

Data driven – how data capability can support better outcomes for PSRBs – the example of GMC's UKMED project

Dan Cook, Head of Policy and Development, HESA

Daniel Smith, Education Data Development Manager, GMC

Denise Jones, Head of Data Intelligence, HESA

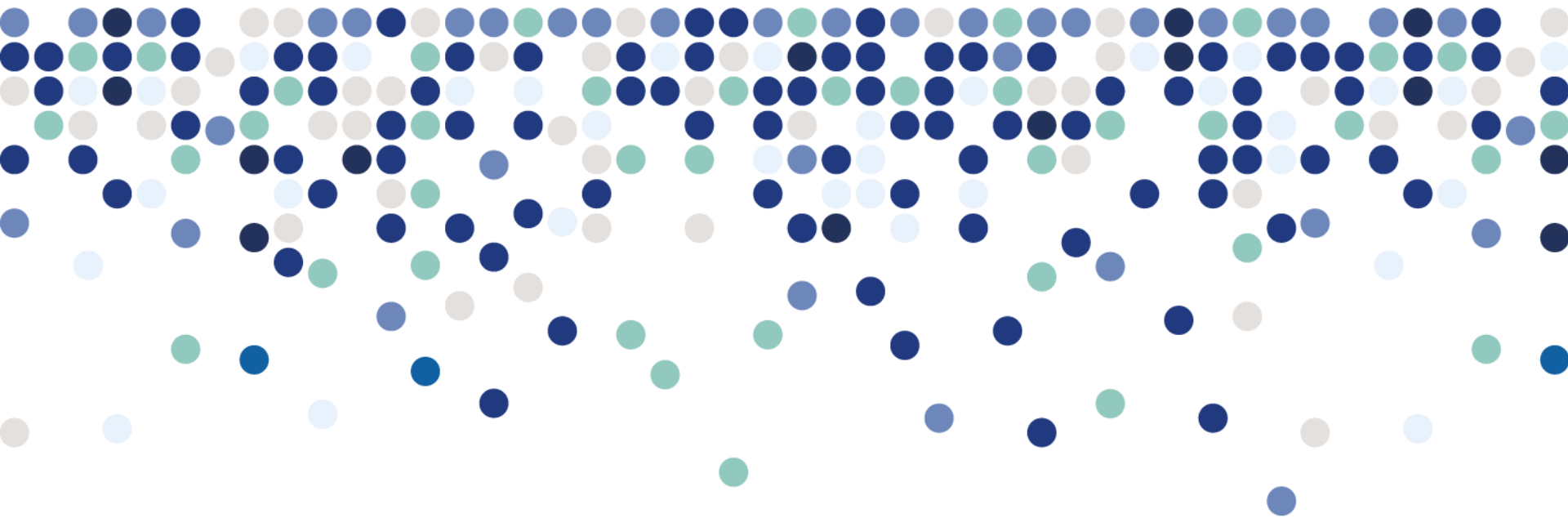
2018-11-28 – QAA PSRB Forum, Woburn House, London



We will cover

- HESA's own data management approach (Dan Cook)
- The GMC's UKMED project – data capability in action (Daniel Smith)
- Your own data capability journey: tools, services, and future developments (Denise Jones)

HESA



**HESA's approach to its own data
capability – how we manage data**



Who are we?

HESA are trusted experts, with 25 years' experience in collecting, processing and publishing data and are a charity operating as an independent organisation.

- We are a sector-owned shared service, set up by agreement between relevant government departments, higher education funding councils and universities and colleges.
- We operate under a statutory framework on behalf of the funding councils and UK government departments.
- We are principally funded by subscriptions from higher education (HE) providers.

Our mission

To support the advancement of UK higher education (HE) by collecting, analysing, and disseminating accurate and comprehensive statistical information in response to the needs of all those with an interest in its characteristics and a stake in its future.



Our vision

To be the analytical powerhouse for the UK higher education (HE) sector

- To be the trusted source of national HE statistics and public information.
- To provide flexible, efficient, high quality data, information, and analytical services to meet user needs.
- To support better decision-making in higher education.
- To promote public trust and confidence.
- To enhance the competitive strength of the UK HE sector.

Subscribers 2016/17

Higher education providers
267

Higher
education
institutions
164

Alternative
providers
100

Welsh Further
education
colleges
3

England
233

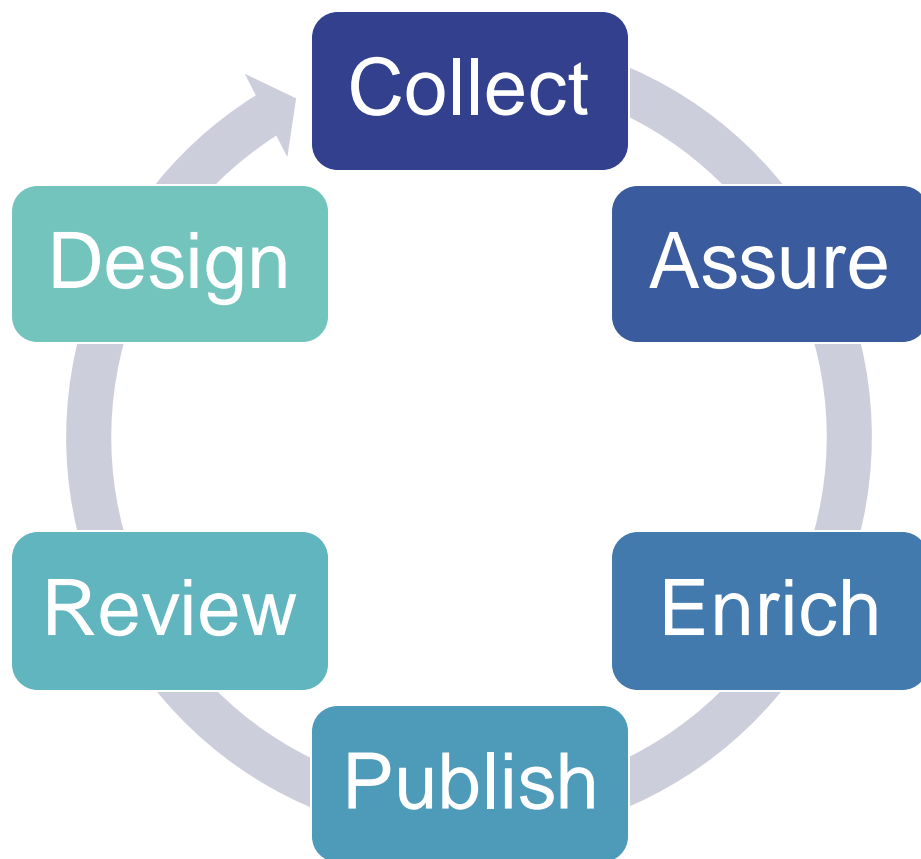
Wales
12

Scotland
18

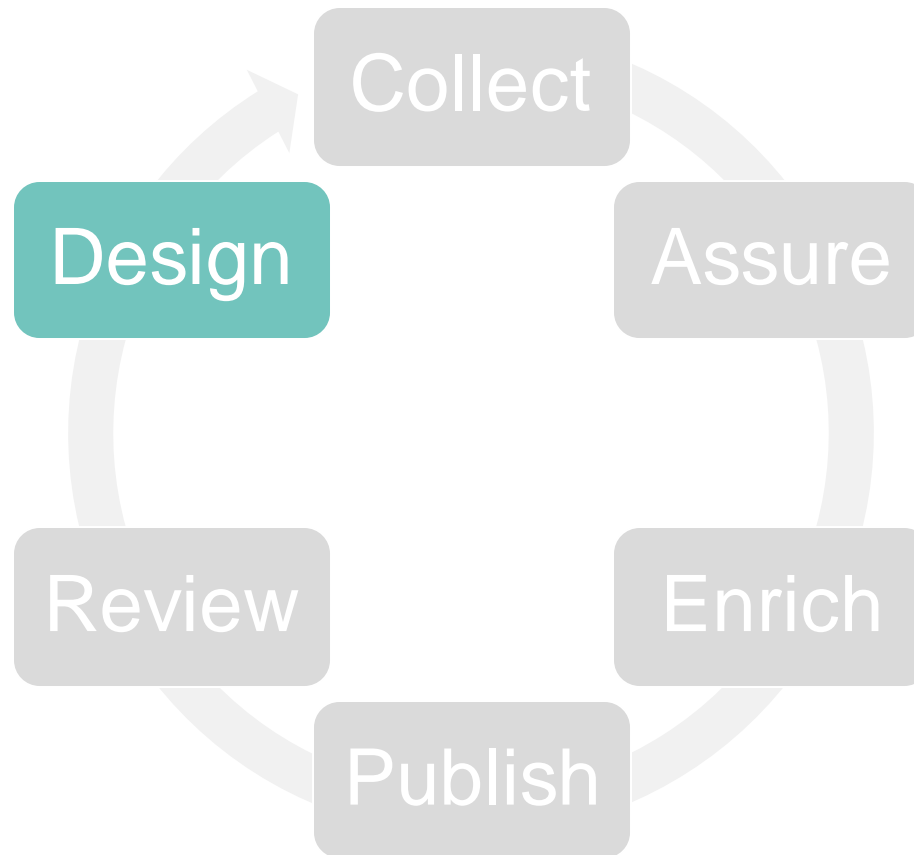
Northern
Ireland
4



HESA's data management lifecycle



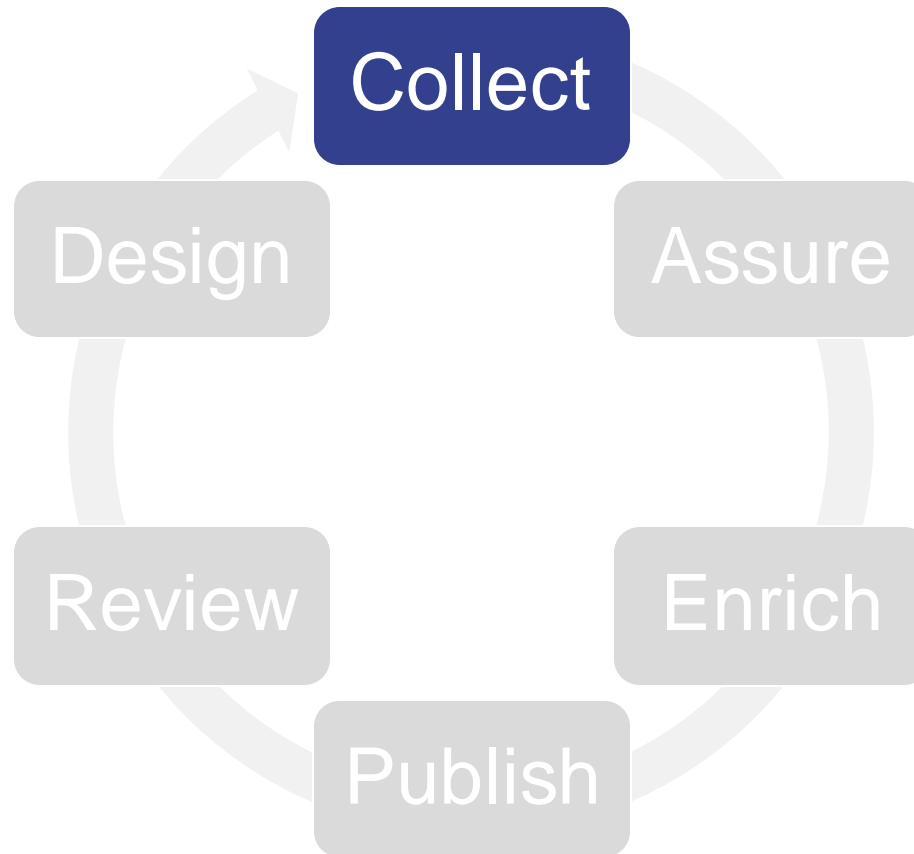
Data management lifecycle



HESA's approach to data design

- Coding manuals, data models and data dictionaries
- Clarity about uses
- Data architecture
- Logic model
- Lead times

Data management lifecycle



How much data does HESA collect?



- 2 million higher education (HE) students per year
- + 50k at alternative providers
- + 200k FE students
- 500k + graduates surveyed



- 200k academic staff
- 200k non-academic staff
- Governing body composition

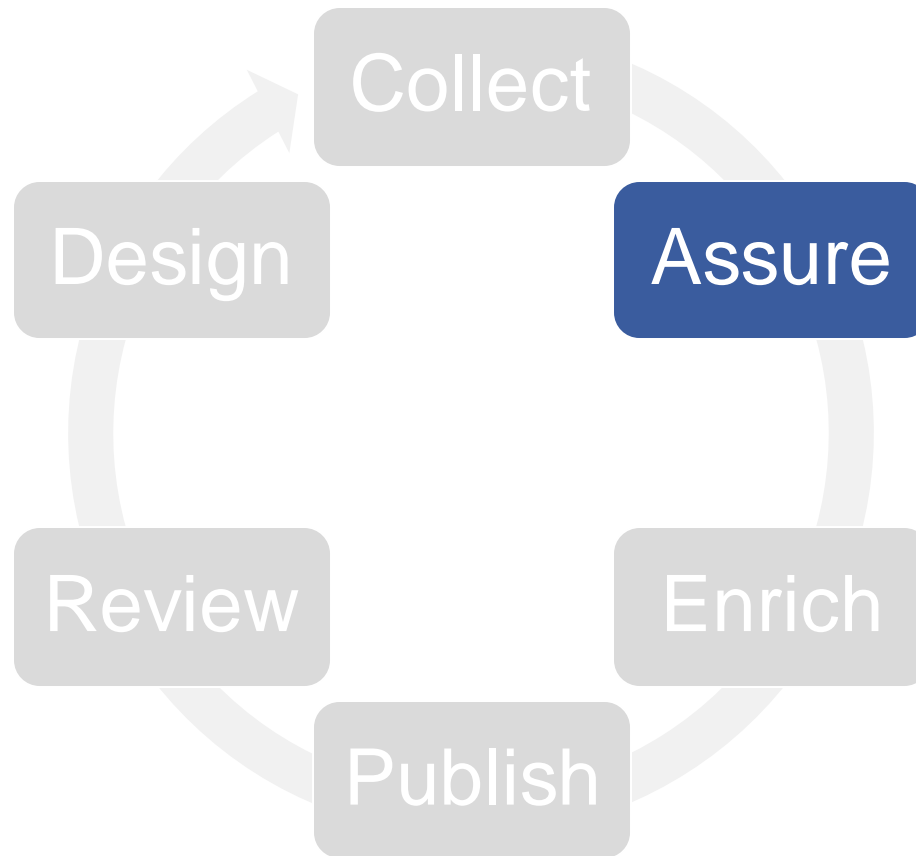


- Accounts of HE providers totalling £35 billion of income
- Data on a physical estate totalling 15,000 ha
- A wide array of HE-Business and Community Interactions data

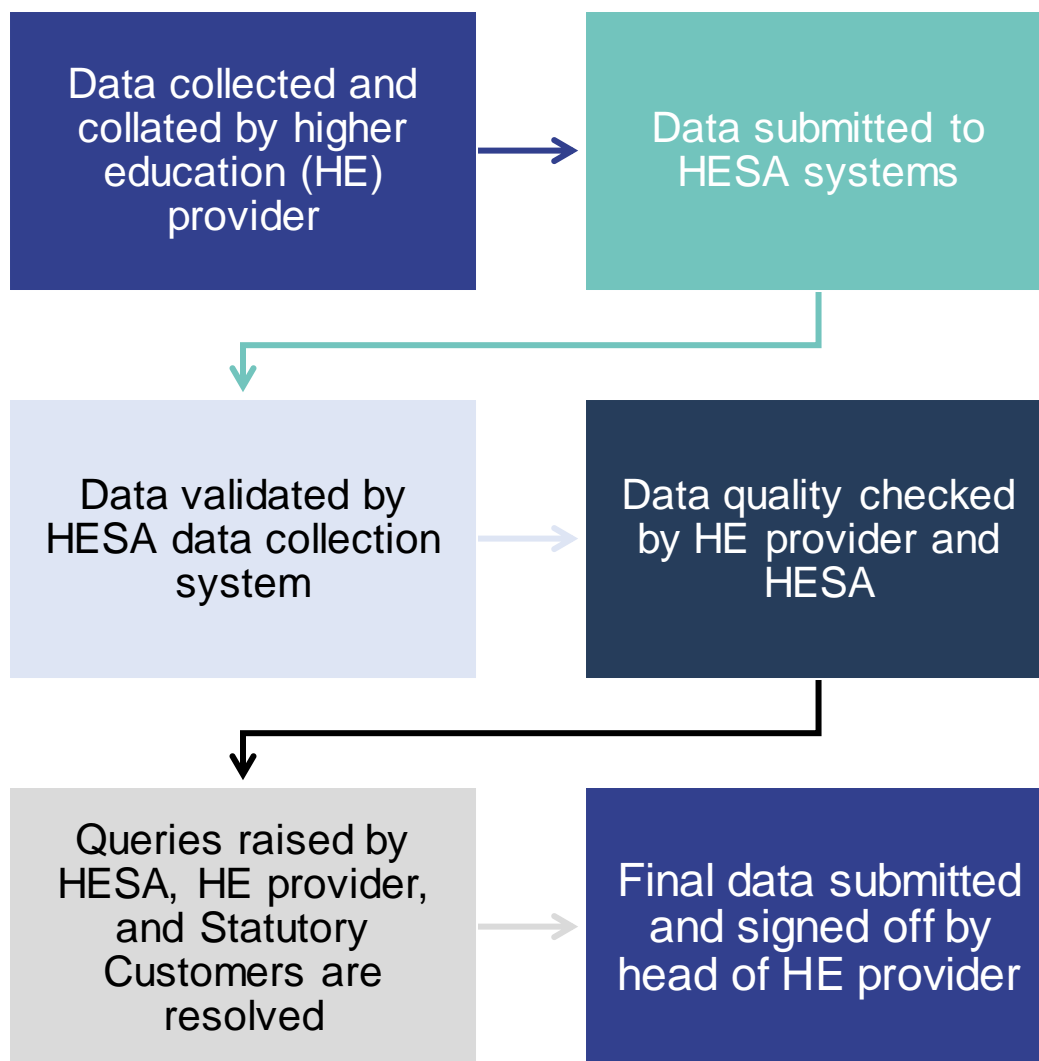
HESA's approach to collecting data

- Systems
- Security
- Support

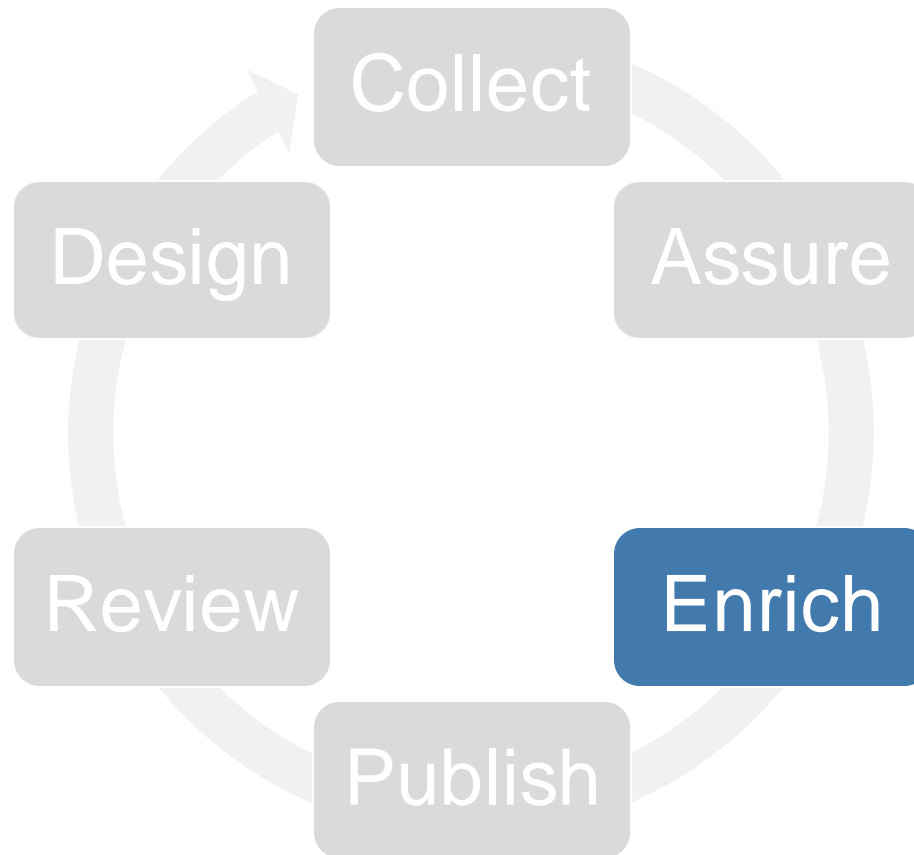
Data management lifecycle



HESA's approach to quality assurance



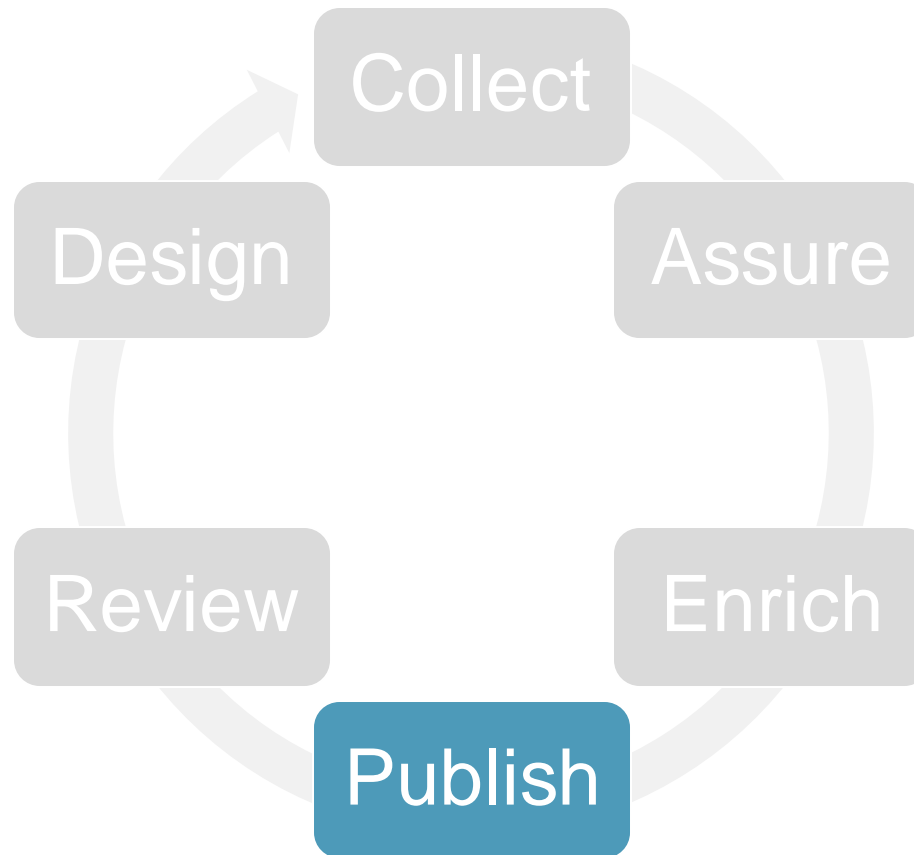
Data management lifecycle



Enriching supplied data

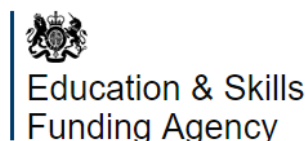
- Derived fields
- Data supply

Data management lifecycle

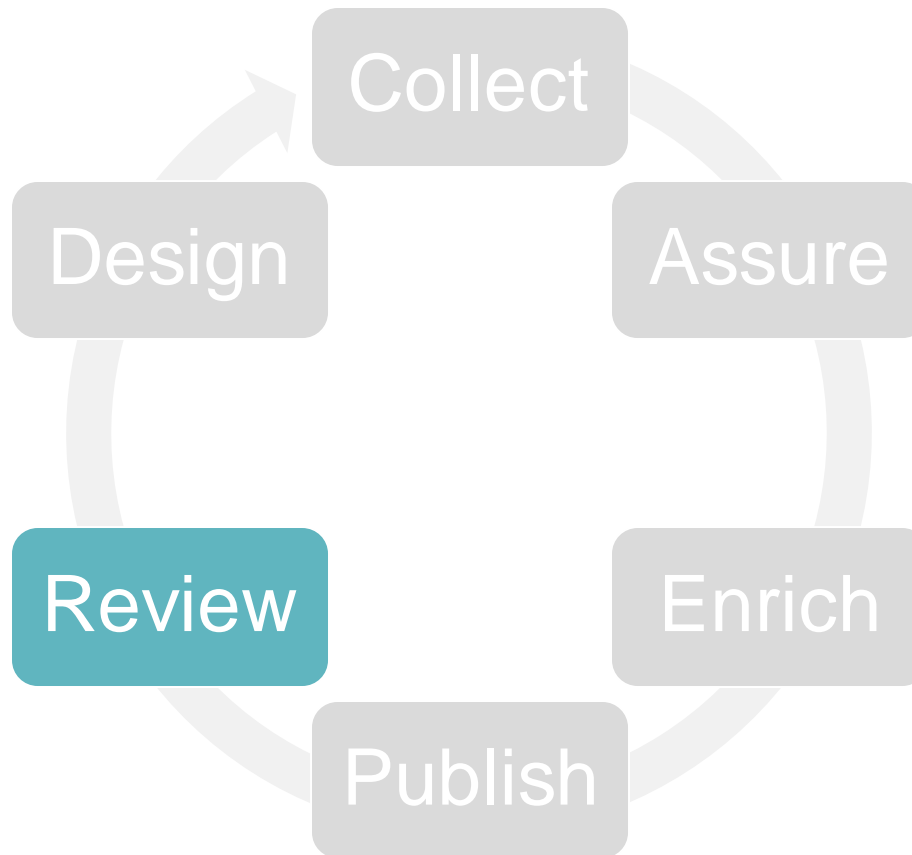


Providing data and analysis

We deliver the data we collect to our **statutory customers** - public authorities with a statutory requirement to receive information from higher education providers

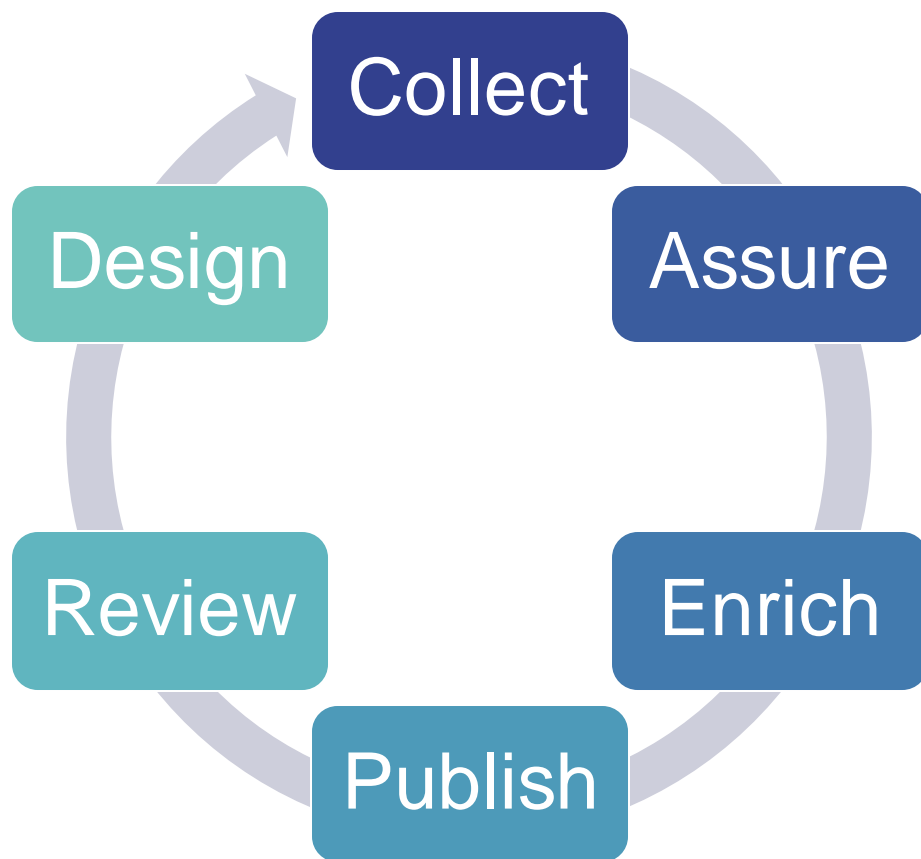


Data management lifecycle



- Reviews – annual updates, major and thematic reviews, post-implementation reviews
- Stakeholder involvement
- Consultations
- Burden assessment
- Notifications of changes

HESA's data management lifecycle



UKMED Introduction

QAA PSRB Forum

Wednesday 28th November 2018 - Woburn House

Daniel Smith



- **Objectives**
- **Data sources**
- **HESA and UKMED**
- **Process – lessons learnt**
- **Outputs so far**
- **Next steps**

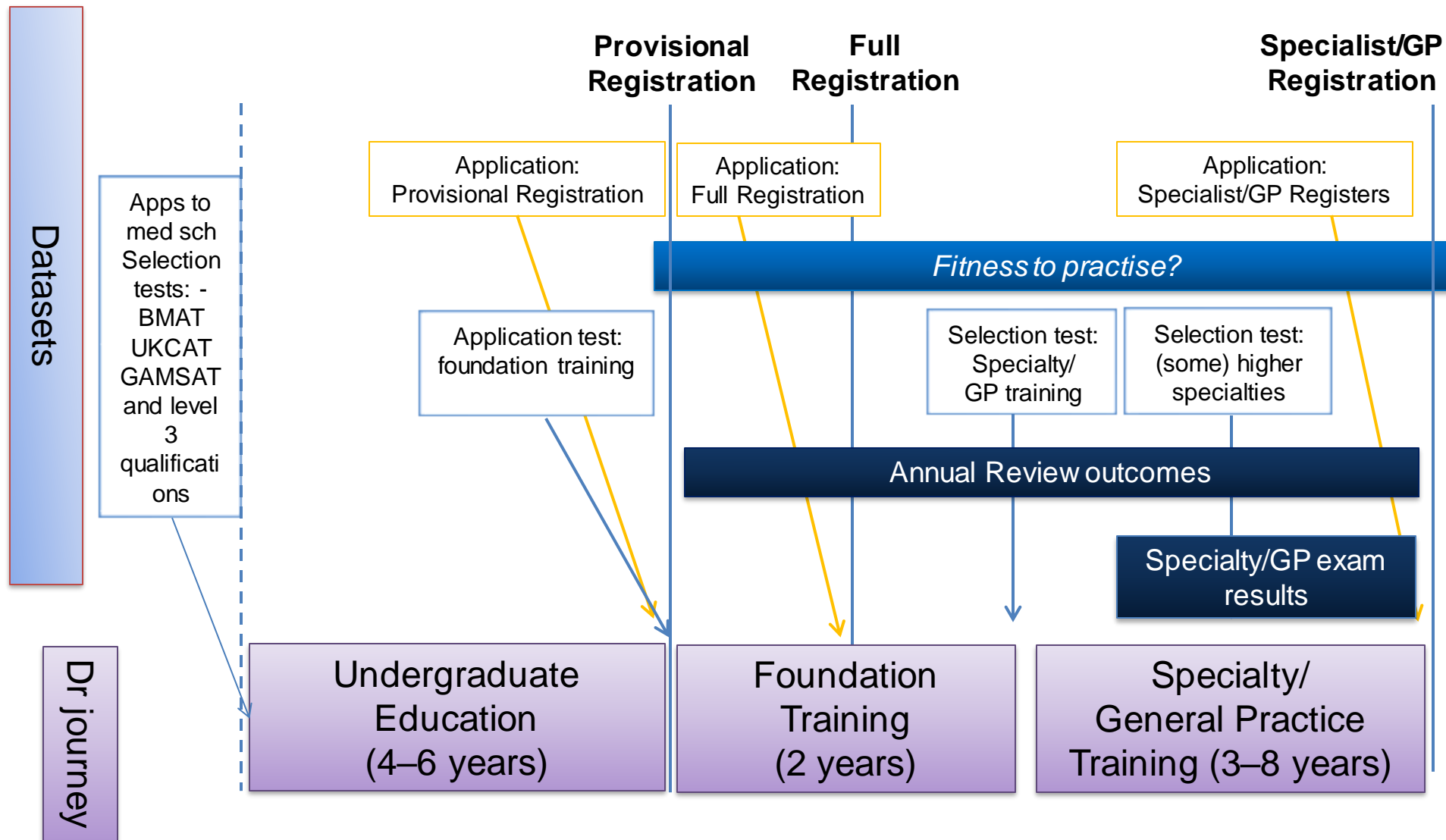
Objectives

- To find out if we were selecting the right people to become the best possible doctors (predictive validity).
- To determine success factors for widening access to medicine for applicants from non-traditional backgrounds
- To explore doctors' career choices to inform workforce planning (who applies to GP)
- To determine the key factors associated with doctors' progression, and the value added by each stage of training
- *And so on* in terms of determining performance beyond training, into fully-qualified posts (consultants and GPs)

UKMED – what it adds

- UKMED is a database that allows us to track Drs' careers from application to medical school through to working as a consultant/GP.
- First time GMC has held identifiable data on undergraduates.
- Linked undergraduate data including data from entry to medical school (test scores and level 3 qualifications) to postgraduate data.
- Agreed shared governance arrangements.
- An infrastructure that allows us to share data with researchers that ensure full compliance with the Data Protection Act.

UKMED – Stage of training covered by the database



Data protection – data in

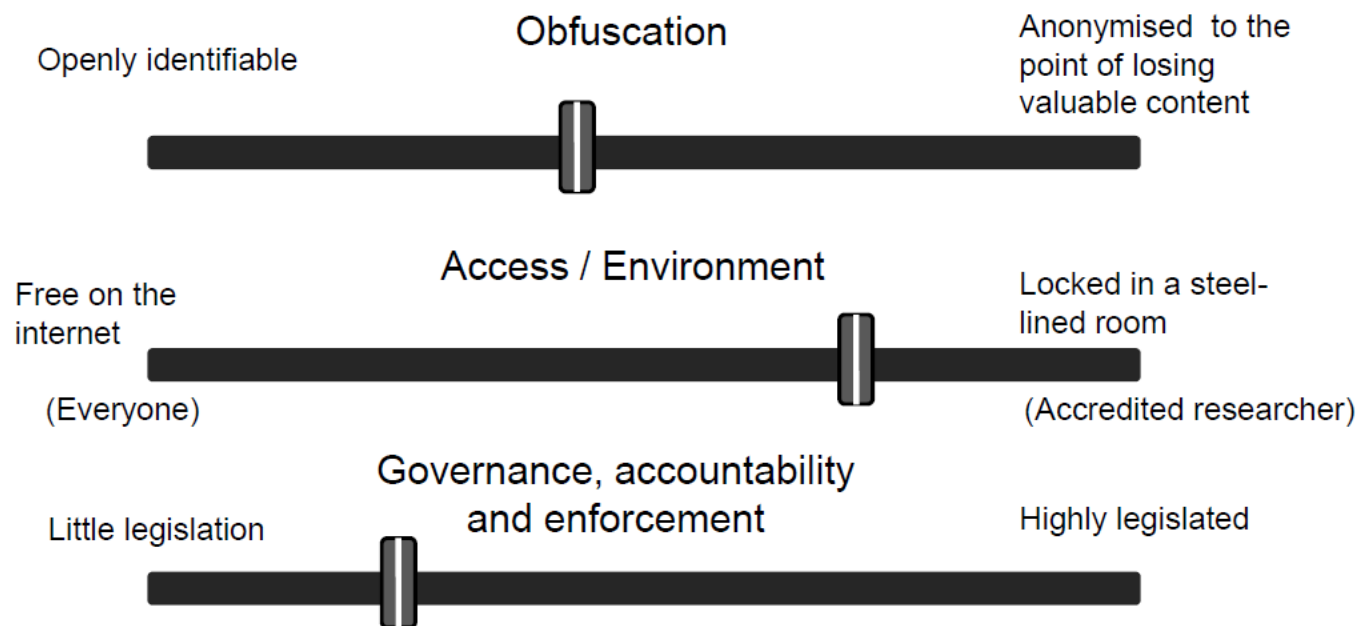
- The database needs to have a legal entity as a data controller. For UKMED there were benefits to the sector in the GMC taking on this role.
- There are 6 bases for processing personal data. UKMED uses GMC's statutory powers (our public task)
 - Section 5(1) of the 1983 Act gives the GMC the general function of promoting high standards of medical education and coordinating all stages of medical education.

Data protection – data out

- UKMED does not provide any identifiable data to potential employers and data are not used to monitor or make decisions about individual doctors.
- Data are held in UKMED for research purposes only. Doctors' information can't be used to make decisions that could impact on their career, either positively or negatively.
- Process of review before publication to ensure no surprises for data contributors.

Data protection – data out

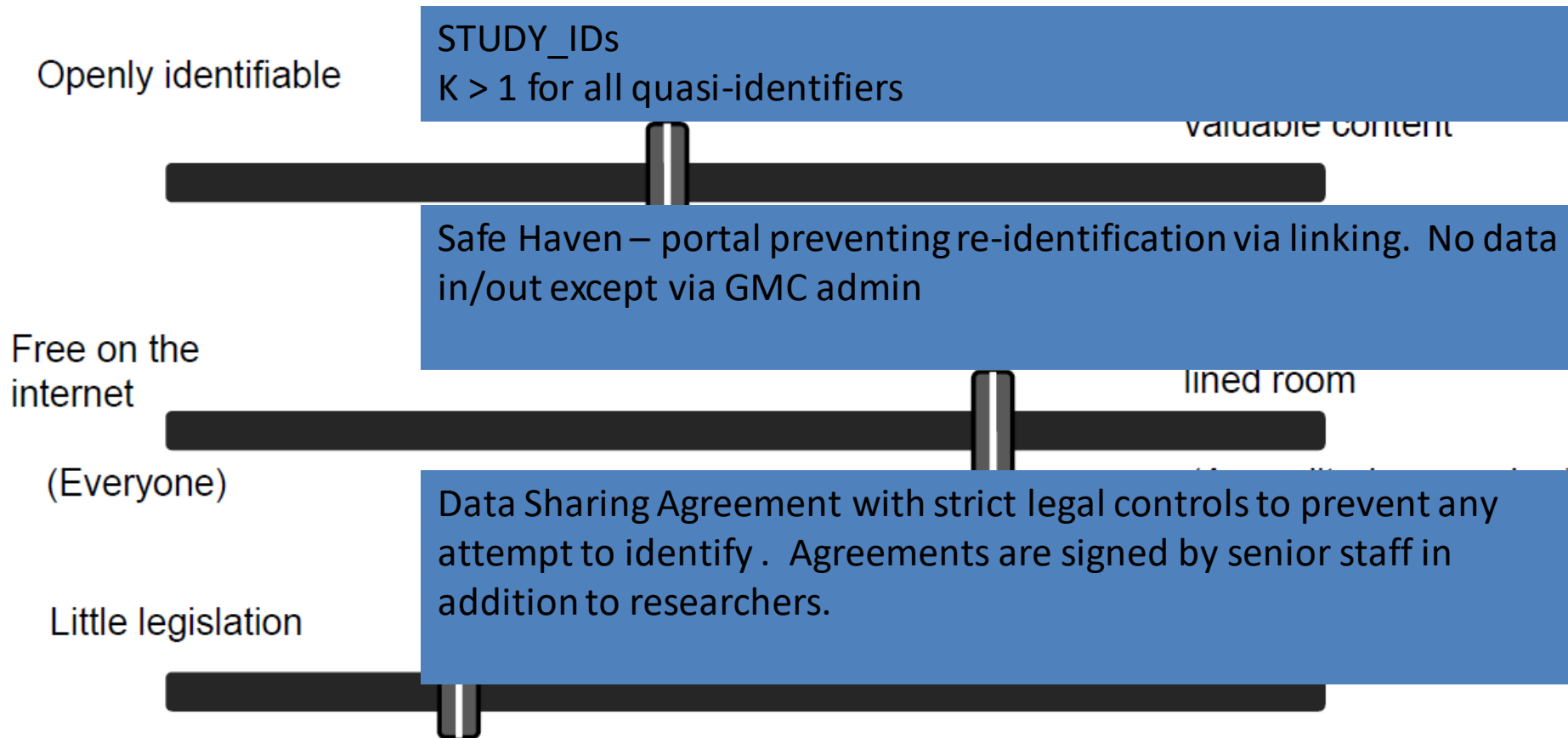
Privacy controls are not binary but fall on a spectra



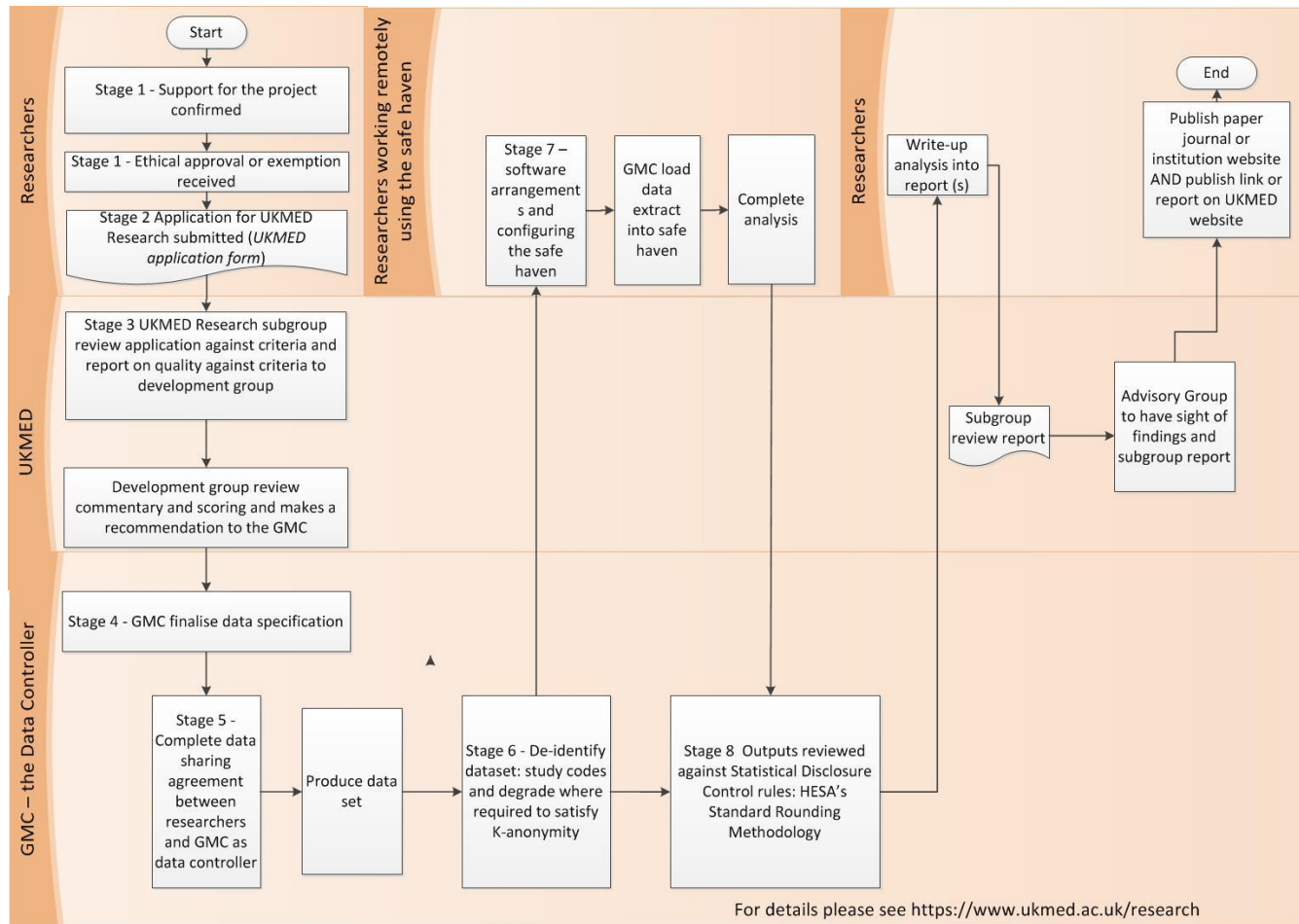
Source: **Data, Anonymity and Consent UKAN**,
September 11th 2014 Sir Mark Walport



Privacy controls are not binary but fall on spectra



The UKMED Research flowchart



Defining the population

- For UG studies: entrants based on (HESA)
- But for PG studies there are non-UK trainees of interest – for example validation of selection methods for specialty training.

UKMED Population

The 2018 V1 of the UKMED database includes:

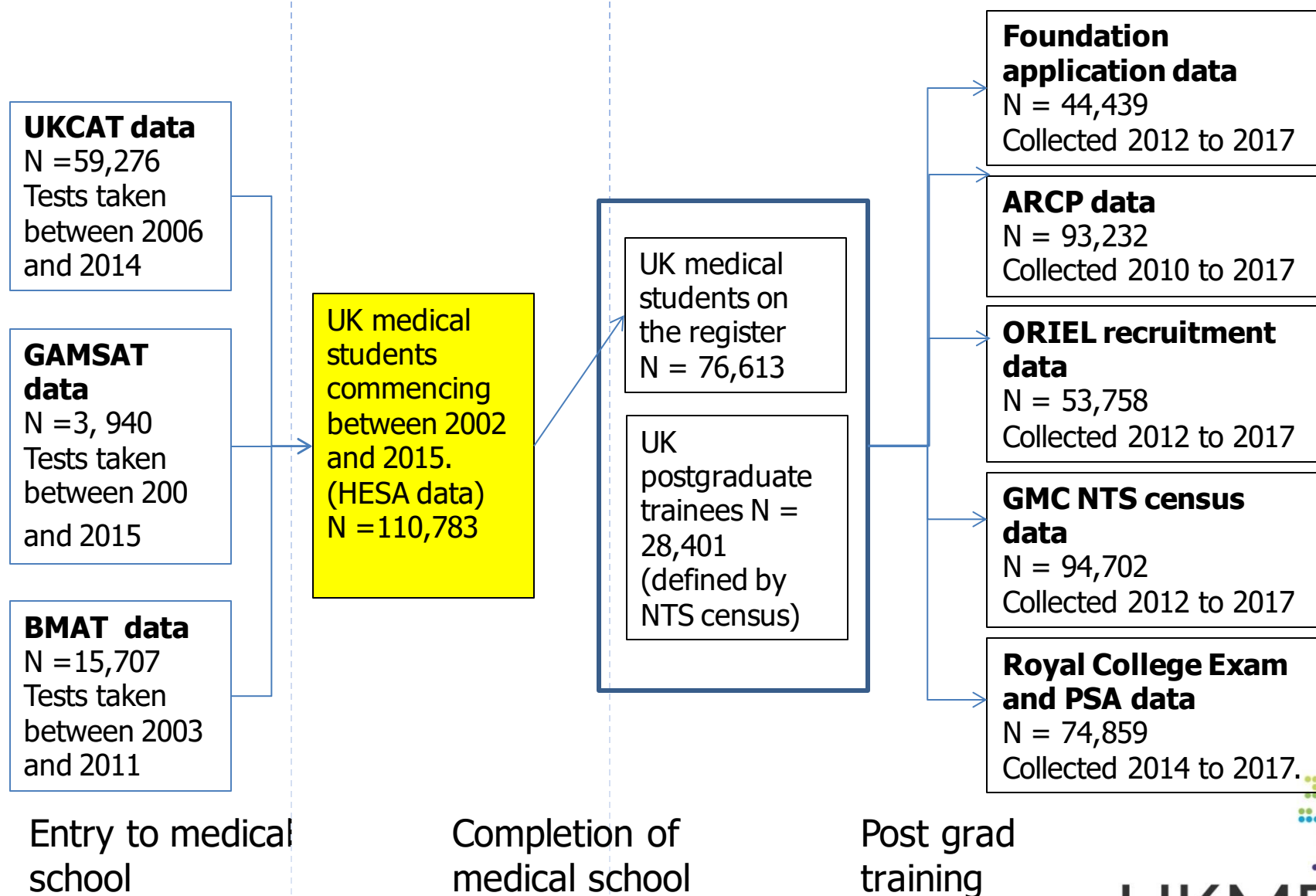
All students who started at a UK medical school from 2002 until 2016 as defined by data supplied to the GMC by the Higher Education Statistics Agency (HESA).

All applicants from 2007 onwards (UCAS)

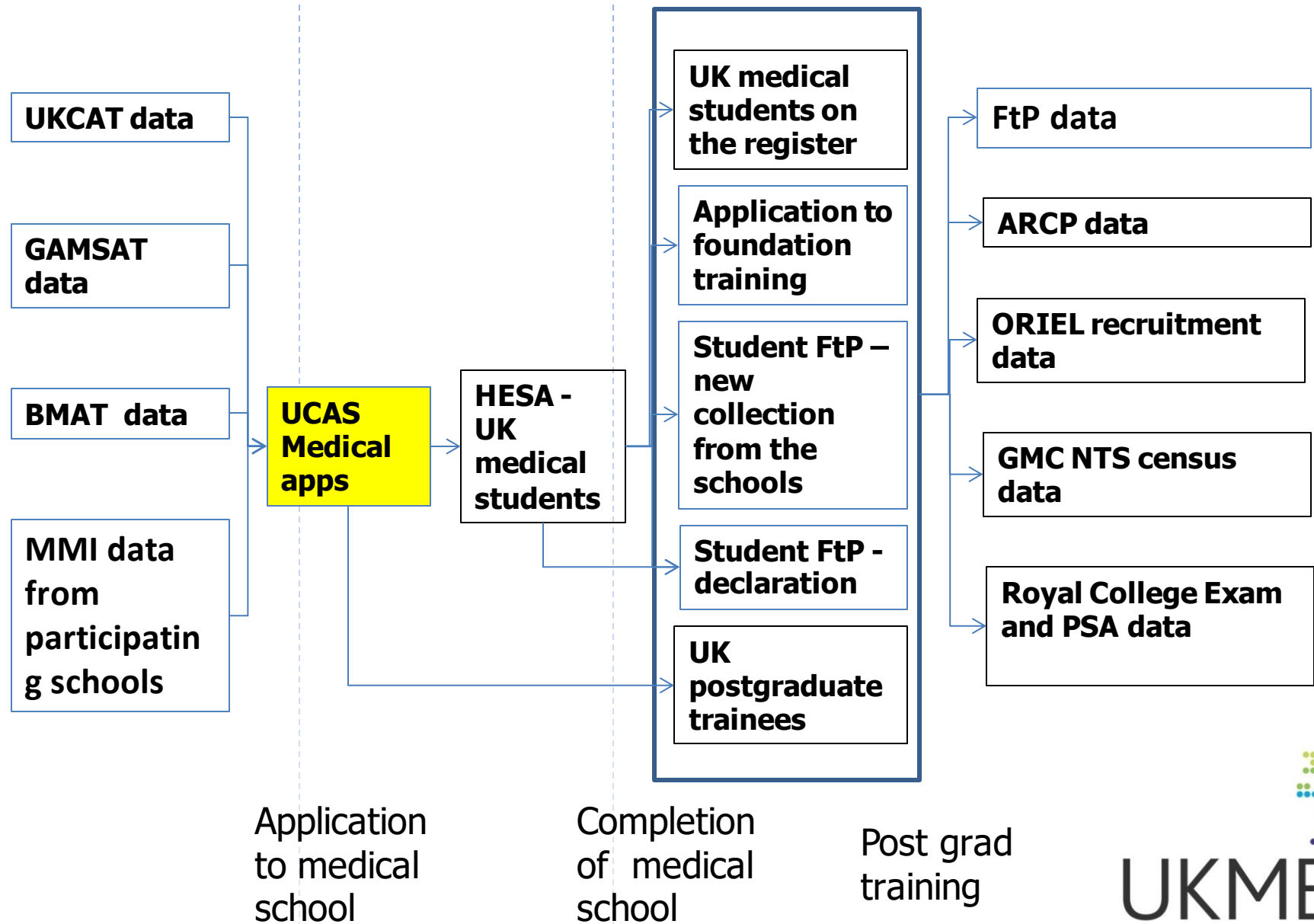
And

All doctors listed in the NTS_TRAINEE table at least once since 2012 (i.e. all those listed as belonging to a UK based training scheme in one or more the annual censuses conducted to administer the GMC's National Trainee Survey.)

Figure 1 - Summary of UKMED available data March 2018



Linkages when population based on applicants from December 2018



UKMED cycle – data loads

There are two data loads per training year:

Undergraduate: March - May

- HESA, test providers (UKCAT, BMAT and GAMSAT)
- **First time in 2018 UCAS. Arrangements for UCAS are more complicated (Hashed linking and the UCAS data only in the safe haven)**
- Multiple Mini Interview Data (MMI) for medical schools that provided it.

Postgraduate annual collections: December – January

- ARCP,
- ORIEL recruitment (FP, Specialty),
- SRA data
- Exam data

UKMED Governance and funding

- Advisory Board includes representatives from Data Contributors (including HESA and UCAS) and academics.
- Costs shared between MSC and GMC. Both organisations contribute staff. MSC pay for website and GMC for data.
- Use extracts from databases used to administer training.
- Piggy back/integrate with existing activities: UKMED uses existing GMC annual collections for PG data and employment.

Lesson learnt

- Produce and publish descriptive statistics and get aggregate statistics confirmed BEFORE moving to research extracts.
- Only way to check completeness of data. We have had cases missing as schools had not coded correctly (REGBODY = 01) and they were therefore not included in the HESA extract we receive
- Try to influence collection e.g. remove ambiguity from CTITLE field

Linkage issues encountered in UKMED

- Entity resolution and IDs: Confidence that ID = one unique person.
- We had to dedup HUSIDs as some students were incorrectly given new HUSIDs when they moved school (e.g. Durham to Newcastle.) So 1 person = 2 HUSIDs
- Beware of deduping people with the same name, DOB and postcode - twins

Checking Links

- 1:1 across the link e.g. a UCASPERID only matched to one test provider ID
- Check for impossible events e.g. same person taking a test twice on same date.
- Plausible chronology: e.g. starting medical school 4 or more years before PMQ awarded date.
- Some data anomalies are real so more clauses need when checking by script: e.g. medical school <> PMQ awarding body because student left school in bad standing and obtained PMQ outside UK so link is correct (need to consider HESA RSNEND)

Checking Links

- ID links confirmed by checking for matching names and DoB on both sides of link.
- Requires detailed contextual knowledge/documentation. E.g. looking at plausible chronology entails considering transfer cases and students joining a course after the first year.
- Good reference data are required. We use MSC's *Entry requirements for UK medical schools*
- The wider literature has a clerical review stage. We use this for checking links that may be incorrect – e.g. where fuzzy matching on strings used (Levenshtein distance).

Deriving variables

We went through a mapping exercise to map the HESA values of COURSEID and CTITLE to the course types described in the MSC's document *Entry requirements for UK medical schools*:

- Standard Entry Medicine
- Graduate Entry Medicine
- Medicine with a Preliminary Year
- Medicine with a Gateway Year

By doing this derivation at source in the database we ensure consistency across research projects and some attempt at efficiency.

The importance of documentation

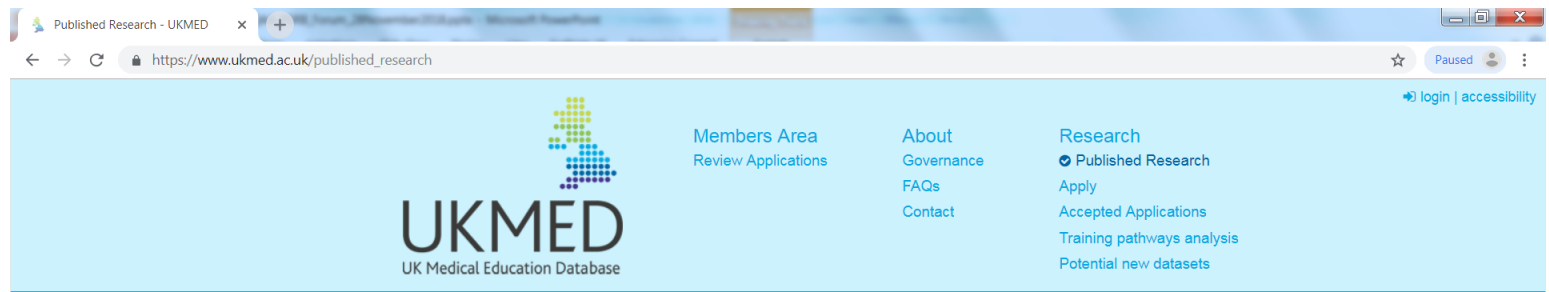
We maintain:

UKMED Data Dictionary which refers back to HESA documentation

We maintain query log and we are grateful that HESA has found the time to respond to all our queries.

UKMED Outputs

Publications emerging – all on the website



Published Research

Title	Publisher	Date
The relationship between school type and academic performance at medical school: a national, multi-cohort study Ben Kumwenda, Jennifer A Cleland, Kim Walker, Amanda J Lee, Rachel Greaux	BMJ Open	September 2017
Factors associated with junior doctors' decisions to apply for general practice training programmes in the UK: secondary analysis of data from the UKMED project Thomas C. E. Gale, Paul J. Lambe, Martin J. Roberts	BMC Medicine	December 2017
The UK medical education database (UKMED) what is it? Why and how might you use it? Jon Dowell, Jennifer Cleland, Siobhan Fitzpatrick, Chris McManus, Sandra Nicholson, Thomas Oppé, Katie Petty-Saphon, Olga Sierocinska King, Daniel Smith, Steve Thornton, Kirsty White	BMC Medical Education	January 2018
Predictors of fitness to practise declarations in UK medical undergraduates Lewis Paton, Paul Tiffin, Daniel Smith, Jon Dowell, Lazaro Mwandigha	BMC Medical Education	April 2018
Relationship between sociodemographic factors and selection into UK postgraduate medical training programmes: a national cohort study Ben Kumwenda, Jennifer A Cleland, Gordon J Prescott, Kim Walker, Peter W Johnston	BMJ Open	June 2018
Evaluating the validity of the selection measures used for the UK's foundation medical training programme: a national cohort study Daniel Smith, Paul Tiffin	BMJ Open	July 2018



UKMED Outputs

Forthcoming work includes:

A comparison of the properties of BMAT, GAMSAT and UKCAT.

Declared disability in the UKMED dataset 2002-2016: an exploratory descriptive analysis

Access to HE qualifications and widening participation in medicine?

HESA other uses

- We have stopped collecting aggregated data from medical school as part of the Medical School Annual Report. Instead we will report on entrants and attrition using HESA data
- We will produce more granular report on progression. Course type not just Medical School. E.g. instead of Southampton we have
 - Southampton - Graduate Entry Programme
 - Southampton - Medicine with a Gateway Year
 - Southampton - Standard Entry Medicine

HESA Future uses

When the GMC introduces the Medical Licensing Assessment (MLA) in 2023. HESA data can be used for reporting, so that only candidate lists need to be provided to the GMC.

More information

<https://www.ukmed.ac.uk/>

Any questions

Contacts form here:

<https://www.ukmed.ac.uk/contacts/new>

HESA



**Your own data capability journey: tools,
services, and future developments**



Introduction to HESA Data Dissemination Services

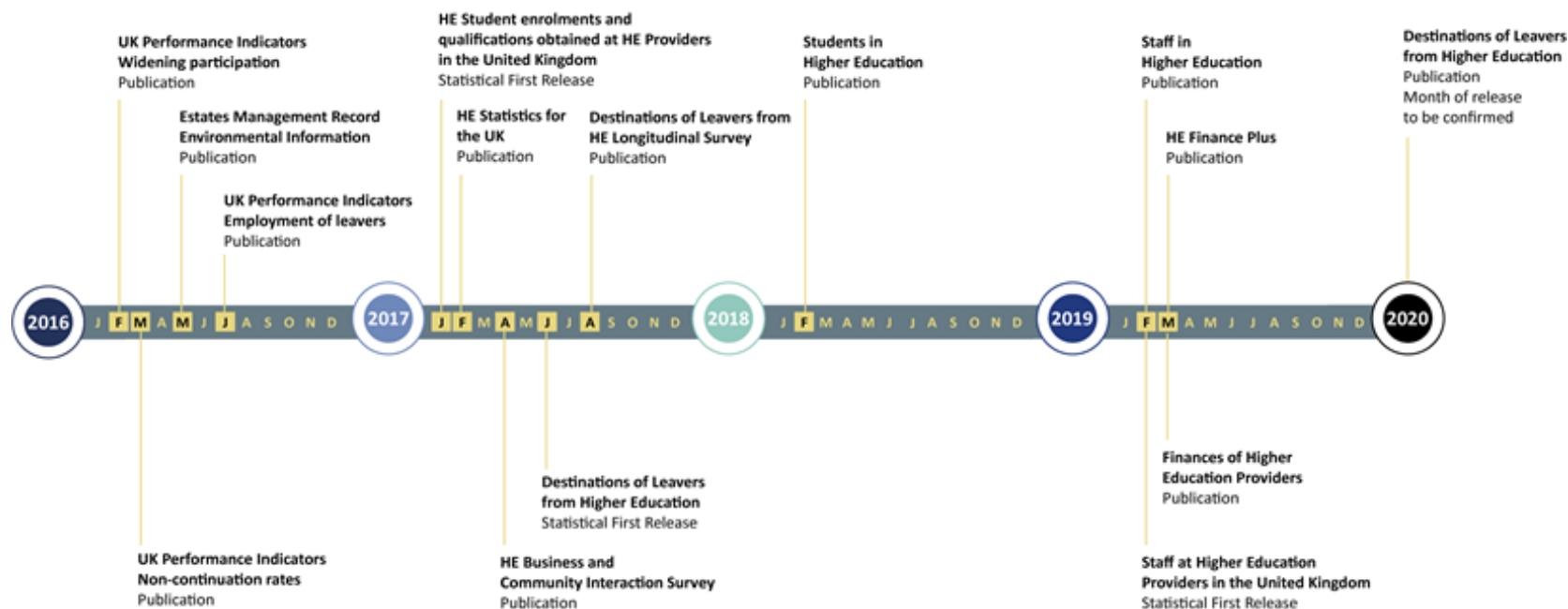
25 years of HE Data intelligence

Data Products and Services:

- Open Data, National and Official Statistics
- Heidi Plus (Data analytics and visualisation platform)
- Tailored Datasets
- Data Consultancy
- Data Collection-as-a-Service



Open Data



National and Official Statistics



Statistical Bulletins (National Statistics)

- First release of Student/Qualifiers Data
- First release of Leavers (Graduates)

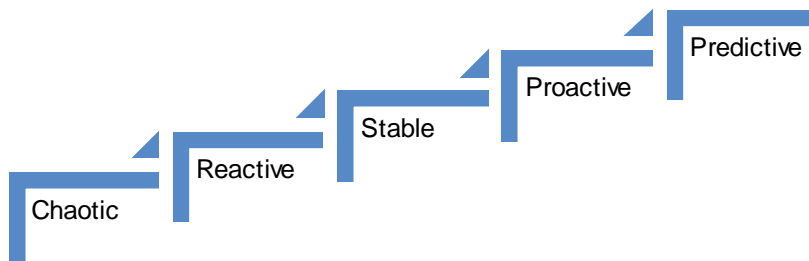
Performance Indicators (Official Statistics)

- Widening participation
- Non-Continuation rates
- Employability

All other published HESA Data is classified as Official Statistics.

Data Capability

- Toolkits available for large and small organisations
- Open license
- Assesses data capability using a maturity model



<https://www.hesa.ac.uk/innovation/capability>



Services: Tailored Datasets

We process approx. 2000 data queries each year for a broad range of organisations

- HE Sector
- Government departments, agencies and public bodies
- Professional, Statutory and Regulatory bodies (PSRBs)
- Academic research and students
- Charities and Not for profits
- Other organisations

For each dataset a specification is agreed with a HESA analyst who has Data intelligence and Data protection expertise.

Data services including:

- Data Extraction, Quality Assurance, Data Cleansing, Data Linking (Student journey analysis or third party datasets)
- Timeseries analysis.
- Multiple delivery options (Excel spreadsheets/PivotTables, CSV, Tableau extract, hosted analytics)

Services: Data Consultancy

Not just a numbers factory, we help you discover the narrative behind the numbers.

- Statistical Reports
- Data interpretation and summarisation
- Data visualisation and infographics
- Data interrogation
- Deep dive analytics
- Data Discovery Labs
- Onsite presentations
- Interactive Q&A

Future Developments

Data Futures

- HESA's transformation programme that will deliver the vision for a modernised and more efficient approach to collecting data, to deliver better output for a wider range of data users.
- Timely - from retrospective to in-year data collections
- <https://www.hesa.ac.uk/innovation/data-futures>

Graduate Outcomes

- Designed a new model to capture rich, robust and innovative data about graduates, using a future-proof and efficient methodology.
- Largest annual social survey in the UK
- Richer insight from more meaningful but subjective questions, 15 months after graduation.
- <https://www.hesa.ac.uk/innovation/outcomes>

Data Collection-as-a-Service (DCaaS)

- Designated Data Body for the HE sector (August 2019)
- Reduction of burden to the sector
- Consistency/ quality assurance
- End to end solution - Data collection to analysis

Contact HESA

Head of Policy and Development

dan.cook@hesa.ac.uk

Head of Data Intelligence

denise.jones@hesa.ac.uk

Head of Sales and Marketing

victoria.atherstone@hesa.ac.uk

Tel.01242 211 495 @ukhesa

