Evaluation Benchmarks

Ambika Suman and Geoff Whitaker

UK Police Information Technology Organisation

Biometrics Consortium 2005
Overview of Talk

• Drivers For Benchmarking
• Test Design
• Results Analysis
• Practicalities and Challenges
• Lessons Learned
The term “mark(s)” will be used to refer to latent finger or palm impressions that are left at crime scenes and used for investigations.
What was NAFIS?

- PNC Records
- National Ten-Print Database
- Print to Print
- Mark to Print
- Print to Mark
- Mark to Mark

- Mark Case Management Records
- National Unidentified SOC Database

6 million fingerprint comparisons every second

10,000 submissions per day

Biometrics Consortium 2005
NAFIS: more than just an AFIS!

- **PEDB** – Police Worker Elimination Database
- **SCC** – Serious Crimes Cache
- **ORD** - Operational Response Database
- Business Continuity Measures
  - Split Central. More than a DRS: data replication *and* load balancing
  - Reciprocal bureau procedures
- **PIDB** – Palm Image Database
- **PIFE** – Police Immigration Fingerprint Exchange
- **GMCI** – Generic Mark Camera Interface [http://nafis.pnn.police.uk](http://nafis.pnn.police.uk)
- **CJ Act** – Detainee processing, requiring 40% capacity increases
- Livescan to Custody Interface
- **Extradition Act** – International Interface
- **LANTERN** demonstrator – remote/mobile identity check service
IDENT1

• Continuity of services delivered by NAFIS
  Plus:
  • Palm print and palm mark searching
    – Including phalanges & hypothenars
  • Inclusion of Scotland
    – Palm-capable Livescan
    – A Unified Collection of print sets and unresolved SoC marks
    – Access same capabilities already available in England and Wales
  • A Strategic Identification Systems Platform
    – Capable of providing biometric identification services across the Police, Criminal Justice & Public Safety domains; e.g. IND
    – Capable of delivering new ID services; e.g. FIND (mugshots)
    – Compliant with broader strategic needs; esp. ISS4PS
      • Information sharing & accessibility; Componentisation/re-use

Biometrics Consortium 2005
Search accuracy is a key factor for user acceptance
‘Operational’ Benchmarking:
• The process used by PITO to determine the ‘end to end’ search accuracy of the system during operational use
Why Benchmark?

• To **differentiate** between Tenderers’ proposed technical solutions
• To determine whether or not Tenderers’ proposed technical solutions are likely to meet operational requirements
• To **quantify a standard** beyond which IDENT1 search accuracy should improve throughout its operational life.
• To give the Police Service the **assurance** that the search accuracy of the IDENT1 System will meet their needs
Test Design

Ambika Suman

UK Police Information Technology Organisation
1M Records

10 rolled finger impressions

80,000 Records

10 rolled finger impressions + both palms

Scope and Scale

- Added Palms capability (marks searching)

500 searches for each test

Scene of Crime marks
Latent fingerprint impressions

Biometrics Consortium 2005
Design Factors

The tests were designed to be:

– *System Independent*

– *Repeatable*

– *Operationally Representative*
Test Controls

• **Test Teams**
  - 10 fingerprint experts per benchmark
  - Equivalent in experience and expertise

• **Training**
  - 2 full days

• **Environment**
  - Reflective of a ‘typical’ bureau

• **Search Specifications/Allowances**
  - Predetermined by “Mark Evaluators”
Data

- Operational data from bureaux throughout UK
  - Randomised data from NAFIS electronic archive used to create 1M dataset
  - 100k fingerprint forms scanned and processed to create 80k dataset with palms
  - Over 4000 seed forms scanned and processed and inserted randomly into datasets

- Enquiries not previously matched on existing systems

- Every seed in datasets manually verified prior to tests
Test Management

- **Communications** with Suppliers during tests
- System / data **integrity** during tests
  - Check that test system matches proposal!
  - Audit trail
  - Spot checks
  - Electronic and hardcopy of results data
- Need to maintain **level playing field** between tests
Summary

• Cost
• Time
• Flexibility
• Trade off between ‘Operational’ and ‘Repeatable’
• Co-operation of User Community
• Identify Benefits to all Parties
  – Value of these benchmarks goes far beyond just obtaining information on accuracy
Thank you