital to move from large, macro-level problems to something narrow and actionable.

Public sector organisations face many large challenges, but some are too broadly defined to invite any particular remedy. For example, the problem statement: "Modern slavery is occurring in the city" is too vague.

With further thought, this could be refined to a narrower problem statement, such as: "We don't know which regulated businesses are most likely to be exploiting victims of modern slavery."

Five specific problem types

Specific Problem

When trying to identify a narrower, more actionable problem, it's helpful to consider the types of problem that data analytics is well suited to address. The analytics team in New Orleans' Office of Performance and Accountability have helpfully outlined these five specific problem types:

5 Specific Problem Types*
Targets are difficult to identify within a broader population
Services do not categorise high-priority cases early
Resources are overly focused on reactive services
Repeated decisions made without access to all relevant information
Assets are scheduled or deployed without input of latest service data

*Derived from NOLAlytics: https://datadriven.nola.gov/nolalytics/

Your specific problem statements should not be in the form of a question, but phrased as follows:

Our problem is that... *[insert specific problem statement]*.

What action do you want to make possible?

KEY QUESTION: What would you do differently if you had all the information you needed about your specific problem?

To be clear, the data analytics process is not the intervention. It's important to identify practical actions and interventions that are within your control to change. For example, no single organisation can 'solve' homelessness – but you might help address a specific aspect of it in your area.

Drill down to precisely who will act, and where and when they will do so.

nesta

What action do you want to make possible?

Defined Action

There are five opportunity types associated with the five problem types we outlined earlier.*

Specific Problem Type	Opportunity
Targets are difficult to identify within a broader population	Identifying specific cases in a wider group
Services do not categorise high-priority cases early	Prioritising cases based on risk or need
Resources are overly focused on reactive services	Creating early warning tools for proactive working
Repeated decisions made without access to all relevant information	Making better, quicker decisions
Assets are scheduled or deployed without input of latest service data	Optimising resource allocation

*Derived from NOLAlytics: https://datadriven.nola.gov/nolalytics/

What action do you want to make possible?

List all the actions or interventions that you would like to put in place to address your specific problem if you had better information:

Our problem is that... *[insert specific problem statement]*.

In response to which we would like to... *[list the different actions you would like to implement]*.

KEY QUESTION: What would a person need to see on a screen in order to enable the actions defined in the previous step?

It's unlikely that whoever is doing the action (e.g. a frontline worker or service manager) will want a spreadsheet or raw data. Instead they will want the data conveyed in a more intelligible way that provides a real insight – that's what we mean by a 'data product'.

A data product could be a map, a heatmap, a prioritised list, an alert, a dashboard, a visualisation, and so on.

What data product do you need?

Clear Data Product

Certain data products are suited to certain problem and opportunity types.

Specific Problem Type	Opportunity	Example Data Product
Targets are difficult to identify within a broader population	Identifying specific cases in a wider group	A graph showing anomalies or outliers
Services do not categorise high-priority cases early	Prioritising cases based on risk or need	A prioritised list
Resources are overly focused on reactive services	Creating early warning tools for proactive working	An alert to flag issues when a threshold has been reached
Repeated decisions made without access to all relevant information	Making better, quicker decisions	A data visualisation
Assets are scheduled or deployed without input of latest service data	Optimising resource allocation	A map or heatmap showing where cases occur

Example data products

Here are four examples of data products used by UK and US public sector organisations:

A PRIORITISED LIST. In London, housing teams in many boroughs rely on random inspections or tip-offs to locate unlicensed 'Homes of Multiple Occupation' (HMOs). The London Office of Data Analytics pilot sought to change this by developing prioritised inspection lists that would lead inspectors to properties most likely to be HMOs, based on the characteristics of known unlicensed HMOs.

AN ALERT. To reduce the risk of excessive force by police officers, the city of Charlotte, North Carolina, combined demographics, training, payroll, internal affairs and other data to develop an early warning system for when an officer was likely to have a negative interaction with the public.

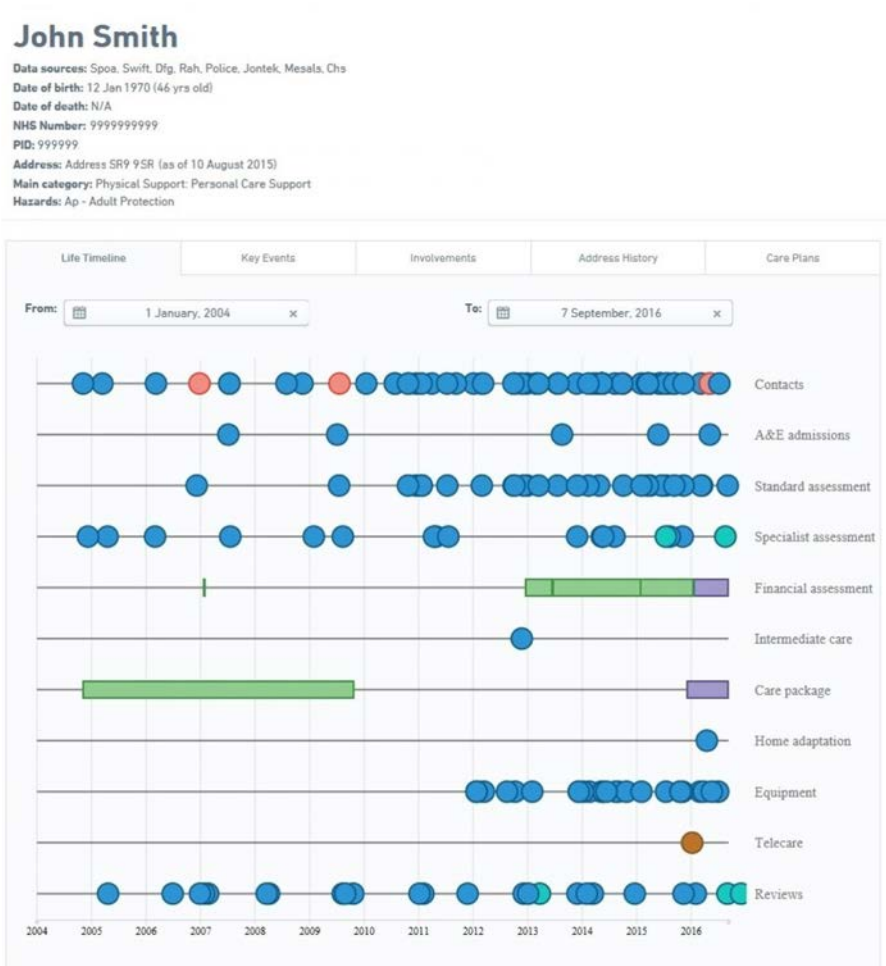
A MAP. In partnership with the SumAll Foundation, New York City is fighting record numbers of homeless by analysing and visualising the patterns of evictions that lead to family homelessness. The project is also improving the targeting of outreach services by predicting the neighbourhoods, buildings and specific addresses where resources are most needed.

A DASHBOARD. Louisville, Kentucky has improved ambulance turnaround times by using data to identify obstacles to speedier response, which have saved the city $1.4 million (USD). Dispatchers are now supported with regular reports from a Computer Aided Dispatch system, which spots hidden inefficiencies and monitors real-time location of ambulances to speed up response times.

Example data products

Sunderland City Council created Adult 360, a project to bring together information about a person and their life from across a number of source systems including Social Care, CES, telecare, intermediate care, city hospitals and the police.

It has helped deliver better and more coordinated care, equipping over 350 health and social care practitioners with a more complete view of all that individual's interactions - as shown in this mocked up version.

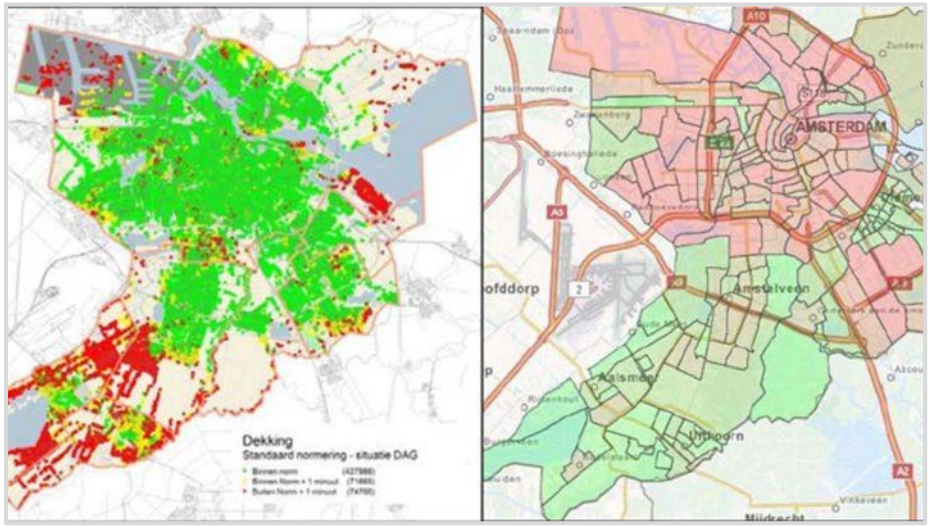


nesta

Example data products

The Amsterdam fire brigade collated data from different sources (information on roads, rails, buildings, neighbourhoods, etc.) and matched them with historical records of previous incidents in the area.

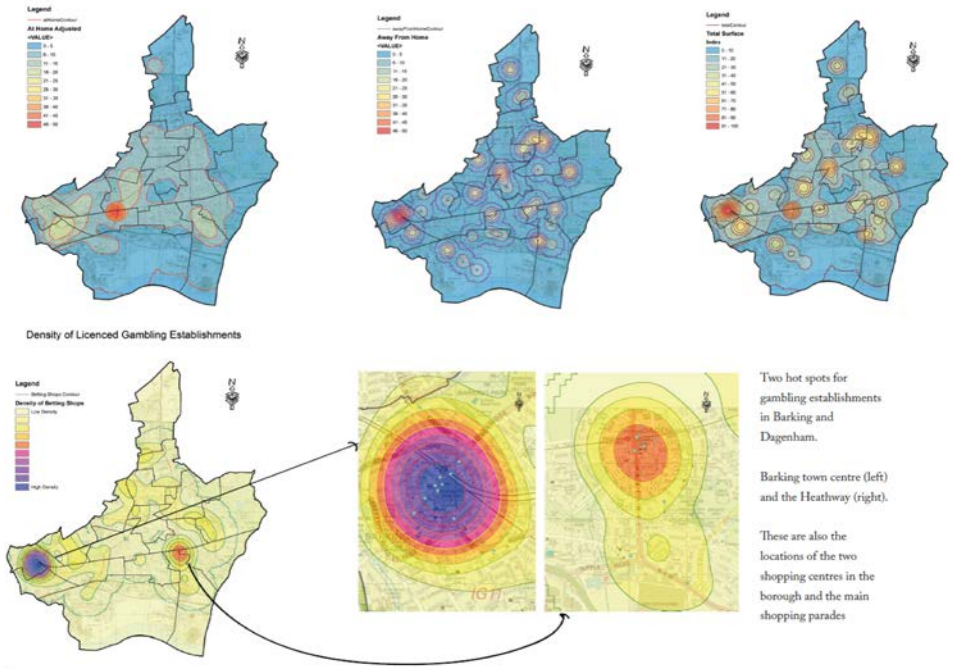
The data was then visualised through maps that the Amsterdam fire brigade use to see where, when and how often fires occur.



Example data products

The London Borough of Barking & Dagenham conducted analysis to identify areas where individuals are more likely to be at risk from gambling-related harm.

The analysis provided context to the local Gambling Licensing Policy revision and helped create a ‘local area profile’ identifying two important clusters, debunking the assumption of vulnerable people and gambling shops being dispersed across the borough.



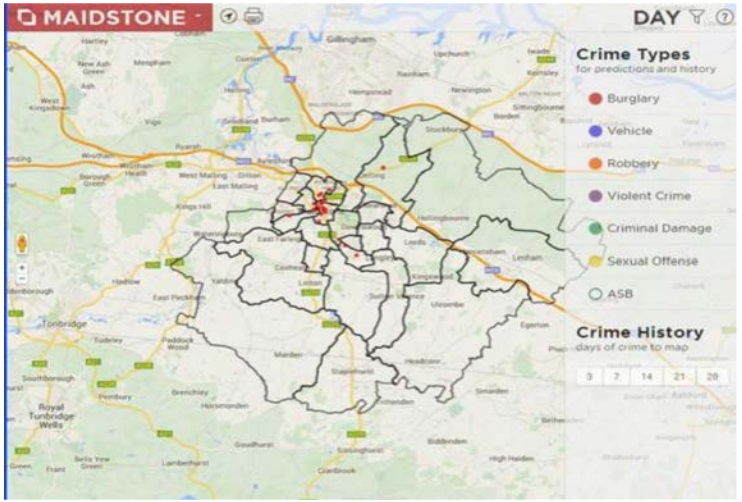
nesta

Example data products

Kent Constabulary has used data on previous offenses in their area to optimise resource allocation and better coordinate police surveillance.

Data includes five years of historical records of crimes committed in the area.

Officers receive daily updates on 180 hotspots in the area and can use them to inform their decisions on what areas to patrol the most.



nesta

Durham Constabulary is using HART, a data tool to support consistency in the decision-making of custody officers when assessing the risk of future offending.

This tool creates a risk score, from high to low, drawing from data that relates to a suspect's previous offending behaviour together with age, gender, residential postcode, and intelligence reports.

What data product do you need?

You can now see whether an insight from a particular data product could enable one or more of the actions you previously outlined:

Our problem is that... *[insert specific problem statement]*.

If we could see / if we knew... *[insert what the data product shows]*

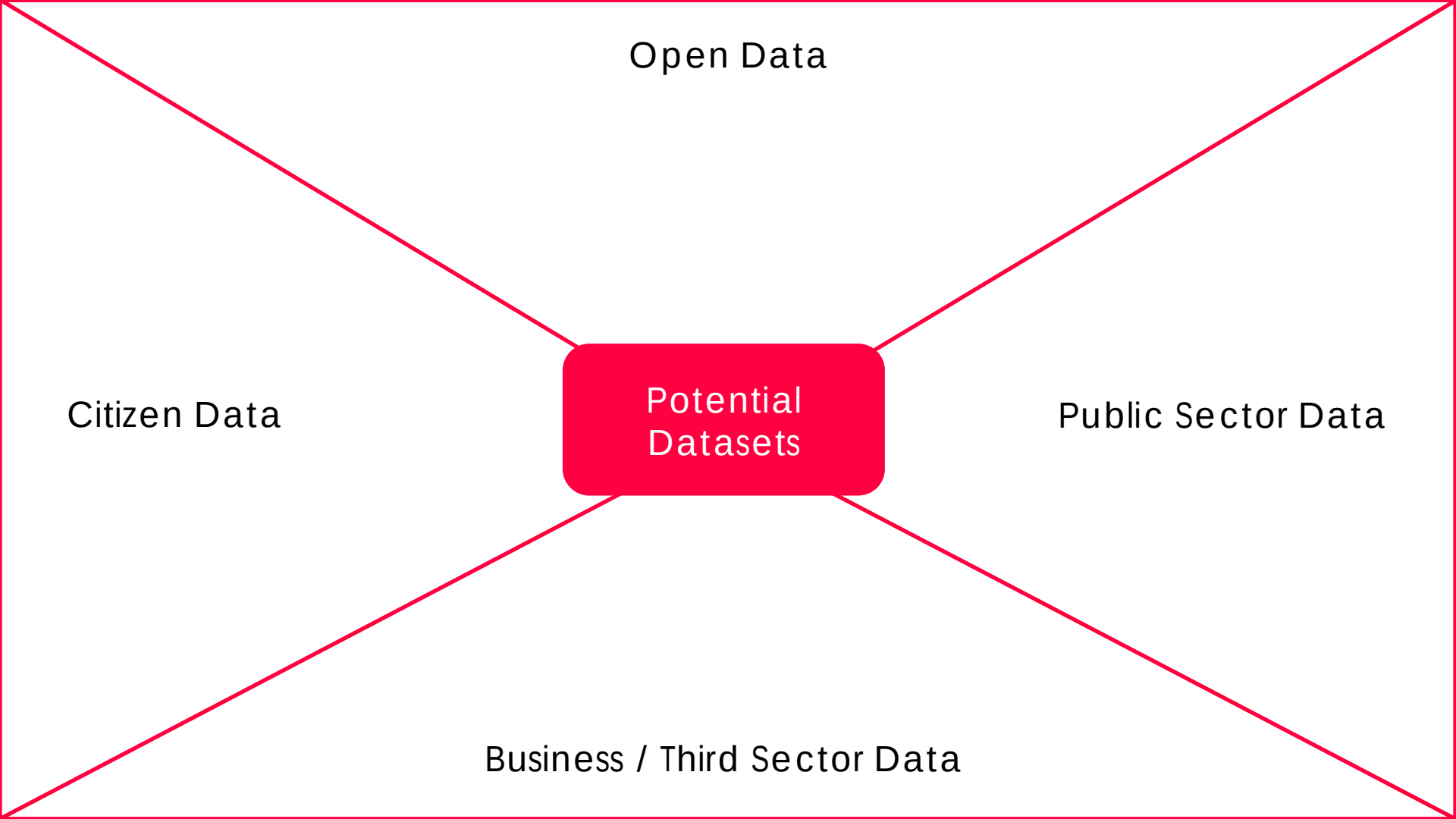
We would... *[insert the action you want to implement]*.

KEY QUESTION: What data do you need to create the data product, does it exist, can you get it, and can you use it?

Data can come from many different sources, such as:

- Open data (e.g. data.gov.uk)
- Public sector
- Businesses & Third Sector
- Citizens

You can use a simple template like the one on the next slide to brainstorm what datasets might be available from these different sources.


Open Data
Citizen Data
Potential Datasets
Public Sector Data
Business / Third Sector Data

If the data you need to create your data product does not exist, you may wish to consider:

1) Are there other datasets that might contribute a similar type of information, or act as a proxy measure?

2) Could you start collecting this data so that analysis becomes more feasible in future? (This is still a useful outcome of the ODA process.)

In Module 5, we'll explore how you can check that you can use and, if necessary, share the data legally and ethically.

For now, it's enough to determine if the data you need to create your data product is *in principle* available.

You should now be left with a four-part statement as follows:

Our problem is that… *[insert specific problem statement]*.

If we could see / if we knew… *[insert what the data product shows]*

using these datasets… *[insert datasets you plan to use]*

we would… *[insert the action you want to implement]*.

Module 4:

The 8 phases of an ODA Project

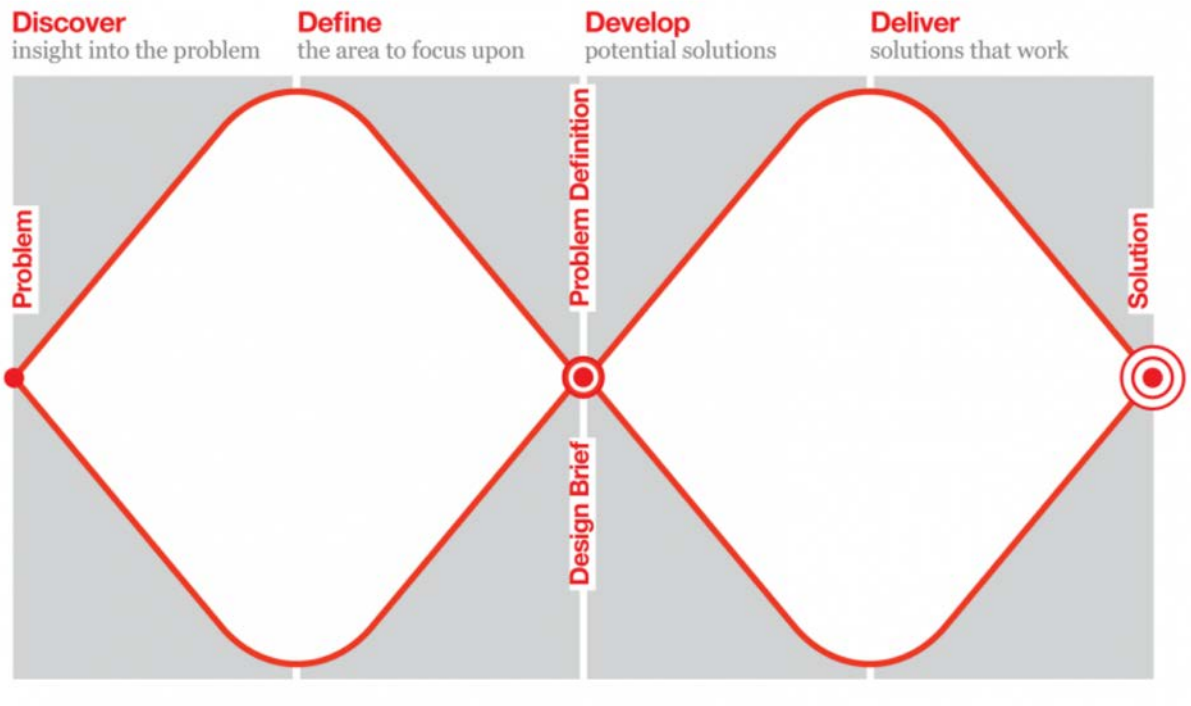
The eight phases of an ODA project

Let's assume you've used the four-step ODA method to identify a challenge you think could be tackled with data analytics. To turn this into a live project, there are eight core phases that should feature in your project plan:

1. Discovery: assessing the project's feasibility and refining its approach
2. Securing the commitment of project partners: identifying who needs to be involved and their roles and responsibilities
3. Information governance: putting in place agreements to share data
4. Data acquisition: getting hold of the required data
5. Data analysis and prototyping: analysing the data and building the first version of the data product
6. Testing and evaluating: trialing the data product in a real-world setting and measuring its results
7. Refining: improving the data product based on feedback
8. Scaling: putting the data product into permanent / wider use

The ODA project lifecycle

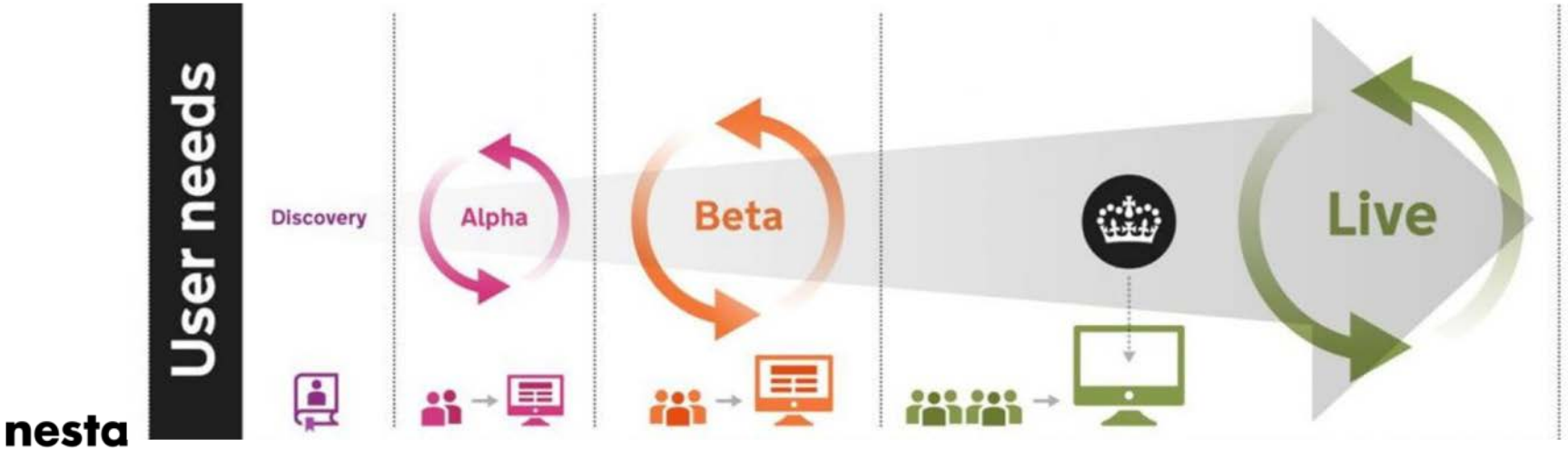
An ODA project entails several stages that vary between thinking very broadly about possibilities before narrowing down to something more specific. The Design Council uses the double diamond diagram shown below.



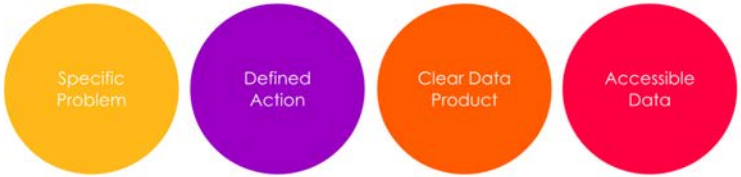
nesta

A discovery phase is commonly used in digital development projects. It's there to ensure that the correct problem has been identified and to verify that the proposed solution is sound and viable.

The UK's Government Digital Service have their own guide to running a discovery phase, and use the diagram below to show how it fits into a project lifecycle.



Phase 1: Discovery - Going deeper on the four steps



During the discovery phase, you should thoroughly check and seek to improve your thinking on each of the four steps.

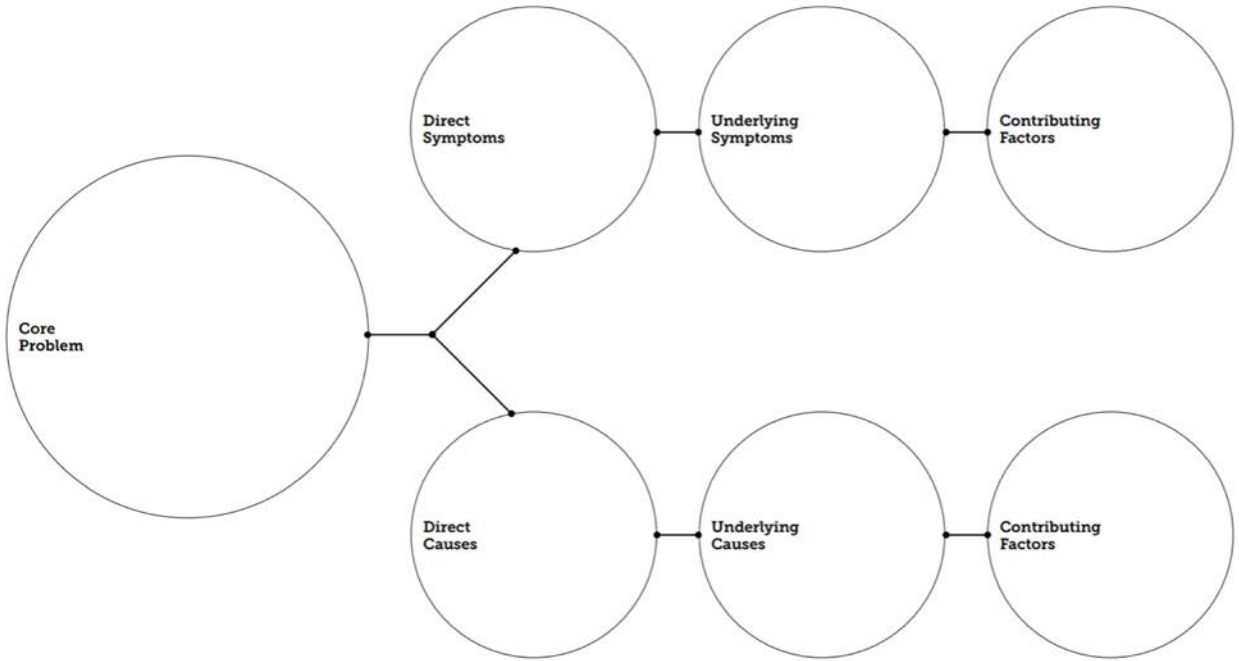
Some of this can be done through desk research, but it should also include interviews and workshops with people whose work the data analytics project is intended to support, such as service managers and front line staff, and those who will be affected, such as specific end users and groups of citizens. Co-designing and testing potential solutions with these groups is a vital at every step of an ODA programme.

You should aim to achieve a deep understanding of:

1) The nature and complexities of the problem to be tackled
2) The range of different interventions available to solve it
3) What data product would serve the needs of those who would use it
4) Whether the data you need is accessible, whether its quality is sufficient, and whether it can legally and ethically be used

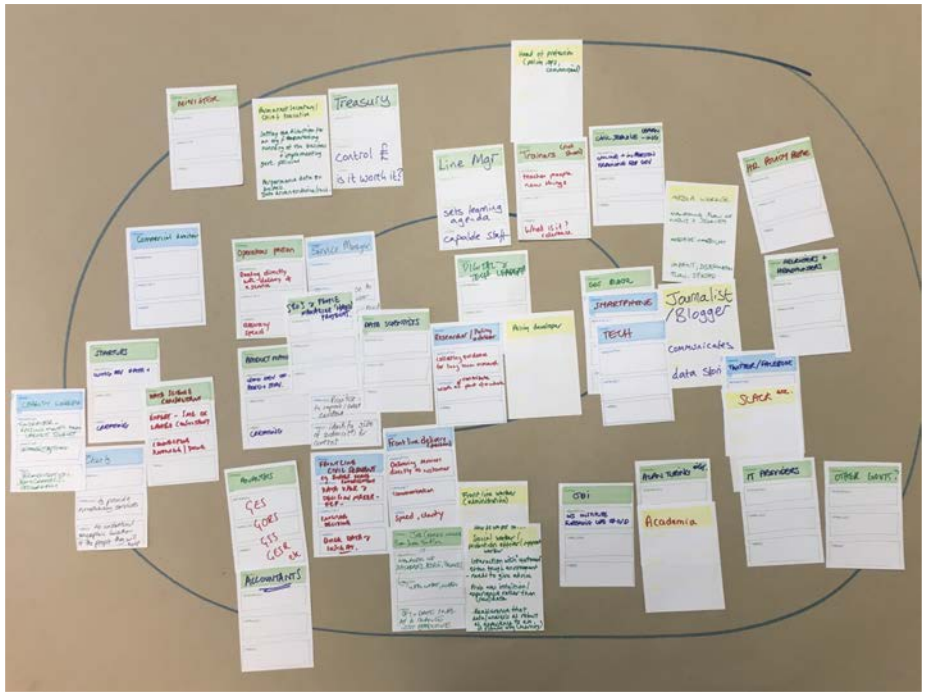
There are a number of tools and techniques you can use to dive deeper in your chosen issue during the discovery phase.

For example, the Five Whys technique, a Fishbone Diagram and Causes Diagram (see image) can help you identify the contributory factors of the problem you are trying to tackle.

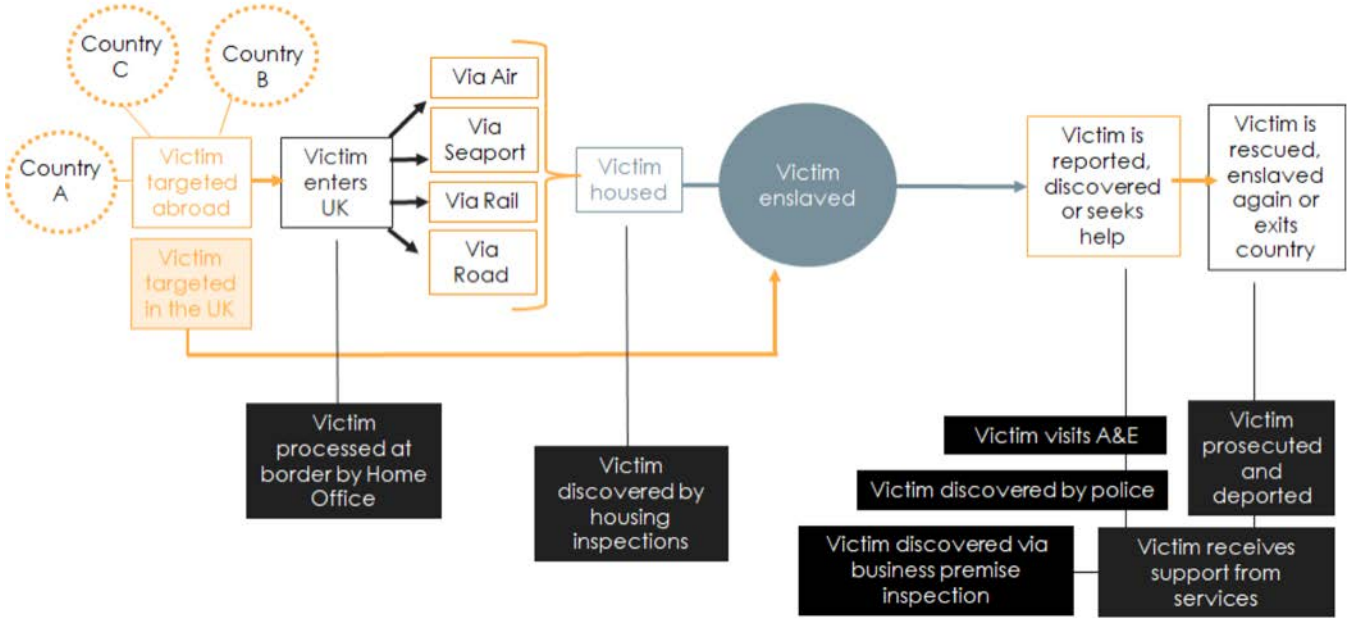


A stakeholder map can help you think about the different organisations who come into contact with the problem you are trying to address.

Those organisations could potentially be sources of expert advice, additional datasets, or even become partners in a data analytics project.



Creating a user journey map can help you understand the touchpoints where the public sector comes into contact with a given issue. This can help you understand what data is collected and see where better interventions could potentially be designed. The diagram below is a hypothetical and simplified view of public sector touchpoints with a victim of modern slavery.



Module 5:
Using data legally and ethically

During your discovery phase, it's vital to check that the data you require can be used, and if necessary shared, legally and ethically.

Most legislation governing the sharing of publicly held data relates to personal data. The UK Data Protection Act defines personal data as "data which relates to a living individual who can be identified from those data" or from those data combined with other information.

The Act also defines sensitive personal data, consisting of personal information on race, ethnicity, political affiliation, religious beliefs, membership in trade unions, physical or mental health, sexual life and criminal background. More exacting conditions must be met to share sensitive personal data.

The ICO has a useful guide to determining what is personal data.

In May 2018, the European Union introduced the General Data Protection Regulation (GDPR), which places greater responsibilities on all organisations who collect and use personal data.

The UK's Information Commissioner's Office provides a useful Guide to the General Data Protection Regulation (GDPR), including a Data Protection Self Assessment Toolkit.

Once you've identified some datasets that you'd like to use, it's best practice to carry out a Privacy Impact Assessment (PIA). A PIA is a standard series of screening questions that guides users through the potential risks and benefits of sharing personal data.

The PIA equally prompts users to develop mitigation strategies to minimise potential downsides of information sharing.

This editable PIA is provided by the Information Commissioner's Office (ICO).

If you must use personal data, an important step is to identify the legal gateways that grant your organisation the permission or authority to pursue certain objectives, which could be supported by the sharing of personal data.

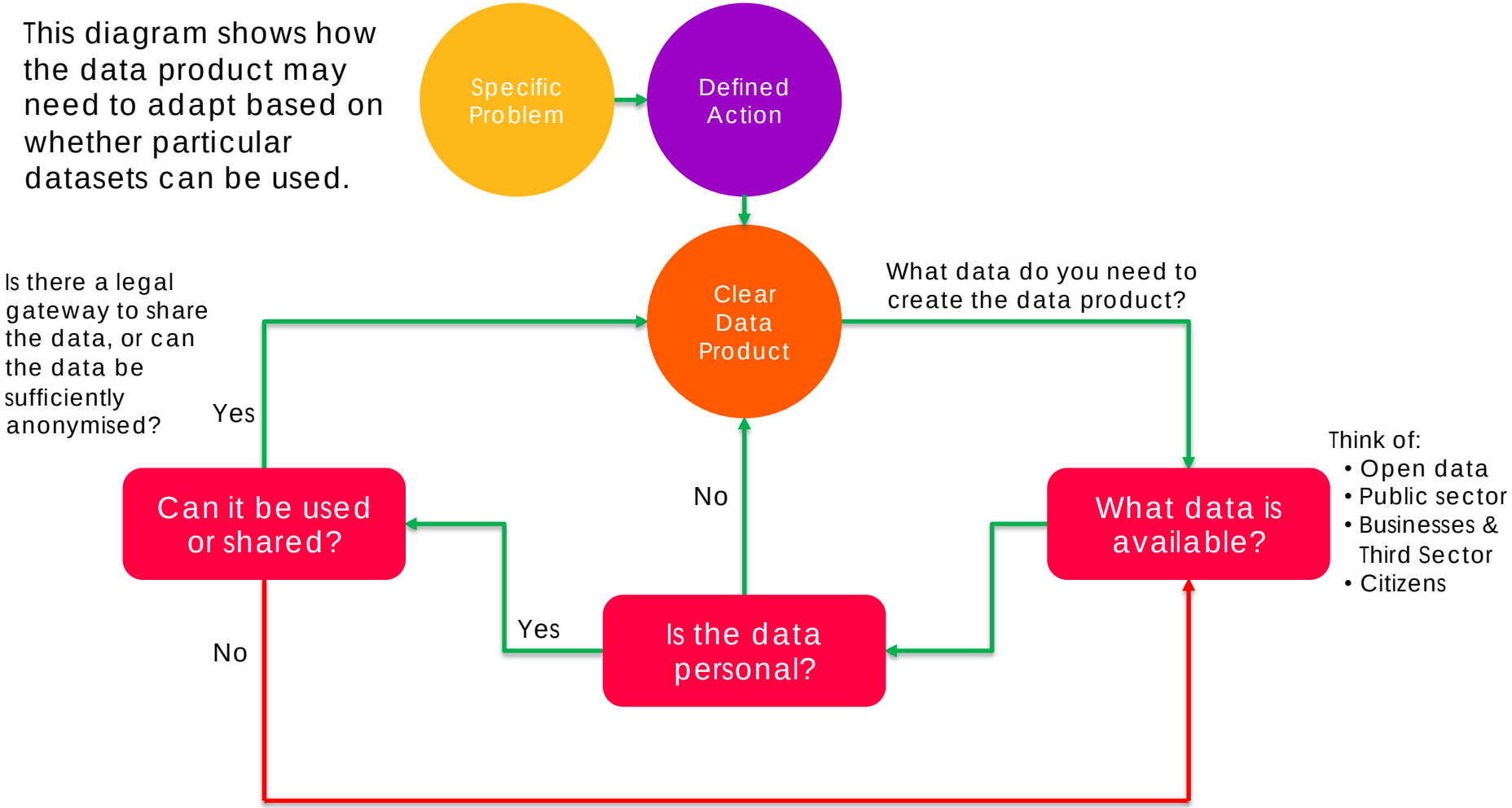
For example, during a pilot for the London Office of Data Analytics that sought to identify unlicensed HMOs, two pieces of legislation - the Housing Act 2004 and the Crime and Disorder Act 1998 - were identified as placing responsibility on local authorities to improve housing standards and to prevent crime and disorder.

KEY POINT: Sharing non-personal data comes with far fewer conditions.

Wherever possible, it's best to use non-personally identifiable data. If the source data is personal, it may be possible to remove names (and other personally-identifiable attributes) and aggregate the data to large enough sample populations that it's no longer personal.

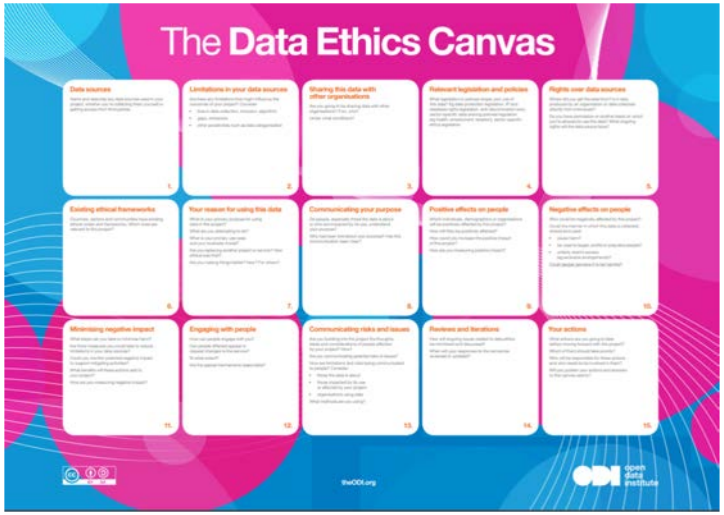
Good guidance on data anonymisation and pseudonymisation is available in the Research Ethics Guidebook.

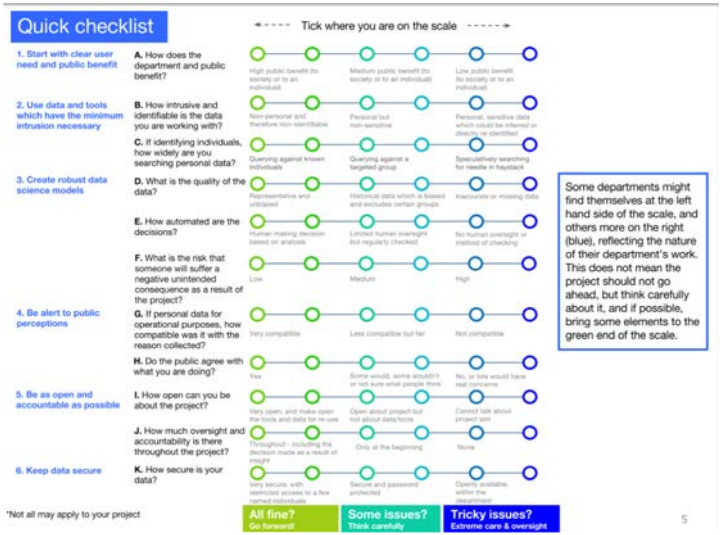
This diagram shows how the data product may need to adapt based on whether particular datasets can be used.



Regardless of whether it's legal to use certain datasets, you must ensure your proposed data analytics project is ethical, too. Ethical considerations apply not just to what data is used, and how it's analysed, but also the *actions* that the data enables.

There are a number of excellent toolkits to help you think about these questions, including the Open Data Institute's Data Ethics Canvas and the Cabinet Office's Data Science Ethical Framework.





nesta

Meanwhile, Nesta is developing a self assessment toolkit specifically for cases where data analytics is used to enable algorithmic decision making.

View the draft toolkit.

Module 6: Designing and running an ODA pilot

Phases 2-8 of an ODA pilot

Having completed the discovery phase, let's briefly remind ourselves of the eight phases of a typical ODA project:

1. ~~Discovery: assessing the project's feasibility and refining its approach~~
2. Securing the commitment of project partners: identifying who needs to be involved and their roles and responsibilities
3. Information governance: putting in place agreements to share data
4. Data acquisition: getting hold of the required data
5. Data analysis and prototyping: analysing the data and building the first version of the data product
6. Testing and evaluating: trialing the data product in a real-world setting and measuring its results
7. Refining: improving the data product based on feedback
8. Scaling: putting the data product into permanent / wider use

Phase 2: Securing the commitment of project partners

By the end of the discovery phase, you should know enough to be able to identify which teams and organisations need to be involved and secure their commitment to the project.

A number of different roles will be required:

1. Project management
2. User research
3. Data science
4. Technical (APIs and MVP build)
5. Legal and information governance
6. Implementation
7. Data Providers

The basic function of each role is as follows:

1. Project management - overseeing the design of the project and the coordination of all partners' activities

2. User research - assessing the real needs of those whose work the data project is designed to enhance

3. Data science - collecting, cleaning, matching and analysing data to produce insights

4. Technical (APIs and MVP build) - putting in place the necessary tools to upload and analyse the data; creating the prototype data product.

5. Legal and information governance - assuring that data is being used legally and ethically.

6. Implementation - the organisations or teams conducting the data-informed action

7. Data Providers - Organisations providing data to create the data product

Phase 2: Securing the commitment of project partners

A Memorandum of Understanding (MoU) should be signed by all participating organisations to confirm their commitment to the data analytics project.

The MoU outlines each organisation's roles and responsibilities, their point of contact, the resources they will assign and how they will participate and communicate within the project.

Once this is done, you can move to the next phase: Information governance.

When sharing data among partners, whether it involves a few teams within a local authority or multiple public sector organisations working together, a common set of rules and conditions should be developed in the form of a Data Sharing Agreement, also known as an Information Sharing Protocol - ISP).

ISPs are necessary whenever personal data is shared, but are equally recommended for the sharing of non-personally identifiable data.

The essential elements to be covered in an ISP are:

- The purpose of the sharing
- The potential recipients and the circumstances in which they will have access
- The exact data to be shared
- Data quality – accuracy, relevance, usability, etc.
- Data security
- Retention of shared data
- Individuals' rights – procedures for dealing with access requests, queries and complaints
- Review of effectiveness/termination of the sharing agreement
- Sanctions for failure to comply with the agreement or breaches by individual staff

Templates are available from the ICO

Phase 4: Data acquisition

Phase 4 entails actually getting access to the data to create your data product. Accessing datasets from multiple partners can take a long time, as each organisation has different procedures for doing so.

During the discovery phase you'd ideally have already sourced some samples of the data to assess its quality and to understand its characteristics.

"Data acquisition" can mean either uploading a dataset once (e.g. a spreadsheet) or connecting your data science platform to each data provider's IT systems to access data continuously through an Application Programming Interface (API).

Phase 5: Data analysis and prototyping

In this phase, the data scientists analyse the data and work with technical developers to build the first version of the data product.

Data scientists can use a wide range of different techniques to analyse the data, depending on the type of data and the insight desired.

Sometimes they will use a process known as machine learning, where computers can “learn” about the nature of a problem based on training data, in order to spot patterns and correlations or predict further instances of the same thing.

Phase 5: Data analysis and prototyping - Machine learning

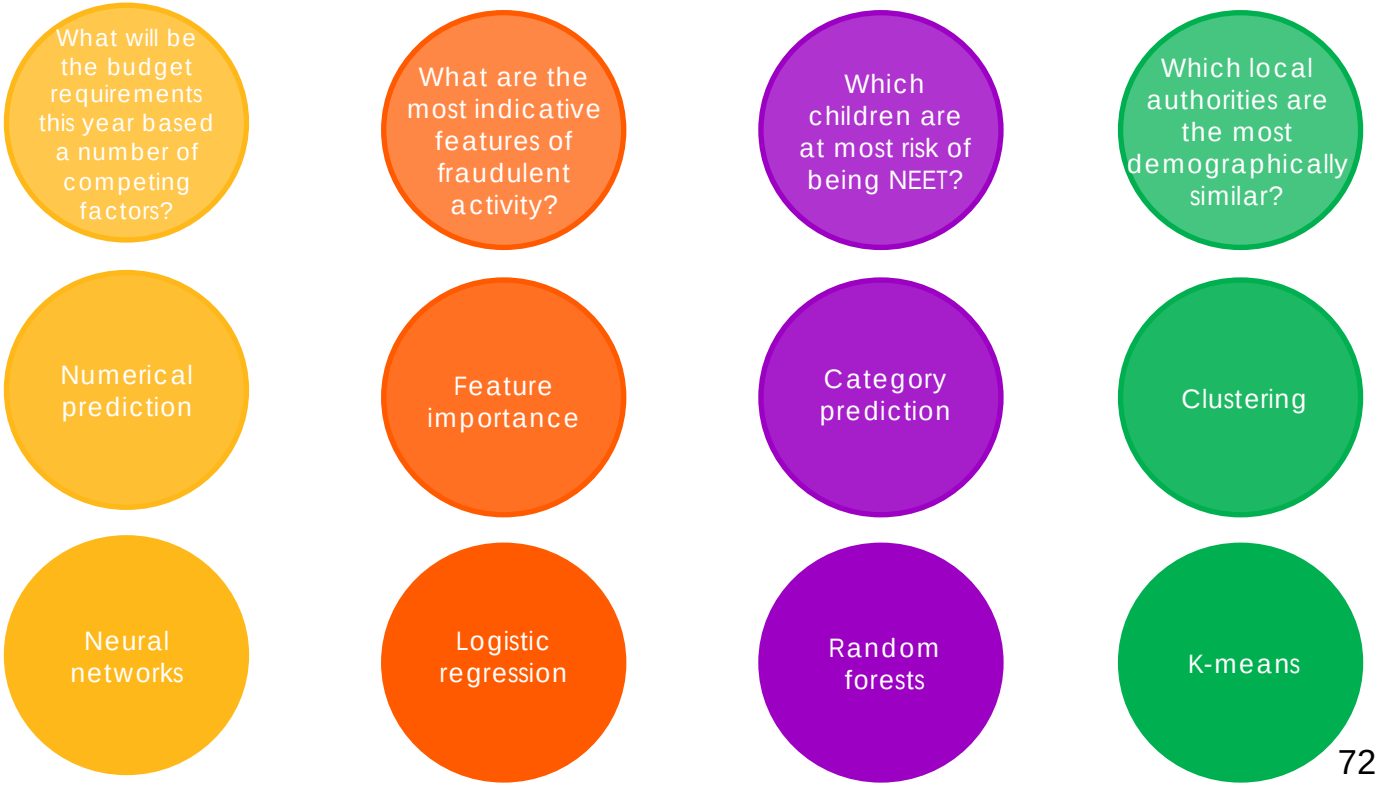
To give an example, a machine learning tool could analyse past financial transactions in order to spot fraudulent activity. In this kind of process, typically you'd need three groups of data:

1) Data on past known cases of the thing you're trying to find (e.g. past known cases of fraud)
2) Data that relates to *characteristics* of the thing you're trying to find
3) Data that relates to characteristics of things you're *not* trying to find

Think of it like a sniffer dog that must be trained in what smells do, and do not, suggest the presence of a particular drug they must locate.

Phase 5: Data analysis and prototyping

Based on the type of problem identified, data scientists can apply an appropriate data analytics technique to address it:



nesta

Phase 5: Data analysis and prototyping

When it comes to analysing data and building useful products, there are a number of resources for guiding good practice.

The Aqua book provides useful guidance on Quality Assurance of data analysis.

Matt Upson has written articles about building transparent and reproducible data products: https://dataingovernment.blog.gov.uk/2017/11/27/transforming-the-process-of-producing-official-statistics/

And how to make them sustainable: https://software.ac.uk/blog/2018-05-21-why-government-needs-sustainable-software-too

Phase 6: Testing and evaluating

The purpose of the testing and evaluation phase is to trial the data product and intervention in a real-world setting and measure their results.

For example, during the pilot for a London Office of Data Analytics, frontline building inspectors were given lists of properties that the data analysis had indicated might be unlicensed HMOs to inspect.

The aim of this process is both to test the usability of the data product, and also to start collecting results on whether it has helped improve the intervention.

Phase 6: The importance of evaluation

Evaluation is often overlooked or conducted hurriedly at the end of a project. This is a mistake.

Good evaluation is vital as it helps you learn about what actually works and increases the chance that the best interventions can be refined and scaled.

In-depth guides on conducting good evaluations are available from Nesta (Research Evidence Practice Guide), from central government (Magenta Book, Green Book), and user-friendly websites, such as betterevaluation.org.

Phase 6: The importance of evaluation

According to the Magenta Book, the UK government's essential guide on evaluation design in the public sector, there are three broad categories of evaluation: process, impact, and economic. The appropriate type will depend on what you wish to learn about your project, as well as the resources and expertise you'll be able to dedicate to the evaluation.

In general, a process evaluation will help you understand how and why an intervention is having an impact, while an impact evaluation will measure the change that has occurred, and whether it has been caused by your intervention. Following a robust impact evaluation, an economic evaluation is possible to monetise the observed outcomes.

As you begin to think about evaluation (which should happen in the early phases of your project), it's worth reflecting on the different Standards of Evidence.

Standards of Evidence are guides to help evaluators determine how confident they can be in their results.

The higher the standard, the more rigorous the methodology and the certainty that your data-informed intervention, and not some other factor, is the cause of the observed change.

Depending on what level you choose, there are different requirements. Let's consider the first three levels of the Standards of Evidence.

Level	Expectation	How evidence can be generated
1	You can explain why your intervention could have an impact and why that would be an improvement	Gathering user feedback or surveying *after* the intervention; reviewing existing research.
2	You can show change has occurred, but not that your intervention caused it	Capturing data *before* and *after* the intervention through surveys or in-depth interviews.
3	You can demonstrate that your intervention is causing the impact	Using robust experimental methods with control or comparison groups, which begin to isolate impact.

Phase 6: The importance of evaluation

Using Experimental Design / Randomised Control Trials (RCTs) is the best way to prove that your intervention *caused* an impact or the change. Participants are randomly assigned to receive an intervention (treatment group) or not (control group).

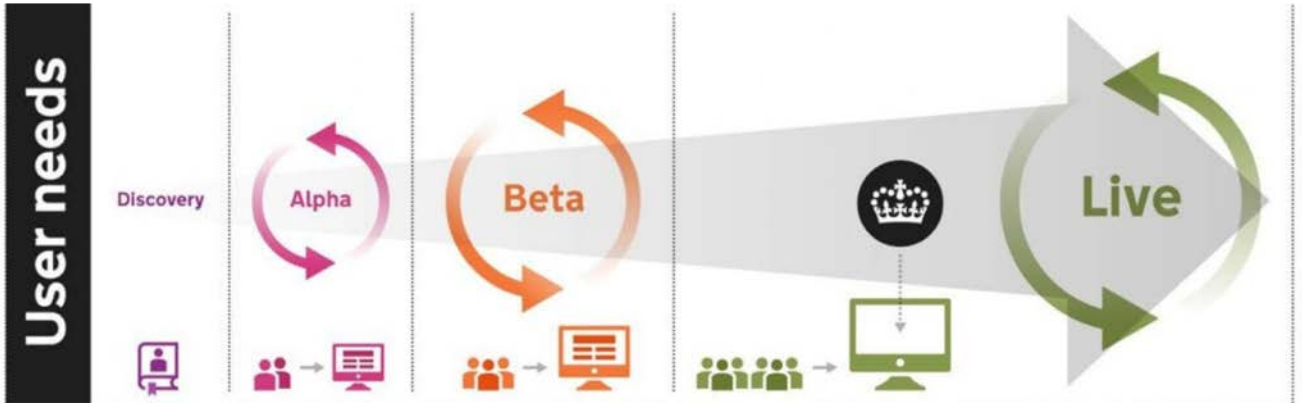
When randomisation is not possible, a quasi-experimental design can be an alternative. This has the same structure as an RCT, but instead of random allocation, you might choose to compare two or more similar groups.

Nesta's Innovation Growth Lab offers the Experimentation Toolkit which explains the experimentation process in plain language. The 'Hypothesis Kit' from the Experimentation Hub is another tool that allows users to quickly formulate a hypothesis, estimate the required duration of an experiment, and determine an adequate sample size.

Phases 7 & 8: Refining and Scaling

An ODA project would ideally take on an agile approach: creating prototypes in short sprints, then testing and improving them based on feedback.

This cycle should be repeated until the data product and intervention are refined to a sufficient point that they can be scaled up or built into everyday practice.



Module 7:
How to make it easier next time round

Module 7: How to make it easier next time round

During Nesta's ODA projects, we've learned a lot about what does, and what does not work with public sector data analytics projects.

We'll now look at some of the big lessons we've learned. Following the advice in this section should help make it easier for you to run data analytics projects in future.

Different organisations have varying degrees of experience in using data. To help you understand where your organisation currently sits on this spectrum, Nesta has developed a data maturity framework.

Before you start a major data project, invest in a new technology, or develop a strategy for getting more out of your data, it's worth first investigating where your organisation's strengths and weaknesses lie.

nesta

Nesta's data maturity framework

The Data Maturity Framework looks at five areas:

1. Data Management: How is data collected, organised and accessed? How complete, accurate and current is the data?

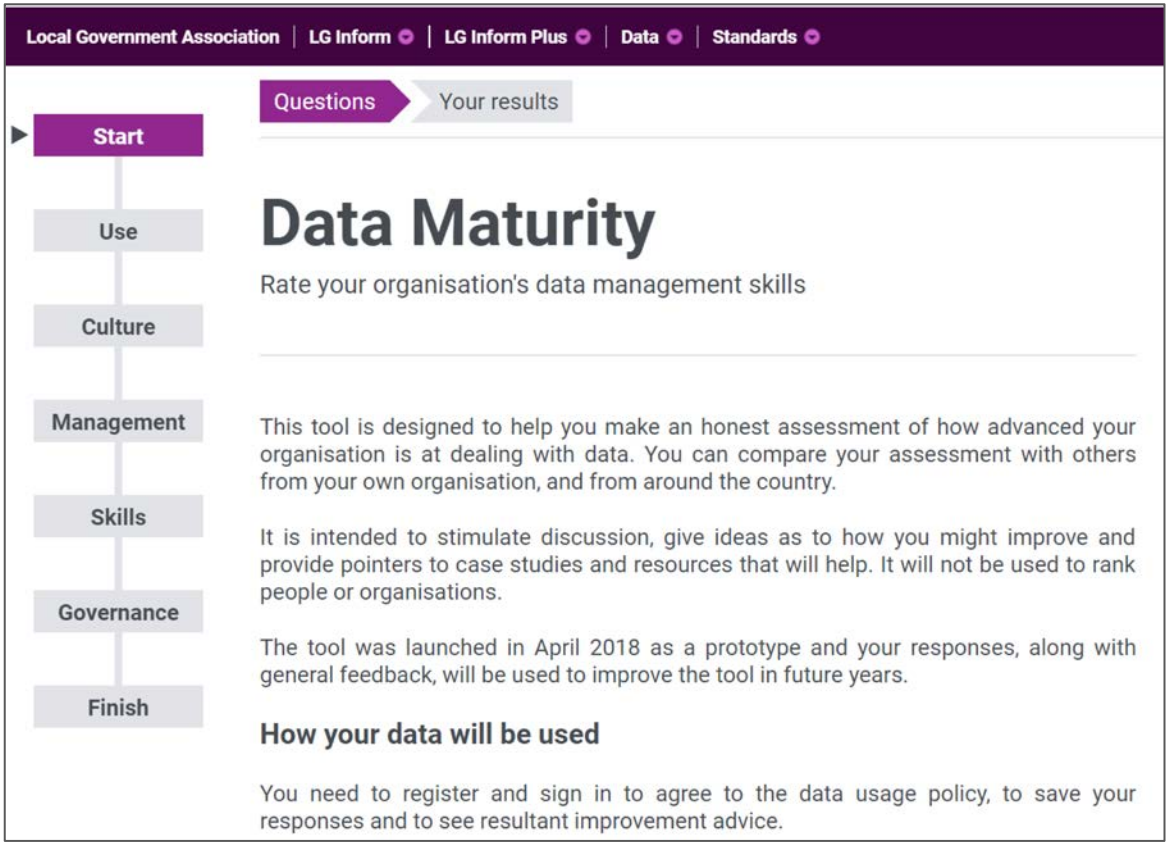
1. Data Governance and Openness: How easy is it to share data? Is data sharing encouraged through clear protocols and leadership? How does the authority handle open data?

1. Data Use: How is data used in the decision-making process? Is it used regularly to optimise services and to track performance?

1. Data Skills and Capability: What is the level of data literacy across the organisation? In which areas of the organisation are skills and capability concentrated?

1. Organisational Culture Towards Data: How much do people in the organisation understand and value data-informed decision making?

You can try a beta version of our online data maturity self-assessment tool, developed in partnership with the LGA



Tips on improving your data quality

KEY POINT: only by trying to use data will it become apparent whether or not its quality is sufficient and whether improving it is worth it. Don't let perfect be the enemy of good!

Here are two tips for improving data quality:

Put in place processes to ensure that clean data is collected in the first place and train staff accordingly. Where possible, capture as much data electronically and from drop down lists rather than free text entry.

Correct erroneous fields in the source data to save yourself from having to correct the same mistakes repeatedly when the data is transferred to other systems. Remember to document where raw data has been sourced from and how it has been cleaned, to allow others to replicate your process when needed.

nesta

Tips on improving your data quality

KEY Point: Organisations that cannot easily match together and link datasets from their own IT systems find it incredibly challenging to collaborate with others on data projects.

Matching and linking data is the process of finding and determining links between individual records across disparate datasets. Once linked, this integrated data is stored in an index or a data warehouse.

For example, the London Borough of Camden's Residents Index links person and place data from different sources within the local authority to create one 'golden' record of local residents. This has improved multi-agency working and helped spot fraudulent activity.

The table below demonstrates how data from five different sources is pulled together into one, accurate ‘golden’ record

FIELDS	Source A	B	C	D	E	GOLDEN
Full Name	**Fred Smith**	Freddie Smith	Frederick Smith	Fred Smith	Fried Smith	**Fred Smith**
DOB	**10-1-1995**	01-01-1900	10-1-1995		10-1-1995	**10-1-1995**
Gender	M	M	M		M	**M**
Address	10 The Grange Camden NW1 0AA	Flat A 10 The Grange Camden NW1 0A	10 The Grange Camden NW1	**Flat A 10 The Grange Camden NW1 0AA**	10 Grange Camden NW1 0AA	**Flat A 10 The Grange Camden NW1 0AA**
Ref	**1007689**		**1007689**	1007699	**1007689**	1007689
NINO	**AB123456B**					**AB123456B**
Home Tel		0231-987-7866	**0205-937-996**			**0205-937-996**
Mobile	**07988674707**	077134563				**07988674707**
Email		**fsmith@v.com**	smithy@bt.com			**fsmith@v.com**
Car Reg	FG12KJK		LA56YBB			**FG12KJK**

nesta

Matching, linking and integrating data

The simplest way to make it easy to match data across different IT systems is to consistently use the same unique identifier.

For place-based data, UK public sector bodies should use the Unique Property Reference Number (UPRN). Using the UPRN:

- Barnsley has generated an additional total rateable value for commercial properties of £170,000 by using data matching enabled by the UPRN.
- Nottingham's LLPG Team identified £40,000 of savings through closer working with the Business Rates team and the Valuation Office Agency in Nottingham
- Savings of £39,000 per election have been identified through the rationalisation of polling stations in Northumberland, together with the identification of 300 addresses that were not listed on the register of electors.

Source: https://www.geoplace.co.uk/addresses/uprn.
See further information about the benefits of UPRNs in this blog by Ordnance Survey.

Person data (or ‘customer’ or ‘client’ data) is more complex to integrate than place data because people move in and out of areas, and may be genuinely known by different names to different people.

Consider a situation where John Nicholson (a child) is known to a local authority social services department and has had interventions from youth offending teams and school welfare teams. John Nicolson (different spelling) is known to the police and to hospital services. If the council does not have an effective identity matching solution, a further referral to social services will not pick up the full picture, because there are separate and incomplete case files. Wrong decisions may be made regarding keeping John safe.

Public sector bodies can try manually matching individuals across different systems based on certain rules (deterministic matching), but this can be very time consuming.

Another approach is to use probabilistic matching tools, which can help save time and improve the quality of matching.

Deterministic systems are well-suited to matching smaller quantities of data with fewer complexities, while probabilistic tools are more powerful and provide greater accuracy with larger datasets.

What's the difference between deterministic and probabilistic matching systems?

	Deterministic Matching Systems	Probabilistic Matching Systems
What is it?	Programmed to match two or more records based on a set of predefined rules. For example, a rule might instruct the system to match two records with the same NHS Number, even if the addresses differ.	Leverages statistical theory and data analysis to establish the likelihood of two records being related. Assigns a percentage, e.g. 80%, to indicate the probability of a match.
Application	Works best with smaller datasets and where relatively fewer records (about 2 million) need to be matched.	Better suited for applications where records require matching across larger quantities of data and multiple databases.
Resourcing and skills	Can be implemented at less cost when the number of datasets, data attributes and matching rules is smaller. Advanced Excel skills required.	May require investment in new software, but potentially more cost-effective in the longer term and for larger or expanding databases. Experience in programming languages like R or Python as well as predictive analytics and machine learning required.
Scalability	Can be labour-intensive and costly to accommodate a growing database. Performance weakens as more customisation and rules are added.	Can more easily absorb, learn from and adapt to greater volumes of data, improving overall performance.

Public sector leaders need to create the space and culture for data to make a difference

Public sector leaders need to create an expectation that it's unacceptable to make a major decision or to reform a service without being aware of what the data shows.

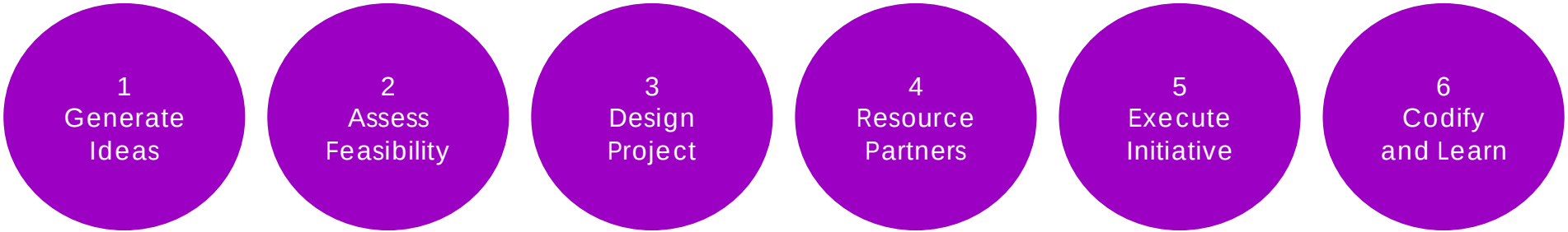
They also need to recognise that within their organisations, they may have data analysts who are frustrated data scientists. Those analysts need to be freed up from working on monthly dashboards and KPI reporting, and instead be given the time to work with service managers and frontline staff on higher value problems.

Module 8:

Setting up an Office of Data Analytics

If you've reached the stage where you'd like to create a permanent function that enables your organisation to collaborate with others on data analytics projects, what would that function look like?

In our view, an Office of Data Analytics (ODA) is not best understood as a team of data scientists, but as a function to oversee and project manage an end-to-end process made up of six steps:



This approach augments the ODA's own capacity by harnessing the talents of a city or region's wider public sector, and digital and tech sectors.

1 - Generate Ideas

Ideas for new data analytics projects can be crowdsourced from a number of different places, for example:

1) Local politicians - e.g. taken from manifesto commitments
2) Local authorities / public sector staff - via workshops
3) The public - potentially sourced via digital democracy platforms like Your Priorities.

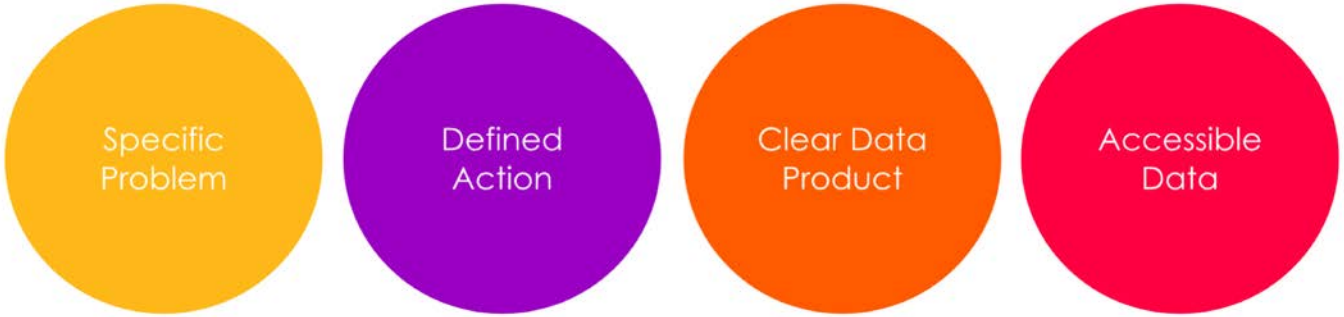
The ODA could offer several windows each year when ideas could be crowdsourced.

This open process would help demonstrate that the ODA is designed to benefit everyone across the region, and not replace the work of individual public sector organisations.

2 - Assess Feasibility

The ideas need to be vetted for their feasibility to be addressed *as data analytics enabled initiatives*.

To do this, the ODA can use the five problem types and four step method outlined in Module 3.



3 - Design Project

Once the problem and desired outcome (including success criteria) have been clearly defined, one of several project methodologies can be selected, including:

1) Experiment: the ODA designs and manages all facets of the project (i.e. data collection, analysis, evaluation, etc.) in collaboration with partners in a pilot.
2) Challenge Prize: the ODA offers a reward to the organisation or individual who provides the best data-enabled solution to a defined problem.
3) Hackathon(s): ODA organises hackathon(s) for smaller scale problems where data is abundant, or as an initial phase of the Experiment or Challenge Prize design methods.

4 - Resource Partners

The ODA assesses the team's internal capacity and identifies gaps for covering the data science, technology, project management and legal expertise requirements of each project.

Using its network with the wider innovation ecosystem, the ODA seeks partners to fill those gaps. Partners could include local authorities, local tech firms, data science SMEs, academic institutions, think tanks, civic hacker groups, etc.

These partners could be secured on a voluntary basis, or hired via a simplified procurement framework, put in place by the ODA.

5 - Execute Initiative

The data project is then executed. For a typical ODA pilot as described in this guide, this would typically entail the eight project phases outlined in Module 6:

1. Discovery: assessing the project's feasibility and refining its approach
2. Securing the commitment of project partners: identifying who needs to be involved and their roles and responsibilities
3. Information governance: putting in place agreements to share data
4. Data acquisition: getting hold of the required data
5. Data analysis and prototyping: analysing the data and building the first version of the data product
6. Testing and evaluating: trialing the data product in a real-world setting and measuring its results
7. Refining: improving the data product based on feedback
8. Scaling: putting the data product into permanent / wider use

6 - Codify and Learn

The final phase is for the ODA to understand the impact of each project and share lessons learned. This is likely to include:

- Assessing impact against the success criteria developed in step 3. This could potentially include rigorous evaluation methods such as a Randomised Control Trial.
- Communicating the lessons learned in regular blogs throughout the project, in a final project report, and through workshops and events.
- Releasing codes, templates, tools, and any other materials that could support others in applying the same work or designing something new.

Functions of the ODA

Overall, an ODA has two essential functions:

1. Doing: conducting 2-3 projects per annum based on the six stage process described in this pack.
2. Templatising: creating reusable codes of ethics, data standards, legal documents, process guides & open source tools.

If resources are available, two additional functions are desirable:

1. Convening & Coordinating: acting as a hub for the region's data science and policy community; supporting and nurturing the ecosystem.
2. Training: running workshops catered for specific needs of public sector leaders and data science practitioners, on legal, data & methods.

Public Sector Data Analytics

A Nesta Guide

Eddie Copeland
Nevena Dragicevic
Hilary Simpson
Tom Symons

nesta