# National Committee on Vital and Health Statistics (NCVHS)

## Ad Hoc Work Group on Secondary Uses of Health Data

# Data Sources and Protections

Glen F. Marshall

July 18, 2007

# Before We Go Any Further...

- The Risk Management Context
  - What data assets are at risk?
  - Who are the stakeholders in the data?
  - What exactly is at stake?
  - What do policies require?
  - What threats exist?
    - What happens if each threat is realized?
  - What controls already exist?

# Data Assets
## (generally)

- Data obtained on a variety of healthcare subjects for the purposes of...
  - Quality measurement
  - Quality improvement
  - Population health assessment
  - Population health improvement
  - Clinical research

# Stakeholders

(a sampling – There may be others.)

- Healthcare subjects
- Healthcare providers who collect data
- Healthcare data repositories
- Consumers of the data
- Public beneficiaries

Every one of these has risks and benefits!

# Policies

## (a sampling -- Do we have enough?)

- HIPAA Privacy Regulations (45 CFR § 160 and 164 Part E)
- HIPAA Security Regulations (45 CFR § 160 and 164 Part C)
- Confidentiality of Alcohol and Drug Abuse Patient Records (42 CFR Part 2)
- Family Education Rights and Privacy Act (FERPA)
- Privacy Act of 1974
- Right to Financial Privacy Act (1978)
- Privacy Protection Act of 1980
- Electronic Communications Privacy Act (1986)
- Communications Assistance for Law Enforcement Act of 1994
- Telecommunications Act of 1996
- Financial Modernization Act (Gramm-Leach-Bliley Act) (2000)
- Emergency Supplemental Appropriations Act for Defense, the Global War on Terror and Tsunami Relief (Real ID Act) (2005)
- State and local laws & regulations
- Individual patient consents

# Threats
## (a sampling of possible ones)

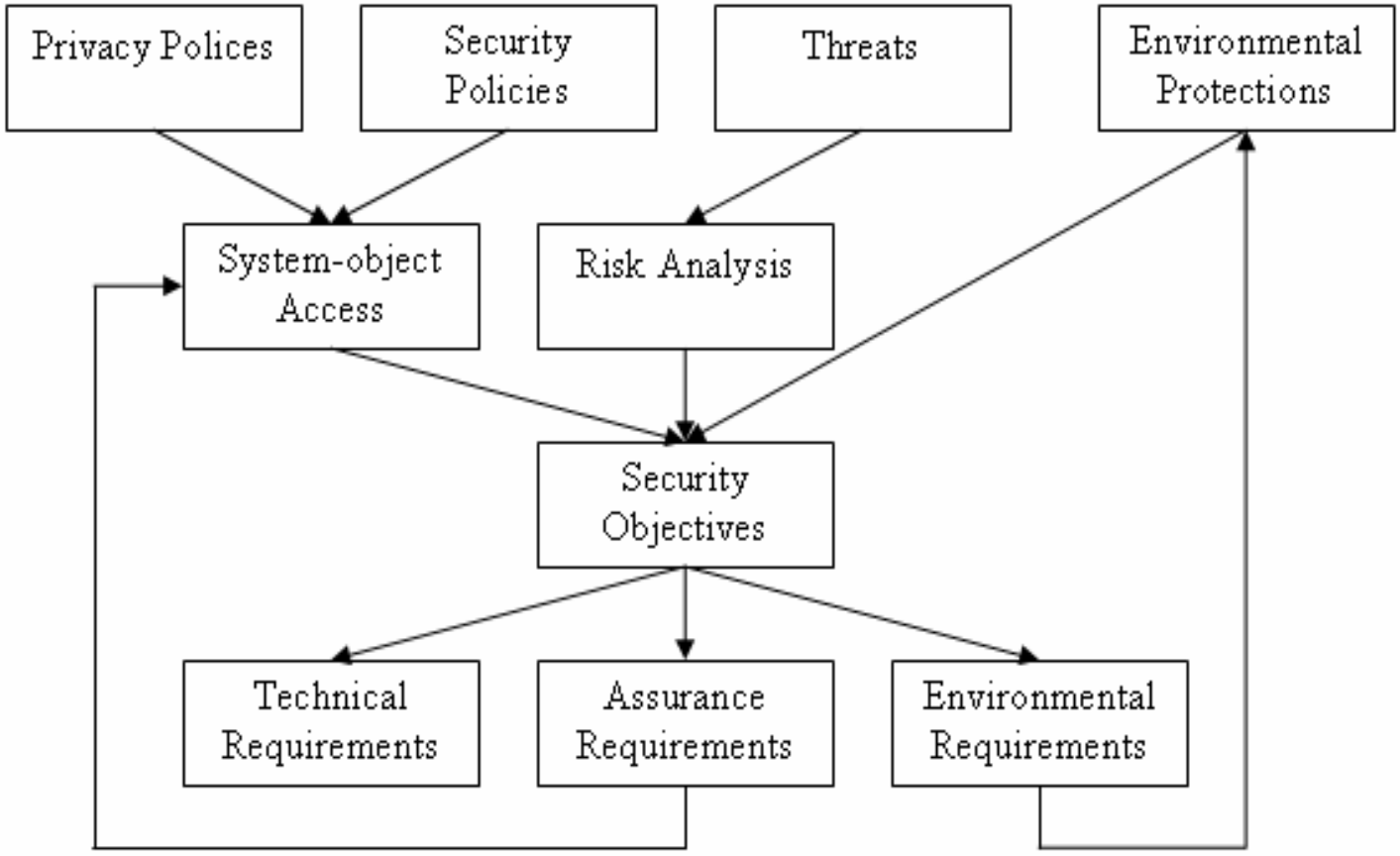
- Loss of data confidentiality
- Loss of data integrity
- Loss of data
- Loss of data collectors
- Loss of data repositories
- Loss of funding for threat mitigation
- Side-effects from threat mitigation
- Terminology overload confusion
- Human ignorance

# Existing Controls

(a sampling)

- Manual protection procedures
- Education for stakeholders
- Physical and network protections
- Penalties for privacy & security violations
- Insurance
- Localized/specialized controls

# A Risk Management Process




From the HITSP Security & Privacy Technical Committee

# Before We Go Any Further...

- Unless risk management is in-place, all proposed controls are speculative and may...
  - Fail to conform with applicable policies
  - Fail to protect against threats
  - Duplicate or conflict with existing controls
  - Incur unreasonable costs
  - Provide no real assurance

# Proposed Controls

- Accuracy for data at rest
  - Using standardized data sets
    - HL7 CDA, in particular
  - Using standardized vocabulary
    - Evolve/mature/harmonize current terminology
  - Providing assurance to data subjects
    - Consents and confidentiality controls
  - Providing incentives for data sources
  - Educating data subjects and sources

# Proposed Controls

- Integrity for data at rest
  - Hashing
    - SHA-256
  - Providing assurance to data subjects
    - Consents and confidentiality controls
  - Providing incentives for data repositories
  - Educating data stewards

# Proposed Controls

- Availability of data at rest
  - Using standardized data sets
    - HL7 CDA, in particular
  - Recruiting data subjects and sources
  - Providing assurance to data subjects
    - Consents and confidentiality controls
  - Providing incentives for data repositories
  - Educating data stewards

# Proposed Controls

- Informed consent from data subjects
  - Standard form for all consents
  - Simple language, in subject's native tongue
  - Verbal and written form
  - Limited authorization for
    - Explicit purpose
    - Explicit duration
  - Explicit accountability
    - For data source, repositories, and users
  - No duress

# Proposed Controls

- Confidentiality for data at rest
  - Standard anonymized de-identification
    Permanent redaction of identifying data to provide assurance of confidentiality for the subject of healthcare information. Re-identification is highly unlikely.
  - Standard pseudonymized de-identification
    Substituting identifying data to provide assurance of confidentiality for the subject of healthcare information. Re-identification may occur in a predetermined manner.

# Proposed Controls

- Aggregation of data at rest
  - Permanent deletion of individual healthcare data, but...
    - A small aggregation group size exposes risk of implicit re-identification.
    - Aggregation exposes a risk to data availability if aggregate information validity is questioned.
  - Only useful for the purposes of aggregation-derived information.

# Questions?