TRANSCRIPTION

I am the High Education Statistics Agency's CEO (HESA) and I am here this afternoon to talk to you about the how, what and why of data for higher education stakeholders, and you'll appreciate from that, that it's quite a broad spectrum of interest that I'm going to cover. To head in, I'll first of all introduce you to HESA as an organization, and then I'll
introduce you to the data.

As an organization, our mission is that we are in existence so that our data and our evidence is there for the advancement of higher education in the UK. We are really interested in being useful and providing information in response to the needs of all those with interest in its characteristics and in its future. That's a broad mission, and again we have a vision looking forward, and the vision is ambitious. The vision is that we take all these interests — that's the public, prospective students, students at university, the universities themselves, colleges, government organizations, funding agencies, politicians, policy makers... a very broad spectrum of people who may be interested in a truly comprehensive body of reliable statistical information — and develop them to the highest standards so that we are trusted as an authoritative, useful and cost-effective organization. That is our vision.

Briefly, this is a picture of where I work, in Cheltenham, in the West of England. It is a small town. We are in a very fine building — I thought it dated from the 18ᵗʰ century, but actually it is a 20ᵗʰ century cinema that was changed ... It fooled me! We have a staff of 90 people, organized, as you might expect, to be concerned with the information and operations, and then we also have finance and supportive staff. Our budget now is 5.5 million a year; important to remember is that the High Education Institution Subscriptions make up 3.9£ million of that, so a big chunk of that. We also have various other income, which I will talk a little bit about, including commercial income, and our main cost, as you might imagine, is the cost of staff.

We are 20 years old in 2013, so that's long enough to really get established, and to develop, and always to be looking at how we are regarding a substainable future, a future that people can rely on, because we have a lot of data that is historic now. Most unusually, perhaps, for you, is the fact that we are an independent company, limited by guarantee, and we are actually owned by the

sector, the higher education institutions themselves, although under the arrangements of the company the membership that own us are the umbrella bodies, in our case Universities UK, and another smaller organization called Guild HE. Those two organizations represent the vast majority of higher education establishments in the UK.

However, in order to collect the data at the level we collect it, we do need to be enabled through legal and statutory requirements to collect that data, which often is detailed, individualized and personal data, and we have those statutory powers by virtue of collecting the data for a specific group of people, our statutory customers. I'll tell you who they are in a minute.

Let's just revise. As I said, we are owned by the universities themselves. In the UK, there are currently 165 higher education institutions across England, Wales, Scotland, and Northern Ireland. We also have in our membership one private university — private universities in the UK are a rather new but growing phenomenon; the current government is very keen to open up higher education to private organizations — and we have another eight independent providers who are just coming to us to provide data. Up until now it's been full entry. Things are changing and there is an expectation that these newer private universities, if they are, or higher education institutions, and colleges in some cases, will also provide some data to us.

I mentioned the statutory customers, and the fact that we are required and enabled, through the Higher Education Act, to provide data. These statutory requirements mean that if a higher education institution is to receive government funding, then they are required to provide their data. It's part of the contract, and that contract is known as the "financial memorandum" between the funding body (government) and the institution. The funding bodies then require information about student numbers (because they exert student number control, they manage the number of students in the system, because they need to manage the cost of those students to the government). They want to verify and check their funding allocations. There is also a requirement by government that there is quality and diversity management and monitoring within the sector — so information about ethnic groups, about the proportion of men and women, about, therefore, the equality of the provision of higher education and those undertaking research.

Then we also have in the UK something that you may have come across, which is "widening participation". That represents the determination of the government to make sure that we have students from poorer backgrounds coming into university. Universities are required to show how many students come to the university from the less well-off areas of the UK. Clearly, another statutory reason the government should have this data, is to develop and monitor government policy. So those are the reasons why, through legal and funding requirements, the institutions are required to provide data to us. Those statutory organizations are across the UK, and in every case, for each of the countries (England, Wales, Scotland), we have a government department responsible for higher education, and we have a funding council. The funding goes from

government through the funding council, to the universities. Perhaps importantly for some of you here, the research councils are also our statutory customers, because they have a requirement for some of the data in order to inform them. I'll give you some examples about that.

In addition the Training Agency (which is all about training teachers) also require information from the universities and colleges about the number of students they currently have in training for teaching in different subject areas.

Of growing importance over the last five years has been the expectation that there is public information provided through us from the institutions. Any expectation that higher education institutions are much more transparent, and that they are able to inform the government about what they are doing and why, and who, and that's, I think, very much the theme, also with research and looking at the impact of research — what is the impact of higher education? How can we challenge? How can we make sure that universities are actually doing what is required by society?

But I think there is a huge payback in collecting this data for the institutions themselves. They are now very dependent on that data, and by virtue of owning HESA, they have ready access to data about themselves (clearly they have that backup base), but more importantly it allows them to look at the activities of all the other universities. One of the particular uses of this comparative data is for benchmarking. Benchmarking means that you can compare your performance with the performance of other universities. You can develop your policy, you can look at the costs in other institutions, and consider the numbers of staff, the research, the number of students, etc. This might seem a bit frightening a prospect for some, because in a way institutions like to remain quite private, and there is the problem of competition, but at the same time there's a huge benefit, over the last few years, of being able to look at and compare, and match your performance with other institutions. Institutions also use it, therefore, for planning and monitoring their own policy and strategy.

We publish a set of performance indicators. One set is about "widening participation" — we compare how different institutions are doing, as far as taking students from poorer areas, but another important set is the research performance indicators, issued July every year. On our website, we are currently showing the research performance indicators for 2 010-11, and a chart that compares data across research institutions. It's not just raw data taken randomly: we actually worked very hard to try and ensure that the performance is managed against the profile of the different institutions. If you have a medical school, your research profile is likely to be very different from if you're an arts and music college, for instance. Those differences are taken into account and the sort of information we provide are the number of PhDs, research grant contracts, staff in different subject areas, specializations, and we look at those as percentages across different benchmarks, which are set against the, different profiles of universities.

Just an example of the way the data can be used and therefore helps institutions to look at, assess and manage their performance: the number of students who

drop out of university. We all very clear now of numbers of students dropping out, and if an institution sees that it has a big drop-out rate compared to others, they'll look quite critically at what they are doing wrong, and how they can improve the experience, and widen participation, as I mentioned. We have something that we call the Quality Assurance Agency, which looks at largely teaching qualities, and again, we publish information about teaching...

So how do we do this? This is the "what" data. I'm just going to run through our main collections of data. The biggest collection is student data. We collect it once a year, at the end of the academic year, and what we do is take the life cycle of the students, so we know how many enrolled, what course they are taking, how many students drop out, how many move on to their second year, we know the degree they complete with; all that information is collected on an individual basis. We don't collect students' telephone numbers, or their address, but we do collect the student name, and the individual record of each student, so you'll appreciate there is quite an issue with data protection that we manage. We also collect data about students that are the responsibility of UK universities but are being taught on campuses abroad. And we also have a specific collection about, as I mentioned, teacher training.

Rather importantly, and I think this is important for those of you who are interested in research data, we collect the destination, the careers and the jobs of students once they have left; we actually carry out a survey called "destination of leavers", and we now hit 80% of the students who completed a degree course or a course at a university and will let us know what they are doing six months after they leave. So we know how many students are in employment, how many are unemployed, we know the jobs that they are doing, we know how many go on to do a master's degree, and because we also collect from masters students, we know how many go on to become research students. Then we do a further collection, which is a longitudinal survey, three years after they've left university, so we can watch their career progression, and interestingly, the research council actually pays for an additional cohort in this survey (because this is a much smaller sample) to try and gain an almost complete picture of the number of students that are actually carrying out research. They are just about to introduce a survey seven years after graduation, to find out what happens to those students that have undertaken research, that have become, perhaps, researchers within institutions going off into industry... And so the research council has got a unit that is very closely tracking the outcome of the research student activity.

Then, in order to provide a much fuller picture, we have an institution profile record; we ask institutions to put their departments into regular cost centres, so we can draw comparisons — we have a "science cost centre", a "chemistry cost centre", a "physics cost centre"... — and then we can put our other data into those cost centre collections to allow comparison. We carry out a well-detailed finance statistics record so that we would know research input in the university, and have a full picture of income and expenditure, which enhances the published accounts.

Again, possibly of interest for some of you is the "Higher education business and community interaction survey". We published a press release, which we do for all

our collections as they are published, about the HE business & community interaction survey, to say "the results are now out there", and the sort of data that we collect in that is, for example, the amount of contract work that universities have — not necessarily pure research, this is beyond that, the work they do with small and medium enterprises, and the interactions they have with industry. This survey will gather together information that is about, if you like, applied research and research as it's taken out into the business community and the social community.

We also collect records about staff, so we know how many staff there are at universities, whether they are teaching and research, only research, only teaching, or administrative and support. I was very interested in a question this morning about identity, and the ability to identify staff and researchers, and their names. We've been struggling, perhaps for the last five years, to decide whether to include the name of each staff member, and so far there's been too much anxiety amongst the unions, the staff themselves, so very interestingly we're looking at Orchid, the international registration of researchers and others against an ID, and it is quite likely that HESA will actually become the registration body for Orchid in the UK. We would then be able to track staff around the system, for one thing, and then actually identify them back with their publications. So at the moment, the staff record is anonymous, but it does collect data at the individual level. Finally, we collect data about estates, and that's actually including, not only detail about research facilities, but a certain amount of data about the investment in research and teaching facilities, but also about number of teaching space, lecture theatres, accommodation and so on. From that, you can tell that it is a very broad spectrum of data we have, that you can then interplay with each other to support a lot of very, very diverse and different interests in higher education.

There is a question: How do we know that the data we collect are the right data? And how do we manage that process of data collection, year on year. We run a structured change process, we run a review of our data collections on a regular basis. We have a round table, which gets bigger and bigger to accommodate all the interests: we have the demand part (statutory customers, government bodies, funding councils) on one side of the table, and we have the supply part (HE institutions) on the other side of the table, and we HESA chair it. So you can imagine it is a long and hard process, because the demand side might say: "we want more information about whatever subject", and the supply side answers: "You've got to be kidding! It would take us twenty weeks of intensive research to find that information". However, we come to a peacefully agreed data set. By virtue of this round table, which can be four or five sessions, we then go out to broad consultations, with the whole sector. This is on our website too. You can see the changes being suggested, we take views and opinions, and then we get around the table again, and say: "For next year, we will be making the following changes". Very interestingly, at the moment, in the UK, there's a lot of policy change for higher education, and whereas we used to do this exercise every four years, we are now finding that it has to be on an annual basis... You can imagine that universities find it

quite hard, not least because the way it works is that we pull the data from their internal information management systems. For those of you who are technical, most are through an xml data pull, with fully defined fields, so they know exactly what they've got to provide. They sometimes have to change their internal system to allow for this data collection. So why do they do it? you might ask. Well partly because if they didn't, they wouldn't get the money in the first place, but also because they now gain so much from the process themselves, that they are willing to input an amount of effort because of the payback, to them, of having that data available to use, and being available to the public, and avoiding multiple demands for data that would otherwise come into their institutions. I did mention the importance of having reliable data. Very importantly, for us, the quality of the data matters: keeping a good quality data that is fit for purpose. We do work hard to identify errors in the data. HESA has a dialog with the institutions during the collection of the data and we have tests on the data; some of them are obviously computerized checks on the data. We get to the point where we're visually checking some of their data and we'll see an anomaly. For example last year they spent 200,000£, this year they are spending two million. Then we say: "What is that all about?" and they'll explain they built a new building or whatever it might be. So, there is this quite intense period of data collection to try and maintain the quality. This is the process we go through. I won't talk about it in detail, but there are five steps. The "sign off" verification is where the head of the institution (in our case it would be the vice-chancellor) signs to say: "This is the data set that my institution is providing". So there's a sense in which they really own that responsibility with us.

What is so important in this process I described is that the higher education institutions are so much HESA customers, are so much the ones that we depend on, we need to have a very good relationship with them. One of the ways we do this is we have a lot of information on the website, we're very, very keen to support and communicate, so there is a helpdesk (telephone, e-mail); I have staff that the institutions really do hold in very high regard, for the support that they can give. We also run training seminars, and that's not just about how you give your data, we also train on how you use the data. So we have this good relationship with many people inside of university: we'll know the finance directors, we will talk to the academic registrars, discuss the collection about student employment with the advisers etc. What is actually helpful for them is that we bring them together to talk about it as well. They really gain from that sense of community, and talking with each other.

I arrived at HESA three years ago. I had been at Bristol University on the executive team, and I knew about HESA that we had to give them our data... Can't say I liked HESA very much, and the next thing I knew was I was going to be running it, so... Interesting decision! What I knew I had to do in moving to HESA, was to put much more effort into access to the data. There is no point in everyone putting effort into collecting that data if it's not been available to use.
There are different routes to accessing the data. There is the very traditional paper publication; some of those were actually national statistics, but on behalf of our national statistics office, we would publish the data; so they don't do a

separate collection, we do it on their behalf, to their standards, and we are inspected from time to time, to ensure that we reach and maintain those standards. We have a bespoke data service, which means individuals can come to us to ask for data. We make available a database of all the data at a certain level for the institutions to use. I'm not going to spend time on this, but if you are interested, you can go onto our website, where you'll find a lot of free statistics available — usually the package is an Excel spreadsheet. We're putting more and more data online. We publish our traditional volumes on an annual basis. In one year after I arrived, we turned it from paper to electronic; we don't do paper anymore, except for one publication for the National Office of Statistics. Otherwise they are all electronic. The institutions get it and share it. For others we charge a fee, because we put quite a lot of effort into putting that data together into the publications, to make it manageable. We also have a mobile app. Since its creation, our heads of institutions certainly realized that HESA was quite useful. When someone asks: "How many students do you have this year? How many staff? How many are international?" instead of a rough guess, people can now do a quick check on their iPhones and find that information. We're putting more and more statistics on that app. You too can download the HESA app for your mobile phone to get the UK data.

I mentioned the bespoke service... We have a lot of inquiries and it's so interesting to analyze such a big number of inquiries and where they come from. We have big employing companies coming to our record to find out where students with a particular degree are to be found; they will do research about where their respective employment targets are. We'll have, of course, organizations wanting to do good university guides, and inevitably, league tables. We do charge, but we charge for the time and effort of actually pulling the data that's required together and making it available. We also make it available with a licence, so that from the point of view of the universities, they are assured that if it's what we call "high profile data", they will see it first, before it goes out. So if *The Guardian* or any other newspaper wants information that we know will be published at a league table or, potentially, in an article, we let the universities see that data. If they aren't happy, then they're able to make a contact with that newspaper and say: "I don't want you to publish that", or "I need to change that figure", or whatever it might be. So this is the information management that we're able to provide, which is one of the paybacks for the fact that they subscribe to us as an organization.

During the twenty years of HESA, and I think in my time never, we very rarely say "no" to a request for data, but we may advise on the information, and the other thing, critical, is that we manage privacy and data protection, so if we give data out, we make sure that they will always round up to five minimum, so you will not be able to identify an individual from data published.

A quick word about Heidi... This is what universities have found absolutely invaluable. It is a querying tool, which can also produce reports, visualizations, etc. across all the universities' data. I mentioned benchmarking, comparison, planning... They are able to

see the data of other institutions, at a certain level. It doesn't go down to the individual student, but it does give them a lot of detail. It is certainly used for example for research profiles. We pull in additional data for Heidi; we have our collections, and we pull in data from the research councils, about libraries and IT, more detailed data, data from our National Student Survey about student satisfaction, so it's a comprehensive resource.

The Unistats website is very new in the UK — I don't know if any of you are aware of this, but it is important if you are interested in teaching and student recruitment: we have started to publish, this year, a dataset which allows prospective students to go into every university and every course that is available, to compare data about student satisfaction on course, fees... There are about sixteen different fields. That is all being put together with a fairly new data collection, and that was government policy. The Minister said "I want this", and it has been done. This is the quality assurance I mentioned earlier. We make data available as public information, which is being seen as a very vital part for a university and for the public confidence in higher education.

I'll just say a few words about research data collections, appreciating it's the area you are most concerned with here. What we've got, as you know, is dual funding: funding council grants and research council awards. The funding councils fund research largely based on the research assessment exercise, which I'm sure many of you are familiar with. Depending on how you perform in the research assessment exercise, a grant will be allocated to your university and allow a grant that you have the choice of spending on research activities. The research assessment exercise that took place four years ago is still relevant, and the funding councils also use the data we collect about input funding. The research councils (I'm sure you must be well aware of the way the UK research councils work) are also our statutory customers. They collect data from us, particularly about staff and post-grads, and also about finance, but they also have their own data collections, and clearly, they have much more detail about their awards, and they collect information about impact and outcomes from those awards.

Then we have the institutional level. Again, institutions will use the Heidi tool to look at the institutional level, but they may have their own management systems. I don't know if any of you have come across Snowball, which is trying to encourage institutions to be more sophisticated in their local performance management through information. Moving from history to current times, the big change in the research excellence framework is that there is now a serious collection on research impact, case studies. They do have a specified data collection about publications and outputs, about inputs, and based on that, will build the next period of funding. Interestingly HESA is being used in particular for the staff data, and whenever there is data that can be pulled from HESA, institutions will not be asked for it additionally. There is a whole series of data analysis around the finance and the research performance indicators.

This is my finale, now. We did a map to show that the higher education institutions, the HE providers really are the source of the data, but there are several demands on them. Here are the demands that HESA makes, and the

whole series of data collections, but there are other organizations demanding data, inevitably. There is the student loan company, the professional bodies (veterinary, nursing, etc.), the UK BA, which is now dealing with international students and the requirements that they have visas when they come and study in the UK. Finally, the funding council asks for their own dataset, since some of the data is rather confidential.

The key to this is that, although there are several data collectors, once we have **the data, it is put through a lot of different uses by a lot of different stakeholders.** We did a survey to try and find out how many demands were made on universities for data. It turned out over 500 different requests for data were captured in this survey; not every institution would have 500 requests for data, some might have one or two hundred, and some others didn't know, because people would come into their different departments and ask for data.

It was quite alarming to find out the numerous demands for data, and therefore, the sense that our education institutions had of the burden of data collection and felt this was getting out of control. We have now undertaken the start of a project to look at the future of data collection across the UK higher education landscape, and we are coming up with various requirements. Timeliness: people want data, and they want it now. They do not want to wait two years, even 15 months is too long, so we've got to look at collecting data in such way that we can get it available much more quickly. They want quality, they want to maintain quality; they want to be able to change their demands on that data; they want it to be open and accessible, and we need to make sure that it's relevant for diverse sectors, and also the private organizations, big and small. This project we are undertaking has got various principles: Collect once for use many times, share true standards (the technical bit most people find hard, dull and boring, but it is quite important, and we need to maintain that), take advantage of the technology that is now available... We are putting in place, looking forward, a new overarching governance process, to try and pull into it all those different demands for information, because it's not just HESA, it's also all these other organizations asking for data. The idea, now, is that they all have to go through, like, a big funnel, so that they have to justify their requirement for that data, and if it's clear that that data is already available, collected by others, then the request and the requirement is that they use the data that's there and don't ask for it again. Likewise, we're working now on a more available common data language, with commitment from different organizations. We're trying to find out, through this inventory of collections, who does, more specifically, ask for what data, when and why. There is for instance the unique learner number — I don't know if you have something similar in France; schools now have now to give every pupil a number, and what we want is for that number to stay with them all the way through the education system, so that they can be tracked. For the moment, they lose it when they enter university.

We're not content about where we are, and we need to make changes and rethink some of what we do. On the other hand, we've got a system that people currently trust. They may grumble a little, but they do trust it, they do engage with it, and that data is highly used.