

# The Biometric Testing Landscape

Patrick Grother and Tony Mansfield

National Institute of  
Standards + Technology (US)

National Physical  
Laboratory (UK)



# Agenda

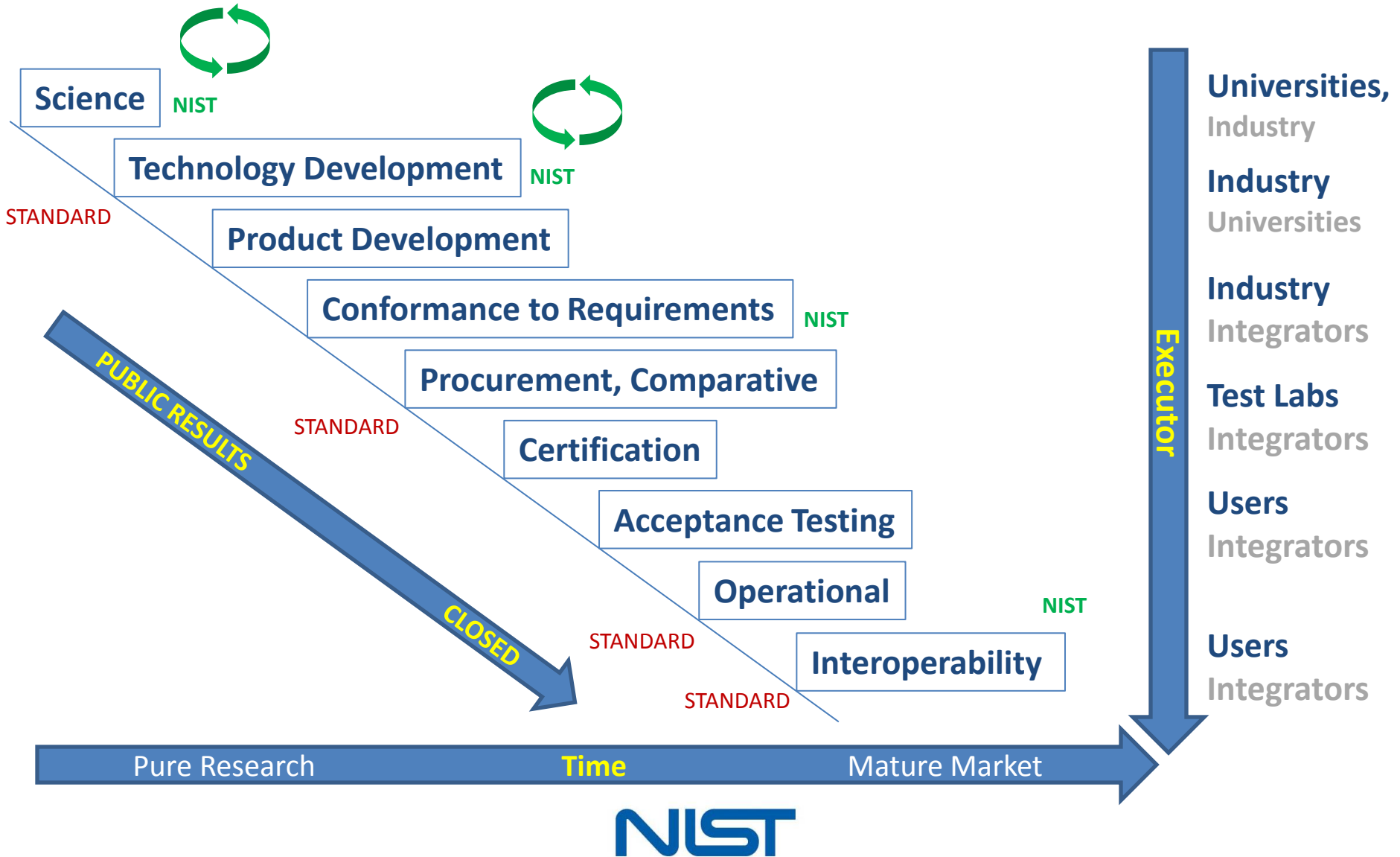
## » Testing

- » Purposes of testing
- » Stages of testing
- » Why bother?
- » Example outputs
- » What should be tested
- » Case study :: A 1:N Face Test

# The purposes of biometric testing

- » Scientific discovery
  - » Basic research, sensors, algorithms,
- » Research and development
  - » Is the method better than before?
  - » Internal commercial, or gov-industry cooperation
- » Capability testing
  - » Is the technology viable?
    - Benchmarking
    - Core capability
  - » Establishing criteria for later use
- » Technical discovery: What's possible
  - » Could requirements be met
  - » Are all requirements met simultaneously?
- » Comparative testing
  - » Which technology?
  - » Which implementation?
- » Conformance
  - » To requirements
  - » To standards
- » Interoperability testing
  - » Can we upgrade, replace, while keeping our data?
  - » Jurisdictional interoperability
- » Certification
  - » Can we leverage others' tests?
- » Regression
  - » Has the update helped?
- » Calibration
  - » Of compression algorithms
  - » Of quality algorithms
  - » Of matching algorithms

# Stages of Testing



# Why test at all?

## » It's about money

### » Biometric errors cause additional cost

- Failure to enroll → additional time, remedial procedures, fall-back modalities, secondary processes, time
- In 1:N false non-match → benefits fraud, immigration fraud
- In 1:N false match → identity resolution processes

### » Can the application requirements be met?

### » Can the technical requirements be met?

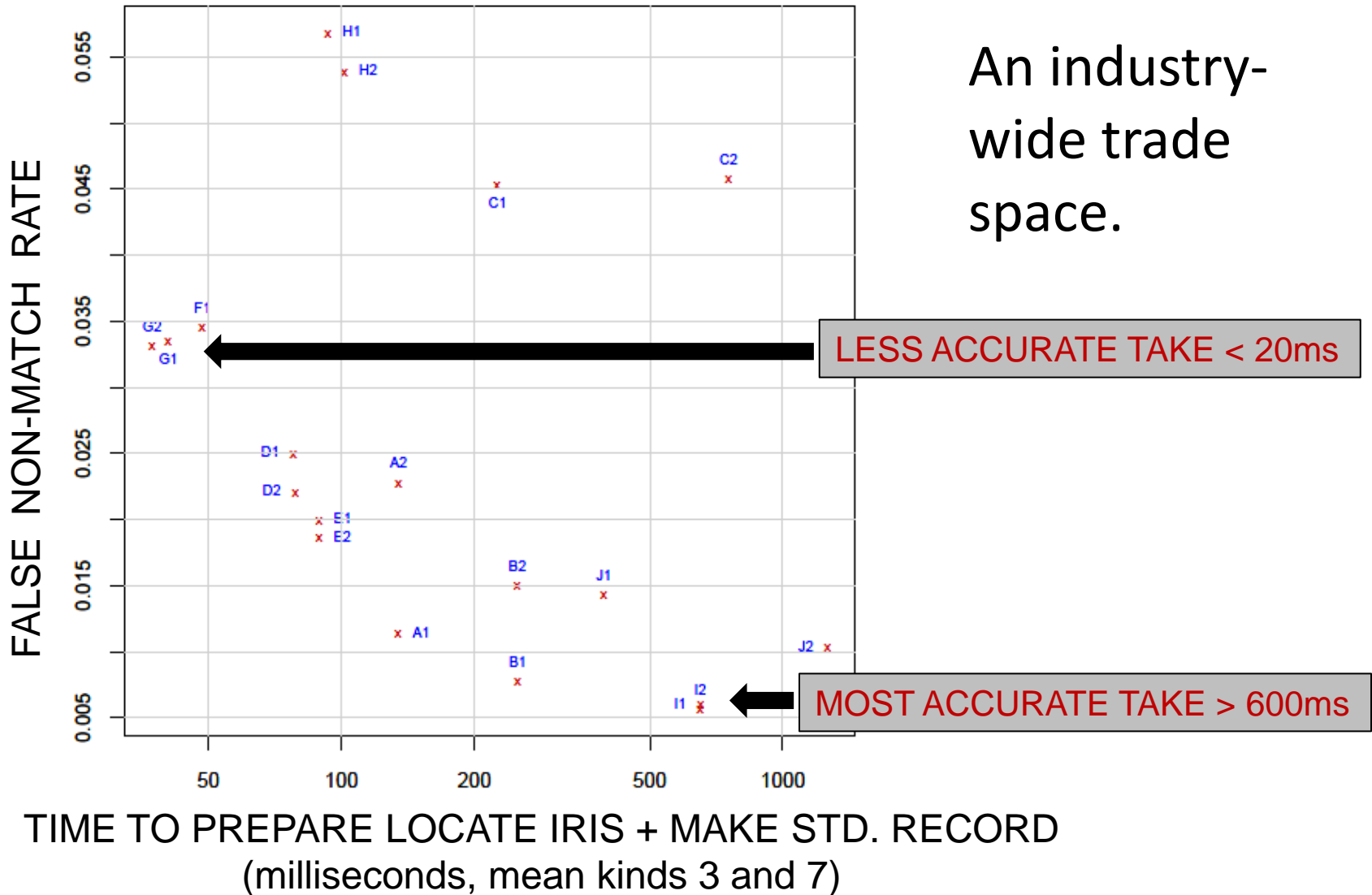
### » Risk mitigation

- Characterizing performance allows procedures to mitigate risk e.g. a finding that  $FTE > 0.03$  prompts environmental redesign to regulate humidity, or ambient light.

### » Conformance, Interoperability

- Entire system needs to be upgraded / junked if it doesn't interoperate with others

# Trading Time for Accuracy



An industry-wide trade space.

# Irrelevant or Inconsequential Tests?

- » Have biometric tests been driven by what can be done, rather than by what should be done.
  - » EXAMPLE: Measure FRR after a data collection vs. observe + count mis-presentation effects (e.g. via video observation).
  - » EXAMPLE: Consider the effects of pose on face recognition accuracy vs. develop, require and test pose estimation built into cameras.
- » Are tests missing the point? Are there more operationally important variables?
  - » EXAMPLE: Multimodal accuracy is excellent. But what's the cost (time and money) of always collecting two modes?
- » Is a test of a technology provider's product's core algorithm capability as important as the provider's ability to
  - » Tune the algorithm
  - » Get version control under control
  - » Do regression testing
  - » Adapt to different samples
  - » Set matcher thresholds to target a specified FMR value.
  - » Execute template update robustly
  - » Interoperate

# Case Study :: MBE-STILL

Objectives of NIST's 1:N Face testing  
tailored to law enforcement needs



# 1:N in Law Enforcement

## Characteristics of the FBI task

- » One-to-many, with  $N \rightarrow 10^8$
- » Aided by multiple arrest records
  - » i.e. multiple enrolled images
- » A federated application
  - » State + local + others  $\rightarrow$  FBI
  - » Heterogeneous images
- » Aided by face standards ... but
  - » Pictures don't conform to standards
  - » Pictures vary despite mug-shot standards
- » Aided by metadata
  - » Age, weight, gender ...
- » Here now, and operationally required



Missouri



Georgia



1. ANSI/NIST Type 10
2. ISO/IEC 19794-5 Full Frontal

ANSI/NIST Type 2 includes, for example,  
Height = 1.67m, Weight=62KG, Sex=M,  
Date=2001-11-27, DOB=1971-12-25

# MBE-2D STILL :: Test Design Objectives I

- » Maintain virtues of technology / offline testing
  - » Level playing field for comparative testing
  - » Repeatability, traceability
- » Reduce cost of testing
  - » Write API, solicit user/supplier comments, publish API, allow global re-use.
  - » Establish fixed test corpora
  - » Automate tests
- » Support algorithm development
  - » Test iteratively :: Provide results + feedback to vendors
  - » Make sample data available in advance
- » Push for greater operational realism, reduce downstream risk to operator
  - » Use operational data!
  - » Extend population size to  $N > 10^6$
  - » Execute in proper identification mode
    - Don't model 1:N as being N 1:1s
    - Open-universe (use impostors, use true impostors) search
  - » Measure duration of all function calls
  - » Measure resource usage

# MBE-2D STILL :: Test Design Objectives II

- » To support face recognition accuracy
  - » Exploit multiple historical images of an individual
    - Allow the implementation to execute fusion
  - » Exploit operational metadata :: Provide to implementation
    - Date of capture | Date of birth | Height | Weight | Sex
  - » Allow implementation to execute post-enrollment processing on the enrolled database
    - Feature space normalization, for example.
- » Support standards
  - » Inform the definition of 1:N and 1:1 transactions.
  - » Support base standards, e.g. on compression limits.

# Thanks

patrick.grother@nist.gov  
tony.mansfield@npl.co.uk