

Data Modernisation Division

A presentation to the NPD User Group.

Sept 2016

Iain Bradley, Deputy Director & Head of Data Modernisation Division



Department
for Education

Structure

- What are the problems we identify within education sector data flows.
- Who am I and what is Data Modernisation Division going to do to improve things?
- Focus on National Pupil Database.
- Where do we go from here?



WHAT ARE THE PROBLEMS WE ARE TRYING TO SOLVE?

1. Inconsistent Standards across the supplier community makes re-use costly.

4. Inconsistent LA role. Schools not owning data as they should in some areas.

6. Data (e.g. census) errors highlighted to schools, that should have been sorted by now, haven't been.

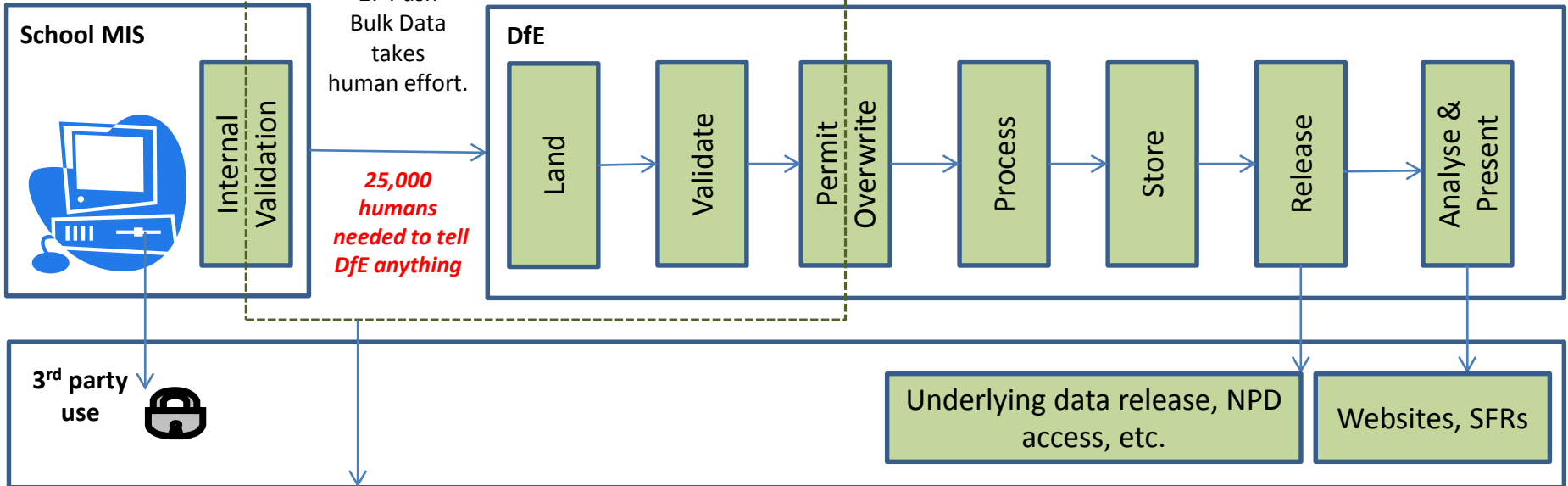
8. Data in disparate areas. Hard to move and blend.

9. Takes too long to reach this point. Perfect is the enemy of the good.

Varied LA role

2. Push Bulk Data takes human effort.

25,000 humans needed to tell DfE anything



3. Data moved even if not changed means unfocused cleaning and wasted bandwidth. 99% of ethnicities don't change / 60% of census data we have already collected.

5. School and DfE data allowed to differ by over-writing data once in DfE systems.

7. Overly complex / costly matching, for example National Pupil Database, to compensate for poor standards upstream.

10. Cost / time of getting hold of data limits 3rd party use.

11. Designed for expert users to engage with via print outs and desktops.

Data Modernisation Division: who are we and what are we going to do about it?

- Work to Iain Bell, Director of Data and Head of Data and Education Standards Analysis Group.
- Borne out of a smaller unit that up until March was focussed purely on Data Exchange/standards and modernising the School Performance Tables website. Working with IT colleagues to establish Agile ways of working and digital delivery. A reputation for user collaboration, iterative design, and regular feedback and test.
- Data Exchange ‘landing’ within DfE, the white paper and now Machinery of Government changes mean that the agenda around data improvement and digital capability has grown. Thus a new division, Data Modernisation Division, was created in June. Our remit includes:
 - Data Exchange & ISB standards – how we can move data more effectively into DfE and around the sector.
 - RAISE online – replacing the existing service in a way built on user need.
 - Teacher Vacancy Service – building/enabling the ‘one stop’ set out in the white paper.
 - Parent Portal – what would it contain and what is our role?
 - National Pupil Database – management of existing process and change for the future
 - Growing digital skills and capability within DfE



Data Exchange : What is it and where are we?

The problem statement

The cost of data movement is high, pace is slow, flexibility is low, and re-use is difficult. More efficient ways of moving data would bring benefits to many across the sector.

What do we think the solution will look like?

- 3 main ways into DfE: MIS system to DfE API, File Transfer and Web Form. They should look and feel similar no matter which bit of DfE you are interacting with.
- Common standards at the point of data movement. Implementing the Information Standards Board standards, but working with suppliers and others as we go to evolve the model.
- Move towards focussing on what has changed within the data, and moving data as frequently as it is needed (not more than once per month into DfE teams).
- A shift away from the concept of 'collections', to a concept of supporting use cases with data, ironing out the duplication as we build.

Where are we now?

- Proof of Concept work focussed on the API element – the most complex (and beneficial) – a need to test and build confidence in what we *think* is the right open technical standard (Systems Interoperability Framework) and what we *think* are the right data standards (Information Standards Board).



RAISE Online Replacement : What is it and where are we?

The problem statement

The cost of exposing detailed performance data to schools can be reduced, and at the same time by rebuilding on modern technology and in collaboration with users we can deliver a better user experience.

What do we think the solution will look like?

- DfE will build a service which allows schools to identify themselves, log in and access data and graphics that help them understand their performance in context.
- We will also make the underlying data available to 3rd parties under a licensing agreement. Subject to security & other capabilities, along with commitment to use a standard set of definitions, 3rd parties will have access to a level playing field upon which to develop value added services to schools.

Where are we now?

- We are almost half way through the Alpha phase. This will consist of prototyping various parts of the solution and conducting further research with schools. We will prototype secure log in, school search, data extraction for 3rd parties, hosting of external report and a set of sample reports & charts.
- Having both the DfE solution and licensing agreements in place for March 2017 is the aim, allowing time for awareness raising and iteration as we move towards peak 'RAISE season' later in the year.



Teacher Vacancy Service

The problem statement

The cost of advertising vacancies can be high for schools. For teachers seeking jobs, knowing where to go to look at vacancies can be difficult. Confidence that *'I'm looking at all the opportunities in area X'* is low. A vacancy site that was free to all schools and contained 100% coverage would be a good thing to have.

What do we think the solution will look like?

- We need a very easy way for schools to get information to us. This might be a web form with certain standards, but we'll explore whether web-scraping alone could work too.
- To enable the wider market, we need to work out how we make the vacancy data we are provided available to credible 3rd parties in a way that supports re-use, but also in a way that schools are comfortable with.
- The big risk is a 'chicken and egg' conundrum. If we don't get nearly all schools using it, then job seekers will disengage. If we don't have nearly all jobseekers using it, schools may not see the benefit of using it.

Where are we now?

- Discovery work confirmed the problem statement. Alpha phase will concentrate on the ways of moving the information (both into DfE from schools, and out of DfE to providers). The need for us to actually build the website will be driven by ongoing user research. At the moment, that doesn't feel like the tricky bit, and hence is not the focus of Alpha work.



Parent Portal

The problem statement

Supporting your child through the English Education System is not as easy as it could be, and a national 'portal' for parents could help address certain elements of this.

What do we think the solution will look like?

- Too early to say. We are currently in 'pre-discovery', with a need to test and build a common picture of what the new teams within No 10 and DfE consider as priorities to tackle within what is a very wide problem statement.

Wider Digital Skills and Capability

Where are we now?

We have been successful in growing pockets of digital capability within DfE. We are currently thinking through the right structures and roles to take that to a higher level as the agenda is growing.

Machinery of Government changes are also providing the right environment for us to refresh the Data Strategy, ensure buy in across the new look department and frame the right priority projects to ensure that the supply chain of data is as efficient as possible.



Department
for Education

Hang on a minute...

Yeah yeah, that's all nice to hear, but what's that got to do with NPD, which is why I came here today?!

- A. As the data we capture changes through Data Exchange, so too will the data we can make available to 3rd parties.
- B. The Agile approach taken on the other change projects will be taken on for our NPD improving access agenda.



National Pupil Database : Why me?!

- The National Pupil Database is a fundamental part of the DfE analytical and operational jigsaw. The production and provision of it is a big 'Business As Usual Activity'.
- But there also several areas of change planned.
- And it was to maximise the chance of delivering the changes & improvements that we have positioned NPD within the Data Modernisation Division

So let's cover two things.

1. The NPD Access Process 'as is'; and my reflections from the first 2 months looking after it.
2. Our plans for the future.



The current NPD Access Process : First Thoughts...

- Large volumes of requests. Roughly 350 per year.
- Citizens can have completely polar opinions. From *'why on earth does my child's data ever leave the school?'* through to *'why isn't more data made available?'*
- Nearly all user requests are to do appropriate things. People know what they can use it for.
- There is risk involved in this. Re-use of public data relies involves trust. DfE is seen as very encouraging of appropriate 3rd party use, but *just one* data loss or mis-use could see decision makers get very jittery.
- Processes are well established, but could be improved.



Improving the current access process – what can we do?

- We are putting **more resource** in so that we can move things through more quickly.
- **Improve guidance** on NPD web page, including clearer explanation of security form, which is so important as part of the process.
- Tighten the input methods and choices on the Application Form. Provide a '**check list before sending**' at the end to ensure that no things get left out.
- Transparency around our '**security red flags**'.
- Articulate the '**what happens once my form is submitted?**' better, including numbered steps in the process, so we can develop performance metrics for how well we are managing applications internally.
- **Trial 15 minute telephone meetings** with the requester / data producer / NPD analyst to manage data requirement queries in a single communication
- **Training and desk aids for caseworkers** on Data Protection Act and Security elements that are relevant. Hope is that by upskilling, we can reduce the frequency that we have to call on legal advisors and security experts, which can take time.



Improving the current access process – what can *you* do?

- Take **Data Protection Act seriously**. It very clearly says what data is sensitive, trying to tell us that certain aspects aren't won't wash.
- Only ask for what you **need**. Can you:
 - Only ask for certain years rather than all years?
 - Aggregate up. Don't ask for Date of Birth if banded age groups will do.
 - Explain really clearly why you need each bit of sensitive data.
- Be really clear on why the work you are doing is going to be of **benefit to children**.
- Keep the ball in our court and **be pro-active**. Investigating slow cases often shows that the to-ing and fro-ing has taken several weeks via email etc. Let's pick up the phone where we can.
- Realise that all requests are treated the same. "*You can trust **me**...*" doesn't quite cut it!



And so what about the other changes?

Problem statements.

- Every time we send data out to someone, there is an element of risk.
- People are accessing sensitive data, but only to then aggregate. The access to sensitive data is a means to an end to produce the higher level findings.
- You have to have a certain degree of analytical clout and sophisticated IT to access the data.

So what can we do to...

- Provide access to NPD data with reduced need to actually send the data?
- Provide ways of interrogating pupil level data, but never actually seeing it?
- Help those who lack the IT and/or analytical data use NPD to drive up the evidence base of what works.



And so what about the other changes?

Problem statements.

- Every time we send data out to someone, there is an element of risk.
- People are accessing sensitive data, but only to then aggregate. The access to sensitive data is a means to an end to produce the higher level findings.
- You have to have a certain degree of analytical clout and sophisticated IT to access the data.

So what can we do to...

- Provide access to NPD data with reduced need to actually send the data?
- Provide ways of interrogating pupil level data, but never actually seeing it?
- Help those who lack the IT and/or analytical data use NPD to drive up the evidence base of what works.



Our aspirations for opening up access responsibly and securely...

Proposed Innovation One : Privacy Controlling API.

- A cut of NPD data sits somewhere within DfE hosting arrangements.
- An application programming interface allows users to interrogate raw data, but not see disclosive data, Possibly by rounding, supressing, or introducing elements of noise to what is returned.
- We are currently doing discovery work with the Open Data Institute to design a competition for building this. That should report by end October.

Proposed Innovation Two: DfE Datalab

- Cuts of NPD sit on DfE machines. External users allowed to come and use NPD for bona-fide reasons.
- Users undergo short training, and strict controls on what is allowed in/and out of the room.
- A model successfully deployed by ONS.
- Aspiration is that we can get one in place by Christmas.



Our aspirations for opening up access responsibly and securely...

Proposed Innovation Three: Education Endowment Foundation Datalab

- EEF provided with cut of NPD
- Small organisations wishing to conduct experiments but without the analytical clout to do so can use the service.
- The organisation would provide the 'treatment group' for EEF to go and match to a control group and detect impact of intervention.
- Results published regardless of what they are.

