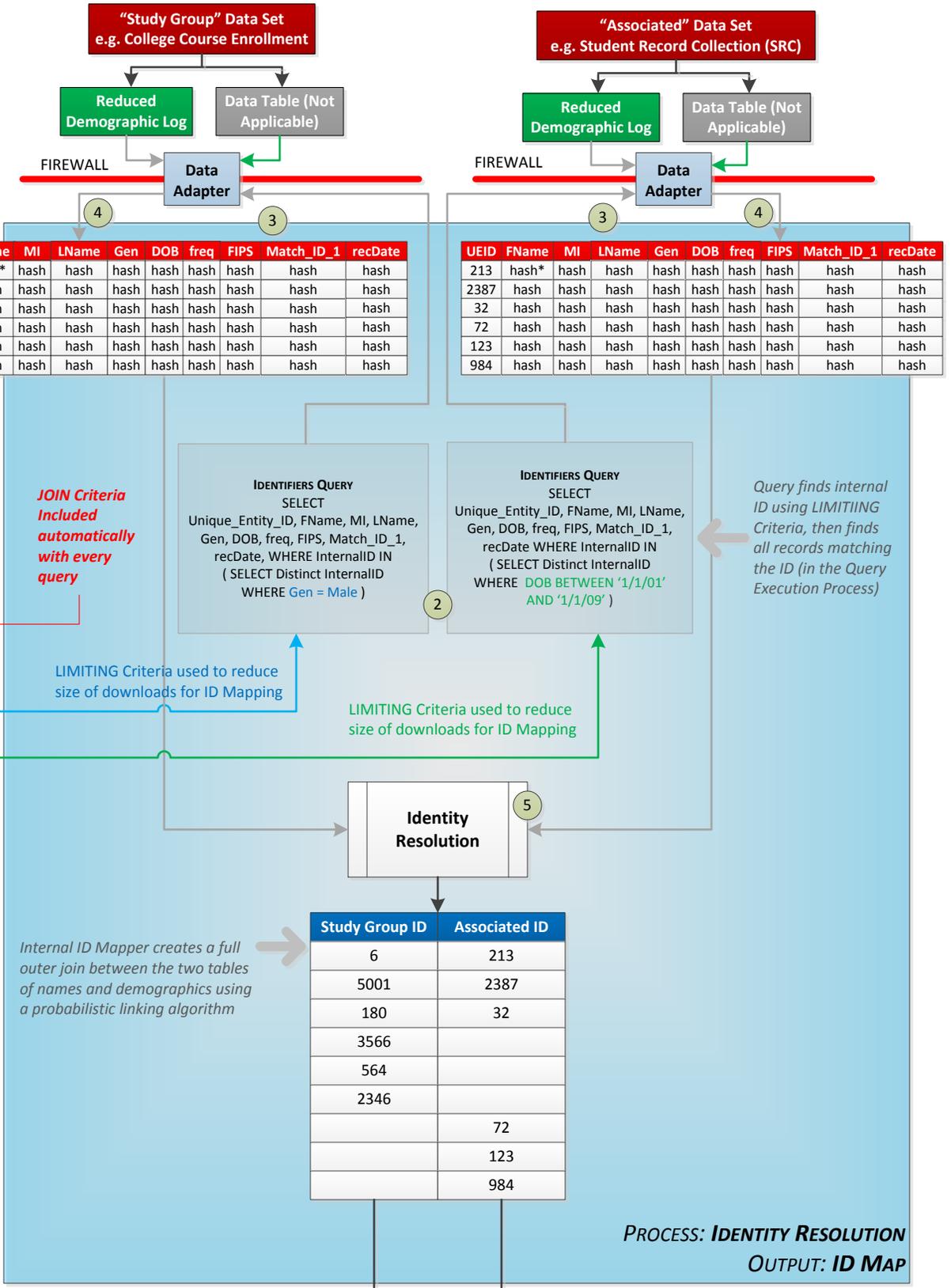


# IDENTITY RESOLUTION PROCESS OVERVIEW

**THE IDENTITY RESOLUTION PROCESS**  
 USES STANDARD DEMOGRAPHIC LINKING FIELDS (IN RED IN THE QUERY BELOW) AND ADDITIONAL SELECTION CRITERIA SPECIFIED BY THE USER (IN BLUE AND GREEN BELOW) TO:

- DETERMINE WHICH RECORDS TO PULL DOWN FROM EACH DATA SOURCE'S DEMOGRAPHIC LOG;
- DETERMINE WHICH IDENTITIES EXIST IN BOTH DATA SOURCES;
- CREATE AN ID MAP TO BE USED BY THE QUERY EXECUTION PROCESS FOR LINKING THE DATA SOURCES.



**JOIN Criteria**  
 Included automatically with every query

**LIMITING Criteria** used to reduce size of downloads for ID Mapping

**LIMITING Criteria** used to reduce size of downloads for ID Mapping

Query finds internal ID using LIMITING Criteria, then finds all records matching the ID (in the Query Execution Process)

1. Query Specification for Identity Resolution Process
2. Identifiers Query to find all matching records
3. Query submitted to Data Adapter web service
4. Chaffer and Hasher creates hashed characters of personally identifying information
5. All IDs that meet criteria sent to through Identity Resolution process.

QUERY EXECUTION PROCESS ON NEXT PAGE

\* Please see Demographic Log Hashing document  
 \*\* Please see ID Mapper and Log Reduction document

# QUERY EXECUTION PROCESS OVERVIEW

THE **QUERY EXECUTION PROCESS** USES THE ID MAP CREATED BY THE IDENTITY RESOLUTION PROCESS TO QUERY THE DATA SOURCE TABLES SPECIFIED IN THE USER QUERY. FOR MATCHED RECORDS, A NEW UNIQUE IDENTIFIER (UNIQUE ID) IS CREATED AND INSERTED IN THE MATCHING ROW(S) OF EACH RETURNED DATA SET

**"Study Group" Data Set**  
e.g. College Course Enrollment

**"Associated" Data Set**  
e.g. Student Record Collection (SRC)

Data Table  
Demographic Log (Not Applicable)

Demographic Log (Not Applicable)  
Data Table

FIREWALL  
Data Adapter

FIREWALL  
Data Adapter

Internal ID Mapper creates a full outer join between the two tables of names and demographics using a probabilistic linking algorithm in the Identity Resolution Process on the previous page

Study Group ID	Associated ID
6	213
5001	2387
180	32
3566	
564	
2346	
	72
	123
	984

Query specification is parsed into separate queries for each data source

All records that have a match in the other data set

```
SELECT * FROM StudyGroup.DataTable a WHERE InternalID = Study Group ID
```

```
SELECT * FROM Associated.DataTable b WHERE InternalID = Associated ID
```

Brth M	Gen	Eval 1	Eval 2	Unique ID	SG ID		Assoc ID	Unique ID	SOL Hist	SOL Math	Gen	Brth M
7	M	88	99	ABC123	6	MATCHED (IN BOTH)	213	ABC123	500	450	M	7
5	F	92	54	DEF456	5001		2387	DEF456	490	320	F	5
11	M	60	26	GHI789	180		32	GHI789	495	410	M	11
22	F	56	22		3566	UNMATCHED IN STUDY GROUP						
105	F	99	79		564							
362	M	86	89		2346							
						UNMATCHED IN ASSOCIATED	72		320	555	F	5
							123		560	350	F	1
							984		454	445	M	8

- 6. Query Specification parsed
- 7. Separated query selecting all records that have the equivalent IDs
- 8. Query returns matched and unmatched results into two virtual tables
- 9. Full outer join created with the matched and unmatched records in one table

**SG ID** – Study Group ID of the starting agency in the data request (not the Unique Entity ID)

**Assoc ID** – Associated ID of the second agency used in the data request (not the Unique Entity ID)

**Unique ID** – generated by the Shaker to indicate only the matches between agencies (not the Unique Entity ID)

Unique ID	SG ID	Assoc ID	Gen	Brth M	Eval 1	Eval 2	SOL Hist	SOL Math
ABC123	6	213	M	7	88	99	500	450
DEF456	5001	2387	F	5	92	54	490	320
GHI789	180	32	M	11	60	26	495	410
	3566		F	22	56	22		
	564		F	105	99	79		
	2346		M	362	86	89		
		72	F	5			320	555
		123	F	1			560	350
		984	M	8			454	445

PROCESS: **QUERY EXECUTION**  
OUTPUT: **DATA SETS**

# Revision History

Version #	Date	Revised By	Comments
1.0	5/2/12	Aaron Schroeder	Original Collection with this document last updated
2.0	5/4/12	Austin Mills	Large revision made including splitting overview into two pages
2.1	5/8/12	Austin Mills	Small revisions and dropping query specification off the 2 <sup>nd</sup> page
2.2	5/9/12	Austin Mills	Specified when Data Table and Demographic Log was "Not Applicable"
2.3	12/3/12	Austin Mills	Further defined IDs returned in the data set by the Shaker
2.4	3/21/13	Austin Mills	Removed Demographic Log Reduction information from Shaker
2.5	3/28/13	Austin Mills	Changed Demo. Log Query to Identifiers Query