



INDEPTH Network

Understanding the data flow from operational database to standard input tables in staging database

Tathagata Bhattacharjee

iSHARE2 Support Team

IDMP – INDEPTH Data Management Programme

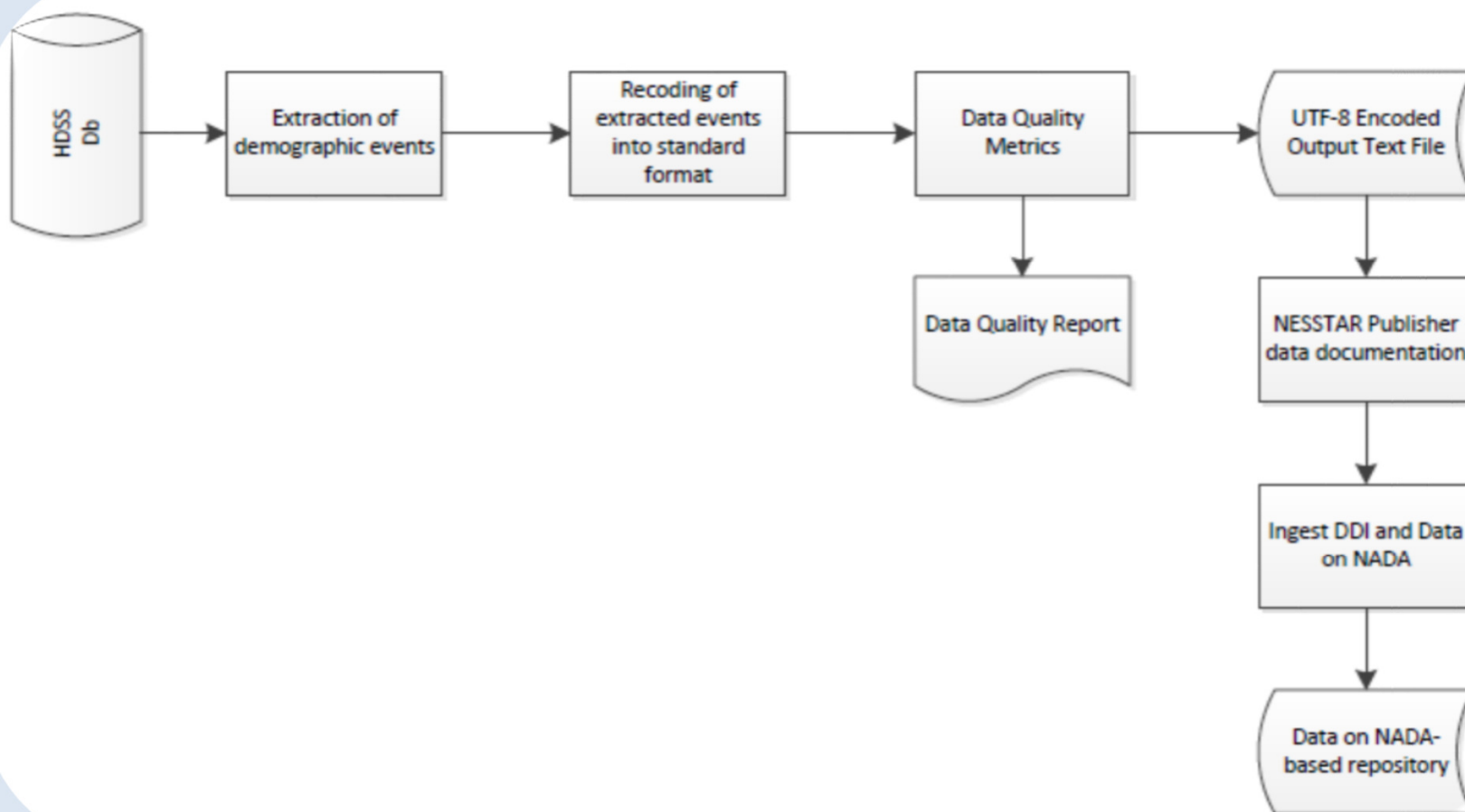
INDEPTH Data Repository

iSHARE 2 
an **INDEPTH** project

Outline

- iSHARE2 process flow overview
- HDSS data structure
- Data flow for generating the core micro dataset
- The Staging data structure
- Understand the transformation steps

Process flow as Identified for the iSHARE2 Project



Core Micro Dataset Variables

Variable Number	Variable Name	Variable Label
1	RecNr	Record Number
2	CountryId	Country Identifier
3	CentreId	Centre Identifier
4	IndividualId	Individual Identifier
5	Sex	Sex
6	DoB	Date of Birth
7	EventCount	Event Count
8	EventNr	Event Number
9	EventCode	Event Code
10	EventDate	Event Date
11	ObservationDate	Observation Date
12	LocationId	Location Identifier
13	MotherId	Mother Identifier
14	DeliveryId	Delivery Identifier

What we try to achieve?

	Individual_ID	Household_ID	Event	DoB	EventDate
A	G0010010010001	G00100100100	ENU	17 Nov 1947	21 Aug 2002
	G0010010010001	G00100100100	OBE	17 Nov 1947	31 Dec 2010
B	G0010010010002	G00100100100	ENU	1 Jul 1976	21 Aug 2002
	G0010010010002	G00100100100	OMG	1 Jul 1976	1 Jul 2007
C	G0010010010003	G00100100100	ENU	23 Aug 1985	21 Aug 2002
	G0010010010003	G00100100100	EXT	23 Aug 1985	1 Jul 2007
	G0010010010003	G00203000104	ENT	23 Aug 1985	2 Jul 2007
	G0010010010003	G00203000104	OMG	23 Aug 1985	10 Nov 2007
	G0010010010003	G00100100111	IMG	23 Aug 1985	30 Mar 2008
	G0010010010003	G00100100111	DTH	23 Aug 1985	15 Oct 2008
D	G0010010010004	G00100100100	ENU	1 Jul 1988	21 Aug 2002
	G0010010010004	G00100100100	EXT	1 Jul 1988	1 Jul 2007
	G0010010010004	G00203000104	ENT	1 Jul 1988	2 Jul 2007
	G0010010010004	G00203000104	OMG	1 Jul 1988	10 May 2008
E	G0010010010005	G00100100100	BTH	1 Jul 2005	1 Jul 2005
	G0010010010005	G00100100100	OBE	1 Jul 2005	31 Dec 2010
F	G0010010010006	G00100100100	IMG	1 Jul 1983	31 Aug 2007
	G0010010010006	G00100100100	OMG	1 Jul 1983	8 Apr 2008

ETL Flow



Initial Staging Tables for iSHARE2

IndividualEvents

- IndividualId
- EventCode
- EventDate
- EventNr
- ObservationDate
- LocationId

ChildDeliveryInfo

- ChildId
- MotherId
- DeliveryId

Deliveries

- MotherId
- DeliveryDate
- LiveBorn
- DeliveryId

Individuals

- IndividualId
- Sex
- DoB

Deaths (no records as of now)

- IndividualId
- Cause1
- Cause2
- Cause3
- Likelihood1
- Likelihood2
- Likelihood3

HDSS Data Structures

- Primarily belongs to two families

- **Family A**

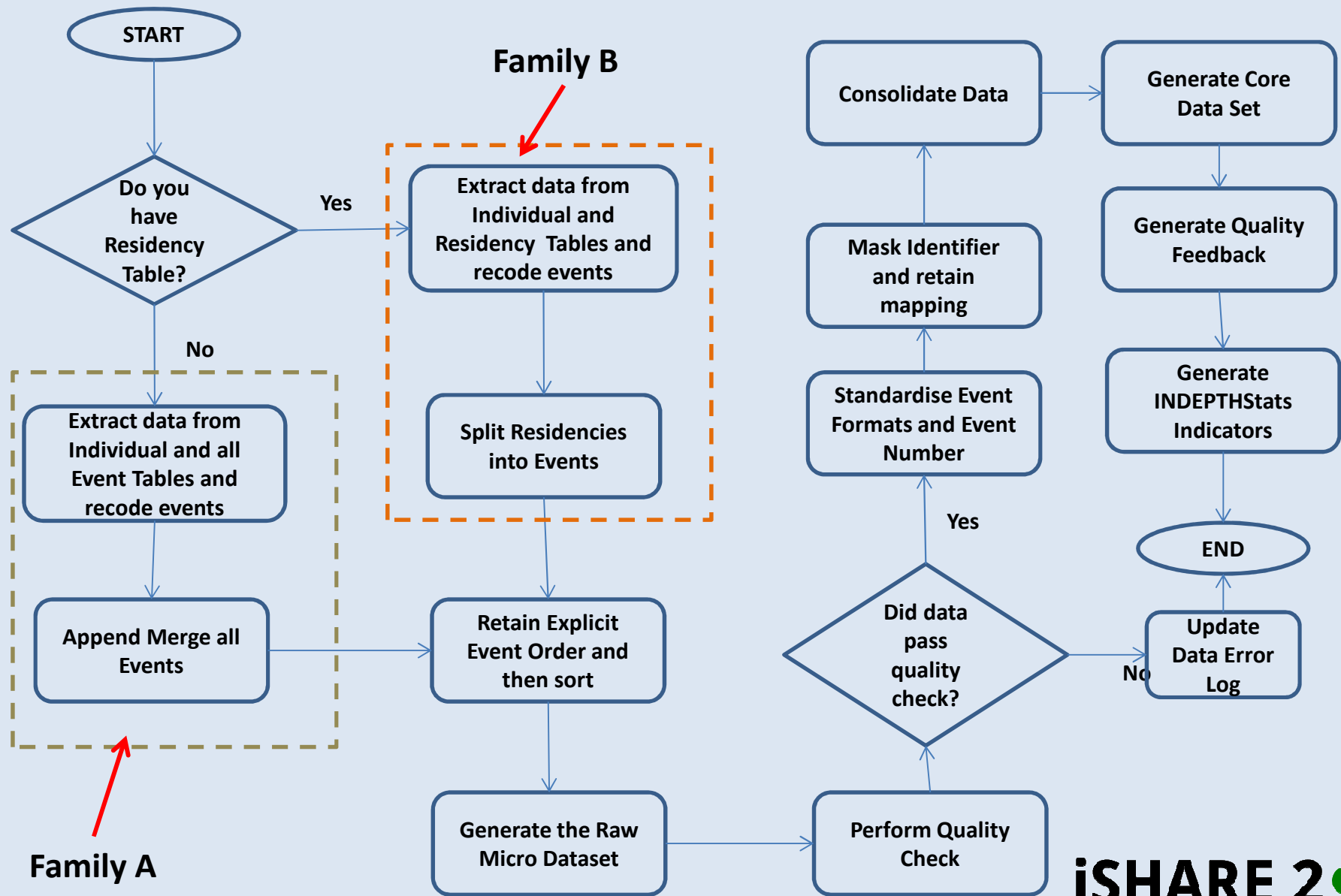
- Follows the HRS I
- Events are stored in different tables
- No in-built episodes stored

- **Family B**

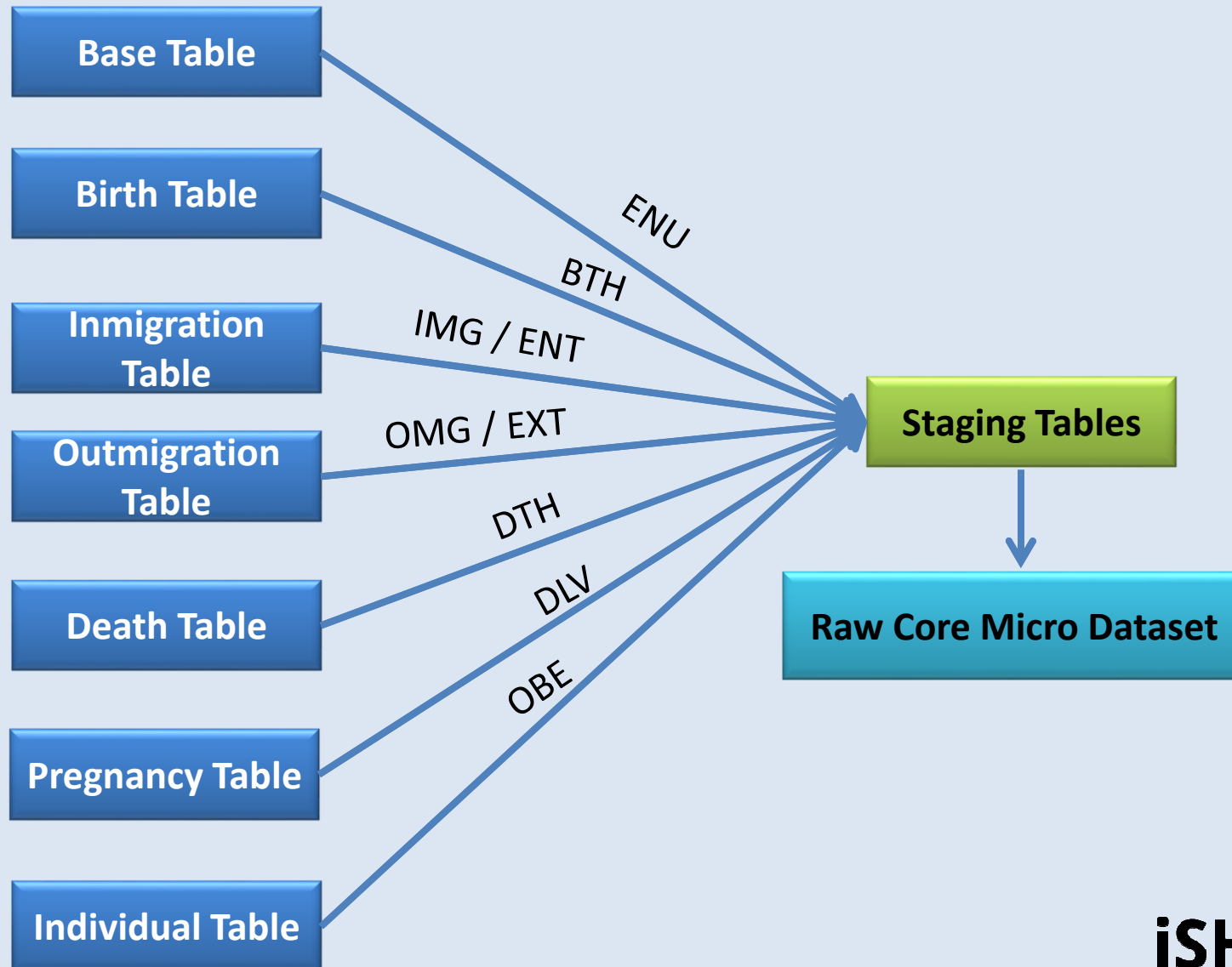
- Follows the HRS II
- Events are stored in Residencies
- Episodes are in-built in the residencies

- **Family C, ...**

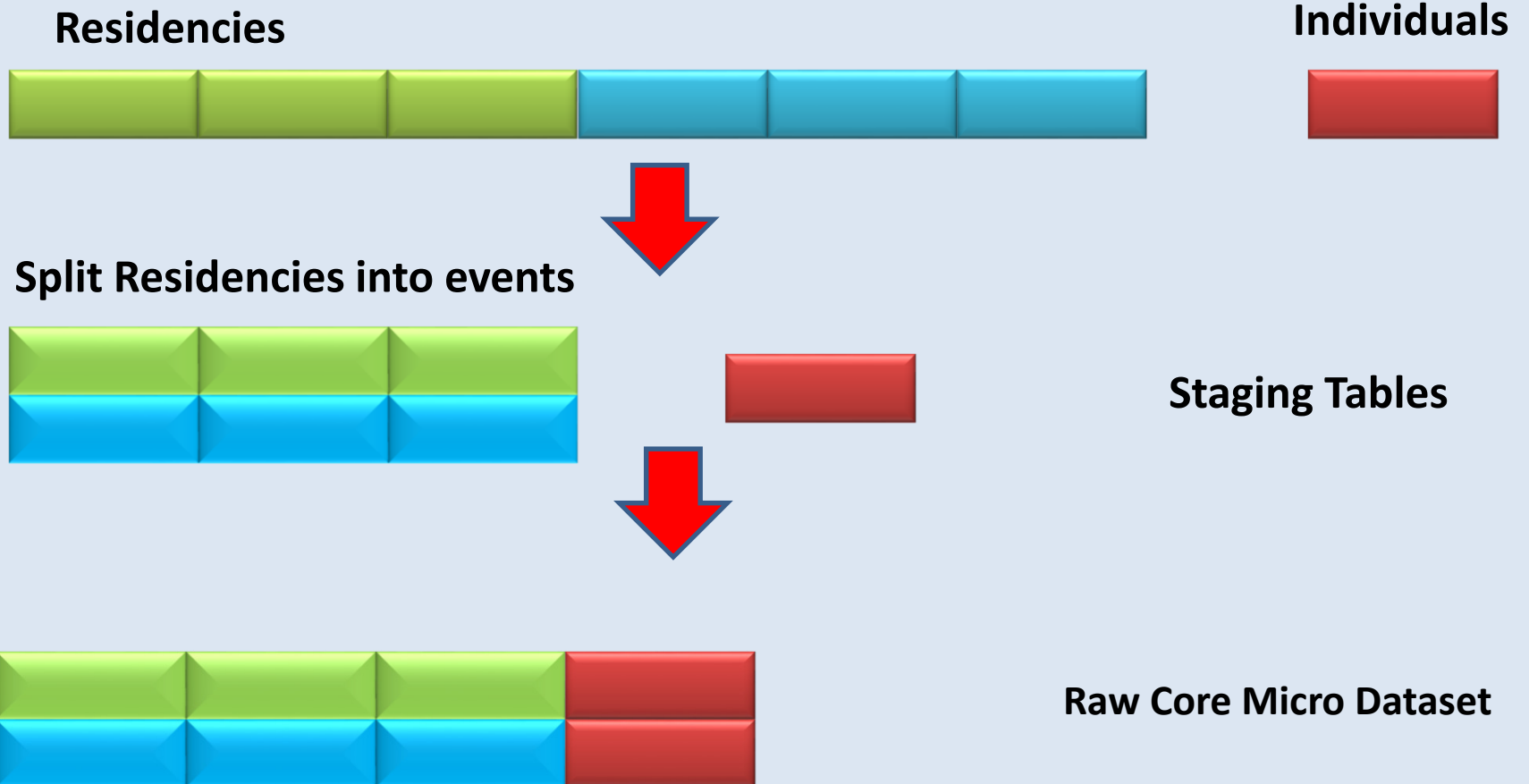
Data Flow for Generating the Micro Dataset



ETL Process For Family A



ETL Process For Family B



Let's Understand our Data Structure

- Identify to which Family does your center's dataset belong to?
- Identify the steps that you would follow to generate the raw micro dataset



Repository

- Repository Name: CiB
- Repository Path:
 - [\\SYSTEM\Archive\ETL](#)
 - Site Specific Transformations are stored here:
 - [\\SYSTEM\Archive\ETL\<sitecode>](#)
 - Common Transformations and Jobs are stored here:
 - [\\SYSTEM\Archive\ETL](#)
 - Master Job is stored here:
 - [\\SYSTEM\Archive\ETL](#)

Staging Database

- The Staging Database:
 - ETLStaging
- Staging Database Connection Name in PDI:
 - Staging

*Thank
you*

