The Nature of the Digital Persona and Its Implications for Data Protection Law

Roger Clarke
Xamax Consultancy, Canberra
Visiting Professor – Cyberspace Law & Policy @ UNSW
and in Computer Science @ ANU

http://www.rogerclarke.com/ID/DP14 {.html,.pdf}

Bahçesehir Üniversitesi, Besiktas
20 January 2014

The Nature of the Digital Persona and Its Implications for Data Protection Law

Agenda

- The Construct
- Its Significance
- Its Context
- The Significance of the Context
- Implications for Data Protection Law

The Digital Persona

A model of an individual's public personality
based on data and maintained by transactions
and intended for use as a proxy for the individual

http://www.rogerclarke.com/DV/CFP93.html (Feb 1993)
http://www.rogerclarke.com/DV/DigPersona.html (Jun 1994)
http://www.rogerclarke.com/DV/HumanID.html (Dec 1994)
http://www.rogerclarke.com/ID/DP12.html (Sep 2014)

The Digital Persona

A model of an individual's public personality
based on data and maintained by transactions
and intended for use as a proxy for the individual

data shadow, virtual identity, epers
construct; recorded personality, partial, ghost, copy, ditto

http://www.rogerclarke.com/DV/CFP93.html (Feb 1993)
http://www.rogerclarke.com/DV/DigPersona.html (Jun 1994)
http://www.rogerclarke.com/DV/HumanID.html (Dec 1994)
http://www.rogerclarke.com/ID/DP12.html (Sep 2014)

The Digital Persona

A model of an individual's public personality
based on data and maintained by transactions
and intended for use as a proxy for the individual

data shadow, virtual identity, epers
construct; recorded personality, partial, ghost, copy, ditto

Anima: the inner personality, turned towards the unconscious
Persona: the public personality that is presented to the world

http://www.rogerclarke.com/DV/CFP93.html (Feb 1993)
http://www.rogerclarke.com/DV/DigPersona.html (Jun 1994)
http://www.rogerclarke.com/DV/HumanID.html (Dec 1994)
http://www.rogerclarke.com/ID/DP12.html (Sep 2014)

Category	Description
informal persona	a persona based on human perceptions
formal persona	a persona constructed on the basis of accumulations of structured data
projected persona	a persona controlled by the individual it is associated with
imposed persona	a persona controlled by someone other than the individual it is associated with
passive persona	a persona that comprises data alone
active persona	a persona that has some capacity to act, on behalf of, or in substitution for, the individual it is associated with, cf. a software agent
public persona	a persona that presents a commonly-held, composite image of a person who is presumed to be well-known (e.g. Marilyn Monroe, Marshall McLuhan) or of an archetype (e.g. 'action man', 'drug mule', 'psychopath')

http://www.rogerclarke.com/DV/DigPersona.html (Jun 1994)

The Digital Revolution 1985-2005

- **Convenient, Inexpensive Creation ('Born Digital')**
  Desktop publishing, graphic design tools, digital music generators, animation, digital-image and digital-video cameras
- **Convenient Conversion ('Digitisation')**
  Scanners, OCR, audio analogue-to-digital converters, digital cameras
- **Near-costless Replication**
  Disk-to-disk copying, screen-grabbers, CD/DVD-burners, networks
- **Very Rapid, Very Cheap Transmission**
  Modems, CDs/DVDs in the mail, emailed attachments, FTP, Web
- **Inexpensive and Widespread Access**
  PCs, PDAs, public kiosks, Internet cafes, mobile phones, tablets
- **Computer-Based Data Analysis**
  Data-matching, profiling, data-mining, pattern-recognition
- **Convenient Data Manipulation**
  Word-processors, sound, image and video-processing tools

Category	Description
open or overt persona	a persona whose existence, content and use are known to the individual who is associated with it
covert or hidden persona	a persona whose existence, content and/or use is/are not known to the individual associated with it

composite persona — a persona that combines data from multiple individuals

http://www.rogerclarke.com/ID/DP12.html (Sep 2014)

Epidemiological and Memetic Analysis

- The primary transmission vectors have been published journal papers, conference presentations
- The main articles have achieved c. 200 citations, and related articles have received only c. 200 citations
- Many citing publications have relatively few citations
- There has been limited complementary commercial activity such as workshops and consultancies
- There has been neither direct (Darwinian) copying nor performative / co-optive (Lamarckian) copying
- The idea has lacked virulence

The Digital Persona Construct
needed to be projected as part of

A Comprehensive Theory of (Id)entification and Authentication

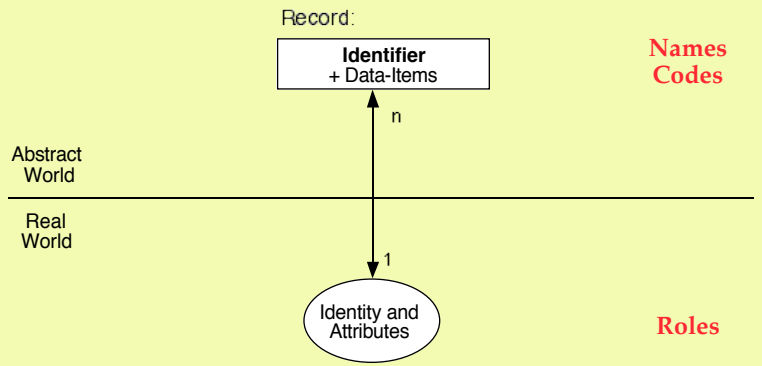
http://www.rogerclarke.com/DV/HumanID.html (1994)
http://www.rogerclarke.com/DV/UIPP99.html (1999)
http://www.rogerclarke.com/II/ECIS2001.html (2001)
http://www.rogerclarke.com/EC/AuthModel.html (2001)
http://www.rogerclarke.com/EC/ACSeAuth.html (2003)
http://www.xamax.com.au/EC/IdMngt.html (2004)
http://www.rogerclarke.com/EC/IdMngt-0804.html (2008)
http://www.rogerclarke.com/ID/IdModel-1002.html (2009-10)

The Nature of the Digital Persona and Its Implications for Data Protection Law

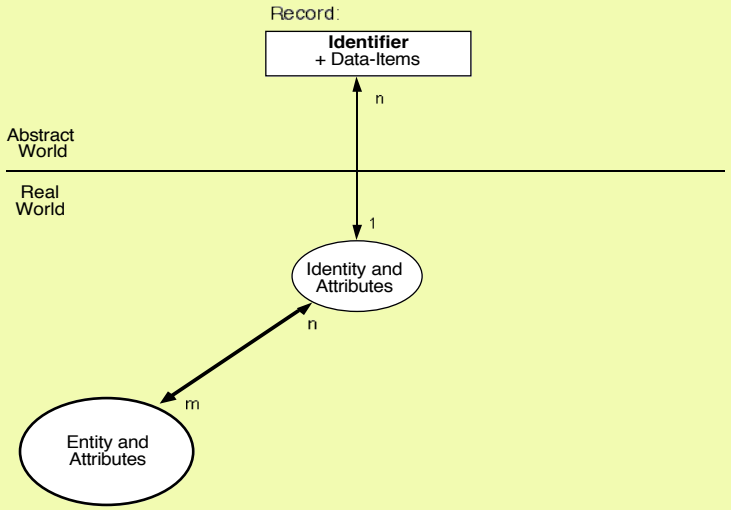
Agenda

- The Construct
- Its Significance
- Its Context
- The Significance of the Context
- Implications for Data Protection Law

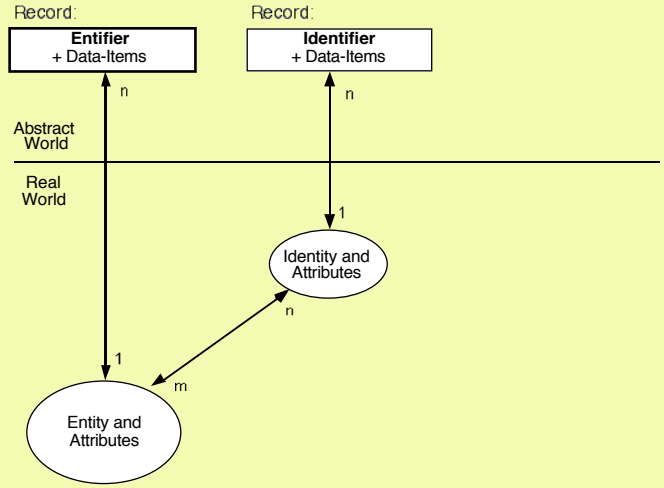
Identity and Identifier



The Entity/ies underlying an Identity



Entity and Entifier



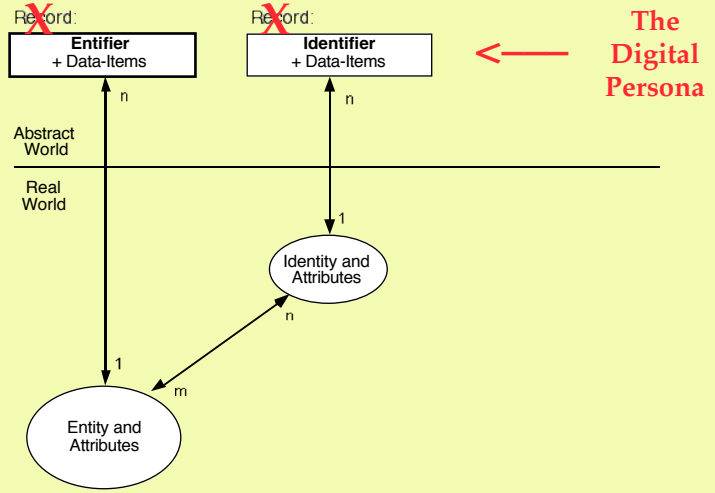
The Digital Persona

A model of an individual's public personality
based on data and maintained by transactions
and intended for use as a proxy for the individual

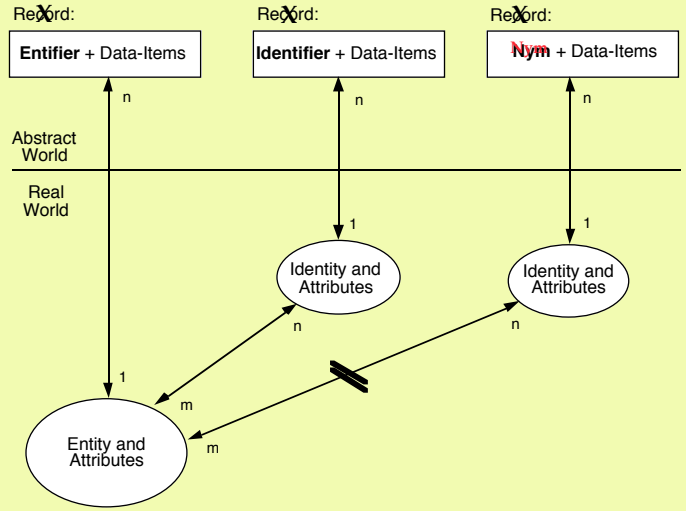
A group of data items that together
form a simplified representation of an identity

http://www.rogerclarke.com/DV/CFP93.html (Feb 1993)
http://www.rogerclarke.com/DV/DigPersona.html (Jun 1994)
http://www.rogerclarke.com/DV/HumanID.html (Dec 1994)
http://www.rogerclarke.com/ID/DP12.html (Sep 2014)

Entity and Entifier



Nymity



The Digital Persona

Nym

A Digital Persona

i.e. a set of attributes of an Identity

that is sufficient to distinguish that Identity from other instances of its class

but

that is not sufficient to enable association with a specific Entity

Pseudonym – association is not made, but is possible

Anonym – association is not possible

Nymality is Normality

aka ('also-known-as'), alias, avatar, character, nickname, *nom de guerre*, *nom de plume*, manifestation, moniker, personality, profile, pseudonym, pseudo-identifier, sobriquet, stage-name

Cyberpace has adopted those and spawned more:

account, avatar, handle, nick, persona, ...

Common Nymous Transactions

- Barter transactions
- Visits to Enquiry Counters in government agencies
- Telephone Enquiries
- Inspection of publications on library premises
- Access to Public Documents by electronic means, at a kiosk or over the Internet
- Cash Transactions, incl. the myriad daily payments for inexpensive goods and services, gambling, road-tolls
- Voting in secret ballots
- Treatment at discreet clinics, e.g. for sexually transmitted diseases

Applications of Pseudonymity

- Epidemiological Research (HIV / AIDS)
- Financial Exchanges, including dealing in commodities, stocks, shares, derivatives, and foreign currencies
- Nominee Trading and Ownership
- Banking Secrecy, incl. 'Swiss bank accounts'
- Political Speech
- Artistic Speech
- Call Centres
- Counselling
- Phone-calls with CLI
- Internet Transactions
- 'Anonymous' re-mailers
- Chaumian eCash™ and Bitcoin?

(Id)entification

- **Identification**

  The **process** of associating a Digital Persona with a particular Identity, by acquiring an Identifier for the Identity

(Id)entification

- **Identification**

  The **process** of associating a Digital Persona with a particular Identity, by acquiring an Identifier for the Identity
- **Entification**

  The **process** of associating a Digital Persona with a particular Entity, by acquiring an Entifier for the Entity

(Id)entification

- **Identification**

  The **process** of associating a Digital Persona with a particular Identity, by acquiring an Identifier for the Identity
- **Entification**

  The **process** of associating a Digital Persona with a particular Entity, by acquiring an Entifier for the Entity
- **Token**

  A recording medium for an (Id)entifier
- **Identity Silo**

  A restricted-purpose Identity, and associated Identifier(s)

Authentication of Assertions

- **Authentication**: A process that establishes a level of confidence in an Assertion
- **Assertion**: a proposition .. ... of some kind ... made by some party
-

Authentication of Assertions

- **Authentication**: A process that establishes a level of confidence in an Assertion
- **Assertion**: a proposition .. ... of some kind ... made by some party

- **Authenticator**: evidence relating to an Assertion
- **Credential**: a physical or digital Authenticator
- **EOI**: an Authenticator for Identity Assertions

Categories of Assertions Relevant to eBusiness

- About Real-World Facts
- About Data Quality (accuracy, timeliness, ...)
- About Value
- About Location
- About Documents

- About Attributes
- About Principal-Agent Relationships

- About Identities
- About Entities

A Defining Aphorism of Cyberspace

The New Yorker
5 July 1993


"On the Internet, nobody knows you're a dog."

Value Assertion

Value is transferred to/from an (Id)entity or Nym

'This bone's been aged in loam for three months'

Value Assertion

Value is transferred to/from an (Id)entity or Nym

'This bone's been aged in loam for three months'

Authentication of Value Assertions

For Goods

- Inspect them
- Get them put into Escrow, for release by the Agent only when all conditions have been fulfilled

For Cash

Release the Goods only:

- For Cash On Delivery
- After Clearing the Cheque
- Against a Credit-Card Authorisation
- After a Debit-Card Transaction

Attribute Assertion

- **An Identity or Nym has a particular Attribute**:
  - Age / DoB before or after some Threshhold
  - Disability, Health Condition, War Service
  - Professional, Trade (or Dog) Qualification

Authentication of Attribute Assertions

- ID-Card and DoB (may or may not record ID)
- Bearer Credential (ticket, disabled-driver sticker)
- Attribute Certificates (with or without ID)

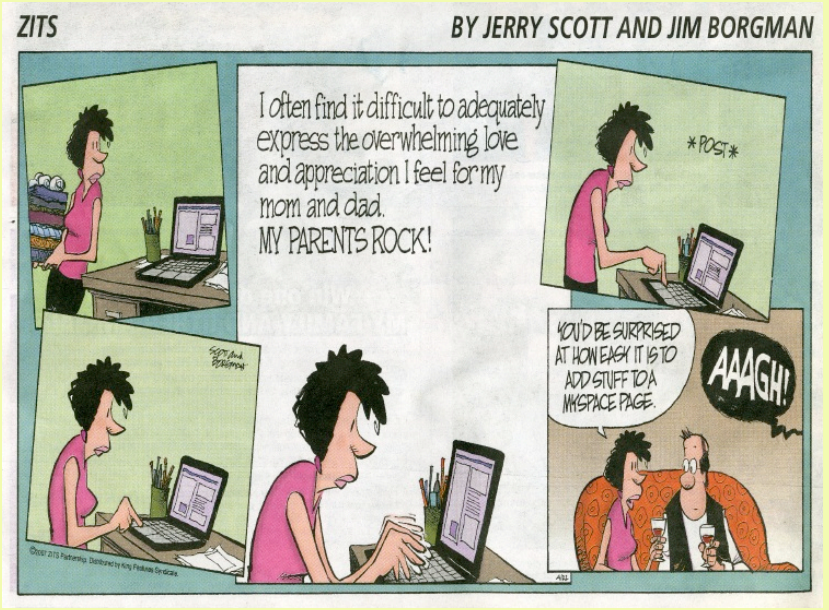
Which Assertions Matter?

- Utilise Risk Assessment techniques to determine:
  - Which Assertions?
  - What level/strength of Authentication?

eVoting
Assertions That Need Authenticating

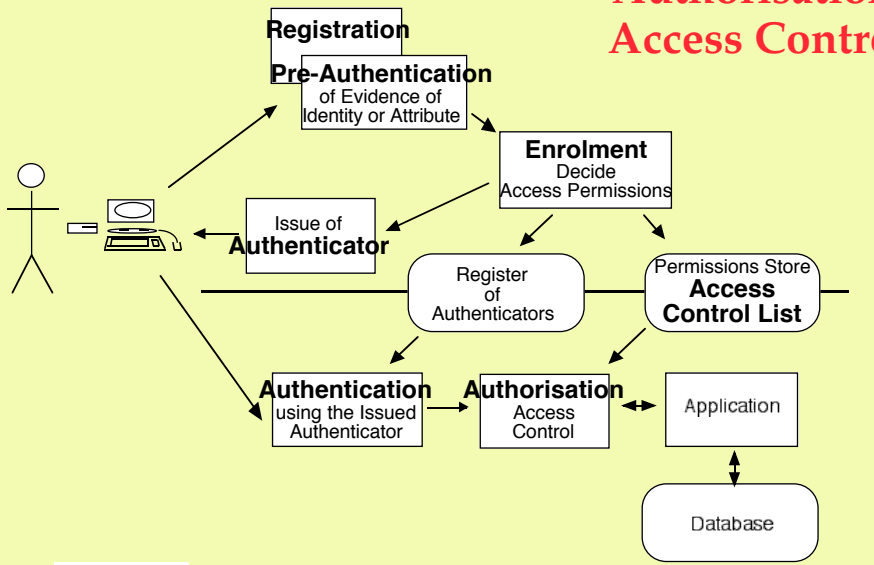
1. That's me on the Electoral Register ('smell this!')
2. Your vote was recorded, as follows ...

   ... by disclosing <u>to the voter</u> the details of the vote that was recorded; but in such a way that:
   - the voter can use it as evidence of malfunction if the voter chooses to do so; and
   - a person other than the voter can <u>only</u> know what the vote was if the voter discloses it
3. Your vote was counted



Authentication and Authorisation / Privileges

Authorisation: Access Control



Human Identification

- **Identification Generally**

  The process of associating a Digital Persona with a particular Identity, by acquiring an Identifier for the Identity
- **Human Identification in Particular**
  - Acquisition of a Human Identifier (Commonly a Name or a Code)
  - High-Reliability Lookup in a Database (1-with-many comparison, a single confident result)

Human Identity Authentication

- **What the Person Knows**
  e.g. mother's maiden name, Password, PIN
- **What the Person Has ('Credentials')**
  e.g. a Token, such as an 'ID-Card', a Ticket
  e.g. a Digital Token such as
  "a Digital Signature consistent with the
  Public Key attested to by a Digital Certificate"

Human Identity Authentication

- **What the Person Knows**
  e.g. mother's maiden name, Password, PIN
- **What the Person Has ('Credentials')**
  e.g. a Token, such as an 'ID-Card', a Ticket
  e.g. a Digital Token such as
  "a Digital Signature consistent with the
  Public Key attested to by a Digital Certificate"

Human Entity Authentication

- **What the Person Does** (Dynamic Biometrics)
- **What the Person Is** (Static Biometrics)
- **What the Person Is Now** (Imposed Biometrics)

The Nature of the Digital Persona and Its Implications for Data Protection Law

Agenda

- The Construct
- Its Significance
- Its Context
- The Significance of the Context
- Implications for
  Data Protection Law

Implications for Data Protection Law
Use & Disclosure of Personal Data

- **Use Personal Data Carefully**, because:
  - data are rebuttable propositions, not facts
  - data is a digital persona
    - not a person
    - not the person
  - decisions about a persona made by algorithms are inevitably dubious decisions, and resented
- **Disclose Personal Data Even More Carefully**, because:
  - all of the above
  - the data has moved beyond its original context

Implications for Data Protection Law
Authentication of Assertions of Fact

- Collect evidence to support assertions
- Ensure the evidence relates to the key quality factors, incl. relevance, accuracy, precision, timeliness, completeness
- The more serious the decision(s), the stronger the obligations
- Retain the evidence as long as the data

Implications for Data Protection Law
Quality Factors in the Digital Persons

- **Projected Personae** involve individuals:
  - painting themselves in a good light
  - omitting data they perceive to be harmful
  - misleading others
  - cheating others
- **Imposed Personae** involve organisations:
  - exploiting personal data
  - exploiting persons
  - misleading data subjects
  - manipulating the behaviour of data subjects

Implications for Data Protection Law
Authentication of Assertions of Identity

- It's expensive, unreliable, error-prone, vulnerable
- Decision-makers must not assume data quality
- Composite Personae arise from m:n Mappings
- The subject must have an opportunity to contest
- The organisation has an obligation to justify

Implications for Data Protection Law
Nymity

- It's vital for individuals, as a basis for:
  - personal safety
  - freedom of thought and speech
  - whistleblowing
- It's vital for organisations, to enable:
  - modelling of identity fraud
  - modelling of actions by agents
  - whistleblowing

Implications for Data Protection Law
Authentication of Assertions of Entity

- Depends on an Entifier / Biometric
- Is highly intrusive, and is an exercise in power, so it invites countermeasures by the public
- Is expensive, unreliable, error-prone, vulnerable
  There are scores of quality factors in biometrics

Quality Challenges in Biometric Applications

Dimensions of Quality

- Reference-Measure
- Association
- Test-Measure
- Comparison
- Result-Computation

Other Aspects of Quality

- Vulnerabilities
- Quality Measures
- Counter-Measures
- Spiralling Complexity
- Consequences

Implications for Data Protection Law
Administrative Fairness and Justice

- A Persona must be contestable
- The organisation must be accountable

- A Covert Persona can't be rebutted
- Declaration of the data relied upon is essential

Implications for Data Protection Law
Authentication of Users

- Access Control
  - requires Authentication
  - involves Authorisation / Privileges
- Privileges must reflect needs
- Corporations are routinely mis-designing and/or mis-deploying their Access Control

- The NSA is just as bad, perhaps worse

The Nature of the Digital Persona and Its Implications for Data Protection Law

Roger Clarke
Xamax Consultancy, Canberra
Visiting Professor – Cyberspace Law & Policy @ UNSW
and in Computer Science @ ANU

http://www.rogerclarke.com/ID/DP14 {.html,.ppt}

Bahçesehir Üniversitesi, Besiktas
20 January 2014

DRILL-DOWN

SLIDES

Applications of the Model

- Goods
- Packaging
- Animals
- Vehicles
- Devices
- Software
- Organisations
- Humans

Applications to Humans and Proxies for Humans

- Goods
- Packaging
- Animals
- Vehicles
- Devices
- Software
- Organisations
- **Humans**

- Personal Goods
- Pets
- Personal Vehicles
- Personal Devices
- Reg-Codes, IP-Addresses
- Organisational Roles
- Biometrics, Embedded Chips

A Sample Personal Device – The Mobile Phone

- **Entifier for the Product** – model-name, model-number

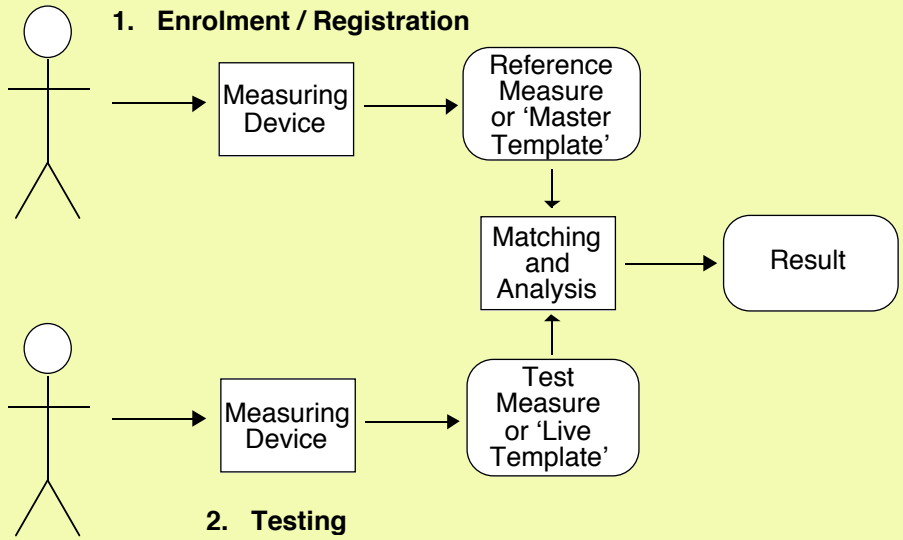
A Sample Personal Device – The Mobile Phone

- **Entifier for the Product** – model-name, model-number
- **Entifier for the Handset** – Serial-Number of the device
  - Mobile Equipment Identity (**IMEI**) – GSM / UMTS
  - Electronic Serial Number (ESN) or
    Mobile Equipment Identifier (MEID) – CDMA

A Sample Personal Device – The Mobile Phone

- **Entifier for the Product** – model-name, model-number
- **Entifier for the Handset** – Serial-Number of the device
  - Mobile Equipment Identity (**IMEI**) – GSM / UMTS
  - Electronic Serial Number (ESN) or
    Mobile Equipment Identifier (MEID) – CDMA
- **Identifier for the Persona** – Serial-Number of a chip,
  the International Mobile Subscriber Identity (**IMSI**)
  - Subscriber Identity Module (SIM) – GSM / UMTS
  - Removable User Identity Module (R-UIM) or
    CDMA Subscriber Identity Module (CSIM) – CDMA
  - Universal Subscriber Identity Module (USIM) – 3G
- **Proxy-(Id)entifier** – MAC Address / NICId, or IP-Address

The Biometric Process



Human Entification

- Acquisition of a Human Entifier
- High-Reliability Lookup in a Database
  (1-with-many comparison, a single confident result)

Human Entity Authentication

- Formation of an Entity Assertion
  ('This is the person who has a specific biometric')
- Acquisition of a Human Entifier
- High-Reliability Comparison with a Prior Measure
  (1-with-1 comparison, a confident Yes or No)

Some Mythologies of the 'Identity Management' Industry

- That the assertions that need to be authenticated are assertions of identity (cf. fact, value, attribute, agency and location)
- That individuals only have one identity
- That identity and entity are the same thing

That biometric identification:

- works
- is inevitable
- doesn't threaten freedoms
- will help much
- will help at all in counter-terrorism

C2K Policy Imperatives

- Maximise the use of anonymous transactions (tx)
- Where anonymity is not an effective option, maximise the use of pseudonymous tx
- Resist, and reverse, conversion of nymous to identified tx
- Preclude identified transactions except where functionally necessary, or meaningful, informed consent exists
- Enable multiple identities for multiple roles
- Enable the authentication of pseudonyms
- Provide legal, organisational and technical protections for the link between a pseudonym and the person behind it
- Resist, and reverse, multiple usage of identifiers
- Resist, and reverse, the correlation of identifiers

Design Factors using Chip-Cards Privacy-Sensitive and Cost-Effective

- 'Electronic Signature Cards' rather than 'Id Cards'
- No central storage of biometrics
- Two-way device authentication
- Attribute / Eligibility Authentication as the norm, identity auth'n as fallback, entity auth'n as rarity
- Nymous transactions unless (id)entity is justified
- Multiple Single-Purpose Ids, not multi-purpose ids
- Multi-Function Chips with secure zone-separation
- Role-Ids as the norm, Person-Ids as the exception

Reference-Measure Quality

- The Person's Feature ('Enrolment')
- The Acquisition Device
- The Environmental Conditions
- The Manual Procedures
- The Interaction between Subject and Device
- The Automated Processes

Association Quality

- Depends on a Pre-Authentication Process
- Subject to the Entry-Point Paradox
- Associates data with the 'Person Presenting' and hence entrenches criminal IDs
- Risk of an Artefact Substituted for, or Interpolated over, the Feature

Test-Measure Quality

- The Person's Feature ('Acquisition')
- The Acquisition Device
- The Environmental Conditions
- The Manual Procedures
- The Interaction between Subject and Device
- The Automated Processes

Comparison Quality

- **Feature Uniqueness**
- **Feature Change**:
  - Permanent
  - Temporary
- **Ethnic/Cultural Bias**

  "Our understanding of the demographic factors affecting biometric system performance is ... poor" (Mansfield & Wayman, 2002)
- **Material Differences** in:
  - the Processes
  - the Devices
  - the Environment
  - the Interactions
- **An Artefact**:
  - Substituted
  - Interpolated

'Factors Affecting Performance'
(Mansfield & Wayman, 2002)

- **Demographics** (youth, aged, ethnic origin, gender, occupation)
- **Template Age**
- **Physiology** (hair, disability, illness, injury, height, features, time of day)
- **Appearance** (clothing, cosmetics, tattoos, adornments, hair-style, glasses, contact lenses, bandages)
- **Behaviour** (language, accent, intonation, expression, concentration, movement, pose, positioning, motivation, nervousness, distractions)
- **Environment** (background, stability, sound, lighting, temperature, humidity, rain)
- **Device** (wear, damage, dirt)
- **Use** (interface design, training, familiarity, supervision, assistance)

Result-Computation Quality

- Print Filtering and Compression:
  - Arbitrary cf. Purpose-Built
- The Result-Generation Process
- The Threshhold Setting:
  - Arbitrary? Rational? Empirical? Pragmatic?
- Exception-Handling Procedures:
  - Non-Enrolment
  - Non-Acquisition
  - 'Hits'

Consequences of Quality Problems

- A Tolerance Range has to be allowed
- 'False Positives' / 'False Acceptances' arise
- 'False Negatives' / 'False Rejections' arise
- Tighter Tolerances (to reduce False Negatives) increase the rate of False Positives; and vice versa
- The Scheme Sponsor sets (and re-sets) the Tolerances
- Frequent exceptions are mostly processed cursorily
- Occasional 'scares' slow everything, annoy everyone

Design Factors Using Biometrics
Privacy-Sensitive and Cost-Effective

Technologies and Products

- A Privacy Strategy
- Privacy-Protective Architecture
- Open Information
- Independent Testing using Published Guidelines
- Publication of Test Results

Application Design Features

- No Central Storage
- Reference Measures only on Each Person's Own Device
- No Storage of Test-Measures
- No Transmission of Test-Measures
- Devices Closed and Secure, with Design Standards and Certification
- Two-Way Device Authentication

Application Design Processes

- Consultation with the Affected Public from project commencement onwards
- Explicit Public Justification for privacy-invasive features
- PIAs conducted openly, and published
- Metricated pilot schemes

Laws, to require compliance with the above

Laws, to preclude:

- Retention of biometric data
- Secondary use of biometric data
- Application of biometrics absent strong and clear justification
- Manufacture, import, installation, use of non-compliant biometric devices
- Creation, maintenance, use of a database of biometrics

Defined Terms in the Model

- entity, identity, anonymity, pseudonymity, nymity, attributes
- record, data item, digital persona, data silo
- (id)entifier, (id)entification, token, nym, anonym, pseudonym, identity silo, multi-purpose / general-purpose identifier
- authentication, authentication strength, assertion, assertion categories, authenticator, credential, (id)entity authentication, evidence of (id)entity, (id)entity credential
- authorisation/permission/privilege, user, loginid/userid/username, account, access control, registration, pre-authentication, enrolment, single sign-on, simplified sign-on, identity management