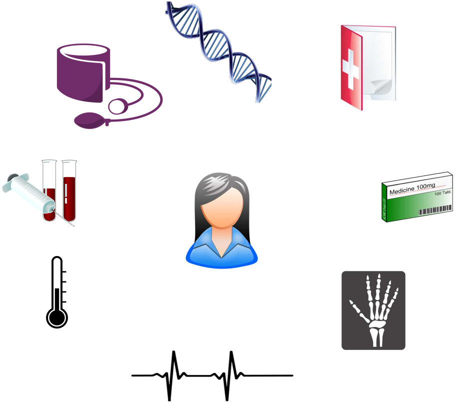
Security and Privacy of Genomic Data

Sara Foresti
Dipartimento di Informatica
Università degli Studi di Milano
sara.foresti@unimi.it

Challenges in Data-Driven Genomic Computing
Como, Villa del Grumello – March 8, 2019

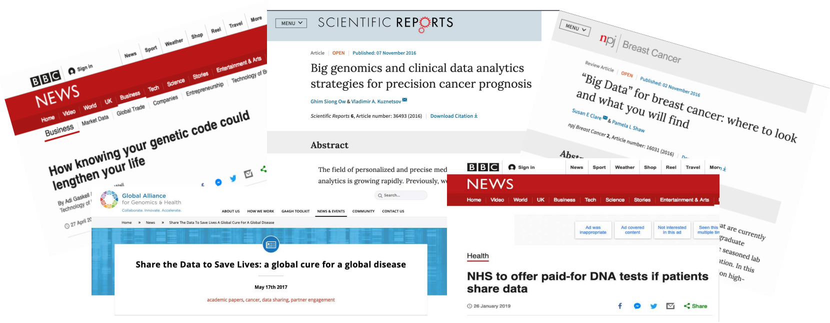
Genomic data (1)

- Large collections generated thanks to reduction of sequencing costs
- Highly related with personal and medical data



Genomic data (2)

- Advantages for research
  - + data analysis for extracting valuable information
  - + sharing for collaborative computation



Business

How knowing your genetic code could lengthen your life

By Adi Gaskell
Technology of

27 April 20

Article | OPEN | Published: 07 November 2016

Big genomics and clinical data analytics strategies for precision cancer prognosis

Ghim Siong Ow & Vladimir A. Kuznetsov

Scientific Reports **6**, Article number: 36493 (2016) | Download Citation

Abstract

The field of personalized and precise med
analytics is growing rapidly. Previously, we

Review Article | OPEN | Published: 02 November 2016

“Big Data” for breast cancer: where to look and what you will find

Susan E Clare & Pamela L Shaw

npj Breast Cancer **2**, Article number: 16031 (2016)

Abst

at are currently
graduate
seasoned lab
ation. In this
on high-

Home » News » Share The Data To Save Lives A Global Cure For A Global Disease

Share the Data to Save Lives: a global cure for a global disease

May 17th 2017

academic papers, cancer, data sharing, partner engagement

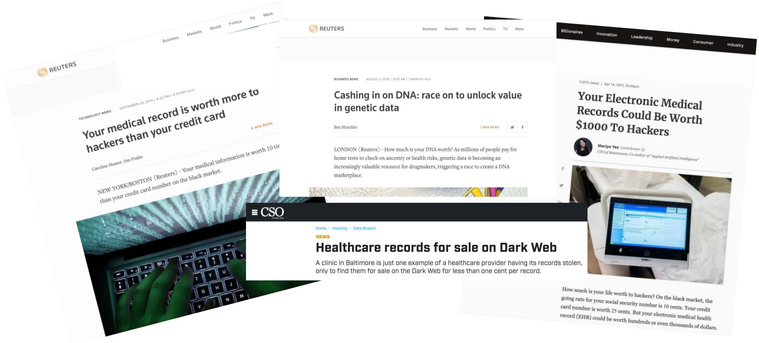
Health

NHS to offer paid-for DNA tests if patients share data

26 January 2019

Disclosure risks

- Considerable economic value
- Highly sensitive content



TECHNOLOGY NEWS SEPTEMBER 24, 2014 / 8:25 PM / 4 YEARS AGO

Your medical record is worth more to hackers than your credit card

6 MIN READ

Caroline Humer, Jim Finkle

NEW YORK/BOSTON (Reuters) - Your medical information is worth 10 tir than your credit card number on the black market.

BUSINESS NEWS AUGUST 3, 2018 / 10:01 AM / 7 MONTHS AGO

Cashing in on DNA: race on to unlock value in genetic data

Ben Hirschler 7 MIN READ

LONDON (Reuters) - How much is your DNA worth? As millions of people pay for home tests to check on ancestry or health risks, genetic data is becoming an increasingly valuable resource for drugmakers, triggering a race to create a DNA marketplace.

11,972 views | Apr 14, 2017, 10:05pm

Your Electronic Medical Records Could Be Worth $1000 To Hackers

Mariya Yao Contributor
CTO of Metamaven, Co-Author of "Applied Artificial Intelligence"

How much is your life worth to hackers? On the black market, the going rate for your social security number is 10 cents. Your credit card number is worth 25 cents. But your electronic medical health record (EHR) could be worth hundreds or even thousands of dollars.

Home > Hacking > Data Breach

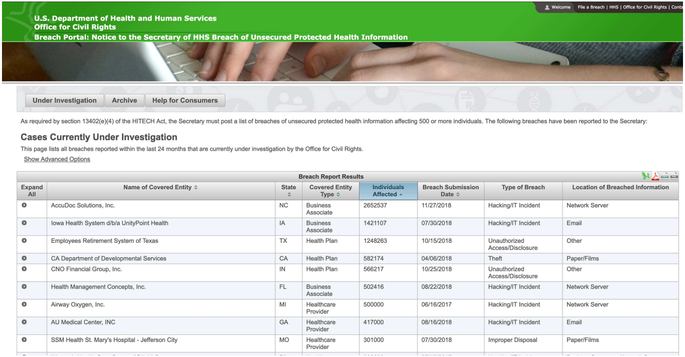
NEWS

Healthcare records for sale on Dark Web

A clinic in Baltimore is just one example of a healthcare provider having its records stolen, only to find them for sale on the Dark Web for less than one cent per record.

Disclosure risks

- Considerable economic value
- Highly sensitive content
- High risk of exposure is case of attacks



U.S. Department of Health and Human Services
Office for Civil Rights
Breach Portal: Notice to the Secretary of HHS Breach of Unsecured Protected Health Information

As required by section 13402(e)(4) of the HITECH Act, the Secretary must post a list of breaches of unsecured protected health information affecting 500 or more individuals. The following breaches have been reported to the Secretary:

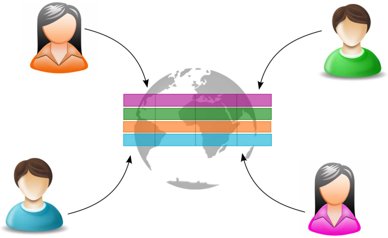
Cases Currently Under Investigation

This page lists all breaches reported within the last 24 months that are currently under investigation by the Office for Civil Rights.

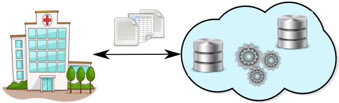
Breach Report Results

Expand All	Name of Covered Entity	State	Covered Entity Type	Individuals Affected	Breach Submission Date	Type of Breach	Location of Breached Information
●	AccuDoc Solutions, Inc.	NC	Business Associate	2652537	11/27/2018	Hacking/IT Incident	Network Server
●	Iowa Health System d/b/a UnityPoint Health	IA	Business Associate	1421107	07/30/2018	Hacking/IT Incident	Email
●	Employees Retirement System of Texas	TX	Health Plan	1248263	10/15/2018	Unauthorized Access/Disclosure	Other
●	CA Department of Developmental Services	CA	Health Plan	582174	04/06/2018	Theft	Paper/Films
●	CNO Financial Group, Inc.	IN	Health Plan	566217	10/25/2018	Unauthorized Access/Disclosure	Other
●	Health Management Concepts, Inc.	FL	Business Associate	502416	08/22/2018	Hacking/IT Incident	Network Server
●	Airway Oxygen, Inc.	MI	Healthcare Provider	500000	06/16/2017	Hacking/IT Incident	Network Server
●	AU Medical Center, INC	GA	Healthcare Provider	417000	08/16/2018	Hacking/IT Incident	Email
●	SSM Health St. Mary's Hospital - Jefferson City	MO	Healthcare Provider	301000	07/30/2018	Improper Disposal	Paper/Films

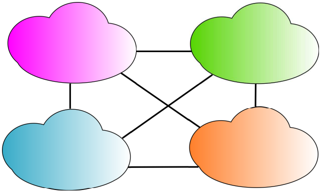
Security: a complex problem



Publication



Outsourcing

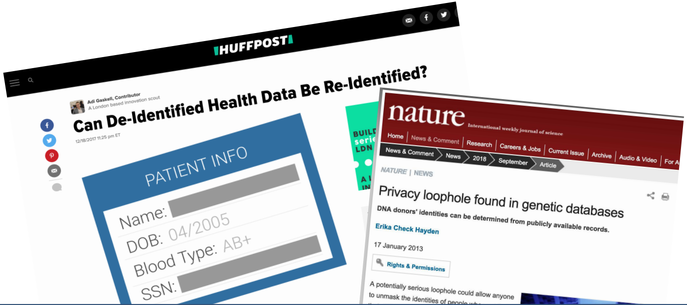


Sharing

Regulations

Data protection – Publication

- Minimize release/exposure
  - correlation among different data sources
  - indirect exposure of sensitive information
  - de-identification ≠ anonymization
  - privacy vs utility



HUFFPOST

Adi Gaskell, Contributor
A London based innovation scout

Can De-Identified Health Data Be Re-Identified?

12/18/2017 11:25 pm ET

PATIENT INFO

Name: [redacted]
DOB: 04/2005
Blood Type: AB+
SSN: [redacted]

nature International weekly journal of science

Home | News & Comment | Research | Careers & Jobs | Current Issue | Archive | Audio & Video | For A

News & Comment > News > 2018 > September > Article

NATURE | NEWS

Privacy loophole found in genetic databases

DNA donors' identities can be determined from publicly available records.

Erika Check Hayden

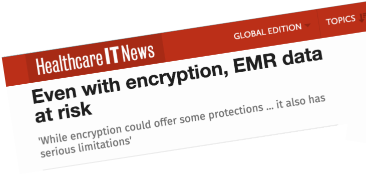
17 January 2013

Rights & Permissions

A potentially serious loophole could allow anyone to unmask the identities of people who

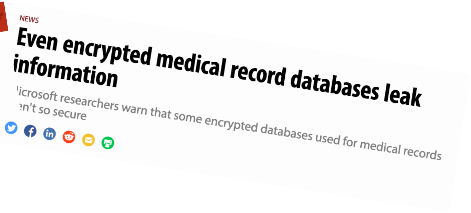
Data protection – Outsourcing

- Encryption protects data confidentiality but
  - limits functionality
  - indirect exposure



Even with encryption, EMR data at risk

'While encryption could offer some protections ... it also has serious limitations'

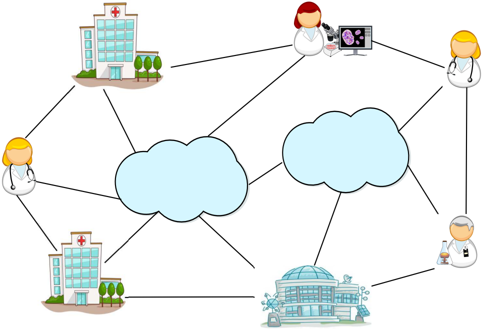


Even encrypted medical record databases leak information

'icrosoft researchers warn that some encrypted databases used for medical records ɜn't so secure

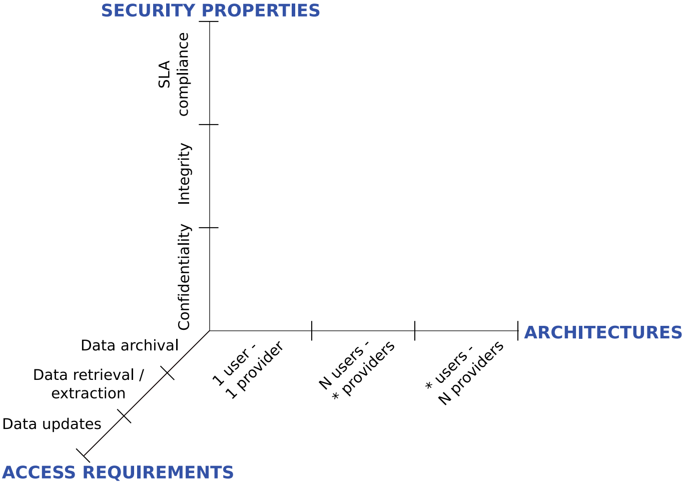
Data protection – Sharing

- Scientific research demands for data sharing
  - + combine data collections owned by different subjects
  - + enables collaboration
  - – requires controlled data release



Characterization of Data Protection Challenges

Scientific and technical challenges



Security properties

Confidentiality

- data externally stored
- users identities
- actions that users perform on the data

Integrity

- data externally stored
- computation and query results

SLA compliance

- assurance and certification

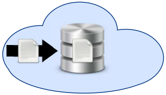
Access requirements

Data archival

- upload/download
- protection of data in storage

Data retrieval/extraction

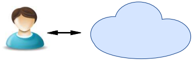
- support for fine-grained data retrieval and queries
- protection of computations and query results



Data update

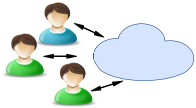
- support for access retrieval and enforcement of updates
- protection of the actions and of their effects on the data

Architectures



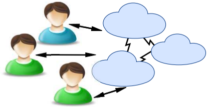
1 user - 1 provider

- protection of data at rest
- fine-grained retrieval
- query privacy/integrity



n users - * providers

- authorizations and access control
- multiple writers



*** users - n providers**

- controlled data sharing and computation

Data Sharing:
Issues and Directions

Today...

- Two extreme solutions
  - Share everything
    - + enables collaboration
    - – requires full trust
  - Share nothing
    - + guarantees privacy
    - – slows scientific research

...Tomorrow

- Selective sharing based on: receiving subject, data sensitivity, context, purpose, . . .
- Requires to study solutions enabling to:
  - identify sensitive data
  - express access restrictions through a simple while flexible language
  - protect data (e.g., encryption, aggregation, obfuscation)

GMQL for enabling sharing

- GMQL is an expressive and flexible query language for genomic data
- Could be extended to:
    - associate protection requirements with the data
    - specify access restrictions
    - enable the enforcement of protection techniques

Conclusions

- Data collection and analysis are vital for scientific research
- Solutions that guarantee data protection are enabling for:
  - data publication
  - outsourcing of data storage and/or computation
  - data sharing and collaborative computations and
  - ...