DfE NPD & Data Sharing Update

NPD User Group

23 January 2020

Gary Connell

Head of Platform Operations and National Pupil Database (NPD)

Data Operations Division | Data Directorate

Department for Education

Agenda

1. Introduction and the case for change

2. Examples of change

3. Update on the ONS SRS

4. Find & Explore

5. NPD Data Packages

6. External access to Longitudinal Education Outcomes (LEO) and other linked datasets

7. Questions & Answers / Discussion

Department for Education

1. Introduction and the case for change

Department
for Education

The National Pupil Database (NPD)

- The National Pupil Database (NPD) is used by external organisations for three high level purposes:

> *"The NPD is one of the richest education datasets in the world and provides invaluable evidence on educational performance to inform independent research, as well as studies commissioned by the department."*

  1. To run operational processes that are needed to *'get the job done'* of running education.

  2. To undertake research and / or analysis to understand what works, what doesn't and how to improve/challenge policy making

  3. To provide externally led information services which add value above and beyond the DfE offer.

- DfE remains fully committed to facilitating external research and recognises the valuable contribution this makes to the evidence base surrounding the education system in England. The NPD holds attainment and characteristics data dating back to 2002 and DfE has tried to support <u>appropriate</u> access to that for many years.

- Over time, the processes to support the NPD were beginning to creak. The measures we have put in place are designed to improve how we work with external researchers whilst continuing to safeguard children's privacy and maintain public confidence in our processes.

Department for Education

The case for change

1. External use of 'education' data beyond the NPD was previously governed with inconsistent processes;

2. The volume of requests to access data continued to grow, but more challenging was the fact that the complexity was also growing, with more focus on linking datasets within, and beyond, education.

3. Expectation and best practice. As technology continues to evolve, so must our ways of enabling access.

4. We now have more sensitive (and potentially powerful) data at our disposal. With that comes greater responsibility and public expectation regarding how we safeguard their personal data.

5. Changes to data protection legislation and, in particular, the introduction of the General Data Protection Regulation (GDPR) and the 2018 Data Protection Act.

Department for Education

2. Examples of changes we have made

Department
for Education

DfE wide governance

In April 2018, we harmonised governance within DfE creating a board with oversight of all DfE data sharing, with appropriate external ethical challenge:
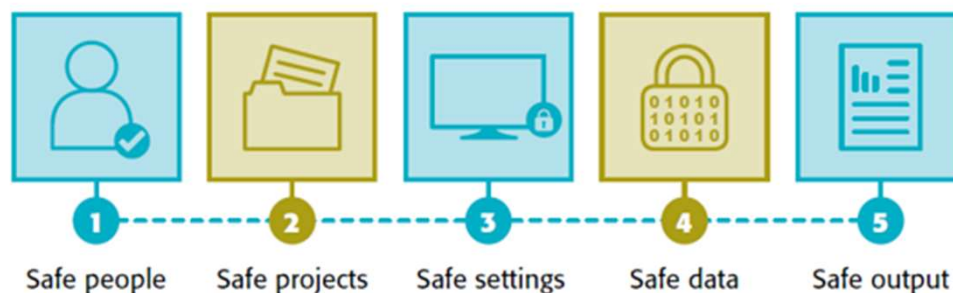
- The Data Sharing Approvals Panel (DSAP) now covers all external data sharing by DfE;

- DSAP membership comprises of senior civil servants in DfE as well 4 external members to scrutinise the ethics of any proposed data share;

- Panel meets weekly so as to not be a 'blocker' for progressing data sharing requests;

- Application process to support decision making aligned to data protection regulations

- Information captured about approved data shares used to inform clearer and more holistic transparency publications

Department for Education

ONS Secure Research Service

Since September 2018, the model for enabling third party access to DfE data is to adopt the '5 safes' model developed by the ONS, and utilise the infrastructure provided by their 'Secure Research Service (SRS)'. The DfE model works by combining 5 key principles of safety:

- **Safe People:** Individual researchers must undertake a formal training event, pass a test, and demonstrate their professional capacity to undertake statistical evaluation of such datasets before being classed as a ONS accredited researcher.

- **Safe Projects**: The projects are evaluated by DSAP to ensure that the project is ethical and beneficial to society, and that appropriate aspects of data protection law are taken care of.

- **Safe Data:** The DSAP panel ensures that the data being shared is proportionate to the project and does not invite unacceptable levels of risk.

- **Safe Settings:** The ONS provide access to the data via a small number of physical office spaces (physical labs) or via vetting the researcher's physical and technical infrastructure and granting them 'connectivity' to access data remotely (remote labs).

- **Safe Outputs:** Prior to allowing a user to take some findings of research out of the SRS, the intended outputs are vetted by ONS against agreed disclosure rules to ensure individuals cannot be identified by any published outputs.

1 Safe people   2 Safe projects   3 Safe settings   4 Safe data   5 Safe output

3. Update on the ONS SRS

Department
for Education

NPD transition to ONS Secure Research Service

- **ONS have made a number of improvements to the current ONS SRS infrastructure which are working well and have brought a period of stability to the service offering.**

  ➢ Storage, File Servers, Computer Servers all significantly increased

  ➢ SQL 2016 Server created and DfE data loaded

  ➢ New process for achieving remote connectivity to the SRS from organisations own premises (now referred to as 'Assured Organisational Connectivity')

- **Following user feedback, DfE and ONS agreed a bespoke statistical disclosure control policy for outputs using DfE data**

- **There are currently 123 projects successfully using DfE data (predominately NPD data) live in the ONS**

Department for Education

Development of a new SRS

- ONS working to develop a 'new' Secure Research Service

  - ➤ New SRS will be Citrix-based virtual desktops, hosted at UK Cloud

  - ➤ No new training required for users

- Proof of Concept delivered on schedule and works as expected

- "Go Live" date scheduled for 31st March 2020, and work on schedule to achieve this

- Significant user testing being planned throughout March, to include NPD users

- Migration from current to new SRS will not happen until ONS can demonstrate consistent, and acceptable, performance levels

Department for Education

4. Find and Explore

Department
for Education

Background and context

- **Initial research with over 50 users helped us identify four broad categories, each with their own pain points, across the 'end to end' NPD process:**

  1. **Planning:** It isn't easy for users to understand and comprehend what data is held within the NPD

  2. **Applying:** Many users find completion of the application form onerous, time consuming and complex

  3. **Communicating:** Users are frustrated at not knowing how their application is progressing

  4. **Accessing Data:** Users are frustrated that they didn't know when they are going to be given access to the data

Department
for Education

The problem we are trying to solve

- The previous method for a user to find what data they need is to access a 'data tables' document which is a 42 tab Excel file

- Completion of the data tables are prone to user error, take users days' to use and are not accessible

- A significant proportion of applications cannot be processed first time due to incorrect information included

- All support is handled by the NPD caseworkers, which takes time away from processing other cases, clogs up the system, frustrates users and makes things inefficient

- The existing process puts off potential 'new' users from applying for NPD data, meaning we aren't maximising the use of our world class public resource
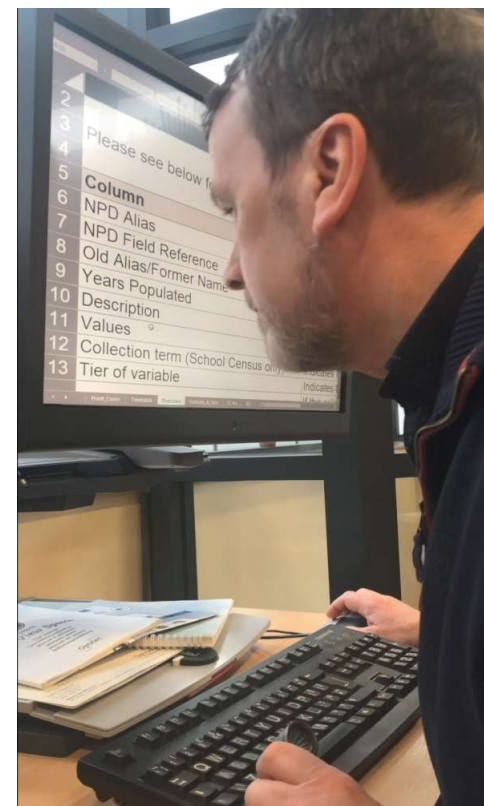
A partially sighted user using the existing Excel Data Tables document with a magnifying glass

Department for Education

What is Find & Explore?

- Essentially a metadata tool that enables all types of users to better understand what data is held within the NPD

- A digital and more accessible version of the Data Tables

- A starting point for further modernisation work around the NPD

- It will provide users with information about the data that is available in the NPD in order to support their applications

- It provides information about next steps on how to apply, but (deliberately…) stops short of starting the actual application.

Department for Education

BETA This is a new service – we are improving how the metadata from the NPD Data Tables can be displayed and therefore the content might change. Your feedback will help us to improve it.

Find and explore data in the National Pupil Database

Search

[   ] [Search]

Categories

Attainment

Information on pupil performance including school and exam registration, foundation stage profiles, value-added calculations, prior attainment, predicted scores, exam results and pupil inclusion.

Demographics

Information about pupils and their school including personal and family data, where they live, the school they go to and when they started there.

Absence and exclusion

Information on all recorded authorised and unauthorised absence and exclusion. Includes reasons, dates and the amount of time missed in sessions.

Children in need and Children looked after

Information on children in need (CIN) referred to children's services and children looked after (CLA) who're in education and looked after by local authorities.

No Category

Department for Education

Current position

- **Find & Explore is already live (public beta) on GOV.UK following a successful testing with users**

- **We are currently exploring ways in which aspects of the NPD application process can be incorporated into Find & Explore**

- **We also close to releasing an alternative 'Dataset View', following feedback from users**

- **This isn't the finished article yet, and it won't be for a while**

- **Please have a look at https://find-npd-data.education.gov.uk/ and take 10 minutes to provide some feedback**

- **Your feedback will help us iterate the tool to better meet your needs!**

Department for Education

Potential future iterations

We are considering a range of options to iterate, and improve, Find & Explore. These include:

- Improve the data descriptions in Find & Explore to make it easier for all types of users to understand and comprehend the data held, and available, in the NPD

- Incorporate ILR, HESA & LEO metadata so that users have a 'one stop' shop for DfE metadata

- Introduce a Data Package view so that users can see what is in, and request, access to specific Data Packages

5. NPD Data Packages

Department
for Education

What are 'Data Packages'

- Data Packages (DPs) are what we previously referred to as 'Standard Extracts'. We have renamed them to avoid confusion with the Standard Extracts that previously existed.

- DPs have been designed as 'bundles' of NPD data items, grouped together to provide the greatest likelihood of approved researchers obtaining the data they require in the most efficient way. Variables are grouped according to:

  1. **Being 'conceptually related' (e.g. demographic, attainment, CIN-CLA)**

  2. **Similar within-group identifiability and sensitivity ratings**

  3. **Educational phase – one of either:**

     - Early years (Age = 2 or 3, tbc)

     - Primary (NC Year Actual = Reception – Y6)

     - Secondary (NC Year Actual = Y7 – Y11)

     - Post-16 (NC Year Actual = Y12 – Y13)

  4. **Academic year that the data relates to**

High level illustration

Children in Need, Children Looked After	Sensitive	Absence and Exclusion	Demographics	Attainment (with prior attainment)
Standard Extract: **CIN-CLA** (Children in Need & Children Looked After) Educational Phases: Early Years, Primary, Secondary, Post-16 Includes information from educational establishments and social services on children who require additional support to reach satisfactory levels of development.	Standard Extract: **Sensitive** Educational Phases: Early Years, Primary, Secondary, Post-16 Includes sensitive data not classed as 'special category' under GDPR but is treated equally (e.g., refugee, service child); and meaningless identifiers (e.g., Unique Learner Number).	Standard Extract: **Absence-Exclusion** Educational Phases: Primary, Secondary, Post-16 Includes all absence and exclusion data (occasions and reasons), as well as demographic information (e.g., ethnicity, special educational needs provision).	Standard Extract: **Demographics** Educational Phases: Early Years, Primary, Secondary, Post-16 Includes all demographic information relating to age, gender, mobility, geographic location, socio-economic status, ethnicity, language, special educational needs, and school registration details.	Standard Extract: **Attainment** Educational Phases: Early Years, Primary, Secondary, Post-16 Includes all attainment and prior attainment measures captured at each educational phase (Foundation Stage Profile, SATS, GCSEs and equivalent, A-Levels and equivalent.

30% of all data requests fulfilled by lowest risk data packages

Department for Education

Progress and next steps

- Development and testing has, so far, initially focused on demographics and attainment data packages.

- We currently have available 12 data packages across two educational phases:

	Secondary		Post - 16	
	Demographics	Attainment	Demographics	Attainment
2015-16	SC, AP, Pupil	KS4, Prior	SC, AP, Pupil	KS5, Prior
2016-17	SC, AP, Pupil	KS4, Prior	SC, AP, Pupil	KS5, Prior
2017-18	SC, AP, Pupil	KS4, Prior	SC, AP, Pupil	KS5, Prior

- As well as internal testing, during 'private-beta' we were able to test usability with DPs being successfully used for two live (real) projects in the SRS

- To build on this learning, we expect to update GOV.UK with further details of the 'public-beta' launch of data packages – expected within the few weeks!!

Department for Education

6. External access to Longitudinal Education Outcomes (LEO) and other linked datasets

Department for Education

What is LEO?

- **LEO is an analytical research dataset created through linking of education data from England, Scotland and Wales to employment, benefits and earnings information from the Department for Work and Pensions (DWP) and Her Majesty's Revenue and Customs (HMRC)**

- **The LEO dataset is an emerging key data source in DfE with the data being used to compare students' levels of education to their levels of employment and earnings in later life in order to:**

  ➢ enable comparisons of the performances of schools, colleges and universities

  ➢ provide statistical information to support education and career decisions

  ➢ evaluate and monitor the impact of education or training on outcomes

  ➢ support government decision making which improve services

Department for Education

Our ambition is to <u>safely</u> maximise the use of LEO

- DfE committed to opening access to data for maximum reuse – but this will only be done where we can ensure we safeguard data subject privacy and maintain public confidence in our processes.

- DfE is working to enable third party research on LEO data via the provision of pseudo-anonymised standard extracts / data packages available via the ONS SRS.

  - ➢ Due to LEO file size and data structures, this will only be done within the new iteration of the SRS to ensure stability of performance

- Whilst we wait for the new SRS to be developed, DfE have approved 8 projects to participate in a pilot exercise where permitted users from third parties will be brought into DfE on secondment to complete LEO research projects on DfE systems

Department for Education

7.

Department
for Education

For more information

Contact us

- By email: data.sharing@education.gov.uk.

Visit the website

- How to access Department for Education (DfE) data extracts: https://www.gov.uk/guidance/how-to-access-department-for-education-dfe-data-extracts

Department
for Education