NRS Project ID: 2024-A002
Attendance 2007-19 and 2020-2022 - data ingest

Input: Attendance 2007-19 and 2020-22

Records for previously recorded pupils 2007-19	86,931,406
Records for previously recorded pupils 2020-22	12,753,694
Records for new pupils 2020-22	4,065,306
Total records 2007-22	**103,750,406**
Previously recorded pupils 2007-19	1,401,239
Previously recorded pupils 2020-22	590,276
New pupils 2020-22	254,646
Total pupils 2007-22	**1,655,885**
Demographic keys for previously recorded pupils 2007-19	86,888,306
Demographic keys for previously recorded pupils 2020-22	12,753,694
Demographic keys for new pupils 2020-22	4,065,306
Total demographic keys 2007-22	**103,707,306**
Storage keys for previously recorded pupils 2007-19	86,888,306
Storage keys for previously recorded pupils 2020-22	12,753,694
Storage keys for new pupils 2020-22	4,065,306
Total storage keys 2007-22	**103,707,306**

Output: Attendance 2007-19*   %Match Rate

Records matched to the spine in Attendance dataset	82,758,780	95.2%
Individuals matched to the spine in Attendance dataset	1,311,913	93.6%

Output: Attendance 2020-22*

Records matched to the spine in Attendance dataset	3,652,140	89.8%
Individuals matched to the spine in Attendance dataset	223,074	87.6%

Output: Attendance 2007-22*

Records matched to the spine in Attendance dataset	98,575,289	95.0%
Individuals matched to the spine in Attendance dataset	1,534,987	92.7%

* These figures correspond to Unique Exact Matches (i.e., the stringent category of matches that is provided by default in research projects)

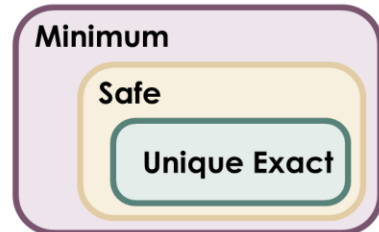## Degree of specificity for EAS datasets *(ADR-S Project Request)*

For Education datasets that are ADR-S ingested, **Exact / Safe / Minimum** are categories referring to the quality of the matches:

Unique Exact is the most stringent category, but with the highest precision. It only includes unique exact matches for sex, postcode, and date of birth (DEFAULT for projects). It excludes same sex twins.

Safe includes also high confidence matches: exact match for sex, exact match for date of birth, and partial match for postcode, this slightly increases the match rate, but also increases the risk of some incorrect matches

Minimum* matches are those which meet the minimal threshold: exact match for sex, exact match for postcode, and two out of three matches from the three fields of date of birth (dd/mm/yy) - this offers the highest match rate, but introduces many false positives.

Competing matches (e.g., same sex twins) are included only as minimum. NRS strongly advises against their inclusion because it introduces many duplicate matches, involving incorrect links, that in many cases will be impossible for the researcher to resolve.

If a match is **Exact**, it'll be included as **Safe** and **Minimum** too; if a match is **Safe**, it will also be considered **Minimum**

* *Minimum* matches have been called *Optimal* in the past. The name was changed as the word *Optimal* was misleading to represent the lowest quality of matches

The following figures correspond to a breakdown of the Unique Exact Matches (i.e., the stringent category of matches that is provided by default in research projects).

Attendance Absence A002 Summary tables

Sex

		Total%
Male	114,345	51%
Female	108,729	49%
Total:	**223,074**	