# Data Quality Video Transcript

## January 2021

### Speakers

Sam Wright (SW) – ESR Central Team

Chris Holroyd (CH) – ESR Central Team

Nick Armitage (NA) – NHS Digital

Introduction (00:00 > 01:13)

Slide 1

SW: Hello and welcome to the ESR Data Quality Video resource. We’re going to use this video to cover a few key topics in relation to data quality and some of the resources that are available across both ESR and NHS Digital that can be utilised to help improve data quality within your organisation.

So a few caveats before we get started; unfortunately due to the circumstances we’ve had to record this video in a slightly differently way to how it would usually in that the sections on the agenda have been recorded separately and we’ve tried to stitch them together as seamlessly as possible to create this video content for you. You may notice as we pass through the agenda that the audio quality slightly changes but hopefully it won’t have too much of an impact on the content that’s being delivered.

So as a result of that I’m not going to introduce Chris or Nick at this stage. I will allow them to do that themselves as we reach their sections on the agenda. But as an introduction for myself I am Sam Wright and I work in the ESR Central Team as a Workforce Information Advisor as part of the NHS Business Services Authority.

Agenda Overview (01:14 > 02:41)

Slide 2

SW: So without further ado I’ll just give a brief overview of the agenda and what we plan to cover today. So I’m going to give a quick introduction as to what ESR is and does in terms of the data we hold and how it can be utilised when we get it right. I’m then going to move onto what ESR currently does to support data quality across the NHS.

Chris will then pick up some of the more specific resources, specifically in relation to ESR BI, so the NHS Data Quality Dashboard and how that can be used by organisations to vastly improve Data Quality. On a similar note the Workforce Information Verifier Dashboard which we’ve developed in conjunction with NHS Digital, I will cover that and the benefits that can bring. At that stage I’ll pass over to Nick Armitage who is going to deliver some really important slides discussing some of the key data quality issues that they face from a national reporting perspective and the importance of data quality as a whole across the NHS. It’s just to give that national perspective on the data that we perhaps don’t always have when you’re just dealing with the data locally within your organisation.

Finally I’ll pick up some of the other resources that we produce on a monthly basis specifically around professional registrations and Unique ID reports and how we monitor that and send individual reports to help improve the data in that space.

Introduction (02:42 > 04:15)

Slide 3

SW: So as an introduction, a lot of this you will obviously be very familiar with but ESR enables organisations to record and report upon a wide range of workforce data. So not only is the data used locally, it’s also used at a regional and national level to manage the NHS. Decision making comes from this data so as a result of that, as a direct consequence, where the data is poor in quality it can impact on every aspect of the workforce.

So decisions around pay, around pensions, around workforce planning, vacancies, etc. may be the wrong decisions if data quality within the system is poor which is being used to inform those decisions. On the flip side of that if we can get it right it will enable us through ESR to analyse and monitor a wide range of policies and activities. You can see some of them listed there (within the slide) again you’ll be very familiar with these but pay & reward, commissioning, statistics for FOI’s and perhaps in the spotlight at the moment, monitoring of diversity and inclusion, especially in response to Covid . . . if we get these things right we’ll have the data at hand to respond to some of these issues that the NHS is facing.

For further information on how improved data quality can benefit your organisation please do visit [www.discover.esr.nhs.uk/](http://www.discover.esr.nhs.uk/). There is plenty information and resources on there as to how changes and improvements in Data Quality can provide a benefit.

ESR Supporting Data Quality (04:16 >09:15)

Slide 4

So ESR, what do we currently do to support Data Quality? So there are a number of things; first we’re going to look at what we do by design. So in terms of functionality there are a variety of methods that we use to try to improve and make the data consistent.

So the first being drop-down list of values. This is instead of free-text fields which we know cause issues with data quality so where possible we try and provide a drop-down list of values which gives us consistent answers or data across the NHS. Following on from that a lot of the data items that we use in those list of values such as Job Roles, Sickness Absence Reasons and Occupation Codes, they comply with the NHS Wide standards such as the National Workforce Dataset and again that enables a consistency of reporting across the NHS and across all workforce systems, not just ESR.

One of the other things we do, we have what we call input masks. You’ll be familiar with this mostly in relation to professional registration PINs. So we can put a mask on the form that only enables a specific combination of characters and text-string to be entered otherwise it will not allow you to save the record. You may be aware that we’ve recently removed this in response to the covid pandemic whereby we’ve enabled organisations to record temporary registrations. But prior to that change, the mask has always been there and you’ll see them in use across other forms in ESR.

There are also context sensitive warning messages. So as and when you move through the forms depending on the data you put in you may be guided or prompted towards entering consistent and accurate and in some cases we do use mandation of fields. We do try and keep that to a minimum where possible for reasons which Nick will explain in much more detail later, but where it is deemed necessary we do use the mandation of fields which don’t allow the user to save the form without entering the data.

Slide 5

So some of the other things we do - you’ll be aware of the interfaces we have with a number of other systems including HMRC, NHS Pensions, Recruitment, Rostering and a whole host of professional registration bodies.

There are clear and obvious benefits that that brings in terms of data quality in ESR. It saves time and effort, it reduces the number of errors through manual entering of data, the data is transferred accurately and on time and once the link is established huge transfers of data are exchanged effortlessly.

Inter Authority Transfer (IAT) – So again, you’ll be very familiar with this but IAT functionality enables the transfer of data from one employer to another. So a lot of the benefits that we’ve mentioned with interfaces are replicated here in that it reduces the time and effort and avoids manual input of data where mistakes may be made. It also enables us to keep the flow of accurate data going through the NHS and get that information from one employer to another efficiently.

Slide 6

Finally, Self-Service – a huge driver in improving data quality specifically in relation to equality monitoring. Empowering employees to access Self-service and keep on top of their own personal information. It also obviously enables manager to monitor the information of their staff much easier than in the past where they would rely on a central HR Function.

So, as we know… Self-Service functionality is available via mobile devices, as well as desktop computers. So users can access whenever and wherever is most convenient to update their data.

We’ve recently undertaken a piece of work to review the correlation between recent log-ins and the likelihood of data completion. (See the table on the slide.) It’s important to note that we only looked at fields that are available to update through Self Service. This relates to logins from January to September 2020.

As you can see the data completion rates for key data items are higher where employee Self-Service log-ins have been recorded in the timeframe. That should be a huge driver to get employees to access their data; because we know once they do that they will update and complete those fields that have been sitting there blank for potentially a number of years.

I’m now going to pass over to Chris Holroyd who will cover some of the resources we have available within ESR BI.

NHS Data Quality Dashboard (09:16 > 29:16)

Training environment on screen (Introduction to BI Dashboard)

CH: Thank you Sam. My name’s Chris Holroyd, I’m the ESR Development Advisor for BI. Today we’re going to look at, as Sam mentioned, some of the resources available within BI for data quality.

So you can see we’ve switched straight into BI. This is a replica environment so any data you see in here is fictitious. We’ve logged in using the BI Admin URP, this is the landing page.

So the first thing we’re going to look at is the Data Quality Dashboard which is available to use from the dashboards menu at the top here. The dashboard is available to HR Admin or BI Admin users or HR Management as well and you can see it here about a third of the way down – NHS Data Quality Dashboard – we’ll click on that and we’ll allow the dashboard to run.

Move to Data Quality Dashboard

You can see we land on a summary page and if we have a look at the top of the dashboard you can see we have a number of pages within this dashboard. We land on part 1, there’s a part 2 page and absence, applicants, assignments, all the way along the top there until you get to supervisors. These are the different pages within the dashboard.

Straight away some of you may notice that you used to land on a page called ‘Summary’. This is a recent development we put in where we split the summary page into part 1 and part 2. This was primarily to help navigation. We reached, I think, 60+ tests within this dashboard. To have 60 tests within one summary page was not particularly useful so we decided to split them out into part 1 and part 2. Part 1 are slightly higher priority tests and part 2 are slightly lower priority tests but both should be run and both should be paid attention to in terms of the number of records returned within each of the tests on these two pages.

The way the dashboard is set up, as we mentioned, it has the summary pages – Part 1 and Part 2 – and these gives you counts of the data quality tests that are included So you can see on screen here, the first test, top left, assignment tests, assignment category should not be NULL. There are 30 records for this fictitious organisation with a NULL assignment category. So, what do we do then do with that number 30? It’s good that we know how many there are but we can then move to, and we can do this later on, to some of the other pages which show the detail behind these numbers.

Part 1 and part 2 give you the summary and then each of these pages are designed to align with each of the areas shown on the dashboard here, assignments, prof reg, employee, position.. and you’re able to go to these detail pages along the top which I’m highlighting here and you can see the detail behind the numbers returned on part 1 and part 2 pages.

Ok, a couple of other things I’d like to point out while we’re on this page. Prompts at the top – So prompts, what do we mean by prompts? Drop-down lists, Effective Date Prompts, that sort of thing, all at the top of the page here. Anyone who uses BI regularly will be aware of dashboard prompts. It’s essentially how we filter the data that is returning within each analysis within the page. So, key to any dashboard or any dashboard set of prompts is an effective date. We can see it here if I just try and highlight that for you. An effective date means when are we returning the data correct as of. So always look for an effective date prompt when you’re running through any of the BI dashboards, but it’s relevant when looking at data quality as well.

The second prompt I’d like to highlight is the exclude employees number prompt which is highlighted there. Just a little on the history of the Data Quality dashboard in BI. It was originally set up to replicate the WoVEn tests that NHS Digital send out. We have now exceeded the number of tests that are within the WoVEn report but we still try to align and make sure we cover all of the tests that are in those WoVEn spreadsheets. We’re aware as part of the WoVEn tests there is the ability to exclude records that are considered data quality issues even though they may be highlighted on the WoVEn reporting. Now, it’s slightly more important in WoVEn because WoVEn provides a score and we’re aware users like to return the best score they can. So we’ve replicated that functionality within the data quality dashboard in BI. So to exclude employee numbers we can use this prompt, we select the employee numbers from the drop-down list here and we select the employee numbers we don’t want including within any of the tests below. Slightly less important in BI because there is no score returned and it won’t really affect anything other than being slightly annoying that the employee number would but it’s actually not an issue. But the prompt is there if users would like to exclude employees from the dashboard page.

OK, the last prompt I’d like to touch on, on this dashboard page, is the actual termination date prompt which is highlighted on the second row of prompts there. The actual termination date prompt doesn’t apply to all analyses as you might imagine but it does apply to some. It applies to.. if I just scroll down the page here, the staff movement analyses. As you can imagine most of the tests refer to employees so actual termination date isn’t relevant but where the test refers to leavers we need to be able to specify the leaving period of the leavers to be returned or to be tested within the analyses and the actual termination date prompt can be used to specify the leaving period if you like. Termination date is between two dates, the leaving period, for any tests that concerns leavers.

I’m going to move onto part 2 page. So I can do that in the top left corner here, so I’ve just moved the mouse to the second page and you can see we’ve moved across a page and it’s highlighted there because there’s one last prompt I’d like to highlight on this page. And you can see we’re presented with a different set of tests now. It’s the email domain prompt, which is here. Again, a bit like the termination date prompt, it’s only relevant to a specific test but still quite important and it is relevant to the test, and I’ll see if I can highlight it here, email address does not contain specified domain names. The idea of this prompt and this test, you are able to enter, select from drop-down lists or type in an email domain and then that will be excluded from the results of the test leaving only records returned that do not contain the email domain that you have entered. For instance you may type in ‘@nhs.net’ because you’re happy with any email address coming from nhs.net and that would leave returned within the test any records where the domain is not @nhs.net.

So just a couple of things we were looking to point out on the summary pages of the Data Quality Dashboard. We’re not going to move on to have a look at the detail page. So we’re going to have a look at the assignment page. I’m going to hover over it then, you can see assignment related data quality issues in the tool-tip. And we’ll have a look at an example of what a detail page looks like. Previously part 1 and part 2 you got the summary figures and now you get the detail behind the summary figures. So these are all the assignment related tests from both the summary pages, so there’s not a part 1 and part 2 detail page. You can see a fairly standard set of fields that are returned for the detail tests/analyses. Some sort of assignment or employee number, some sort of unique identified and then information relevant to that test. In the top left you can see assignment category should not be NULL and we return assignment category to show that assignment category is NULL in these instances.

On the top right, a person’s total combined contracted FTE should not exceed 1.28 and we return the combined total fte to show that these are records where the combined contracted fte does exceed 1.28. So we try to keep it as simple as possible, you get the summary figures and then you move to a detail page and you get returned the detail which enables you to go and investigate if required.

The third page we’re going to have a look at is the applicant page. This was recently introduced as part of a review we did into the data quality dashboard this year and that works slightly differently in that the applicant page contains summary and detail analyses. It’s sort of a one-stop-shop if you like for any applicant data quality issues that you may come across. You can see we have the summary figures at the top and the detail below. Now we’re in a replica environment so we actually have no applicant records so nothing is returned in here but you can see summary figures on the top and then if we had detail then they would be returned below. Most of the applicant tests do mimic the employee tests because in most cases they are relevant in terms of what should be correct for an employee record should be correct for an applicant record. There are a couple of applicant specific tests in there but in general they replicate across employee and applicant.

OK, that covers the first sort of 10 pages we have in this dashboard. The summary part 1 and part 2 and the relevant detail pages. We then have three pages on the end; system DQ, trending and supervisors. So if we have a look at what these pages do, they’re slightly different. We’ll launch the system DQ page and you can see the table returns here under the system DQ page with four records returned. So what the system DQ page is designed to do is return records where there are some sort of system data quality issue resulting in any of these records returned here not being in any of the other BI reporting. So if we were to report staff lists, compliance, appraisals, whatever it is, these records that are returned under this page will not be included in any of the BI reporting. The way to resolve that normally would be to raise a P4 SR on the service desk and IBM will be able to advise how the records can be resolved and then they will automatically be included back into BI Reporting.

You can see with these 4 rows the object name and the error type fields, they have issues with their budget values and the error type is that they have over-lapping date-track records and so they cannot be included in the BI warehouse and so they are returned in this page so organisations can take action against any of the records returned in this page to re-introduce them then back into BI reporting.

Ok, we’re going to move onto the next page which is trending. And you can see a summary analysis is returned within the trending page. Now what the trending page is designed to do, fairly self-explanatory, it’s designed to return summary figures across a period, a long term period, so you can see how your organisation’s data quality is trending. Whether it’s up, down or staying the same, whatever. You can see we return a line-graph with each of the summary figures for each of the business areas if you like (absence, assignment, employee etc.) and then below is a more detailed which groups by each of the business areas of the functionality areas of ESR and then returns figures per test.

The trending data is summary data and a snapshot is taken on the last day of every month. So where we see monthly figures here we cannot drill down any further than monthly figures. A snapshot is taken on the last day of every month and that is the number that is returned against each of the months.

You’re able to select a point on each of the lines within the line graph, click on it, and that will result in the detail table below updating to return figures for that business area that you clicked on. So we can see here, I clicked on the green line which is Equality & Diversity and you can see below has been updated to Equality & Diversity. Now we can also do the same thing using the prompt here. So the test category we can change manually there if you want to. We’re in a copy environment so the data is very static, so the all-important change column returns 0 for all of these tests.

Ok, now we’re going to move onto the last past of the Data Quality page which is the supervisor page. The Supervisor page is designed to return supervisor hierarchy loops. So an assignment becomes its own supervisor. An example of it is there for you. Any assignments that are part of a supervisor loop are listed, and you can see one is listed here, 90214658. We’re able to click on that assignment number and details of the loop will be presented to us. We can’t do this in this copy environment as the navigation isn’t working but we’d presented with a new window and details of the loop with all the assignments that are relevant to that loop would be included within that new window.

Ok, that’s all we wanted to go through in terms of the Data Quality Dashboard today. We are going to have a quick look at the payroll dashboard because the payroll dashboard Data Quality tests were something we introduced this year as well for out payroll colleagues so if I open the dashboard menu here and select NHS Payroll Dashboard 2/3 of the way down and we’ll click on that. You can see the Payroll Dashboard launches. The last page, if we open up the rest of the pages, you can see we have a large number of pages across the Payroll Dashboard but the last page is called Data Quality and we’re going to launch the Data Quality page of the Payroll Dashboard.

Moves to NHS Payroll Dashboard

You can see the Data Quality page of the NHS Payroll Dashboard returns and much like the NHS Data Quality Dashboard itself we return a number of detail analyses, no summary within this, but we return detail analyses of various data quality tests. So the first one you can see ‘NHS Assignments with missed pay-steps’… and there are a number of tests within this page relating to Payroll. It’s something we introduced this year following feedback from our Payroll colleagues and it’s something we’re looking to expand on as well.

OK, the last dashboard I want to touch on, and it’s not one that I’m actually going to launch here, but it’s the NHS Recruitment Dashboard. This is a copy environment so the changes we made to the recruitment dashboard are quite new so they’re not reflected in this environment but I can show it you in the dashboard list. It’s towards the bottom – ‘NHS Recruitment Dashboard’ there.

I only point it out because there were a number of applicant tests as part of the NHS Data Quality Dashboard; there is the applicant page. But recruitment URP users don’t have access to the Data Quality dashboard in BI so we’ve made all the tests that are under the applicant page in the Data Quality Dashboard available within the NHS Recruitment Dashboard which the Recruitment URP users do have access to. So something to look out for if your Recruitment URP users are looking to go through those tests but don’t have access to the HR URPs, they can check the recruitment dashboard and the same applicant tests are available within the Recruitment Dashboard.

Ok, that’s everything that I was hoping to cover within ESR BI within Data Quality so I will hand you back to Sam.

NHS Workforce Information Verifier (29:17 > 33:13)

Slide 7

SW: Thank you Chris. I’m now going to cover a further resource available within ESR BI which is the NHS Workforce Information Verifier dashboard. This dashboard replicates the Verifier Tool which was initially created back in 2013 in partnership with NHS Digital. So the purpose of the tool was to assist organisations in improving their data quality in relation to positons and assignments. The way it did that was that it reviews the combinations of Job Roles, Occupation Codes and Pay Bands used against the positons to flag any potential or actual invalid combinations to organisations to enable them to either investigate or correct where necessary.

So since 2013 and the introduction of ESR BI, we’ve since replicated the tool within BI which hopefully makes it a lot more accessible to users and enables organisations to run it on a much more regular basis, whether that’s monthly, weekly or whatever works best for you. But it enables you to have an up-to-date overview of all your active positons and current assignments to see if there are any issues that need addressing.

So the formulas within the dashboard replicate the combinations that were within the tool and are within the Excel resource that is available that is available on the Hub. So these combinations were agreed and approved by the Workforce Information Review Group (WIRG) which is chaired by NHS Digital. So I would just say if there are proposed changed to the combinations then please do raise them with the email address there, [enquiries@nhsdigital.nhs.uk](mailto:enquiries@nhsdigital.nhs.uk), as soon as they are approved by WIRG we’d look to reflect them in ESR as soon as possible.

I’ll just give an overview of how the tool works with some really obvious examples. For example if you had a Job Role of ‘Staff Nurse’ within the Nursing & Midwifery Staff Group but it had a Pay Band of Band 2 against the position; that would flag as an invalid combination and show as red in the Job Role and Pay Band column you can see down the right hand side there. So as we know all Nursing & Midwifery staff should be band 5 and above so ‘Band 2’ would immediately flag as red. Likewise if you had the same Staff Group and Job Role of ‘Staff Nurse’ with an occupation code of 001, again that would flag as red as ‘001’ is a Medical & Dental Occupation code. Finally, another really obvious example but the Occupation code of ‘N6A’ which relates to a Nursing role; if you were to have Pay Band of Pay Band 2 that would also flag as red as nursing roles again should be band 5 and above. So they’re really obvious examples which hopefully you wouldn’t need the tool to identify but it’s just to give you an idea of what the tool is doing and what the data is showing.

So it’s regularly updated after each National Workforce Dataset uplift so where there are new Job Roles, Occupation Codes or where we retire Job Roles and Occupation Codes we ensure that the combinations reflect the current values. We ensure that where roles have been retired that all of those flag as red as they should no longer be in use. So it is regularly updated but we are aware that organisations do sometimes spot combinations that should be valid or contradictions within there that have perhaps been missed when the tool was initially set up. So as I said earlier, please do raise those with enquiries and we’ll look to correct those as soon as we possibly can.

That was it for that tool, so I hope you found that useful. I’m now going to pass over to Nick Armitage at NHS Digital for his section on the agenda.

Importance of Data Quality/Key DQ Issues (33:14 > 59:28)

Slide 8

NA: Hello with thanks very much to Sam and the ESR team for inviting me to speak about Data Quality and Workforce Information and why that is important to both us in Arms Lengths Bodies and to yourselves out there in trusts and other organisations using ESR.

So my name is Nick Armitage and I’m an Analytical Section Head at NHS Digital and for more years than I care to remember I’ve been involved in Workforce Information and it’s fair to say I really have a passion for Data Quality and hopefully I can pass that passion on to you.

Slide 9

So, why is it important? Well, I think this quote here from Ben Goldacre is a good example to start with. That’s to really emphasise the fact that data is not just data, it’s used for lots and lots of purposes and that bad information can cause all manner of chaos and I think it’s fair to say that this year has been a real exemplar for when you need to have good data on all sorts of different topics.

Slide 10

So briefly I’m going to go through a quick introduction of what is data quality? Or, what is quality data? I’m going to talk a bit about improving data quality and how you can measure that, as that’s often the most difficult part, and then to think about how those data quality improvements and better data quality can benefit you. Some concrete examples around what we’d like to do, what we can do and what we think we should do in future and then a little round up at the end. Hopefully I won’t take too long.

Slide 11

So, what is data quality? Well, it may mean different things to different people. There are many different aspects of data quality: relevance, timeliness, completeness, validity, accuracy and comparability which are complex and interrelated. Often, one aspect may be improved but that may have to be balanced with an impact on another. For example, to increase the accuracy of the data it may require that more time is spent on it thus the timeliness element is made slightly worse.

There are systematic and people dimensions to data quality. It’s about the systems and data standards but it’s also about peoples behaviours and the way that they interact with the system and their understanding of the information that they’re inputting or indeed, as you’ll come to see later on, that they’re trying to analyse. It’s really important to understand your data and to know your requirements both at the start of the process in terms of getting data into systems like ESR for operational purposes and also for those people that are ultimately doing the analysis at the end.

The improvement in data quality relies on collaborative working. This is something that I come back to over the years’ time and time again and it’s very true now as it always has been; this is a collaborative process. You cannot achieve data quality alone. You need service managers to tell you what the values are, you need service managers to give you information about their staff for you to be able to record it in ESR. Equally the people that are doing the analysis of the data, they need to understand what it is that they’re looking at and they need to go back to people like yourselves to get a feel for what that data is.

Equally the data must be available. The most validated and accurate data is not quality data if it isn’t made available to anybody or those that need it. That’s where ESR I think is really powerful because of the Data Warehouse and the fact that people like ourselves are able to get that data, publish it and make it as widely available as possible.

Slide 12

So, how can you improve data quality? Well, there are various strands to this but the thing that cuts across all of it is that you need to have a clear purpose and to understand why you’re doing it and what you’re attempting to achieve. Different elements that link into it in a classification and standardisation of the values that are in the system and the ability to provide education and information guidance and things such as this video about data quality. To provide mechanisms for validation of the data, to provide feedback on the data. It’s no good somebody coming and saying something needs to be changed but not explaining why and providing information about what has been changed and what has been achieved.

And as I’ve already mentioned you need to have a visible and useful resource for it to be considered to have achieved quality improvement. So how can we turn that into practice? Well, you need to involve individuals. They need ownership of the data. People like yourselves, people like the service managers need to work together to ensure that all the people that need to play a part in this have got ownership of the data. This is where things like self-service and manager self-service are really powerful in ESR because they help people to take ownership of their data and by doing that you’re encouraging improvements in Data Quality.

Equally people like myself and others in central organisations need to understand the barriers to data quality. There’s awareness obviously about what the data is being used for at all levels. Things like this video are there to try and raise awareness of data quality and the importance of workforce information but equally within your own organisation you need to think that people that are inputting the data into ESR, do they understand what that data will do? It’s very easy to focus on the task in hand without thinking about the wider implications of things downstream. If there’s some piece of data that you’ve added to ESR for a specific purpose might also become relied upon for something else.

There are resource and capacity issues; we’re all very much aware of this. The NHS has for many years needed additional funding and I think finally as NHS E/I have got a people directorate and long-term planners this has spawned the NHS People Plan and I think more realisation has been given to the fact that if you want to know more about the workforce then you need to provide more resources upfront to be able to pay for people to take the time to care about the data and to ensure that it is good data quality. So I think that is something that after a long time of hoping, that we might be able to push on more to help with.

And technology. . . Again, ESR is much better than the systems that I remember from the early 2000’s when I used to have to take spreadsheets from a plethora of different HR and Payroll systems and have to try and stitch those together to try and create the annual census. Technology has improved but there is always room for improvement and that’s why we want to speak to people and try and think where further improvements can be made. On that front, we need people like yourselves to work with us to look at data standards, to think about systems, to ensure that data standards that are implemented in systems like ESR reflects the real world and is responsive to change. On that note, the next uplift to the National Workforce Dataset is currently going through approval so we’re hopeful that some of the values that we’ve been asked for, for several years, for example; to better improve information for AHP’s, will become available in ESR sooner or later.

There needs to be validation at source, we need to be to provide positive and negative feedback on data quality. As I’ve mentioned there needs to be feedback and visible results. There also needs to be some idea of consequences for people to understand that it’s not just good enough to put the data in and then forget about it. And obviously the data needs to be available.

Slide 13

So, how you then measure improvements in Data Quality. . . This I think is harder to get a real hold on but there are various different ways that you could think to do it. So against what do you measure improvement in data quality?

* Is it previous data standard quality?
* Is it some idealised gold standard?
* Is it difference between different fields in the system? For example reducing inconsistences between data that’s in ESR and a professional register.
* Or is it between different organisations? Do you look at ways in which different organisations are doing things?
* Or perhaps the field level metrics?

This is something that people will have to consider themselves depending on the specific task and the specific elements of data quality that is being considered. I think it’s fair to say that one of the best ways to make a start on this is to utilise the available resources.

For example, the NHS Digital WoVEn process that I hope you’re all familiar with, the ESR Data Quality dashboard and verifier tool that provides lots of additional resources to help you with your data quality and a key theme to all this really; it’s all about communication and working with others. So consult the users and providers, develop your own means of assessing quality and understand the ask that people have for the data quality. So if someone in your trust is repeatedly asking for the same piece of information, is there something that you could do to help improve the data quality to help them with their work to get them to help you.

And as I’ve said, different aspects of data quality need to be measured in different ways. For example, validly and accuracy . . . whilst some data may be 100% populated because the values are valid values it could also be 100% inaccurate if the values do not represent the actual values of what you‘re trying to code and have merely been added to try and satisfy the requirements of the system. This is a really important aspect of ESR that people repeatedly ask for fields in ESR to be mandated and I think the ESR Team have very rightly left it so that as few fields as possible are mandated because not only does it risk people getting stuck and not being able to complete the data but it risks people inputting something simply to get past the stage that they’re at, rather than leaving it as being blank and then being able to fill it in later when they’ve got the correct information. It’s much harder to check data quality for something that’s been filled in with the wrong value than it is when the value has just been left blank. So that’s a really important consideration and as I say, an important way that ESR works.

So on that front is it easier to see or measure poor data quality and its impact? And perhaps think about reducing such issues as a proxy for the value of improved data quality. Again, it’s an idea and for one you to consider.

Slide 14

So, having said that . . . We’ve talked about how to measure data quality and how to improve data quality, but what’s the benefit? One of the key things really is time savings for yourself. A reliable central source of data means that you don’t need to collect the same data over and over again and do further self-validation in response to multiple similar data requests. I think we’re all aware of the ever growing plethora of spreadsheet related data collections from different organisations and different parts of your own organisations about different aspects of your workforce. If that information can be captured once in ESR then it would really mean that you would be able to reduce the effort made to complete those returns. You would yourselves have greater confidence in the data that you hold on your workforce and similarly others looking at it would have greater confidence and again that would mean that you wouldn’t need to do further validation.

I think one of the important aspects of this is that it may be difficult to quantify the effect of improving data quality and understanding how close you need to be to the data for that improvement to benefit you, but I thinks it’s fair to say that efficient management of your HR management processes depend upon good quality data. Patient care and safety depends upon good quality data. Poor quality data can damage the reputations of organisations and individuals, and poor quality data leads to flawed clinical, administrative and planning decisions. These are all facts. Additionally, poor quality data costs money. There are many occasions where whilst it costs money to improve data quality, by not doing that you’re actually spending ultimately more money and time by dealing with the issues of poor quality data.

Slide 15

So, some examples rather than just talking theoretically… Well as we all know no doubt from your own experience, there is an increasing desire to know about locations of staff and not just who they’re paid by. ESR includes information such as the position workplace organisation code; a list which is now regularly refreshed alongside other elements of the system. This also includes generic values for GP Practices, Local Authorities and more different areas where your staff may be employed but that is quite difficult for you to capture elsewhere in ESR. So I think it’s important to say that in national level statistics produced by NHS Digital we rely heavily on the Position Workplace Organisation Code field to show where your staff are actually working rather than who is paying them. This is an important thing for you to think about and we’ll come back to it in a moment.

Additionally there are other fields in ESR that are related to the location that staff may be working in.

One of these is Area of Work, now that’s not specifically just about location and we’ll come onto Area of Work quite a bit shorty because I think that’s absolutely key fields at the data quality end where things could be done differently to improve things for us all.

Site Code, Description, Post Code information; all of these aspects of ESR allow more granular location details to be captured. Trust level is certainly not enough anymore and I think people need to think about how they improve location information in ESR.

So, why? Data quality in these location fields is increasingly important as I’ve said. Examples of why is because staff who are co-located or hosted staff. For example, improving access to psychological therapy staff who are delivering services at GP practices but are employed by community or mental health trusts, I know there’s a lot of interest in this that that has led to separate data collections.

There are staff employed by NHS Trusts but working across PCN. This is only going to increase in the coming years as more of the roles are delivered across PCN’s. Also, a key issue that’s come out of the Covid crisis is a renewed desire to understand the true size of the community workforce and linked to that to understand workforces working across integrated care systems and other new ways of working, as it currently stands without good quality location information it’s not possible to be able to tease out these important things and there will definitely be additional data collections as a result of the lack of this information.

Having said that, I think it’s important to reiterate the fact that the people that are asking questions of the workforce data need to understand their questions. Poor data quality can actually be good data quality to somebody else but it just depends upon your perspective and the reason you want to use it.

Slide 16

For example in this front, the possible issues are around the ways things have been specified in terms of the data standard and it’s guidance that’s associated by it and interpreted by analysts. For example, going back to Area of Work, unfortunately it’s fair to say that Area of Work is trying to provide both the area or location that the activity takes place in and also the specialism of the workforce and it’s often not possible in the same field. Sometimes it might be fine if the role is one where you don’t need the additional detail of the Area of Work to define the specialism. But for many other roles because of the level of the Job Role and the Occupation Code, Area of Work is trying to do two things at once. So this, amongst other things, really hampers efforts to und4rstand the true size of the community workforce and it increases the need for additional data collections.

Ultimately where will that data be coming from? People will no doubt be asking for spreadsheet collections but the data will ultimately come from ESR and by having to then look into different aspects of ESR in order to be able to answer these questions it just adds additional burdens to yourselves.

Need for changes to the standards or guidance and systems to try and capture the data – so on this front, the fact that the Area of Work is currently trying to do two different things at once – it’s not something you can yourselves improve by simply improving your data as it stands. So thought needs to be given by people like ourselves and the ESR Central team whether there are other aspects of the system or the standard that could be used to help to capture the data elsewhere. For example, consideration is being given at a very high level at this moment in time to consider utilising the ‘Care Group’ field within ESR.

Could this in future with the right effort to correct this and the right guidance and communication, could this be a means of capturing this sort of data and no longer requiring Area of Work to do two things at once? I think there are other examples on the horizon for things that are going to become increasingly important which relate again to the NHS People Plan. So NHS Talent Management for example – There’s a desire across the NHS to reduce the cost of head-hunters and to reduce the risk of lacking long-term succession planning. So at present NHS Talent Management is very piecemeal, there’s a real desire to be able to do it nationally and to do it in a more efficient way but at present this requires burdens such as data collections. Potentially improving systems and processes would be a much better solution for the future.

Source of Recruitment within ESR – This links to the commitment for 50,000 more nurses – This field at the moment is populated sporadically and there is no guidance provided nationally. So when people come to do the analysis there is a muddled picture. In the coming months there is highly likely to be some much more focused guidance developed to try and explain what the different values mean and how they should be used to help to understand the background of different nurses that are being added to the NHS Workforce or not lost from the workforce in order to meet the commitment for 50,000 more nurses. To be able to do that through ESR would definitely reduce any need for an additional data collection and mean that it could be used in many other ways as well.

Slide 17

So that being said, behaviour change is often needed. We can ensure that the system is as good as possible, we can look to improve the standard and guidance but people need to think about how they interact with the systems and the guidance. This is especially in relation to fields that may not be relied up on locally where alternative sources may exist.

Again, I’m coming back to Area of Work; this has been an issue, more so in the past I feel personally, but now Area of Work is becoming more important for Model Hospital and for an ever growing range of data settings including the Mental Health definition which is used to define the full scale of the mental health workforce. I think there was a time when people really did not believe that the data in Area of Work was believable but I think those days are gone but there are ways in which it could be improved. Certainly for some organisations, it’s not as used as well locally as it might be and that’s because there are other fields within ESR locally where this information may be held. For example, trusts may use work structures or subjective codes but unfortunately these have no national definition and therefore this information is for local use only and Area of Work is what has to be relied on by regional and national bodies such as NHS Digital and Department of Health.

In terms of things relating to pay and establishment control which obviously have a great deal of uses – Both for workforce planning and vacancy information and the like. Reconciling the general ledger and not ESR is a real problem. It works locally but not nationally; there is no clear national picture. If Finance and HR workforce teams in trusts could work in partnership it would enable the development of a single source of truth which would be useful locally, and at a regional and national level. And as I say, it’s important for a number of different aspects such as understanding capacity and vacancies.

Having said all of this, behaviour does need to change, everybody needs to play a part in that but I don’t want to forget about the fact that we need to play a part in this too. The nationally defined standards, the guidance, and the systems such as ESR must support users. I feel that over the last decade or so that much has been done to really improve things but more still could be done. Could systems be improved to facilitate better working between finance and HR to help reconcile the general ledger? Could Area of Work be better defined? Could it be split in the data standard? Or could more use be made of other fields such as Care Group? Could guidance and communication from national bodies be improved? For example at this moment in time we’re developing some guidance specifically about how to code the improving access to psychological therapy and ultimately the entire mental health workforce. Could your voice be heard more clearly in decisions about the development of workforce information? That is definitely the case.

There is a Workforce Information Review Group that is chaired by NHS Digital that is responsible for the up-keep of data standards and responsible for considering the ways that systems Like ESR and NHS Jobs capture information. In order for that group to function correctly we need representation from trusts and other people involved in the process. If you want to be involved in the Workforce Information Review Group then please get in touch with NHS Digital or the ESR Central team and we’d been more than happy to add you to the list and in that way your voice will be heard.

Slide 18

So, you’ll be glad to know we’re coming the end now and I‘ve got a few little thoughts I’d like to leave you with if I can. To reiterate, improving data quality relies upon collaborative working, you cannot achieve it alone. There needs to be individual and collective responsibility for data quality with an understanding of how issues flow from the start to finish of HR processes and how the impacts can be felt locally, regionally and nationally. This is something that I can’t overstate because it may seem like a little thing at the time but we need to all understand what the data is used for.

On that side, national and regional organisations need to give improving workforce data quality the priority it deserves. There needs to be more funding to ensure that local organisations understand the importance of improving data quality and have the resource to make an impact. We’ll certainly be pushing for that as best we can with the likes of NHS England and Improvement in relation to the implementation of the NHS People Plan. Good data quality in ESR means that there should no longer be a need to conduct further data collections for the service. This would save time and money that could be better applied elsewhere.

We all aim for ESR data to be accepted as authoritative by highly engaged users who know what they’re asking for and why. Again, I want to reiterate the fact that it’s not just about the data it’s about the people that are trying to use the data for analysis and their understanding of what they’re doing with the data and why and how.

Slide 19

So finally, ultimately data quality will save the NHS money by improving decision that are made. For example, reduce in locum and agency spend, ensuring that NHS Litigation Authority premiums are correct and this is where the position workplace org code field is so important because NHS Digital official statistics are used as part of the process for setting the NHS Litigation Authority premiums. So if you have staff that you host on behalf of somebody and you don’t want to pay their litigation premium then you have to ensure that you put the correct organisation in the workplace organisation code field.

Also, more positively we need accurate workforce data to be able to produce workforce plans that properly reflect your needs. HEE can only go on the information that they’re presented with when it comes to workforce planning.

Improved clinical care and better patient experience can come from having better workforce information by ensuring that we get the most from the long-term plan and the people plan.

You need to speak out about what you know to be important. Get others to listen and work with colleagues. Don’t just have a great idea and keep it to yourself, share it with others and try to make a change. You never know, it might really help and might work well for you and you might find it a very positive experience.

Help is available. Make the most of resources that are available through ESR and NHS Digital. Push for change where it’s needed. Challenge us if there’s something that you think is missing or can be improved. And remember, be positive: Data Quality is often much better than people think.

So on that positive note I’ll say a thank you very much, enjoy improving your data quality, all the best, thank you.

Data Team Report (59:29 > 1:01:30)

Slide 20

SW: Thank you very much for that Nick. I’m now just going to briefly cover some of the additional resources that we make available to organisations. So as well as the data quality reports that Chris and myself have covered within ESR BI, there are also some additional reports that we produce on a monthly basis within the data analysis team.

So, three of these relate to registration monitoring; so NMC, GMC and HCPC. These reports flag issues or potential issues with mismatching data in relation to the interfaces we have with those registration bodies. Finally, the unique ID monitoring which assists in monitoring records across multiple organisations.

These are sent to designated contacts at each organisation on a monthly basis. I’m not going to go into too much further detail on these reports you’ll be pleased to know but further information on all four of them can be found in the HR Best Practice guide which is available on the ESR Hub.

Just to flag though that we do have some organisations that don’t have a designated contact for one of the reports or in some cases all four. So can you please check that within your organisation that somebody or the correct person is receiving these reports and hopefully actioning them if there are issues to be address.

If you would like to make change to who receives it or would like to make an addition then please contact the ESR data team on the email address you an see on your screen there.

Thank you very much for listening; I hope you’ve found the resources useful and thank you very much to Nick Armitage for delivering his slides which again I hope you found very informative.

Slide 21

END