Knowledge Management for Electronic Health Information Security Management - Towards a Standard Approach

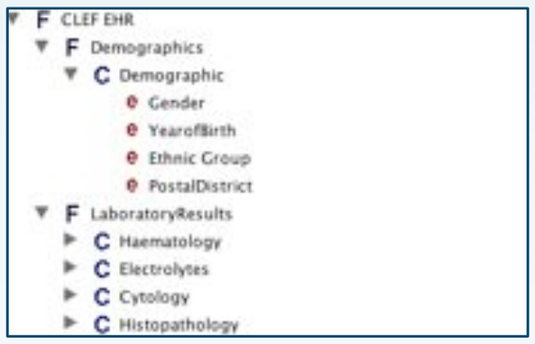
Nathan Lea
Centre for Health Informatics and Multiprofessional Education
UCL

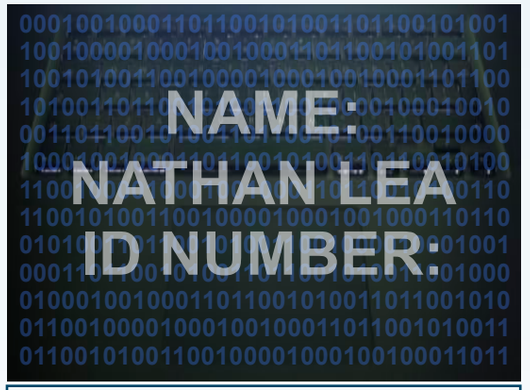
ETSI 6th Security Workshop
19th January 2010

Sharing Healthcare Information

Health care data are progressively being collected in a shareable format according to internationally recognised Electronic Healthcare Record (EHR) standards to provide a fuller picture of a person's health for all the professionals responsible for their care. The data could now be more easily shared to support biomedical research, clinical trials recruitment and disease surveillance.

Data Capture and Sharing Overview





Clinical data is collected from an individual during routine care delivery

→

The data is put into a structured format according to EHR standards

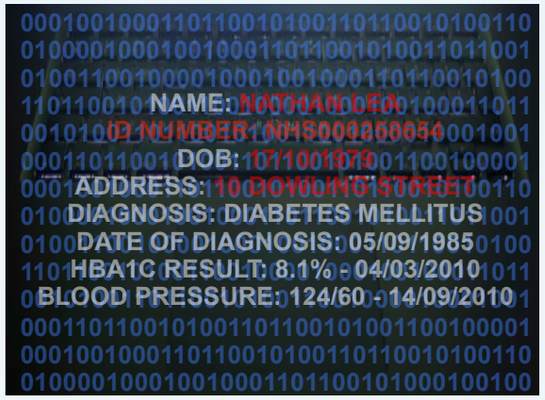
→

The data is stored in a repository to support care and other purposes

Information Security Challenges and Solutions

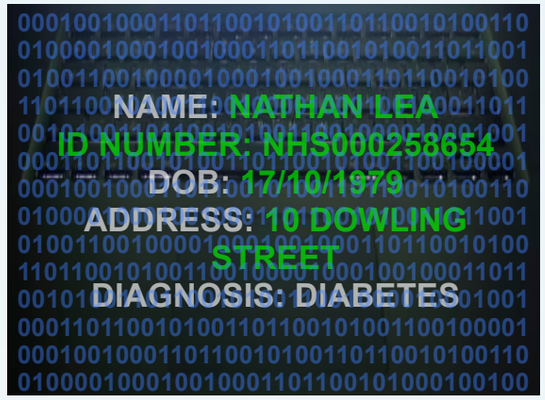
- Data protection legislation requires the sharing of identifiable data only for purposes for which they were originally collected.
- Legal mechanisms can permit the sharing of these data for other purposes, sometimes without consent, provided that the identity of the data subject can be protected and the data rendered anonymous.
- The reuse of the data must be in the public interest, or subject to ethical approval.
- The data must also be de-identified to assure anonymity and the protection of the individual.
- De-identification is achieved using anonymisation, pseudonymisation and query result aggregation to provide a non identifying picture of the relevant data for those purposes.

Example of Some Protection Methods



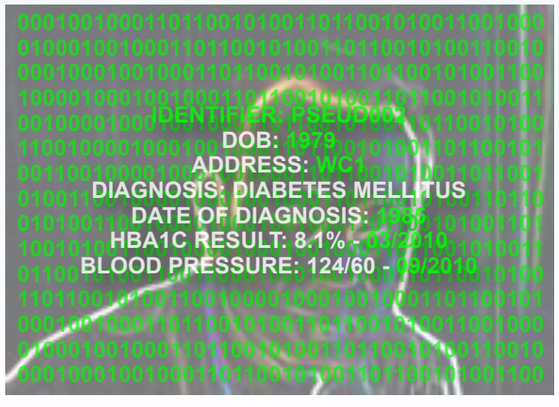
0001001000110110010100110110010100110
0011011001010011001
0001000100100011011

ANONYMISATION removes all identifying attributes. This arguably offers the greatest protection against individual identification, but renders the data less useful for other purposes.



0001001000110110010100110110010100110
0011011001010011001
0001000100100011011

PSEUDONYMISATION obscures data items, limiting the possibility of identifying an individual, including releasing only years of birth. It provides less assurance than anonymisation but more useful data.



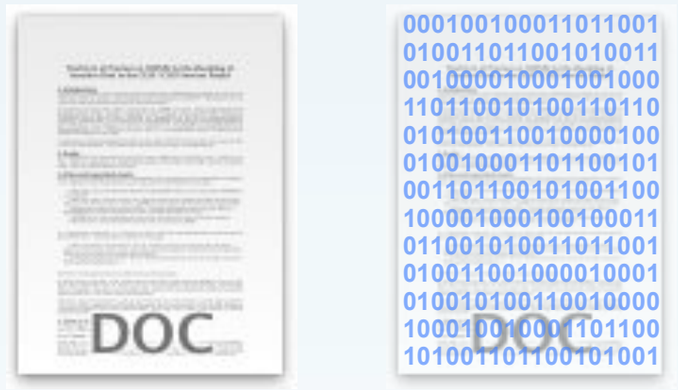
0001001000110110010100110110010100110
0011011001010011001
0001000100100011011

RESULT AGGREGATION ensures that only numeric or graphical results are returned for a query across a dataset containing thousands of records. No actual data about individuals is released.

Issues with this Approach

The methods for de-identification are currently applied manually by data handlers after the guidelines and stipulations have been assessed and interpreted.

Issues with Current Practice



Data release controls are specified by policies in narrative

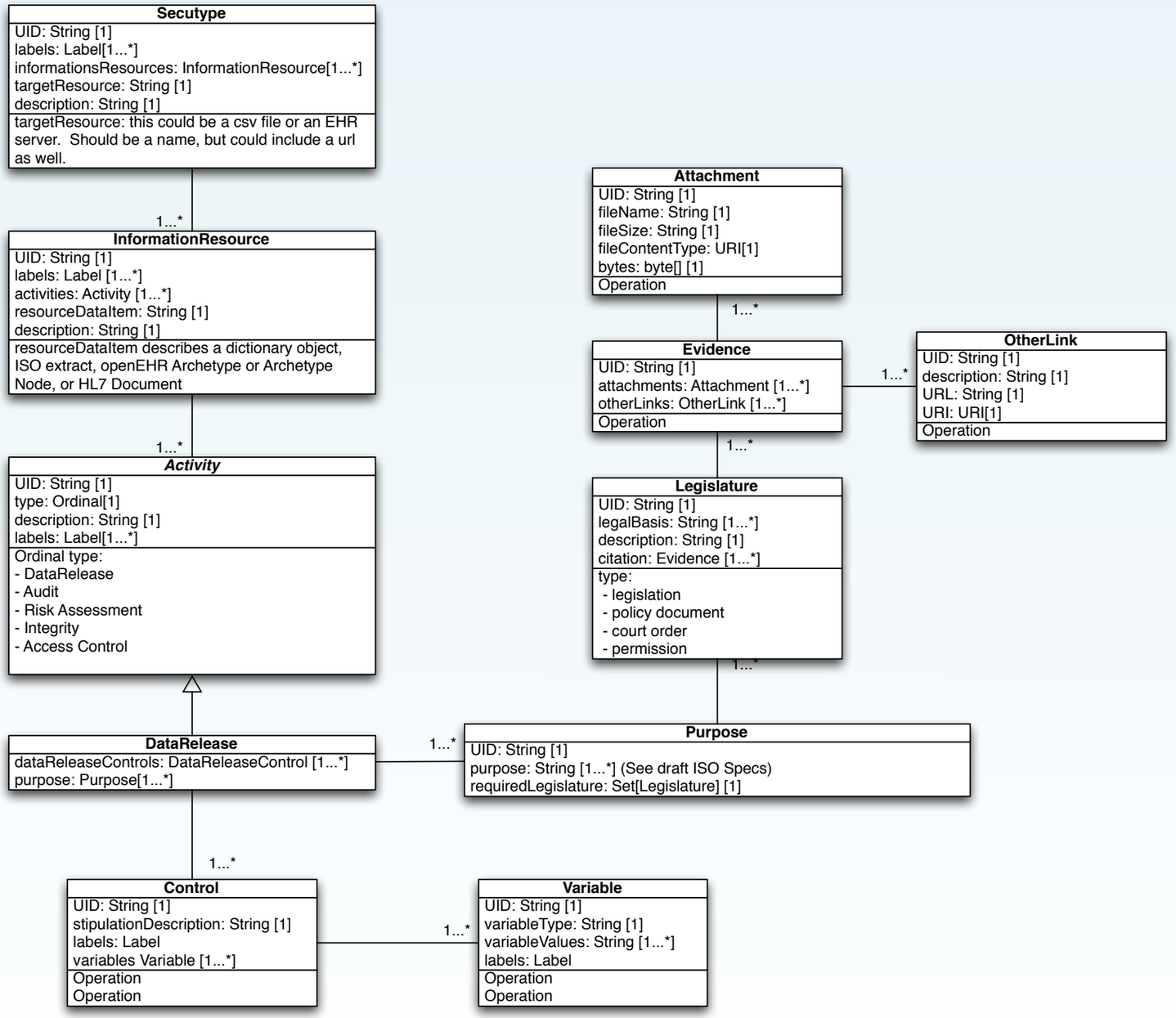
Narratives must be interpreted and manual methods devised to apply controls computationally

Knowledge Management as a Solution

CHIME is researching the use of knowledge management to capture these policy details and automatically apply them when data is shared.

A new formalism called the *Secutype* has been established to apply protection measures in a consistent and automated way.

The Secutype



Demo!

A demo will now be presented (please contact me if you would like details about this as it was given on the day).

Related links - for interest

- Example of de-identification algorithms: http://www.biomedcentral.com/1472-6947/8/32
- The National Information Governance Board for Health and Social Care: http://www.nigb.nhs.uk/
- National Research Ethics Service: http://www.nres.npsa.nhs.uk/
- Health Ethics - Centre of Research Ethics Campaign: http://www.corec.org.uk/

Thank You!

Nathan Lea
Centre for Health Informatics and Multiprofessional Education
UCL
n.lea@ucl.ac.uk
http://www.chime.ucl.ac.uk/~rmhincl