

Data Use Community and OpenHIE Community Patient Identity Management – Biometrics

Lessons learnt on fingerprint quality from Côte d'Ivoire August 24th, 2022

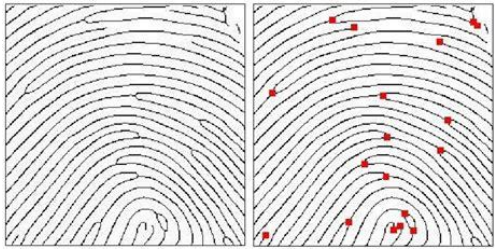
Franck Agah – Software engineer SEJEN-Côte d'Ivoire

Nicolas de Kerorguen – Contractor for CDC/CGH/DGHT/HIDMSB/Health Informatics Team

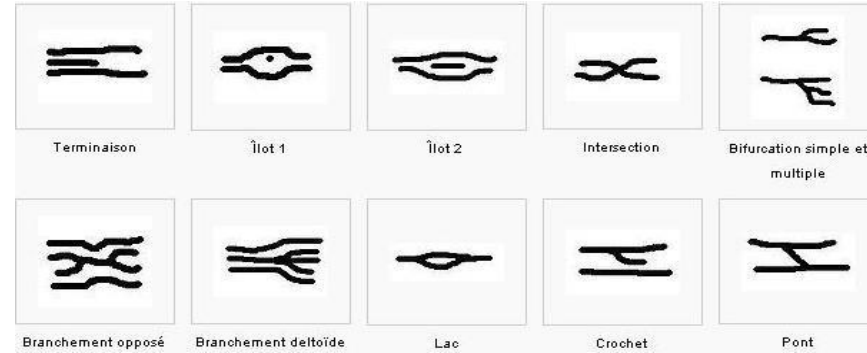
Overview

1. Background: Basic fingerprint concepts
2. Background: Fingerprint activity in Côte d'Ivoire
3. Process for fingerprint in CI
4. False positive challenge in CI
5. Proposed solutions for CI
6. Observations
7. Questions for the group

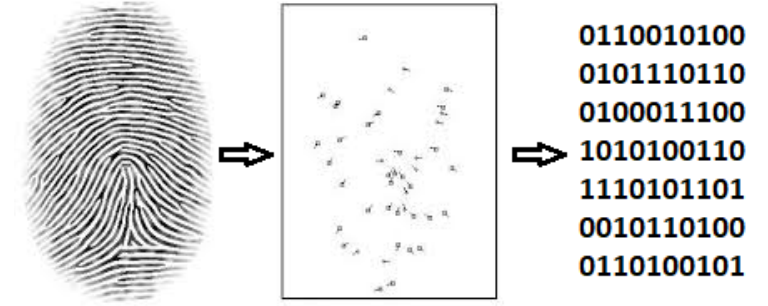
1 - Background: Basic fingerprint concepts



Minutiae in a fingerprint



Examples of minutiae types



From fingerprint image to template

Digitalization

- **Minutiae** are characteristics found in the FP
- Image of the FP is digitalized (+binarization & thinning) and transformed into a **template**
- => A template contains information on a certain number of minutiae (type, position)

Storage format

- Usually, the FP image is not stored (ethical/security consideration)
- The template is stored, either in an ISO or a proprietary format

1 - Background: Basic fingerprint concepts (cont.)

Quality of a fingerprint template

- Number of minutiae, extraction algorithm, number of views
- The size of the template file is not a good indicator for the quality
- There is a **quality score** that can be returned by the system (also green light on some devices/screens)
 - There is a quality threshold, under which a fingerprint cannot be validated. It is configurable.
- On the ISO template it is possible to count the number of minutiae in each template
- => Quality depends on configuration, process and training

Matching

- There is a **matching score** (likelihood of matching) between 2 templates
 - There is a matching threshold, above which 2 templates are seen as potential match. It is configurable.
- False positive: 2 different individuals are seen as a match based on their fingerprint template
- False negative: 2 templates belonging to the same individual but seen as non-match
- => There is always a balance to find between risking to have more false negatives or more false positives

Other considerations

- Ethical aspect, sensitivity, stigma
- Approval to collect biometric-related information

2 - Background: Fingerprint activity in Côte d'Ivoire

Goal of the activity

- Provide a unique ID for ART patients in a country where there is not yet a unique citizen ID or health ID covering the population. PEPFAR funded.
- Deduplicate patients seen at several ART facilities, with the following benefits:
 - Get more accurate active file count and other indicators
 - Optimize drug dispensing
 - Improve services for patients
- Match records between database. Extend beyond ART patients.
 - Improved services

Local company SEJEN

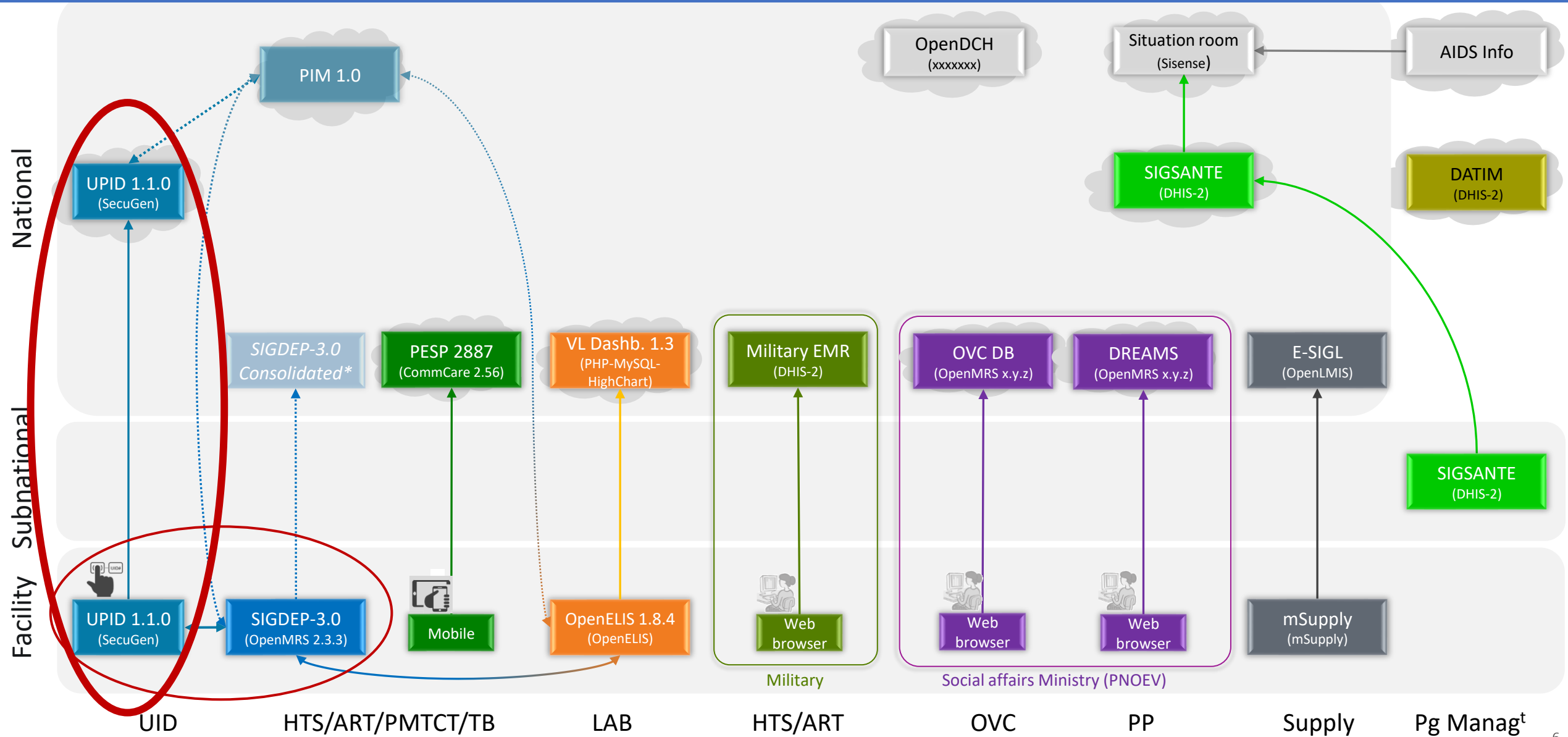
- 2019: subcontracted through CDC/UNAIDS CoAg for the UPID solution
- 2020: piloted in 10 facilities
- 2021: rolling out; 6 clinical Implementing Partners
- False positives identified in 2021
- As of Aug. 2022: 100 ART facilities equipped with UPID, out of 400+ ART facilities with EMR (SIGDEP2, based on OpenMRS)

Other information

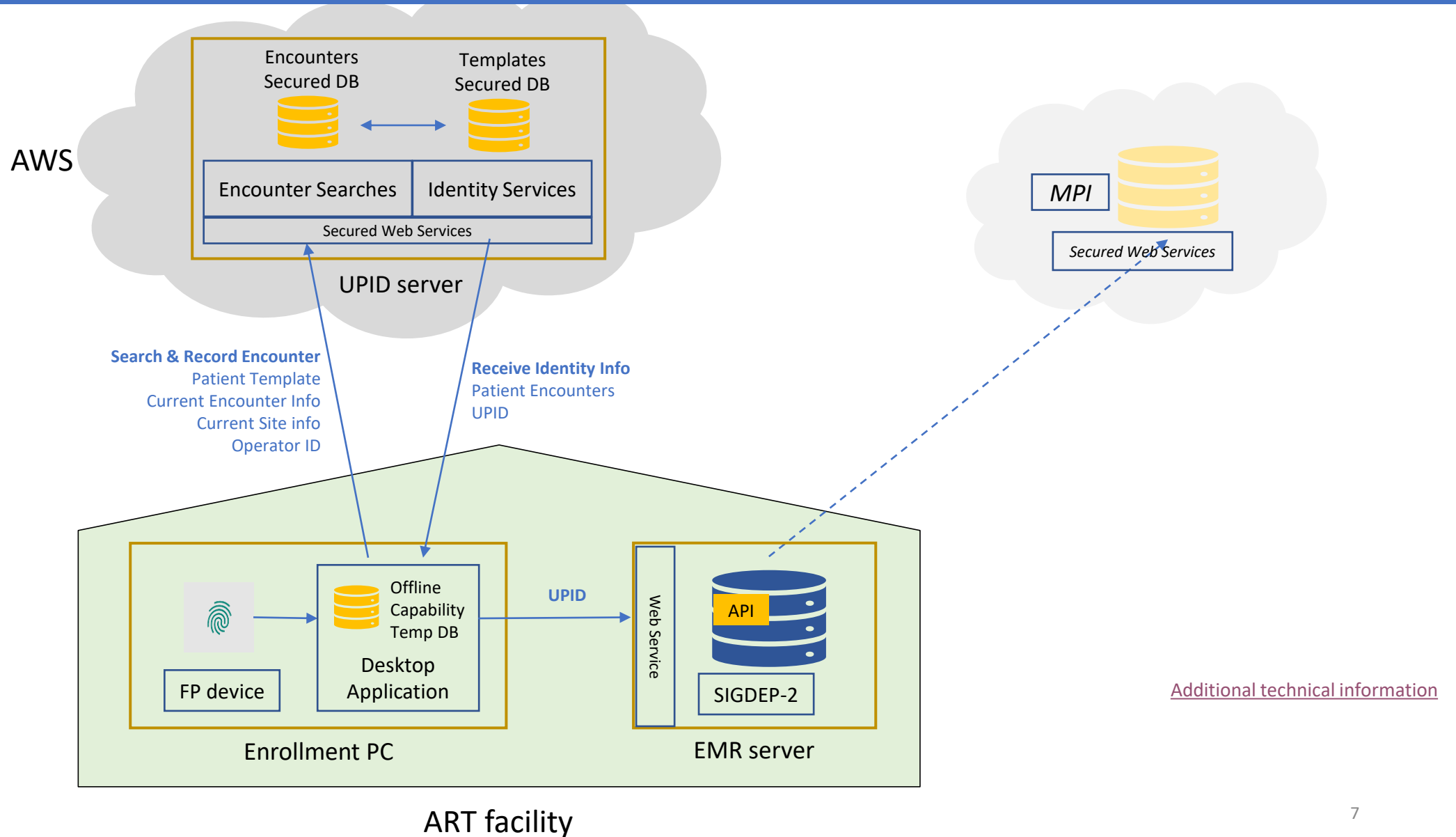
- Solution: SecuGen (proprietary) + Optical fingerprint device SecuGen Hamster Plus (& Hamster Pro 20TM)
- Management: Weekly TWG with SEJEN, MOH, CDC-CI, CDC-HQ, IPs (ACONDA, ARIEL, EGPAF, ICAP, FHI360, SEV-CI)



2 – Background: HIV HIS PEPFAR landscape as of end FY22



2 – Background: UPID architecture



3 - Process for FP in CI

SOP

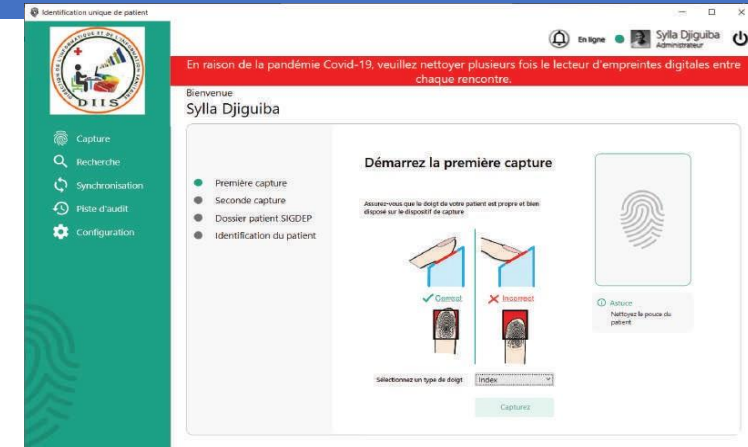
- For patient registration and identification, used at facility level

Registration

- Right and left index (if no index, then the following finger)
- Default threshold for data capture: 50/100 *
- No visualization screen for the image of the fingerprint taken *
- FP -> template stored with proprietary SecuGen format (and ISO), along with date, facility id, and HCW id.
No Patient socio-demographics or clinical information
- Local encounter DB-> central UPID database -> local server
- Generation by the central DB of a 9-digit unique ID (called **UPID**; format **nnn-nn-nnnn**)
- UPID sent to the EMR, via API (EMR = OpenMRS)

Identification

- Right or left index
- Default matching threshold: 5/10 *
- Central database returning a list of potential matches (previous encounters)



Example of snapshot in the SOP

4 - False positive challenge

- Observations
 - Initially, false positives were due to errors with facility codes, hiding the false positives issue
 - Merged securely records from all IPs (UPID; Facility code; Facility name; ART Patient code; Sex; Age; UPID creation date)
 - (March 2022) Found **6% duplicates** with same UPID but different ART #, out of 10,000+ records.
 - 10,850 records: 1 quintuple, 5 quadruples, 46 triples, 566 duplicates => 6.24%
 - (Aug. 2022) Found **10% duplicates** with same UPID but different ART #, out of 40,000+ records.
 - 41,556 records: 4 sextuples, 20 quintuples, 63 quadruples, 390 triples, 3,237 duplicates => 10.36% (corr. 10%)
 - **A vast majority were false positive**, based on quick analysis on age, sex and location; few were duplicate records
 - UPID goals cannot be met (*more accurate active file and other indicators, optimal drug dispensing, improved services for patients*); It is not yet easy to identify the 'real duplicates' (one patient with several ART#)
- Investigations by SEJEN and the national HMIS TWG (including IPs, MoH and CDC)
 - Code review, configuration review
 - Analysis of some known false positive and true positive duplicates (looking at all encounters, etc.)
 - Scripts developed to perform other analysis; log review
 - Little literature available on FP implementation; Discussed with CDC-Haiti team
- Conclusion
 - Main issue: the **poor quality** of the templates stored
 - Second issue: **matching threshold too low**
 - Other issues: algorithm for generating a new UPID, handling of false positive at facility level, - see next page

4 - False positive challenge (cont.)

- Additional issue:
 - There was also a false negatives challenge:
 - The system was initially built to help quickly enroll patients with a minimum set of steps. When the system did not recognize a patient, it would automatically create a new identity. However, in some cases a patient is returning to a facility where he has already been enrolled for fingerprint; for some reasons, such as bad quality of previous or current fingerprint, the system sees him as a new patient and generates another UPID; now the service provider realizing that this is a mistake takes again the fingerprint, and that would eventually match with a previous template. But the newly created UPID will remain in the database as a false negative.
=> Additional templates with low quality in the DB may generate more false positives
 - Handling of false positives by IPs:
 - A service provider seeing a list of potential matches (encounters), can check some information with the patient. Not all the facility providers may proceed the exact same way, the SOP being not precise enough.

5 - False positives: Solutions proposed

- Principles

- It is essential to reduce the number of false positives (high risk on patients)
- It is OK to increase the number of false negatives, if it cannot be avoided
 - However, increasing template quality will reduce both false positives and false negatives
- It is essential to optimize the quality of templates stored

- Fingerprint quality

- Increase the capture threshold from 50 to 80
- Visualization of the fingerprint image during capture
- *Replace the finger several times on the optical reader until complete capture (?)*
- Update the SOP and retrain people at facility

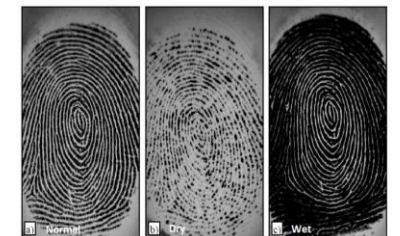
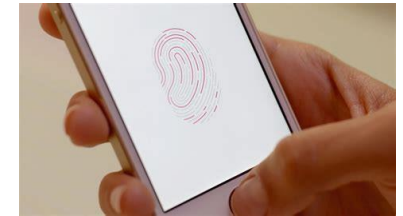
- Matching

- Increase matching threshold from 5 to 7

- Replacement

- Retake fingerprint for all patients coming back to facility

Quality score ≥ 80



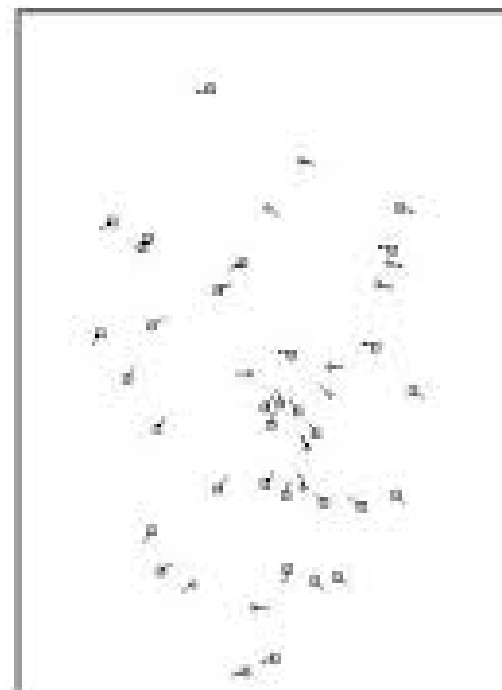
Matching score ≥ 7

6 – Observations

- Need to have a clear SOP and training to capture FP
 - Fingers sequence *- not always OK at the beginning*
 - Finger position and pressure on the device *- not always OK*
 - Cleaning the optical reader *- usually OK, with COVID-19*
 - Cleaning/moisturizing/drying fingers *- usually not OK*
- Need support from the solution company
- Fine tune the parameters and assess results at the beginning
 - Automated tests, reference dataset
- Little amount of literature available online
- Acceptance from patients, some resistance with service providers

7 – Questions for the group

1. Do other countries face similar challenge?
2. FP quality
 - a. How to count the number of minutiae in a template?
 - b. What is the minimum amount of minutiae (8, 12, 15, 20, other?) that should be considered for a valid FP?
 - c. Are there regional disparities with FP?
 - d. What about replacing the finger several times to complete the capture?
3. Matching
 - a. How many/which fingers for optimizing the result?
 - b. Increase matching threshold from 5 to 7: reasonable?
 - c. How to deal with people with no valid fingerprint (manual workers, elderly people)?
 - d. How to deal with infant/babies?
4. Replacement
 - a. Is it better to restart from scratch with a new database, or replace as patients return to facilities?
 - b. How did other countries manage this?
5. Overall
 - a. How to do the fine tuning (any best practices?)
6. What are the lessons learnt from other countries?



011001
010111
010001
101010
111010
001011
011010

Thank you

CDC-Côte d'Ivoire: Nicoué Aimé

SEJEN-Côte d'Ivoire: Alain Kouyaté, Franck Agah – www.sejen.ci

Other: all the members of the national HMIS TWG

Appendix 1 - Technical information on the UPID in Côte d'Ivoire

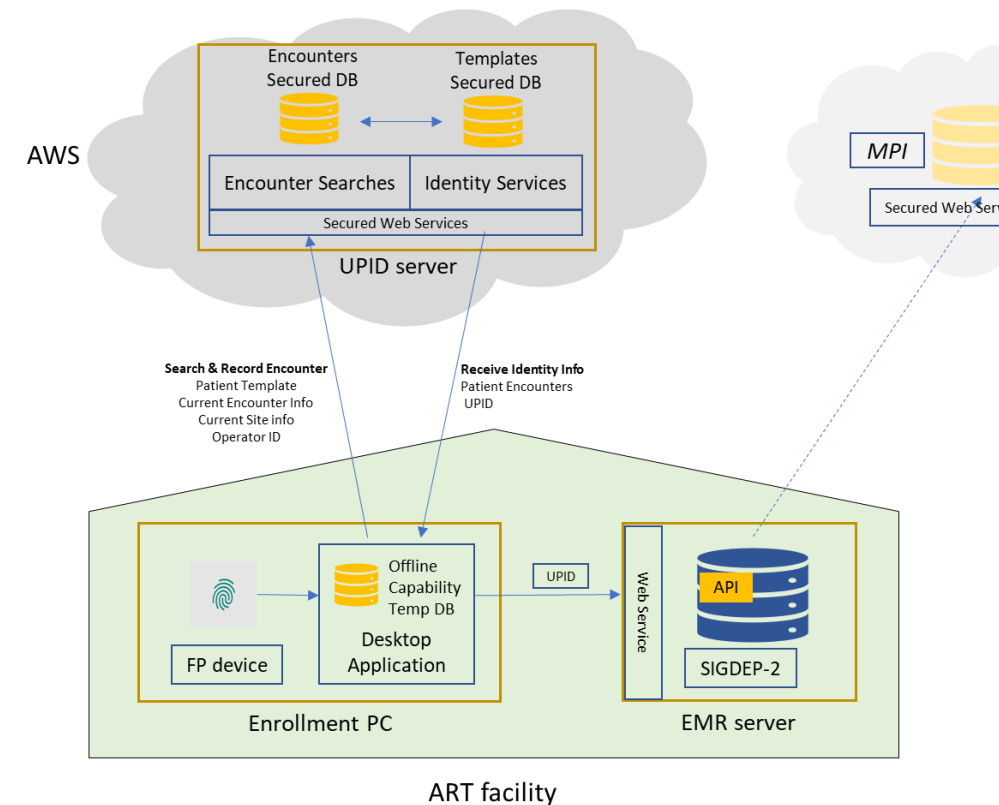
UPID solution implemented by SEJEN:

- Facility level

- Run on Windows
- PostgreSQL version 12
- PgAdmin 4
- Optical reader SecuGen Hamster Plus + SecuGen WBF driver 2.5
- Offline mode and synchronization
- *SIGDEP 2.2 (OpenMRS 1.9.8)*

- Central level

- AWS Amazon API Gateway
- AWS EC2 instance with MS Windows
- AWS Relational Database Services (RDS) with PostgreSQL
- AWS CloudWatch for monitoring
- SecuGen DB



Appendix 2 - Identifiers in Côte d'Ivoire

- UPID
 - XXX-YY-ZZZZ: 9 digits
 - Generated by the UPID system (central level)
 - UPID are unique
 - One patient shouldn't have more than one, except in the case of false negative
 - Two patients shouldn't have the same UPID, except in the case of false positive
- ART number (aka SIGDEP #)
 - NNNN/MM/YY/SSSSS/(E)
 - NNNN: facility number, 4 digits – national unique facility code
 - MM: service number within the facility, 2 digits (or 2 letters) -
 - YY: year of enrollment, 2 digits
 - SSSSS: sequential number, 5 digits
 - E (optional): 'E' for infant
 - Generated manually on the paper-based ART register and patient chart
 - One patient can have several ART#, for instance in the case of a silent transfer, of ART re-initiation, or for 'shopping'
 - Two patients shouldn't have the same ART#

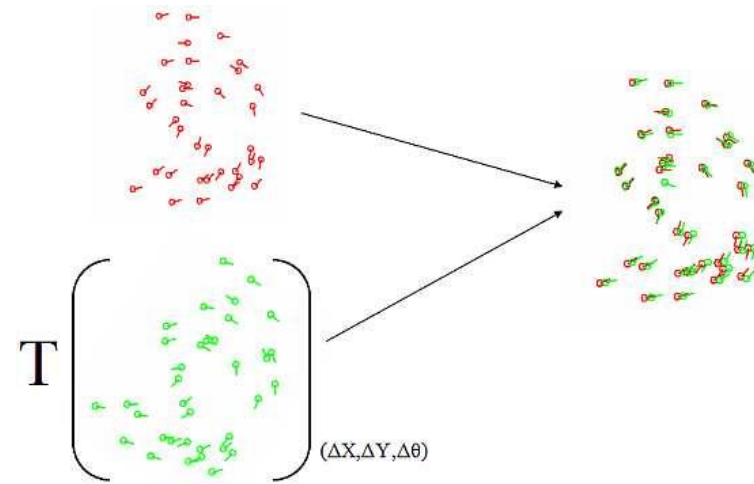
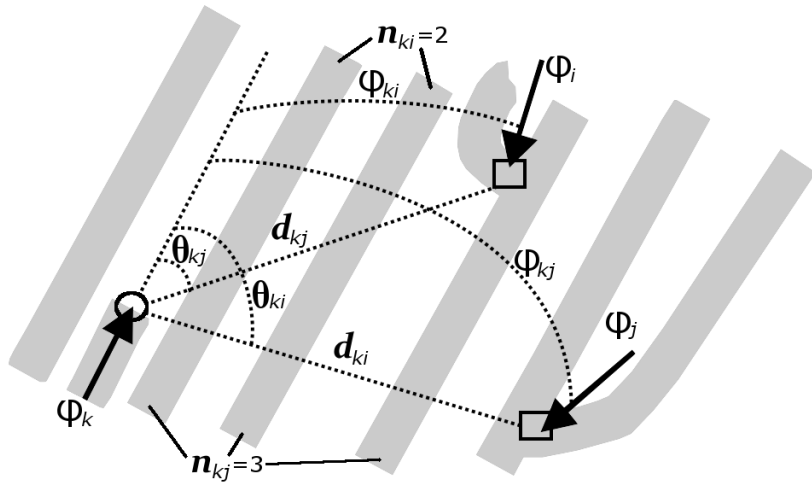
Appendix 3 – Investigation details

- Code
 - Client Side:
 - Check Capture threshold
 - Templates Sizes
 - Server Side:
 - Templates sizes
 - Search Engine settings: Matching threshold
 - Search steps that merges (Secugen DB & Business Logic Database)
 - Add new end point for better Matching Score Analysis
- Log added
 - Make the server more verbose:
 - Enable info traces
 - Add more traces at specific points of the process (capture, search)
- Data Analysis
 - First Understand the given use cases(False Positive/Negative)
 - Try to detect correlation between records on the server's database

Appendix 4 – Process off-line

1. Scan Fingers
2. Extract Munitiae
3. Delete FP Image
4. Generate Encounter ID
5. Attempt to reach AWS Server for Identification
6. No connection? Keep the encounter in Sync Menu : *No UPID generated until connection appears*

Appendix 6 – Minutiae, location and orientation



Some Resource

- Links

- ISO/IEC 19794-2:2011(E). Information technology — Biometric data interchange formats — Part 2:Finger minutiae data
<https://www.iso.org/obp/ui/#iso:std:iso-iec:19794:-2:ed-2:v1:en>
- SEJEN www.sejen.ci
- <https://www.biometrie-online.net> (French)