Requirements for Maximizing Privacy Protection in Identity Management Systems

W3C Workshop

Giles Hogben
Joint Research Centre
European Commission
/ENISA (in 2 weeks)

Key Points I'd like you to come away with

- **Identity is**

  Assertions + Evidence + Audit Trail

- **Restricting identity to binary assertions is bad for privacy and business.**

- **Evidence (not only PKI certs) needs an abstraction layer.**

- **Idemix + RDF + OWL + SPARQL can solve these problems.**

What I'm going to talk about

- Requirements for Privacy in IDM
- Solutions
  - PRIME architecture
  - Using the semantic web for Privacy Enhancing IDM
  - Describing Minimizeable Assertions
  - Describing Evidence about Assertions

Requirements - EU Directive

- Minimum amount of data should be collected for the specified purpose.
- Is this realistic any more?
  - Myspace
  - Blogs
  - Gmail …

What is Identity management

Assertions

What I and others claim about me (AKA claims)

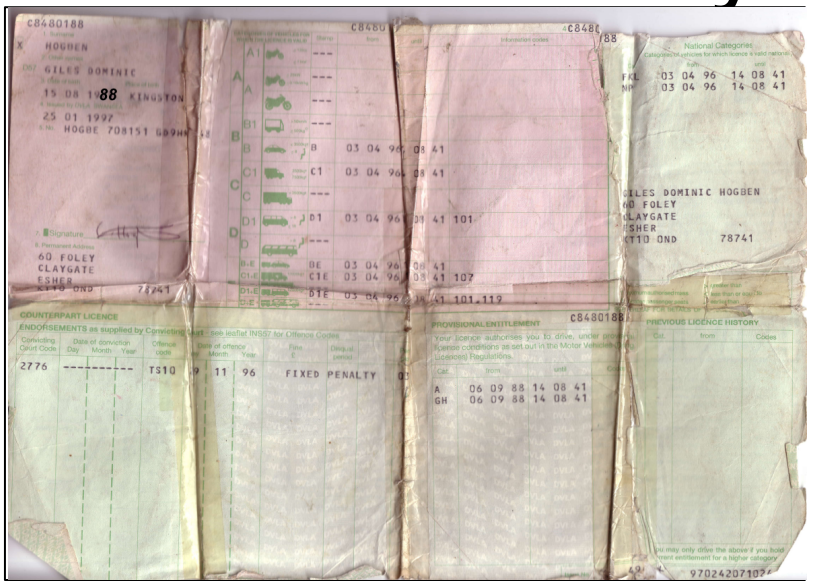
Evidence

Why you should believe what is claimed (AKA Credentials/Proofs)

Audit Trail

What happened to your claims after you made them => (ENFORCEMENT, PURPOSE LIMITATION)

Identity Management: the Old Way



- Assertions

  "I am 18"

- Evidence

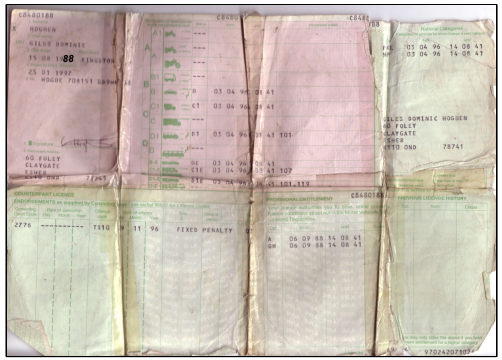
  Because my (slightly smelly) token c8480188 says so.

  (and here's all the other data on my drivers' licence too – why not take a look, while you're here...)

- **Audit Trail**

  Your caution for drink driving which led to a prison sentence was deleted as of 2005 (NOT REALLY)

Why this is bad – user/legal requirements



- Assertions

  "I am 18"

- Evidence

  Because my token c8480188 says so

  Audit Trail

  Your caution for drink d[illegible] prison sentence was deleted as of 2005 (NOT REALLY)

You didn't need to know that – only that I was over 17

What is **c8480188**?
Can I trust it?

I just wanted it deleted, I didn't want it announced to the world at the same time

Requirements -Inference management

- I am a holder of UK drivers' licence => I am over 17
- My first name prefix is miss => my civil status is "unmarried"
- User holds SwissPassport => User nationality Swiss

- Decision engines need to understand inferences from data release.

Business Requirements

- IDM policies for complex relational DB's (not just flat tables).
- Communicate IDM policies to other businesses with different data models.
- Describe reputation.
- Automated handling of evidence (rule systems).

Solutions

PRIME Identity Management Semantics

- **Assertions**

  Someone who submitted their data is over 18

- **Evidence**

  I can prove this using an OECD Government certified Electronic ID card

(*BTW I know that the use of this card also gives away something about my nationality*)

- **Audit Trail**

  Assertion ID 5021312 was used for marketing
  Assertion ID 5021312 was deleted

Components to achieve this 1

- **Assertion/Request language**
  - Minimizeable assertions (>18 not age)
  - Matching over inferences

    (Possession of **UK Drivers licence => age >17**)
  - Talking about data without revealing its semantics

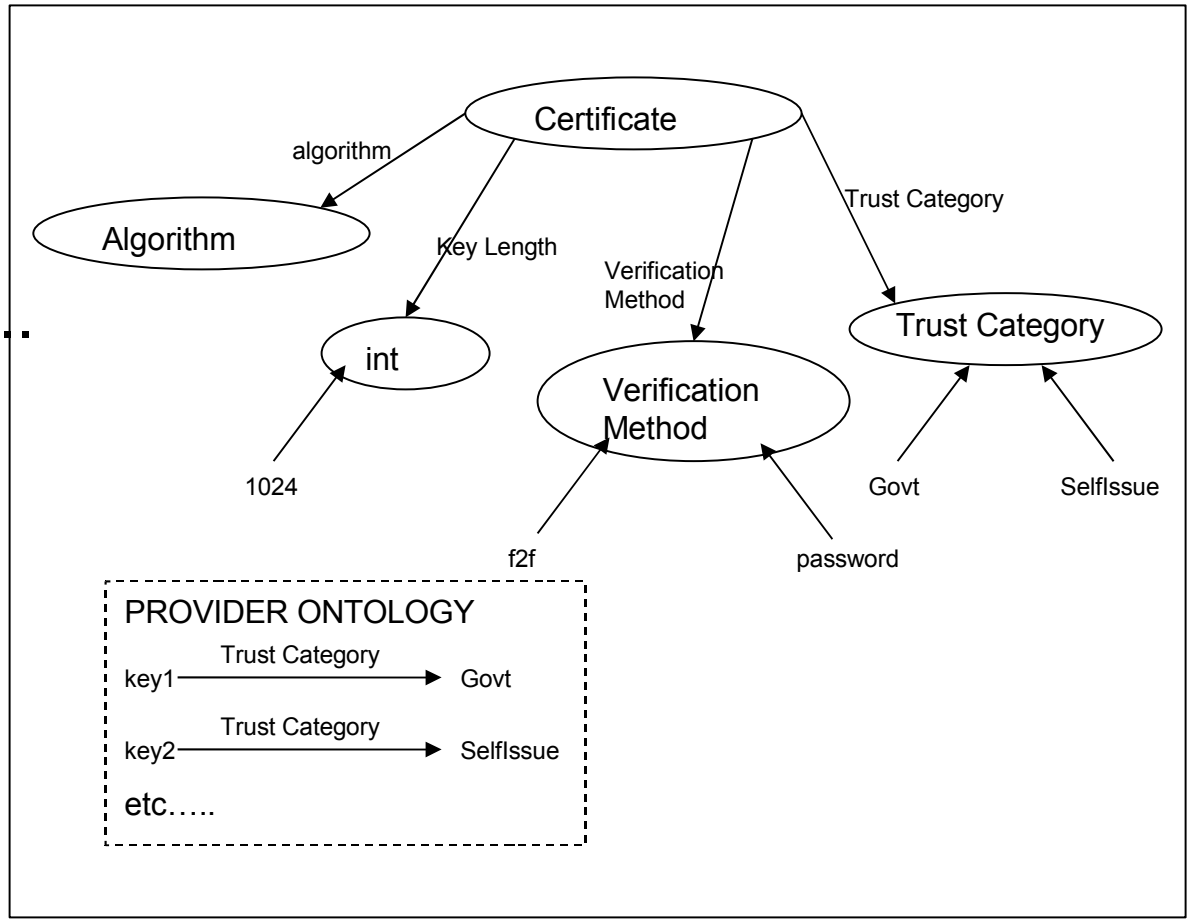
    (NOT – we deleted your AIDS test result)

Components to Achieve this 2

- **Certification language**
  - User-friendly semantics for Evidence.
  - E.g. for certificates - describe properties of Certificates, which are relevant to trust decisions.
  - Security model for attribution of properties to evidence

Evidence/Certification Ontology

- Not just string-matching but properties
  - OECD
  - Idemix/DSA etc…



Certification Ontology, top level

Provider Ontology

- Map of which certificates hold which properties

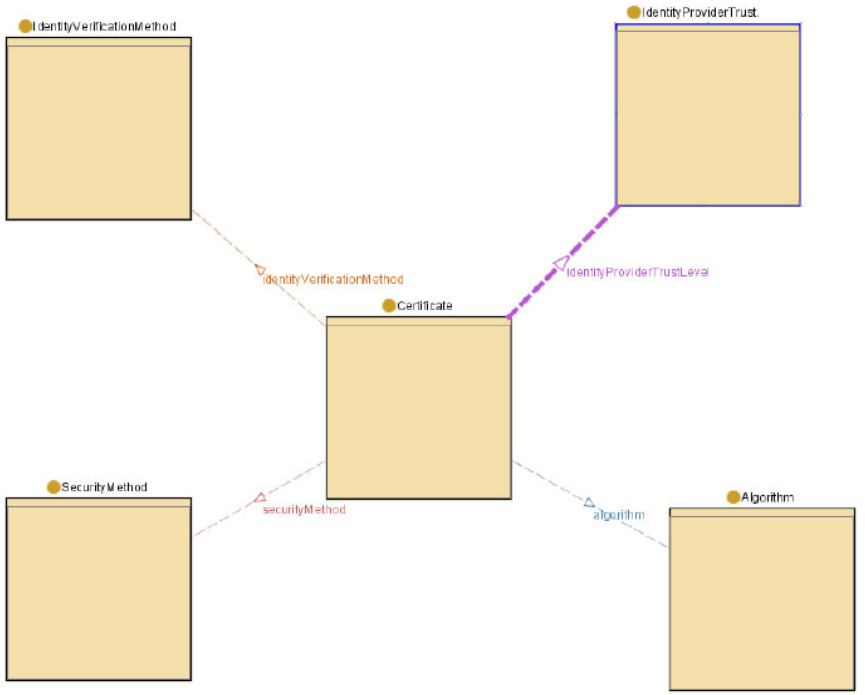
PROVIDER ONTOLOGY

key1 —Trust Category→ Govt

key2 —Trust Category→ SelfIssue

etc…..

Certification Ontology, top level



Ontology top level

Security Model

- Map of certification properties is Extremely Vulnerable (similar to Common Criteria PP)
- Who certifies the trust properties of a certificate?

Other Benefits of Evidence Ontology

- Factoring out the trust => extensibility to other types of trust
  - Reputation
  - X509, Idemix ….
  - TCG (some work on a trusted platform ontology done in PRIME)
- User friendly and consistent management of trust attributes.

Technology to implement this - Idemix

- Provides evidence for truly minimizeable assertions
- Can prove a space of queries rather than fixed attributes.
- Provides true unlinkability

Technology to implement this - RDF

- Minimizeable assertions
- Data about data without referring to semantics
- Interoperate with other data models using OWL
- SAML assertions with variables can be mapped?

Technology to implement this - OWL

- Provides inferences
- Harmonizes heterogeneous data models
- Describe context based user-models
- User-friendly certification model
- Isolate IDM layer from Business Logic layer. E.g. obligations can be applied to PRIME types, not enterprise types.

Technology to implement this - SPARQL

- Query RDF
- Query inferences
- Existing standard

Summary

- PRIME Model provides semantics for Assertions + Evidence
- RDF allows N-ary assertions, which is good for privacy and business.
- PKI certs need an abstraction layer
- PKI certs aren’t the only form of evidence
- Idemix + RDF + OWL + SPARQL can solve these problems.

Reference

- For complete architecture

See IBM Report number RZ 3674

Out end of October