

Modernising 3rd party access to DfE data.

**Including Partnership with the
ONS Secure Research Service**



**Neil McIvor and Richard Machen, Data Group, Department for Education
Nick O'Donnell and Sarah Fisher, Research Services and Data Access,
ONS**

19 Sept 2018



Department
for Education



Office for
National Statistics

Background

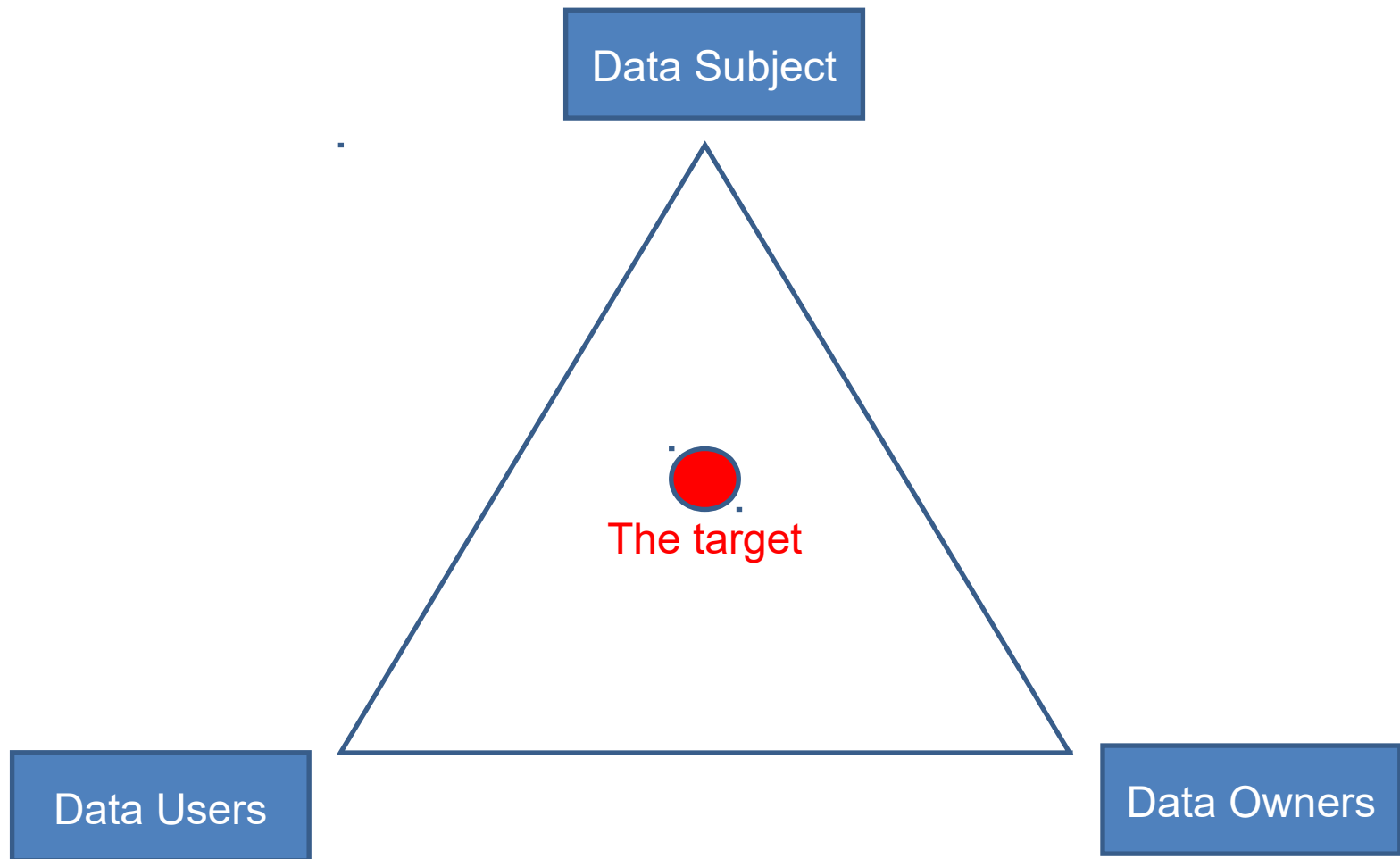
For many years, DfE has supported external research and re-use of centrally owned data for public benefit. That principle remain unaltered.

The mechanism by which that aspiration was enabled has become increasingly creaky over time. Namely:

- NPD sits across 350 organisations running 480 projects. That is costly and difficult to govern, and has risks.
- Those risks have materialised in the form of data breaches, as well as the form of significant challenge in the public domain. Public trust in such a disparate model is low.
- Each request that is met involves a bespoke cut of data being created, in part a result of the necessary governance associated with physically moving data. This is highly inefficient.
- The process is entirely demand led, with no controls over supply. It leaves users frustrated, with long waits.
- Over time, research has become more complex and DfE's data assets more sophisticated.



Modernising 3rd party access: balancing different perspectives.



So what did we set out to do?

In the last 12 months DfE's "modernising 3rd party access" project has spoken to many key stakeholders, and within the stakeholder community there can be divided opinion.

- Our users typically hate the slow governance model, but like the fact that they get the data at the end in the way they want to store locally.
- Privacy and information rights groups campaign hard that we need to do more to reduce risk – these sometimes land in the form of Office for Stats Regulation (OSR) or Information Commissioner's Office (ICO) formal correspondence.
- Ethics experts, anonymization experts and leaders in communicating data usage in public ways have helped develop our thinking about sensitivity and identification risk, drawn out the risks of current modes of operation

In drawing the issues together, and assessing solutions in Spring 2018, the investment the ONS is making in the Secure Research Service (SRS) positioned it as a leader in terms of its potential to meet significant proportions of user need as well as meet DfE strategic objectives. The two departments have been working in partnership to deliver an SRS service for education research built upon the **'5 safes' principle of safe people, safe projects, safe data, safe settings and safe outputs..**

Overview of the Modernising 3rd party access project

The work with ONS is one key strand of a wider set of modernisation changes DfE are delivering, as summarised below.

Strands of activity

1. Harmonised board for all DfE 3rd party data sharing decisions
2. Grow external ethics and scrutiny in DfE data sharing
3. Ensuring decisions and processes in respect of Data Sharing are GDPR and DPA 2018 compliant
4. Overhaul DfE's understanding and assessment of identification risk and sensitivity within personal data.
5. Standardise and improve the language of data sharing and the clear public narrative
6. Improve the user experience of applying for and using DfE data
7. Deliver a 'distribute access, not data' model of data access for the vast majority of research use of DfE data

The remainder of today focusses on the ONS Secure Research Service to ensure you are up to date on the element that affects you most.

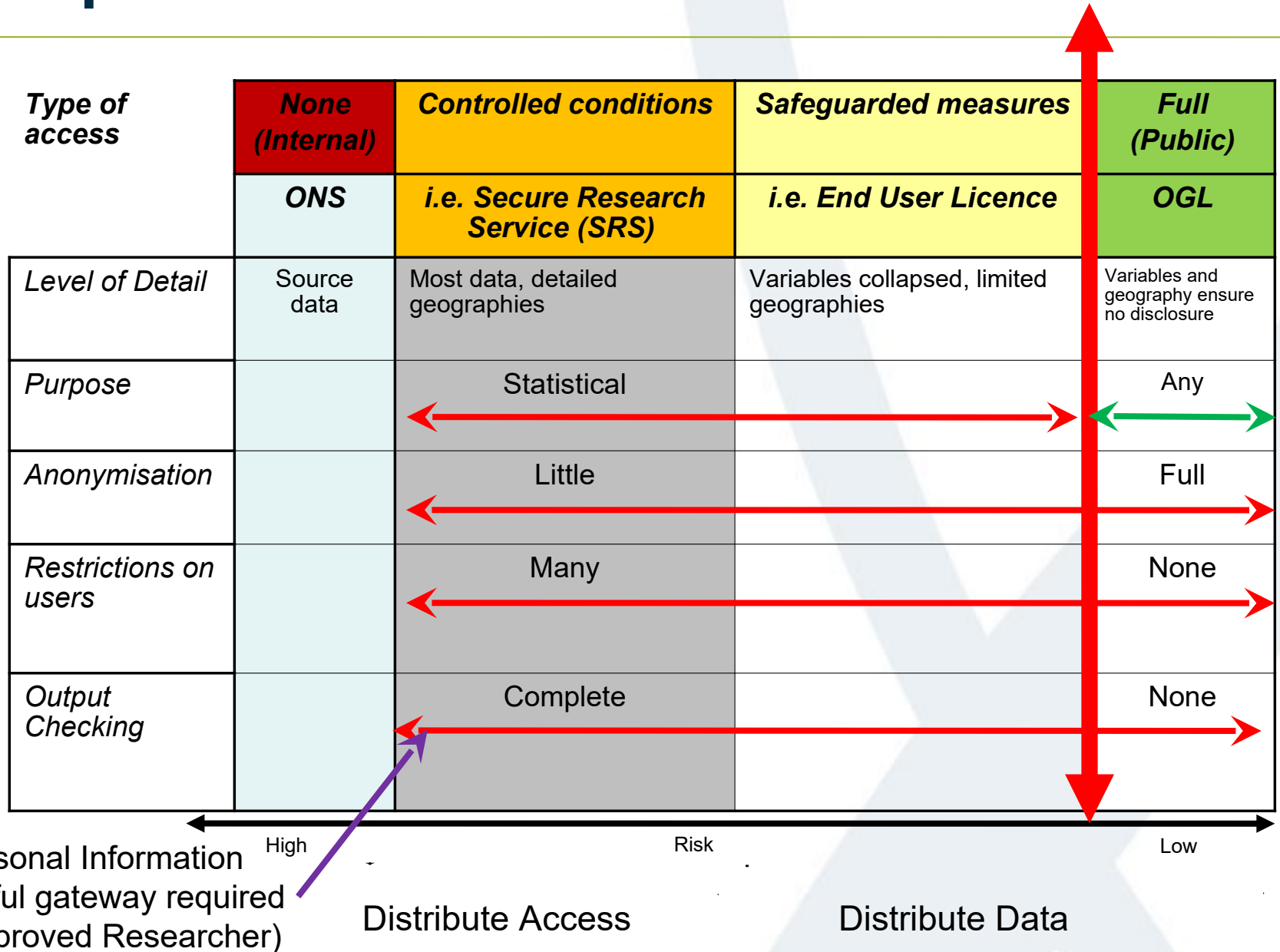
Secure Research Service: Context and Demonstration



Research access to Data

- The UK Statistics Authority (and the Office for National Statistics) has the statutory objective to promote and safeguard the production and publication of official statistics to “**serve the public good**” ...
- ...And to **promote and assist statistical research**, in particular by providing access to data.

Spectrum of Data Access

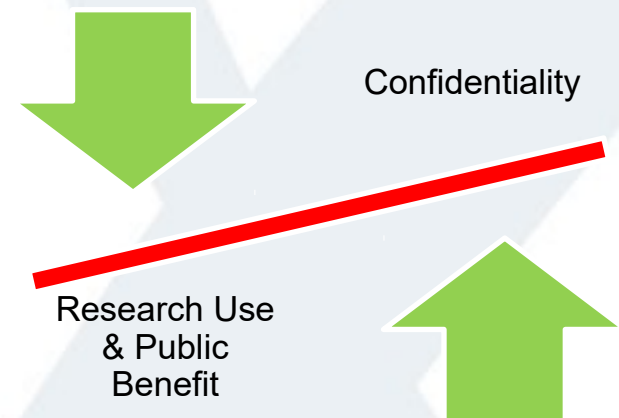


Stopping the Distribution of ONS Special Licence data

- ONS review in 2016/17
- Intruder test of Labour Force Survey and Living Costs and Food Survey data confirmed they were personal information
- Nine positive identifications in each test, using a range of variables together with published content
- Personal information (PI) is protected in law
- Public consultation concluded that PI should only be accessed in a safe setting, such as the Secure Research Service (SRS)
- Have stopped the distribution of these data

Our Strategy

- To seek to maximise Government and researcher use of microdata holdings, to deliver clear and significant public benefits
- Whilst protecting the confidentiality of data-subjects at all times



The Five Safes

Ensuring Safe Use of Data



Safe People

Trained and Accredited Researchers,
trusted to use data appropriately



Safe Projects

Data only used for valuable, ethical,
research that delivers clear public benefits



Safe Settings

Access to data only possible using our
secure technology systems



Safe Outputs

All research outputs checked to ensure
they cannot identify data subjects

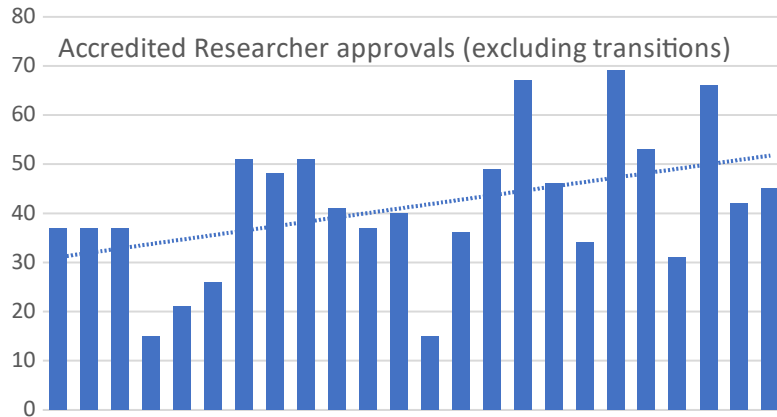


Safe Data

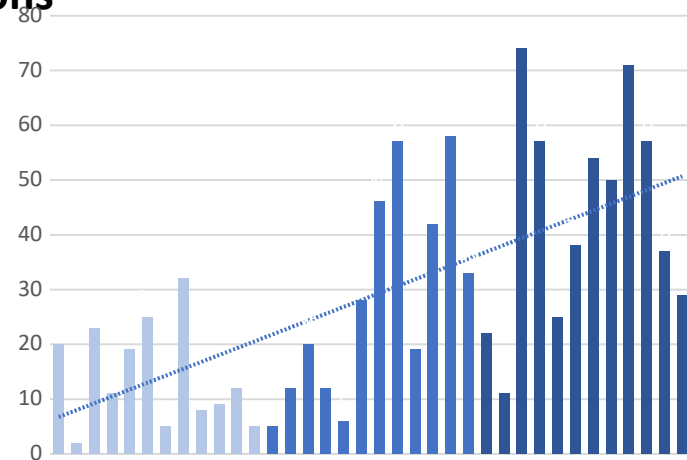
Researchers can only use data that have
been de-identified

Increasing Demand for access to ONS research data

Number of Accredited Researcher applications



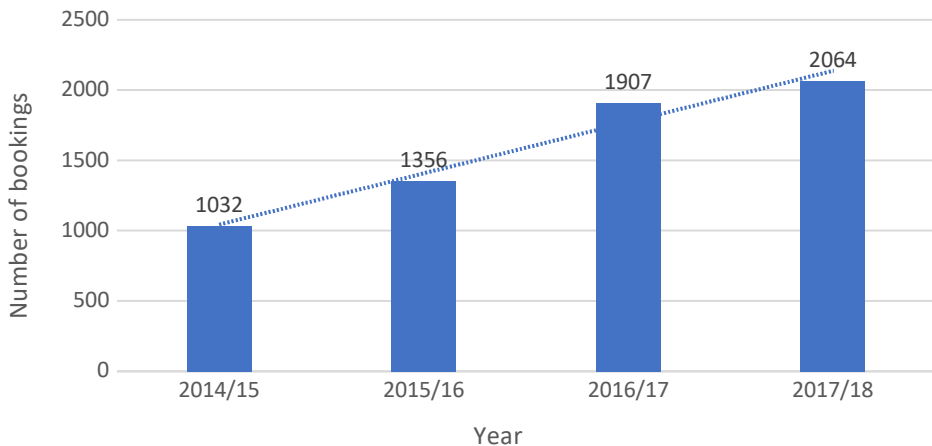
ONS Trained Researchers 2016-18



- There are almost 2,000 accredited researchers accessing ONS data for statistical research. Approx 600 researchers every year.

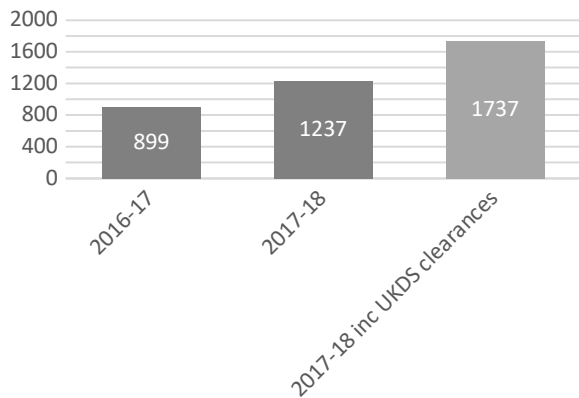
Increasing demand on using the Secure Research Service

Number of SRS bookings, 2014 - 2018

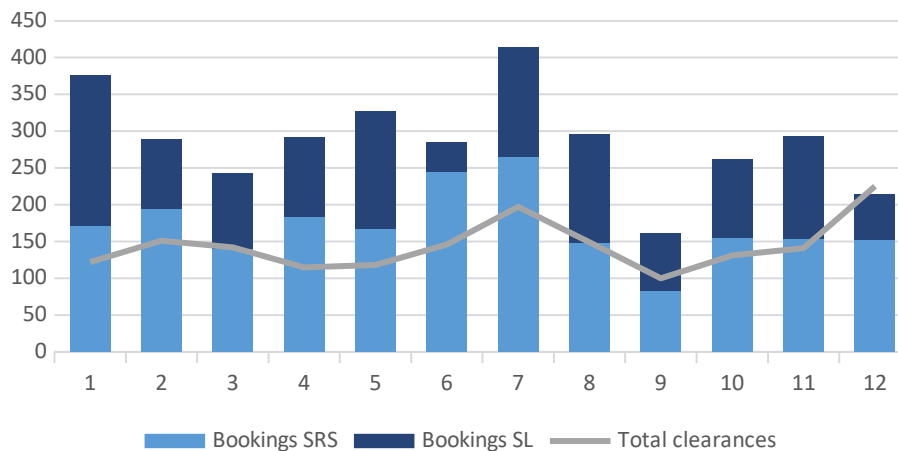


- Increasing number of bookings into SRS and Secure Lab (almost 1,400 bookings in 2017/18).
- The number of SRS bookings are expected to increase further once Digital Economy Act and new SRS developments implemented
- Increasing number of requests to clear outputs or code

Total output/code clearances



2017-18 SRS & Secure Lab bookings & data clearances



Moving Forward – Our Challenge



Our challenge:

Maximise public benefit (impact) by **increasing** use of government data for **research**.

- **ONS** Research data Government Research Data
 - Next phase of investment supporting Administrative Data for Research

We will deliver:

- EXCELLENT**
- GOOD**
- AVERAGE**

Awareness

Connectivity

Data

Capacity

Research Community: What do I get out of this?



What doesn't work:

Platform:

- ☒ Limited geographic availability
Only Govt. Depts can access remotely
- ☒ 'Noisy Neighbours'
lack of processing power

Application Process:

- ☒ 'Paper based' – very 'manual'
- ☒ Takes too long
- ☒ Cannot track progress

More Data:

What we are delivering

Secure Cloud Technology (AWS)

- ✓ Connect from any 'approved' **location**
- ✓ Scalable compute power
- ✓ Support for Data Science tools

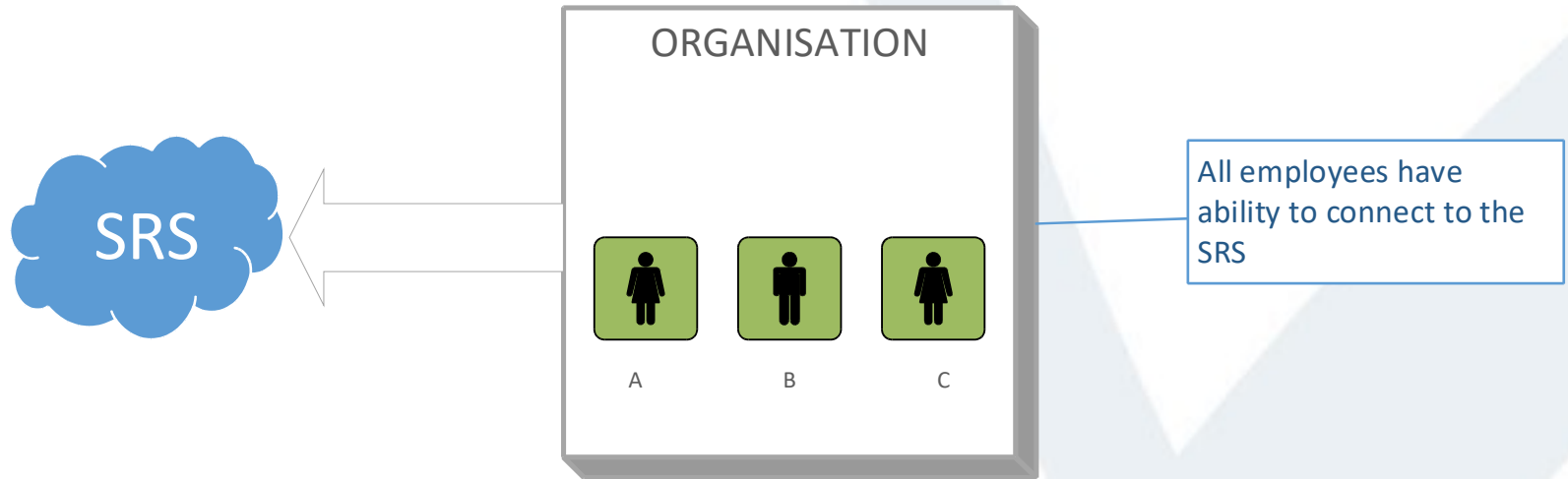
Online Application & Workflow

- ✓ Efficient
- ✓ Reduce bureaucracy
- ✓ Greater transparency

- ONS
- X Government

SRS Connectivity

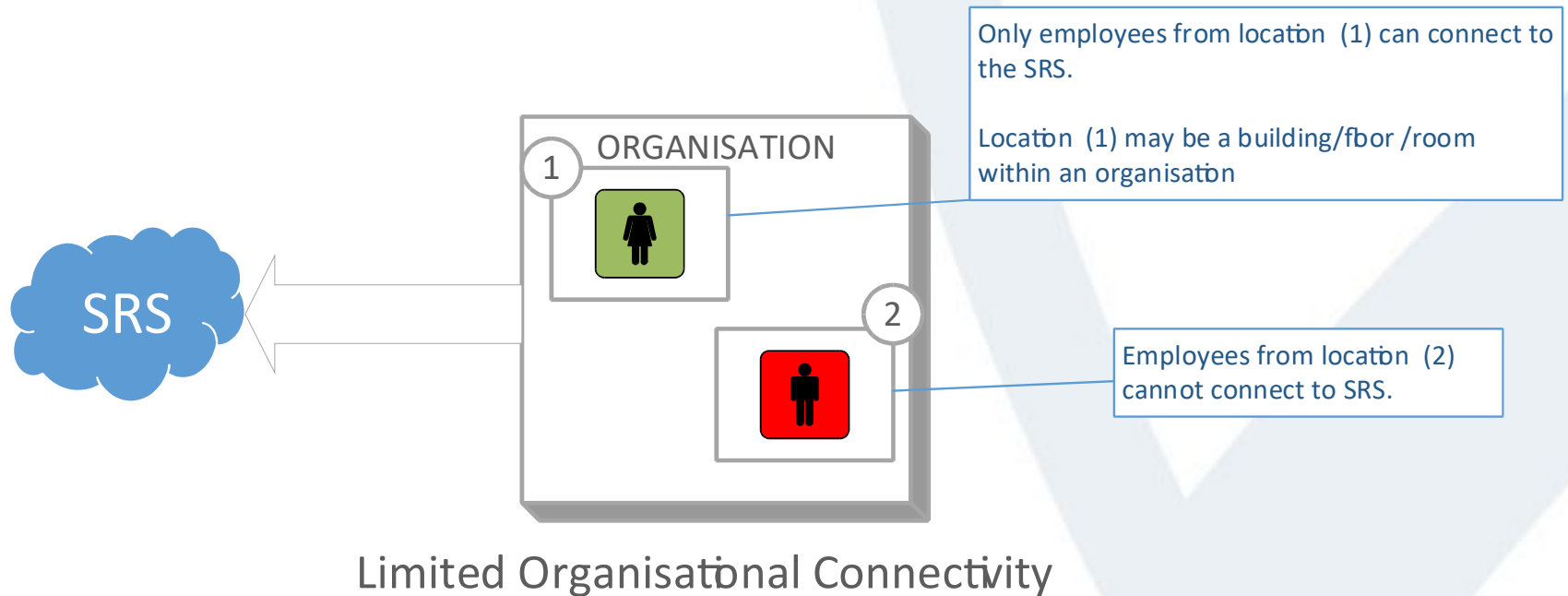
Full Organisational Connectivity



Full Organisational Connectivity

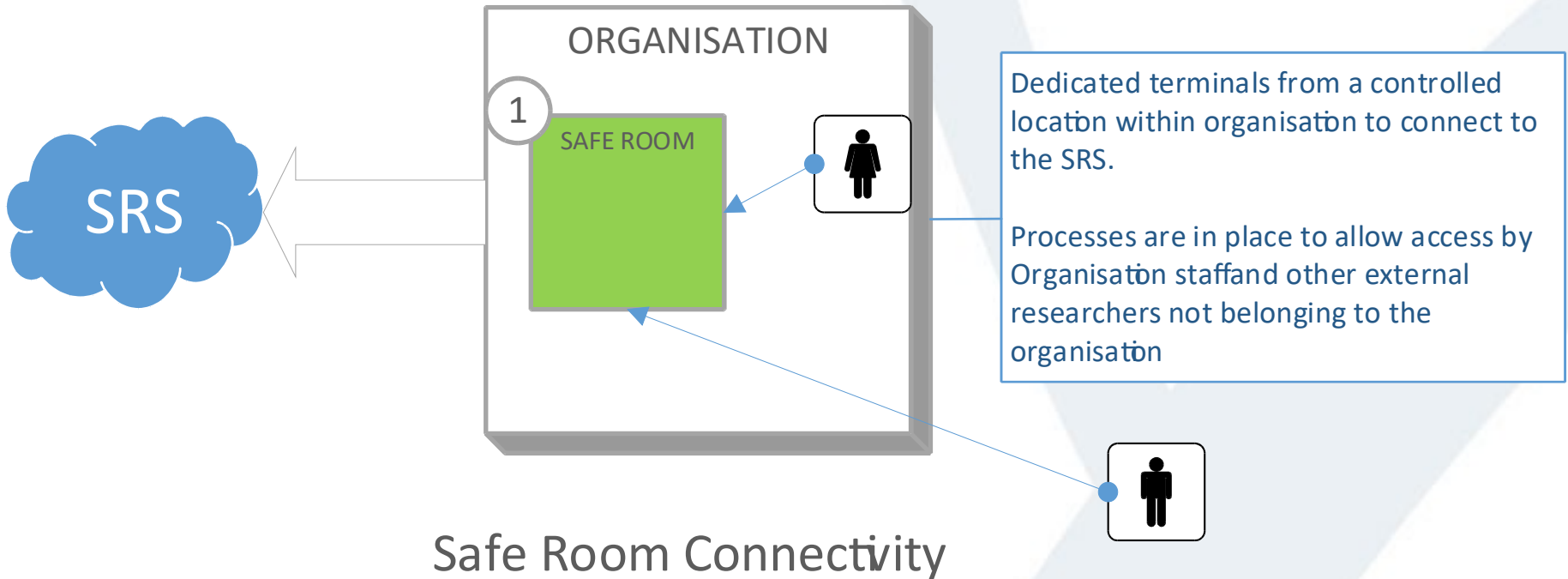
SRS Connectivity

Limited Organisational Connectivity



SRS Connectivity

Secure Room / Safe Setting Connectivity



SRS Connectivity - Criteria

Organisational Criteria:

Does the organisation fully control access to **all** its buildings and premises?

Can the organisation ensure members of the public or people not working on your behalf are escorted?

Does the organisation have policies in place to stop tailgating?

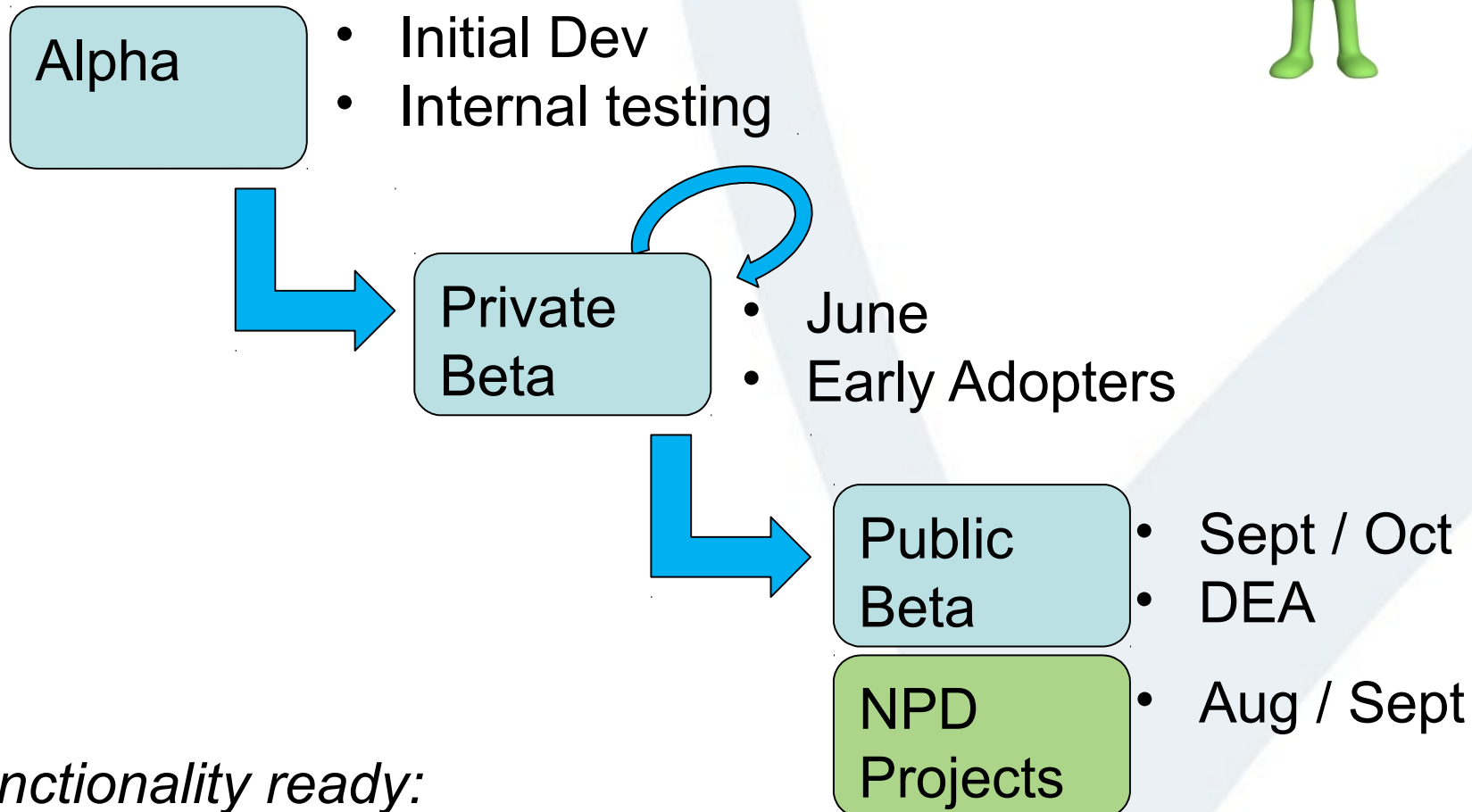
Does the organisation have procedures in place to know who is within any buildings at any given time?

Does the organisation ensure all people within its buildings are visibly identified? (e.g. name badges)

Are there policies and procedures in place to manage the physical security of the organisation?

Are there policies and procedures in place to manage personnel recruitment within the organisation?

Roadmap & MVP



MVP: *Functionality ready:*

- ✓ Allow first research project to be approved
- ✓ Research can start in new environment, in time for DEA
- Connectivity

SRS Connectivity - Criteria

Technical Criteria:

Does the organisation have a recognisable, public facing unique Internet Protocol (IP) address?

Does the organisation have an Information & Communications Technology (ICT) policy, covering device management, network management, patch management?

Can the organisation ensure equipment used to connect to the SRS will be managed under an organisational ICT policy or equivalent?

Can the organisation ensure it does not share its IT Infrastructure with other organisations?

Can the organisation confirm that those machines accessing the SRS meet the following specification?

Operating system must be either Windows 7, Windows 10 or MAC OS X.

4GB RAM;

1.4Ghz CPU: and

50GB free HDD space.

Internet connection – uses organisation network or secure organisation WiFi.

Minimum speed of internet connection; 2.8 Mbps

Use of a Web browser Chrome or Firefox, (not Internet Explorer)

SRS Connectivity - Criteria

Can the organisation ensure its employees only use IT equipment that belong to the organisation for the purposes of conducting business or access to the SRS?

Can the organisation ensure workstations or laptops are assigned to specific individuals who work for the organisation?

Can the organisation confirm that no public access machines will have the ability to connect to the SRS?

Can the organisation confirm that equipment belonging to the organisation used outside of the organisational premises (e.g. home working etc.) will never be used to access the SRS environment without express permission from ONS?

For those machines used to access the SRS, can the organisation confirm it has measures in place to ensure they cannot be overlooked?

Does the organisation protect, log and monitor access to its computer systems and networks?

Can the organisation ensure any laptops used to access the SRS have appropriate disk encryption?

If required, can the organisation confirm in principle that it will allow a simple security inspection carried out by ONS of the premises, systems and services to perform compliance or audit checking?

SRS Connectivity - Criteria

Governance Criteria:

Can the organisation nominate a Senior Responsible Owner with appropriate authority to take ownership of the agreement between the organisation and ONS / SRS?

Can the organisation nominate a person as a single point of contact for all matters relating to SRS operational access?

Can the organisation nominate a person as a single point of contact for all matters relating to SRS technical and infrastructure connectivity tasks?

Can the organisation nominate a responsible person with appropriate authority as a single point of contact for all matters relating to SRS security and assurance?

SRS Platform Demonstration



NPD – Training Process:

For those researchers who are not trained:

Research.support@ons.gov.uk

Attend training (sign SRS agreement)

Pass online assessment

Demonstrate suitable experience / qualifications

(Projects can start before all researchers are trained)

Forthcoming Training Dates

4th September, London

10th September, Sheffield

14th September, Durham

24th September – Titchfield

2nd October – Drummond Gate

10th September, Titchfield

25th October, Leeds

26th October, Leeds

7th November, London

10th November, London

Impact of Research

During 2017-18 we have worked with the UK Data Service to increase the publicity of research carried out in the SRS. We published 10 case studies via the [UK Data Service website](#)

“ONS research data are essential in allowing us to understand what’s happening to living standards, and to use that insight to inform the policy debate. The event in February was an excellent showcase of the opportunities the data offer”

Conor D’Arcy, Resolution Foundation

“The Tech Nation reports could not have had the impact they have without the use of ONS data. The public benefit of access to data for research from the ONS is clear to most of us, but as a research community, we have a responsibility to continue raising awareness about the role of data driven evidence in informing policy and decision making.”

George Windsor, Tech City

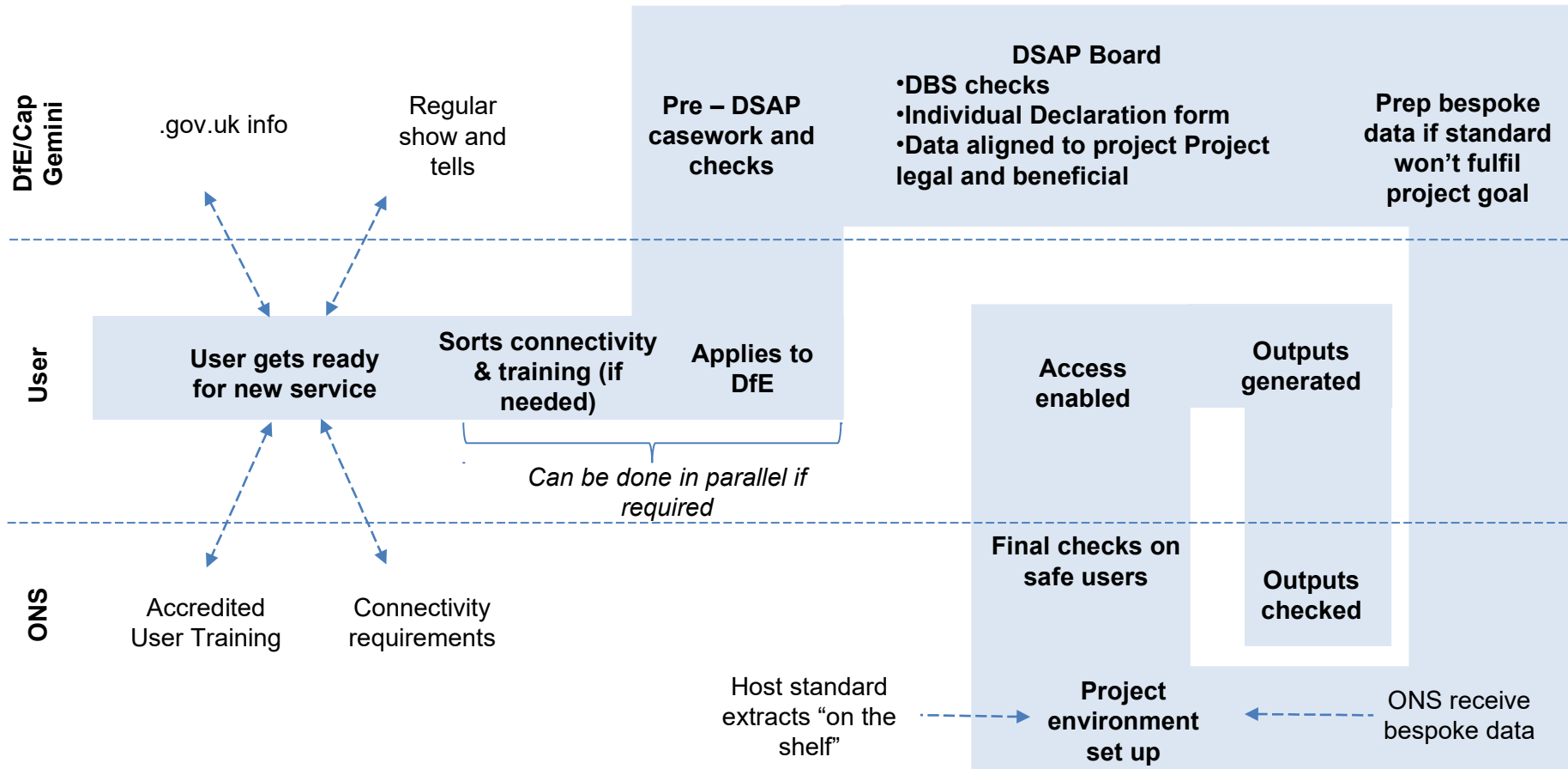
“It was great to participate in an event that brought together the producers and users of data. The CBI’s success in influencing regional policy development was made possible through our access to the ONS’s microdata. We all have a stake in shaping the UK’s future and high quality, accessible data helps us understand the problems and design the solutions.”

Anna Leach, CBI

[ONS Research Excellence Award inaugurated in 2018](#)



Summary : piecing the components together : The new business process



Our first two projects (Education Policy Institute and Durham University) have everything ready for them once they can provide their connectivity application. ONS standby to support them in that process.

Several other projects for SRS will likely have data created w/c 18th September.



End September 'switch on'.

It is our intention to have a soft launch of the new service by the end of September.

This will involve updating .gov.uk in relation to the new ways of working.

We will deploy User Researchers to gain feedback from early adopters.

Early life:

1. Skype walk throughs of new service / chance to ask questions. 3 times a week (find on .gov.uk).
2. Get yourself trained, accredited and connected.
3. Lots of standard extracts: Early Years Census, School Census, absence and exclusions, EYFSP, Key Stage 2, Key Stage 4 and Key Stage 4 all linked to prior attainment: 5 years worth of data, all 'Tier 2 and Tier 4' data items in old money. Access for these is faster.
4. Bespoke extracts can be requested still : please prioritise your needs!
5. Application forms and data tables available from data.sharing@education.gov.uk once .gov.uk is updated, as will the timings / details for the walk through of application stage.

Developing on the initial service: Our plans to March 2019

- A Digital Front End based on GDS standards, that meets user needs in respect of applying for access data.
- Dummy / Synthetic data, allowing education researchers to better plan for projects, develop code whilst waiting for real data, and to prepare code in advance of using physical datalabs if that is their choice.
- More standard extracts. Serving up the data most commonly sought, in the structure people want to consume it, with sensible pre-processing applied to save users time. This allows us to fully realise the efficiency angle of working with SRS
- A solution to make additional proprietary software available, as needed, either through ONS-provision or introduction of a “Bring your own Software” mode
- Increased advice and support from ONS team, when conducting research, as they become more familiar with the NPD