

EXPLAINING DATA LINKAGE QUALITY TO THE PUBLIC

GEORGINA FERGUSON-GLOVER
PATIENT AND PUBLIC INVOLVEMENT MEMBER



DARE UK



Why are we looking at data linkage?

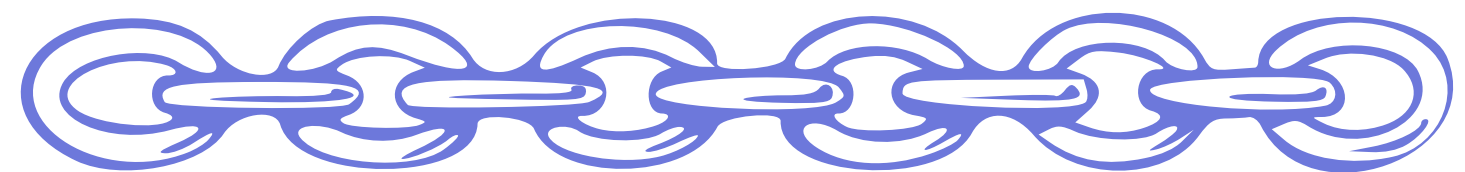


**National Health Service in
the UK - free healthcare**

NHS Single Patient Record

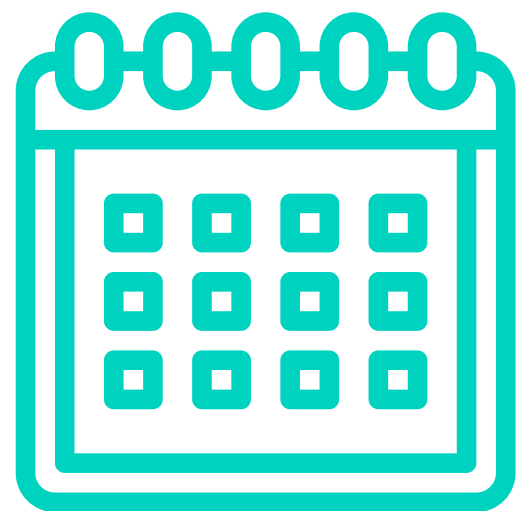
Linked healthcare data in one place

**But what do the public think and know
about the linkage?**





Workshops with members of the public to educate them on data linkage and gain their views



14 participants, 18-65+, ethnically diverse from four different backgrounds



3 workshops over 3 months

See if their thoughts and views on data linkage change



Pre-meeting glossaries defining key terms

- **Accredited Researcher**
- **Administrative Data**
- **Data Linkage**
- **De-identified Data**

- **Digital Economy Act (DEA) 2017**
- **National Data Opt Out (NDOO)**
- **Research Accreditation Panel (RAP)**
- **The Five Safes**
- **Trusted Research Environment (TRE)**

We separated data linkage for operational use vs research use

**Single
Patient
Record**



Due to UK laws DEA-Accredited linkage projects do not need consent



It's a Different Legal Route

The Digital Economy Act 2017 allows public bodies to share **de-identified data** for important research that is in the public good.



Your Privacy Remains Protected

Personal identifiers are removed before data is accessed, and researchers can only use the data **in secure, accredited** environments.



Better Research, Better Outcomes

Linking data helps us understand society better and improve public services, health, and outcomes **for everyone.**

Health data does not fall under the DEA and so in UK people can opt out of their health data being used for research.

Starting attitudes: If I was told my data had been linked I would feel:

Worried

Confused

Cautious

Unsure

Curious

Excited

Intrigued

Most wanted more information:

By who

Why

How



Mixed views

Task 1: What ethical issues can you see with linking data?

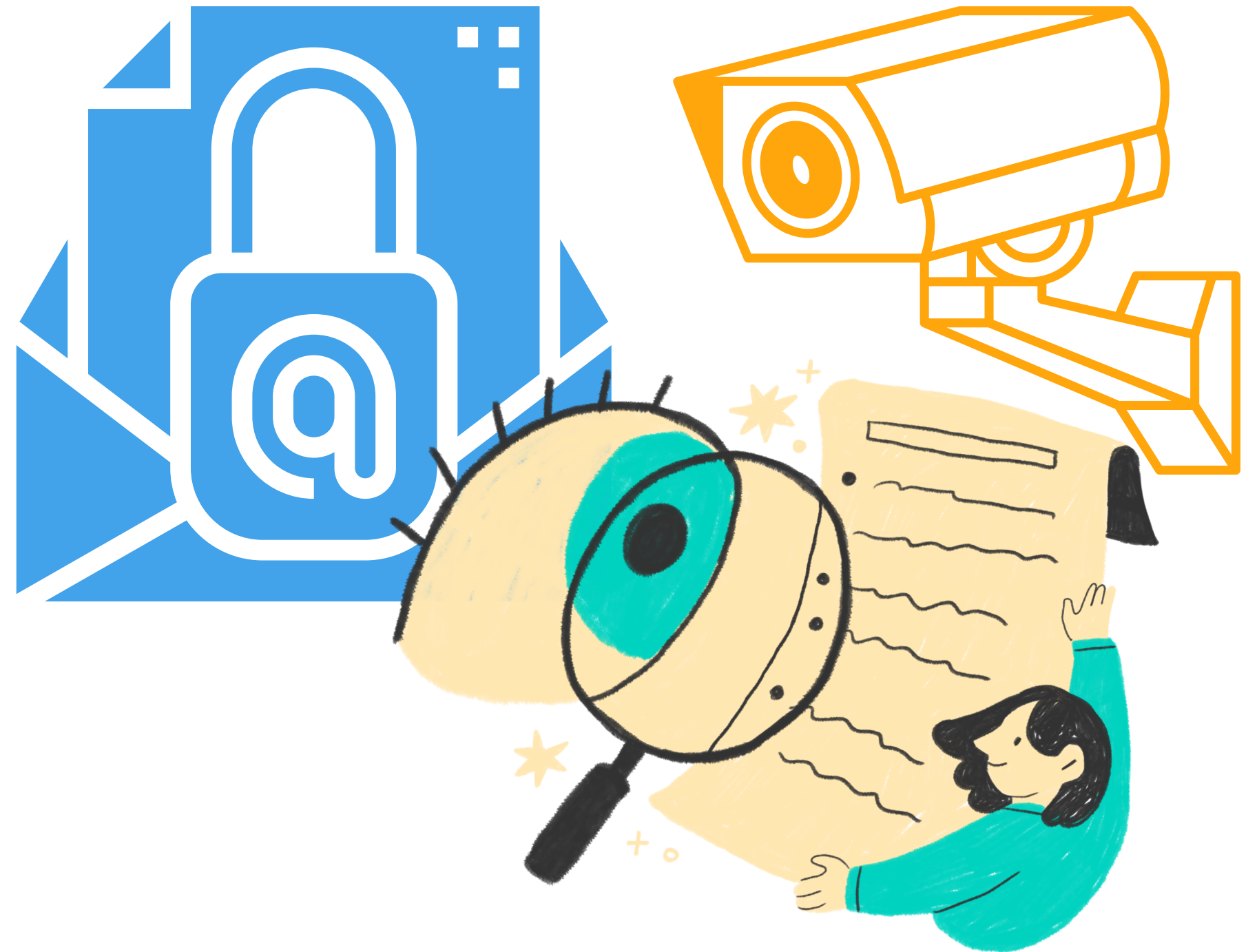
1. Confidentiality

2. Consent

3. Privacy

4. Usage

5. Who owns the data



Workshop 2:

5 Safes

De-Identified & Security

Opt out

Trusted Research Environments

Consent

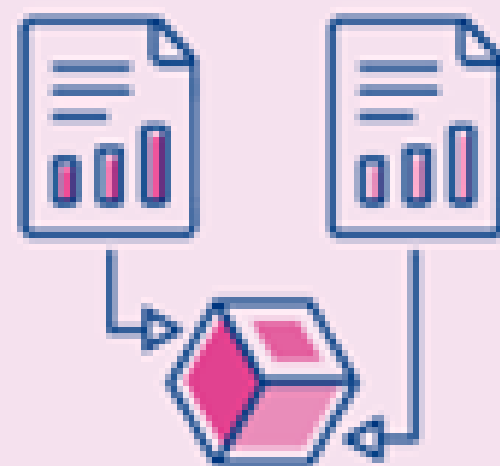
Five Safes: The framework data services operate under – they provide safe access for data research.



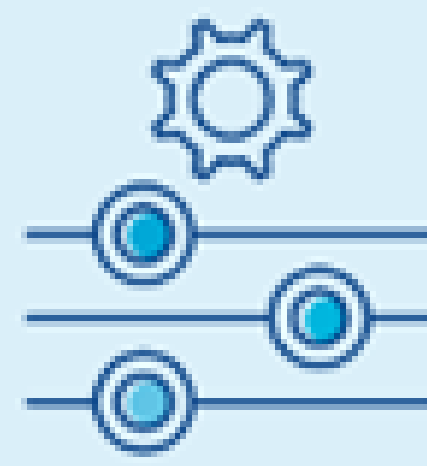
**Safe
People**



**Safe
Projects**



**Safe
Data**



**Safe
Settings**



**Safe
Outputs**

UK BioBank leak came just before workshop 2



**De-identified medical data of around
500,000 people was listed for sale on
the Chinese e-commerce site Alibaba**

**Incorporated it into the workshop: breaches
in security can happen but reassured that
all the data in this case was de-identified**

Task 2: Scenarios

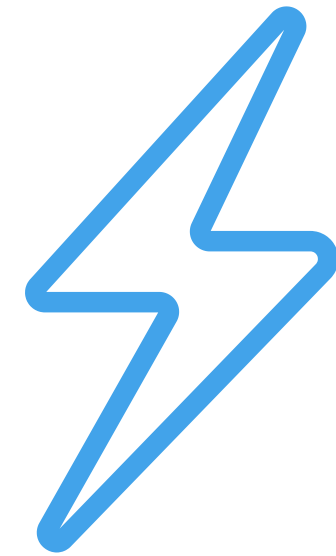
**1: Health-
education**



**2: Financial-
Social care**



**3: Police-energy
consumption**



**What information do you need to decide
if a particular linkage of your data is
approved/ethically sound**

“Understanding children’s health and school outcomes”



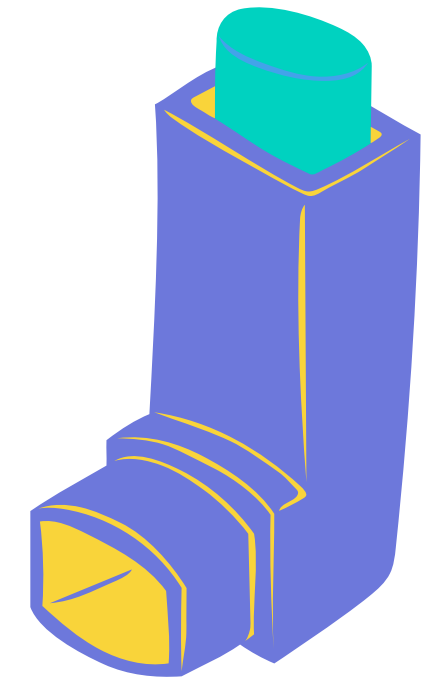
Children’s hospital records with school records.

Understand long-term health conditions, hospital admissions, or mental health difficulties = more absences, needing extra support, or having different educational outcomes.

Could help schools and health services identify where children need better support.

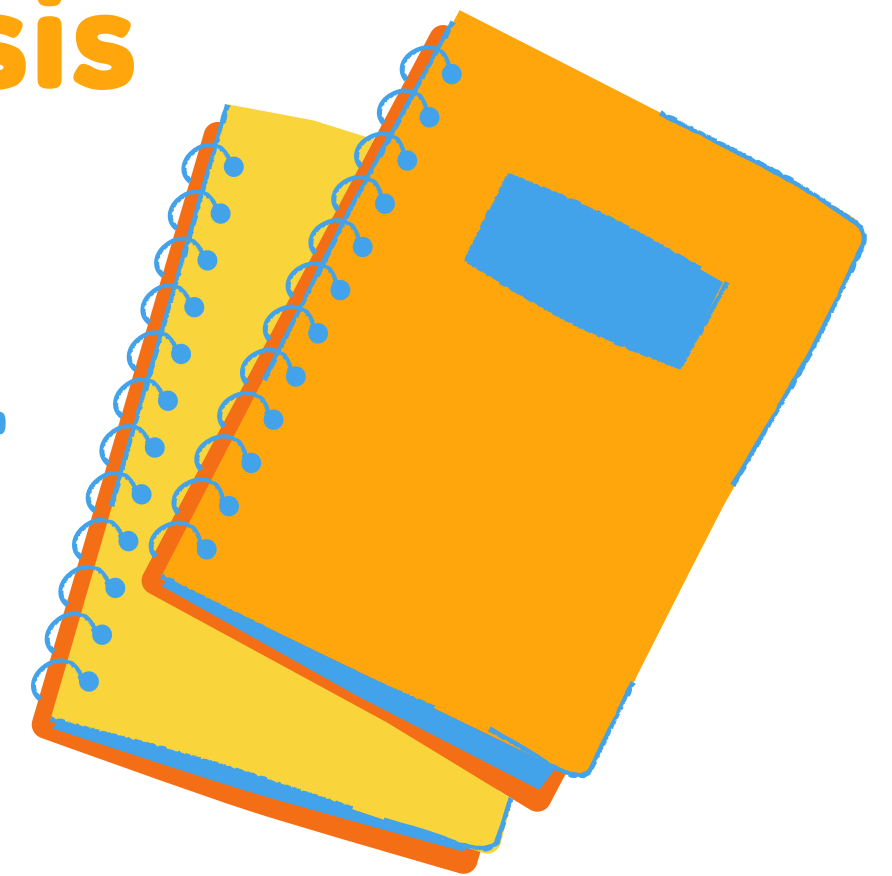


For example: children with asthma, epilepsy, anxiety, or long hospital stays are being disadvantaged in education.



The researchers say names and addresses would be removed after linkage is complete and before analysis

It would include: hospital visits, diagnoses, school attendance, special educational needs, exam results, age, sex, ethnicity, and area-level deprivation.



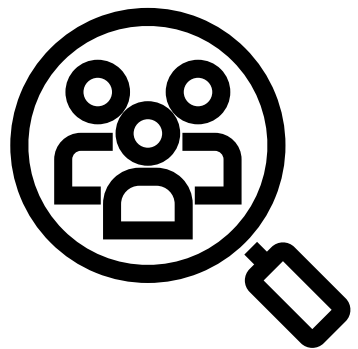
Task 2: Scenarios

**1: Health-
education**

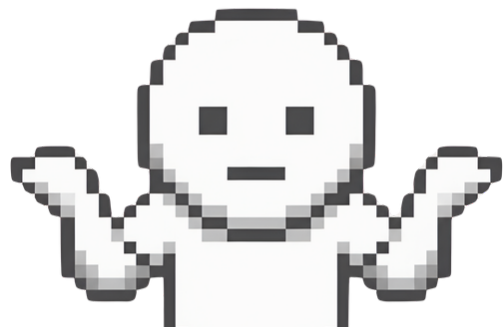
**2: Financial-
Social care**

**3: Police-energy
consumption**

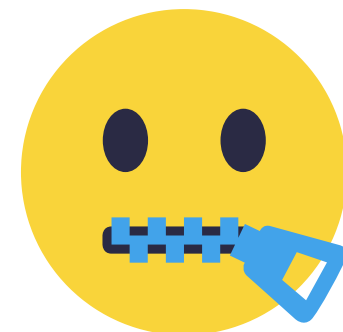
What information do you need to decide if a particular linkage of your data is approved/ethically sound: Subtle differences depending on the linkage



To who



Why



Confidentiality

+ vs -

**Benefits
vs risks**

Workshop 3:

Data linkage is like matching pizzas

We do not rely on one clue. We compare several features, then combine the evidence.



Two records can look similar in some ways and different in others.

The question is not “is every part identical?”

It is “is there enough evidence that they refer to the same thing?”



Matching on pizza shape

Broad pattern: does the record look like the same kind of record?



Matching on crust

Context around the edge: address, place, organisation or setting.



Matching on toppings

Distinct details: names, dates, IDs and other memorable features.



Matching on sauce base

Supporting fields that make the match more or less plausible.

Linkage decision

Each feature adds weight. Agreement strengthens confidence; conflict increases uncertainty.

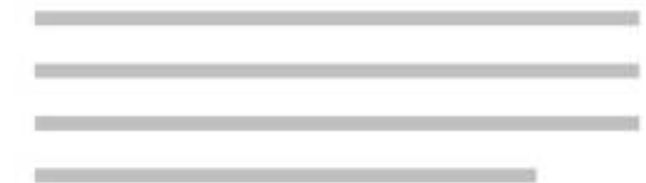
likely same

uncertain

likely different

Takeaway: data linkage is evidence-weighting — not a single exact-match test.

MISSING



Contact : xxx-xxxx

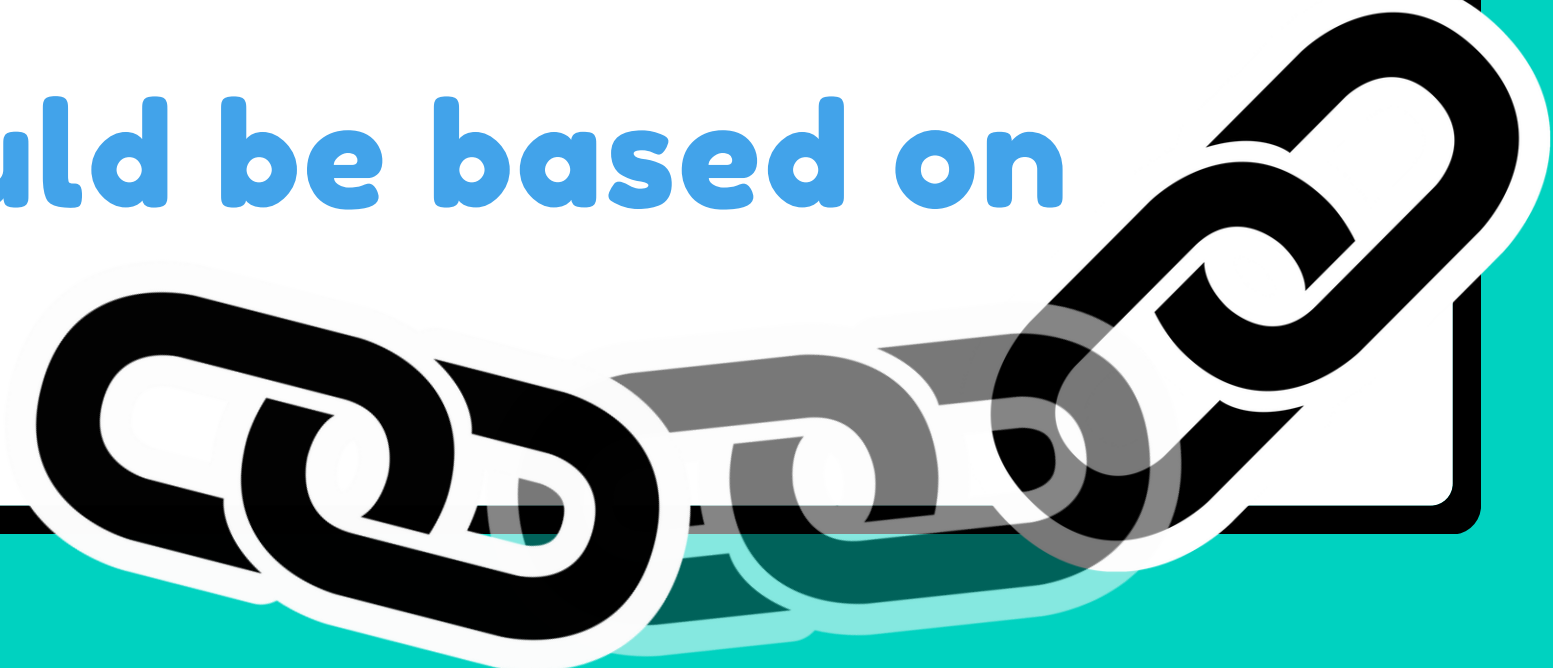
Workshop 3: Differential Bias

How would missed links and false links impact the scenarios from Workshop 2?

Missed links: People with greatest need / most complex circumstances are more likely to be missed

False links: lead to inaccurate conclusions

Policies or interventions could be based on incorrect evidence



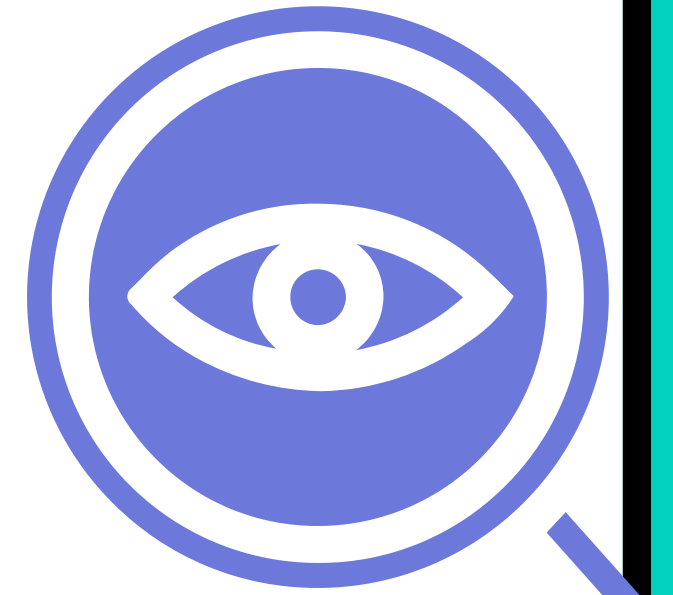
Workshop 3: Differential Bias

Transparency and quality assurance:

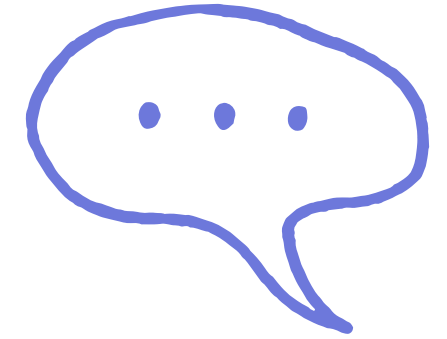
They wanted to be informed on what could go wrong and felt linkage should not be presented as a perfect process

They also wanted to know how linkage errors are identified, measured and corrected

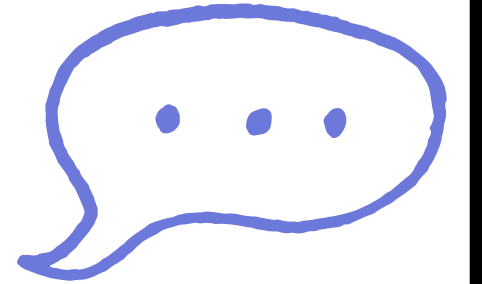
Concern that small errors can accumulate and lead to larger problems



Workshop 3: Differential Bias



Communicating uncertainty:



The public should be informed that data linkage involves some uncertainty and that it is less precise than people assume

Again the balance of the benefits and risks came up and they wanted these communicated better to the public



Recognise that no system is perfect

Final attitudes:

If I was told my data had been linked now I would feel:

Curious

Interested

By who

Excited

Worried

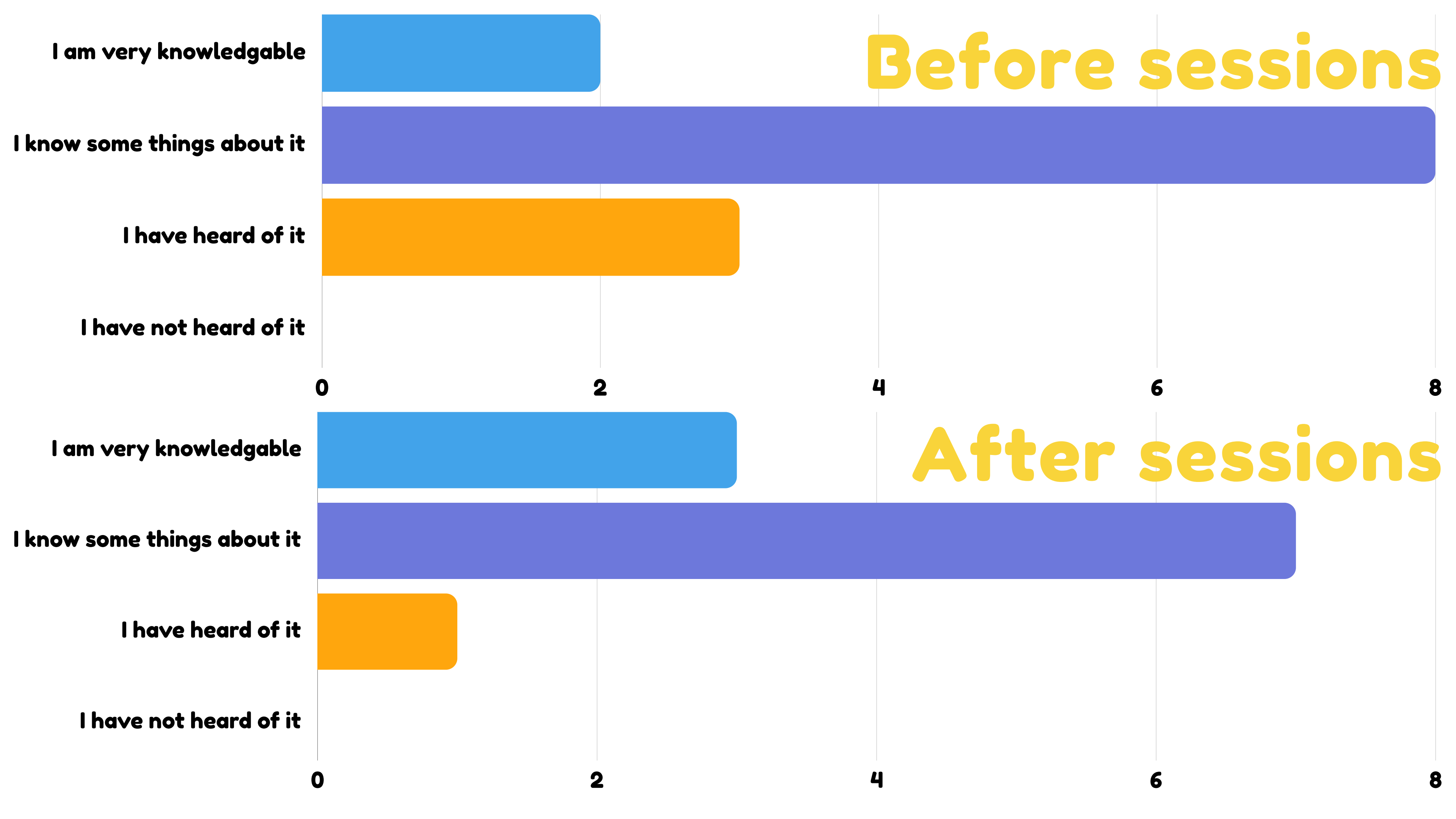
Cautious

Why

**What was
linked**

Suspicious

**More
inquisitive
than
before
and less
worried**



How can we better communicate data linkage to the public?

We need to start education aimed at young people in schools

Educate adults in different places - doctor's offices, hospitals, through media and social media

Need to share the benefits

Target people who have opted out



Final thoughts:

**What words come to mind
when we say “linked data”:**

Accuracy

Error prone

Needs governance

Relevant

Connecting the dots

Important

Less concern

Danger of misuse

Not enough public involvement

Needed

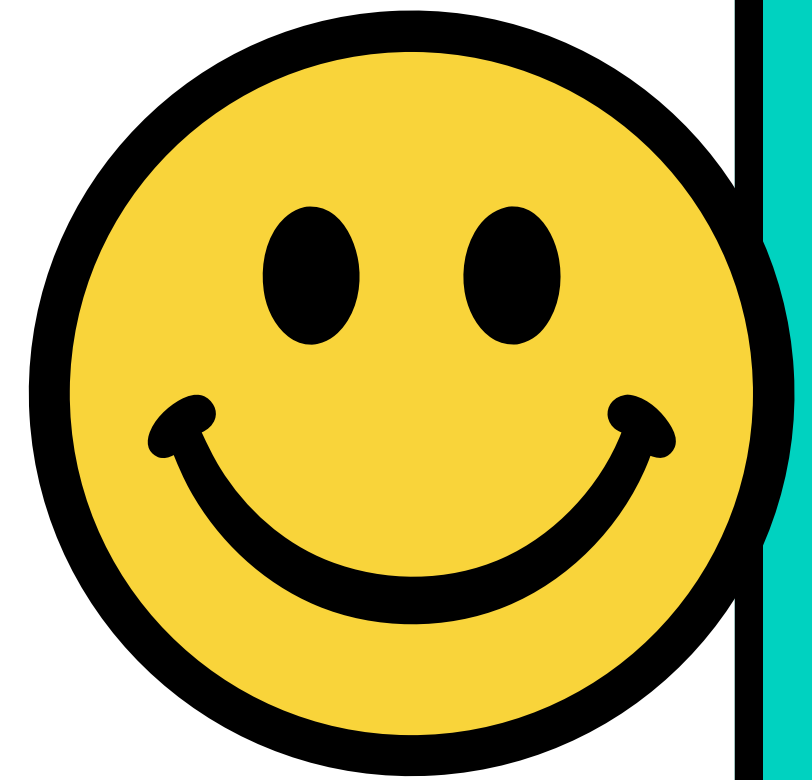
Takeaways from participants:

Enjoyed discussions, hearing from others, scenarios and learning more

Very interesting

Made it easy to understand a complex topic

Don't believe everything in the news

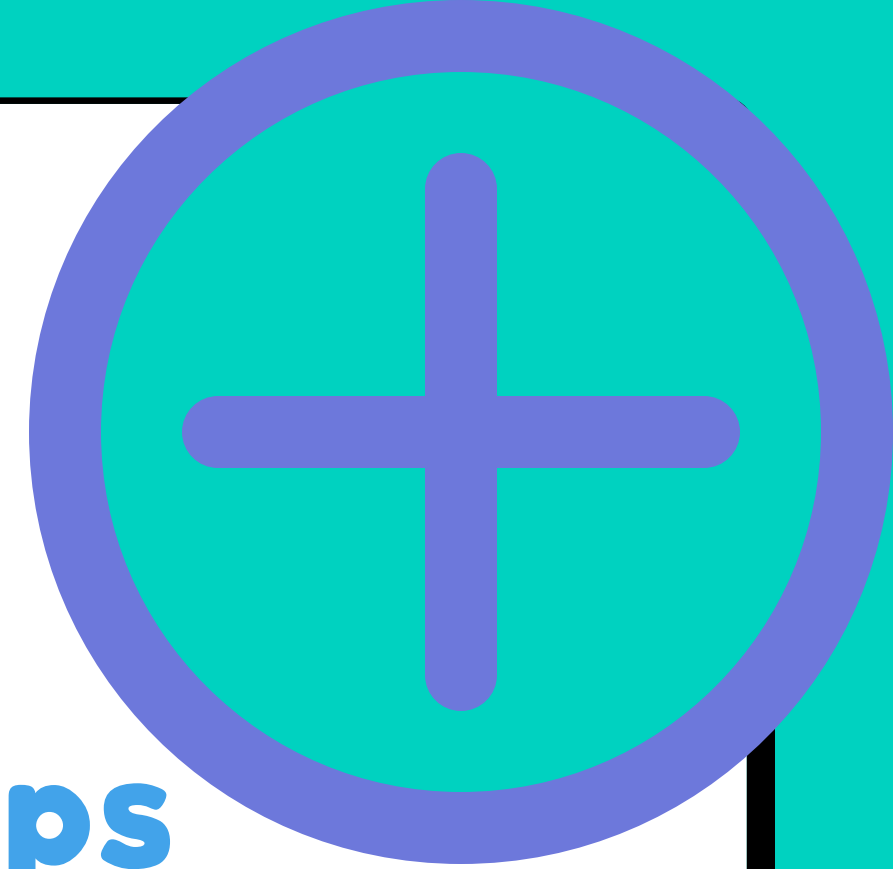


8/10 felt it was pitched at the right level

Wanted to know what our next steps are

Very keen for more public involvement in data linkage

Some wished there had been more sessions and that we'd spoken about AI's impact



Takeaways:

The public are not objects to be converted - they need to be informed and come to their own conclusions

The public are very interested in data linkage and want to learn more but needed more information on specifics of it

The main things the public want are accuracy, confidentiality and trustworthiness

