

# Ochrona poufności cyfrowych danych medycznych

Problemy i rozwiązania

**Interdyscyplinarne Seminarium Przetwarzania,  
Analizy i Interpretacji Obrazów w Medycynie**

5 grudnia 2008

**Bartosz Borucki**

Interdyscyplinarne Centrum Modelowania  
Matematycznego i Komputerowego  
Uniwersytet Warszawski



# Plan prezentacji

- Dane medyczne a dane osobowe
- Konieczność ochrony
- Prawo
- Dane identyfikacyjne
- Anonimizacja
- DICOM
- Systemy bezpieczeństwa



- **Dane medyczne a dane osobowe**
- Konieczność ochrony
- Prawo
- Dane identyfikacyjne
- Anonimizacja
- DICOM
- Systemy bezpieczeństwa

# Dane medyczne a dane osobowe

- Dane osobowe – informacje pozwalające w sposób jednoznaczny zidentyfikować daną osobę
  - Zdefiniowane przez ustawę
- Dane medyczne – informacje o stanie zdrowia osoby – pacjenta
  - Brak prawnej definicji
- Medyczne dane osobowe
  - Przykład danych osobowych
    - Dane identyfikacyjne + dane medyczne
- Cyfrowe dane medyczne = Elektroniczna dokumentacja medyczna – dane medyczne przechowywane w postaci dokumentacji elektronicznej
  - Zdefiniowane w prawie



# Plan prezentacji

- Dane medyczne a dane osobowe
- **Konieczność ochrony**
- Prawo
- Dane identyfikacyjne
- Anonimizacja
- DICOM
- Systemy bezpieczeństwa



# Konieczność ochrony

- Po co chronić dane osobowe?
- Prywatność
  - Prawo zachowania prywatności – prawo do uniknięcia niechcianych i/lub nieprzyjaznych ingerencji osób trzecich lub instytucji w sferę życia prywatnego osoby
- Stopniowanie ryzyka
  - Imię i nazwisko
  - PESEL, adres zamieszkania
  - Dane ubezpieczeniowe, bankowe, ....
  - Dane medyczne
- Pojęcie danych wrażliwych



# Konieczność ochrony

- Po co chronić medyczne dane osobowe?
- PHI (Protected Health Information)
- Sprzeczność celów
  - Oczekiwania instytucjonalne – najszerszy możliwy dostęp do danych
  - Oczekiwania indywidualne – minimalizacja rozprzestrzeniania się danych
- Konieczność zapewnienia ochrony z uwzględnieniem zachowania równowagi
  - Które dane są identyfikujące?
  - Jakie metody i mechanizmy?
- Bezpieczeństwo vs. poufność

- Elementy ryzyka
  - **Ryzyko utraty prywatności** – ingerencja w życie prywatne pacjenta
  - **Ryzyko utraty bezpieczeństwa** – nieautoryzowany dostęp do danych
  - **Ryzyko utraty poufności** – nieautoryzowane udostępnienie danych

# Plan prezentacji

- Dane medyczne a dane osobowe
- Konieczność ochrony
- **Prawo**
- Dane identyfikacyjne
- Anonimizacja
- DICOM
- Systemy bezpieczeństwa

- Konstytucja RP 1997

(art. 47.)

Każdy ma prawo do ochrony życia prywatnego, rodzinnego, czci i dobrego imienia oraz do decydowania o swoim życiu osobistym.

(art. 51.)

1. Nikt nie może być obowiązany inaczej niż na podstawie ustawy do ujawniania informacji dotyczących jego osoby.
2. Władze publiczne nie mogą pozyskiwać, gromadzić i udostępniać innych informacji o obywatelach niż niezbędne w demokratycznym państwie prawnym.
3. Każdy ma prawo dostępu do dotyczących go urzędowych dokumentów i zbiorów danych. Ograniczenie tego prawa może określić ustawa.
4. Każdy ma prawo do żądania sprostowania oraz usunięcia informacji nieprawdziwych, niepełnych lub zebranych w sposób sprzeczny z ustawą.
5. Zasady i tryb gromadzenia oraz udostępniania informacji określa ustawa.

- Ustawa z dnia 29 sierpnia 1997 r. o ochronie danych osobowych  
(Dz.U. 1997 Nr 133 poz. 883)
  - Poprawki z roku 2001 – efekt wdrożenia dyrektywy UE 95/46/WE
  - Definicja danych osobowych  
(art. 6.)
    1. W rozumieniu ustawy za dane osobowe uważa się wszelkie informacje dotyczące zidentyfikowanej lub możliwej do zidentyfikowania osoby fizycznej.
    2. Osobą możliwą do zidentyfikowania jest osoba, której tożsamość można określić bezpośrednio lub pośrednio, w szczególności przez powołanie się na numer identyfikacyjny albo jeden lub kilka specyficznych czynników określających jej cechy fizyczne, fizjologiczne, umysłowe, ekonomiczne, kulturowe lub społeczne.
    3. Informacji nie uważa się za umożliwiającą określenie tożsamości osoby, jeżeli wymagałoby to nadmiernych kosztów
  - Dane medyczne  
(art. 27.)
    1. **Zabrania się przetwarzania** danych ujawniających pochodzenie rasowe lub etniczne, poglądy polityczne, przekonania religijne lub filozoficzne, przynależność wyznaniową, partyjną lub związkową, jak również **danych o stanie zdrowia, kodzie genetycznym, nałogach lub życiu seksualnym** oraz danych dotyczących skazań, orzeczeń o ukaraniu i mandatów karnych, a także innych orzeczeń wydanych w postępowaniu sądowym lub administracyjnym

- Dokumentacja elektroniczna

- Rozporządzenie Ministra Spraw Wewnętrznych i Administracji z dnia 29 kwietnia 2004 w sprawie dokumentacji przetwarzania danych osobowych oraz warunków technicznych i organizacyjnych, jakim powinny odpowiadać urządzenia i systemy informatyczne służące do przetwarzania danych osobowych (Dz.U. 2004 r. Nr 100 poz. 1024)

- Dokumentacja medyczna (w tym elektroniczna)

- Art. 27. Ustawy z dnia 29 sierpnia 1997 r. o ochronie danych osobowych
- Art. 6. Konwencji Nr 108 Rady Europy
- Art. 8. Dyrektywy 95/46/WE
- Dział VIII ustawy z dnia 27 sierpnia 2004 roku o świadczeniach opieki zdrowotnej finansowanych ze środków publicznych
- Art. 18. ustawy z dnia 30 sierpnia 1991 roku o zakładach opieki zdrowotnej
- Art. 41. ustawy z dnia 5 grudnia 1996 roku o zawodach lekarza i lekarza dentysty
- Rozporządzenie Ministra Zdrowia w sprawie zakresu niezbędnych informacji gromadzonych przez świadczeniodawców, szczegółowego sposobu rejestrowania tych informacji oraz ich przekazywania podmiotom zobowiązanym do finansowania świadczeń ze środków publicznych
- Rozporządzenie Ministra Zdrowia w sprawie zakresu niezbędnych informacji gromadzonych i przekazywanych przez apteki podmiotom zobowiązanym do finansowania świadczeń ze środków publicznych
- Rozporządzenie Ministra Zdrowia w sprawie rodzajów i zakresu dokumentacji medycznej w zakładach opieki zdrowotnej oraz sposobu jej przetwarzania

- Rozporządzenie Ministra Zdrowia w sprawie rodzajów i zakresu dokumentacji medycznej w zakładach opieki zdrowotnej oraz sposobu jej przetwarzania
  - par.52 ust.2 – Udostępnienie dokumentacji następuje w trybie zapewniającym zachowanie poufności i ochrony danych osobowych
  - par.57 ust.2 – Dokumentację prowadzoną w postaci elektronicznej udostępnia się z zachowaniem jej integralności oraz ochrony danych osobowych
  - par.60 – Dokumentację prowadzoną w postaci elektronicznej uważa się za zabezpieczoną, jeżeli w sposób ciągły są spełnione łącznie następujące warunki:
    - 1) zapewniona jest jej dostępność wyłącznie dla osób uprawnionych;
    - 2) jest chroniona przed przypadkowym lub nieuprawnionym zniszczeniem;
    - 3) są zastosowane metody i środki ochrony dokumentacji, których skuteczność w czasie ich zastosowania jest powszechnie uznawana
- HIPAA - Health Insurance Portability and Accountability Act of 1996
  - Uchwała kongresu Stanów Zjednoczonych 1996
  - Szczegółowa dokumentacja wyznaczająca reguły i zalecenia przetwarzania i udostępniania danych medycznych
  - The Privacy Rule – dokument uzupełniający HIPAA z 2002 roku – federalne prawo ochrony prywatności danych zdrowotnych

- Medyczne dane osobowe a nauka

- UODO, art. 27. ust. 2 pkt 9

*(Przetwarzanie danych, o których mowa w ust. 1, jest jednak dopuszczalne, jeżeli) jest to niezbędne do prowadzenia badań naukowych, w tym do przygotowania rozprawy wymaganej do uzyskania dyplomu ukończenia szkoły wyższej lub stopnia naukowego; publikowanie wyników badań naukowych nie może następować w sposób umożliwiający identyfikację osób, których dane zostały przetworzone*

- Udostępnianie osobowych danych medycznych w celach naukowych jest zgodne z prawem
- Jednostka naukowa
  - Przetwarza dane osobowe
  - Podlega UODO
  - Musi zapewnić ochronę i bezpieczeństwo danych
- Zaleca się jednak postępowanie zgodne z HIPAA
  - zredukowanie danych tak aby nie były danymi osobowymi

# Plan prezentacji

- Dane medyczne a dane osobowe
- Konieczność ochrony
- Prawo
- **Dane identyfikacyjne**
- Anonimizacja
- DICOM
- Systemy bezpieczeństwa



# Dane identyfikacyjne

- Kiedy dane są danymi osobowymi?
- Które informacje mogą doprowadzić do identyfikacji osoby?
  - Identyfikacja bezpośrednia – jednoznaczne zidentyfikowanie osoby wprost na podstawie posiadanych danych (np. imię, nazwisko, adres)
  - Identyfikacja pośrednia – jednoznaczne zidentyfikowanie osoby na podstawie relacji wewnątrz wszystkich posiadanych danych
  - Dane obrazowe – odrębny problem

# Dane identyfikacyjne

- Identyfikacja bezpośrednia – które pola danych niosą informacje identyfikujące?
- HIPAA
  - Imię i nazwisko – włączając aktualne, wcześniejsze oraz panieńskie nazwisko matki
  - Adres pocztowy – wraz ze wszystkimi geograficznymi regionami mniejszymi niż województwo, włączając powiat, miasto, dzielnicę, kod pocztowy i odpowiadające im kody geograficzne
  - Wszystkie elementy dat (z wyłączeniem roku) dla dat bezpośrednio powiązanych z jednostką, włączając datę urodzenia, ewentualne daty przystąpienia/odstąpienia (organizacji, leczenia itp.), datę śmierci, oraz liczby określające wiek powyżej 89 lat wraz ze wszystkimi elementami dat (włączając rok), które mogą na taki wiek wskazywać, z wyłączeniem zbiorczej kategorii wiekowej „90 i powyżej”.
  - Numery telefonów i numery faksu
  - Adresy poczty elektronicznej (e-mail)
  - Numery ubezpieczeń społecznych
  - Numery dokumentacji medycznych
  - Numery kont bankowych
  - Numery certyfikatów, licencji, uprawnień
  - Numery identyfikacyjne pojazdów, włączając numery tablic rejestracyjnych
  - Numery seryjne i identyfikacyjne urządzeń
  - Sieciowe adresy URL (Universal Resource Locator)
  - Sieciowe adresy IP (Internet Protocol)
  - Identyfikatory biometryczne, włączając odciski palców i głosu
  - Obrazy fotograficzne pełnej twarzy lub inne porównywalne obrazy
  - Inne unikalne numery identyfikacyjne, charakterystyki lub kody



# Dane identyfikacyjne

- Identyfikacja pośrednia – jak zapewnić niejednoznaczność danych?
  - Pomimo usunięcia danych identyfikacyjnych istnieje możliwość zidentyfikowania jednostek na podstawie relacji wewnątrz danych oraz z wykorzystaniem dodatkowych informacji zewnętrznych
- Każdy rekord w zadanym zbiorze danych musi być nierozróżnialny od co najmniej jednego innego rekordu
- Pola zawarte w danych nie powinny być dostępne publicznie w innych zestawach danych
- Problem: zapewnienie równowagi pomiędzy usunięciem wystarczającej ilości informacji, a zachowaniem wystarczającej ilości potrzebnych danych przenoszących informacje
- Prawo – pojęcie „nadmiernych kosztów lub działań”
- Przykład:

Nazwisko	Adres	Wiek	Diagnoza
Kowalski	Warszawa	26	Zapalenie płuc
Nowak	Kraków	42	Nadciśnienie
Iksiński	Zakopane	47	Cukrzyca
Igrekowski	Radom	28	Astma



# Dane identyfikacyjne

- Identyfikacja pośrednia – jak zapewnić niejednoznaczność danych?
  - Pomimo usunięcia danych identyfikacyjnych istnieje możliwość zidentyfikowania jednostek na podstawie relacji wewnątrz danych oraz z wykorzystaniem dodatkowych informacji zewnętrznych
- Każdy rekord w zadanym zbiorze danych musi być nierozróżnialny od co najmniej jednego innego rekordu
- Pola zawarte w danych nie powinny być dostępne publicznie w innych zestawach danych
- Problem: zapewnienie równowagi pomiędzy usunięciem wystarczającej ilości informacji, a zachowaniem wystarczającej ilości potrzebnych danych przenoszących informacje
- Prawo – pojęcie „nadmiernych kosztów lub działań”
- Przykład:

Nazwisko	Adres	Wiek	Diagnoza
***	Warszawa	26	Zapalenie płuc
***	Kraków	42	Nadciśnienie
***	Zakopane	47	Cukrzyca
***	Radom	28	Astma



- Identyfikacja na podstawie danych obrazowych
  - Zdjęcia
    - rozpoznawanie twarzy
    - porównywanie zdjęć
    - osoby publicznie znane
  - Obrazowe dane medyczne
    - rekonstrukcja 3D
    - CT/MRI głowy
    - charakterystyczna cecha
- Problem jest zauważalny, ale
  - mało istniejących metod
  - „nadmierne koszty lub działania”

# Plan prezentacji

- Dane medyczne a dane osobowe
- Konieczność ochrony
- Prawo
- Dane identyfikacyjne
- **Anonimizacja**
- DICOM
- Systemy bezpieczeństwa

- Co to jest anonimizacja?
- Zróżnicowanie pojęć w literaturze
- Znaczenie słowa „anonymous”
  1. nienazwany lub niezidentyfikowany
  2. nieznanego autorstwa lub pochodzenia
  3. nie posiadający indywidualności, niewyróżniający się, nierozpoznawalny
- 1 – usuwanie informacji identyfikacyjnych
  - **deidentyfikacja** (ang. *de-identification*)
  - depersonalizacja (ang. *de-personalization*)
  - **pseudonimizacja** (ang. *pseudonymization*)
- 3 – zapewnienie niejednoznaczności danych
  - **uniejednoznacznianie** (nadawanie niejednoznaczności, ang. *ambiguation*)

# Anonimizacja - deidentyfikacja

- Deidentyfikacja to proces usuwania z danych informacji identyfikacyjnych
  - Pola jawnie identyfikacyjne (lista)
  - Informacje zawarte w polach opisowych – tekst dowolny – szczególnie w dokumentacji medycznej
- Proces deidentyfikacji
  - Pola dobrze zdefiniowane można bezpośrednio usunąć
  - W polach dowolnych trzeba znaleźć
    - Powtarzające się informacje z pól jawnie identyfikujących
    - Inne dane identyfikacyjne (nazwiska, numery, itp.)
- Problem tekstów dowolnych
  - Jakiego rodzaju informacji szukać?
  - Jak szukać tych informacji?
  - Zależność od języka
  - Zależność od dziedziny

# Anonimizacja - deidentyfikacja

- Metody
  - Wyszukiwanie wprost
    - Wyszukiwanie informacji usuniętych z pól jawnie identyfikujących, wraz z odmianami (np. Kowalski, Kowalskiego, Kowalskiemu,...)
    - Wyszukiwanie znaczników tożsamości (Dr, Pan, Pani, p., ...)
    - Wyszukiwanie informacji geograficznych (np. adresy)
    - Wyszukiwanie numerów
  - Wyszukiwanie odwrotne
    - Klasyfikowanie poszczególnych słów jako odpowiednich części mowy i zdania
    - Klasyfikowanie sensu poszczególnych terminów specjalistycznych
    - Usuwanie pozostałych, niesklasyfikowanych informacji
    - Słowniki
    - Computational Linguistics
    - Natural Language Processing

# Anonimizacja - pseudonimizacja

- Pseudonimizacja to proces zastępowania pól identyfikacyjnych pseudonimami
- Pseudonim może być generowany dla pojedynczego pola lub dla zestawu pól
- Najistotniejsze cechy pseudonimizacji
  - Różne zestawy danych dotyczące zadanego pacjenta zawsze posiadają jednakowy pseudonim – możliwość grupowania danych pacjenta
  - Dane identyfikacyjne są zastępowane pseudonimami (kodami), a zatem zestaw danych przestaje podlegać kategorii danych osobowych
- Zadania stawiane przed algorytmami pseudonimizacji
  - Powtarzalność procesu – zawsze jednakowy pseudonim dla danego pacjenta
  - Unikalność pseudonimów – inny pseudonim dla każdego pacjenta
  - Odporność na ataki słownikowe (*brute-force*)



# Anonimizacja - pseudonimizacja

- Podział
  - Pseudonimizacja odwracalna (dwukierunkowa)
  - Pseudonimizacja nieodwracalna (jednokierunkowa)
- Metody
  - Pseudonimy losowe
    - generacja unikalnego losowego kodu
    - skojarzenie „pacjent  $\leftrightarrow$  pseudonim” zapamiętywane w postaci listy mapującej
    - odwracalność za pośrednictwem listy
    - bezpieczeństwo – lista mapująca
  - Symetryczne i niesymetryczne algorytmy szyfrujące
    - pseudonim generowany poprzez zaszyfrowanie danych identyfikujących
    - odwracalność poprzez deszyfrację
    - bezpieczeństwo – klucz szyfrujący
    - DES, AES, RSA, ...
  - Funkcje hashujące
    - pseudonim generowany w postaci nowego ciągu znaków
    - w oparciu o funkcje hashującą
    - niemożliwe skojarzenie pseudonim  $\rightarrow$  pacjent (algorytm nieodwracalny)
    - możliwe skojarzenie pacjent  $\rightarrow$  pseudonim (ponowne zakodowanie)
    - SHA, MD5, ...

# Anonimizacja - uniejednoznacznianie

- Uniejednoznacznianie zestawu danych polega na przetworzeniu danych w taki sposób, aby przy jednoczesnym zachowaniu jak największej informatywności danych, zapewnić wystarczający poziom niejednoznaczności
- Klasyfikacja oceny stopnia niejednoznaczności
  - niejednoznaczność zadanego zestawu wartości cech na przestrzeni populacji (jak również w danym zestawie danych)
  - efekt dodania konkretnej wartości cechy do zadanego zestawu wartości cech
  - dostępność danej cechy na zewnątrz danych (tzn. potencjalny „koszt” wejścia w posiadanie wartości danej cechy)
- **Pojęcie *k-anonimowości***
  - Zestaw danych jest *k-anonimowy*, gdy każdy rekord w obrębie tych danych jest nierozróżnialny od co najmniej  $k-1$  innych rekordów

# Anonimizacja - uniejednoznacznianie

- Uniejednoznacznienie przez grupowanie
- Przykład

Nazwisko	Adres	Wiek	Diagnoza
***	Warszawa	26	Zapalenie płuc
***	Kraków	42	Nadciśnienie
***	Zakopane	47	Cukrzyca
***	Radom	28	Astma

- Wprowadzamy grupowanie na polach adres i wiek
  - Adres jest zastępowany wyższą jednostką administracyjną
  - Wiek jest przedziałowany

Nazwisko	Adres	Wiek	Diagnoza
***	woj. Mazowieckie	20-29	Zapalenie płuc
***	woj. Małopolskie	40-49	Nadciśnienie
***	woj. Małopolskie	40-49	Cukrzyca
***	woj. Mazowieckie	20-29	Astma

- Osiągnięty poziom 2-anonimowości

# Anonimizacja - uniejednoznacznianie

- Uniejednoznacznianie przez usuwanie części informacji

ID										matches
0										4-6-8
1										2-5-8
2										0-3-11
3										1-5-11
4										1-9-11
5										1-9-11
6										4-6-8
7										4-6-8
8										0-1-5-7-9
9										0-1-5-7-9
10										4-6-9
11										1-9-10

(a) – początkowa baza danych z 5 cechami: włosy, kolor oczu, kolor skóry, wynik badania hemoglobiny, wynik badania moczu

(b) – baza niejednoznaczna o poziomie 3-anonimowości

(c) – losowe przemieszanie rekordów

Liczby wskazują początkowe rekordy do których pasuje zanonimizowany rekord

# Plan prezentacji

- Dane medyczne a dane osobowe
- Konieczność ochrony
- Prawo
- Dane identyfikacyjne
- Anonimizacja
- **DICOM**
- Systemy bezpieczeństwa

- Digital Imaging and Communication In Medicine (DICOM)
- Standardowy format danych obrazowych opracowany przez *American College of Radiology* oraz *National Electrical Manufacturers Association*
- Nagłówek
  - Grupy
  - Elementy
  - Np. (0010,0030) – grupa 0010 to dane osobowe pacjenta, element 0030 to data urodzenia pacjenta
- Dane obrazowe
  - Binarne dane obrazowe zależne od urządzenia diagnostycznego
  - Dodatkowe obrazy (np. zrzuty ekranu)

- Standard podaje zalecenia które znaczniki powinny podlegać procesowi deidentyfikacji – suplement nr 55.
- W zależności od zastosowania pola podane w standardzie mogą być niewystarczające
- **Zalecana procedura postępowania** (R.Noumeir, A.Lemay, „Pseudonymization of Radiology Data for Research Purposes”, J. of Digital Imaging, Vol.20, No 3, 2007)
  - Atrybuty prywatne mogą zostać skasowane, gdyż wykorzystywane są jedynie przez sprzęt diagnostyczny na którym powstały dane
  - Atrybuty powiązane z instytucją, technikami i lekarzami mogą zostać usunięte, gdyż przeważnie nie są wymagane do celów innych niż wewnętrzne
  - Opis badania, opis serii oraz informacje o protokole muszą pozostać, gdyż zawierają istotne informacje np. do przetwarzania obrazów
  - Historia i komentarze pacjenta powinny zostać wyczyszczone, gdyż jako pola wolego tekstu mogą zawierać informacje identyfikacyjne, natomiast ich znaczenie naukowe jest niewielkie
  - Wszystkie unikalne identyfikatory (UID) powinno się zastąpić pseudonimami – nowymi unikalnymi identyfikatorami. Formaty poszczególnych identyfikatorów są opisane w standardzie DICOM
  - Identyfikator badania oraz numer przystąpienia (*Accession number*) powinny zostać usunięte
  - Imię i nazwisko pacjenta, godzina urodzenia, inne identyfikatory pacjenta, inne nazwy pacjenta, MRL (*medical record locator*), grupa etniczna oraz zawód są usuwane, podczas gdy data urodzenia jest uniejednoznaczniata poprzez usunięcie dnia urodzenia. Waga pacjenta, płeć oraz wzrost są zachowywane lub również uniejednoznaczniata, gdyż mogą być kluczowe dla algorytmów przetwarzania. Identyfikator pacjenta jest zastępowany pseudonimem.

- 0008,0012 - Instance Creation Date
- 0008,0013 - Instance Creation Time
- 0008,0014 - Instance Creation UID
- 0008,0016 - SOP Class UID
- 0008,0018 - SOP instance UID
- 0008,0020 - Study Date
- 0008,0021 - Series Dale
- 0008,0022 - Acquisition Time
- 0008,0023 - Content Date
- 0008,0024 - Overlay Date
- 0008,0025 - Curve Date
- 0008,002A - Acquisition Date/Time
- 0008,0030 - Study Time
- 0008,0031 - Series Time
- 0008,0032 - Acquisition Time
- 0008,0033 - Content Time
- 0008,0034 - Overlay Time
- 0008,0035 - Curve Time
- 0008,0050 - Accession Number
- 0008,0080 - Institution Name
- 0008,0081 - Institution Address
- 0008,0082 - Institution Code Sequence
- 0008,0090 - Referring Physician Name
- 0008,0092 - Referring Physician Address
- 0008,0094 - Referring Physician Telephone Num
- 0008,0096 - Referring Physician ID Sequence
- 0008,1000 - Network ID
- 0008,1040 - Institutional Department Name
- 0008,1048 - Physician(s) of Record
- 0008,1049 - Physician(s) of Record ID Sequence
- 0008,1050 - Performing Physician Name
- 0008,1052 - Performing Physician ID Sequence
- 0008,1060 - Name of Physician(s) Rending Stud
- 0008,1062 - Physician(s) Reading Study ID Sequence
- 0008,1070 - Operator Name
- 0008,1072 - Operator ID Sequence
- 0010,xxx - Related to Patient Information
- 0010,0010 - Patient's Name
- 0010,0020 - Patient ID
- 0010,0030 - Patient's Birth Date
- 0010,0032 - Patient's Birth Time
- 0010,0040 - Patient's Sex
- 0010,0050 - Patient's Insurance Plan Code
- 0010,1000 - Other Patient IDs
- 0010,1001 - Other Patient Names
- 0010,1005 - Patient's Birth Name
- 0010,1010 - Patient's Age
- 0010,1020 - Patient's Size
- 0010,1030 - Patient's Weight
- 0010,1040 - Patient's Address
- 0010,1050 - Patient's insurance Plan ID
- 0010,1060 - Patient's Mother's Birth Name
- 0010,1080 - Patient's Military Rank
- 0010,1081 - Patient's Branch of Service
- 0010,1090 - Patient's Med Record Locator
- 0010,2152 - Patient's Region of Residence

- 0010,2154 - Patient's Telephone Number
- 0010,2160 - Patient's Ethnic Group
- 0010,2180 - Patient's Occupation
- 0010,21B0 - Additional Patient History
  
- 0012,xxxx - Related to Clinical Trial Information
  
- 0020,xxxx - Related to Study Information
- 0020,000D - Study Instance UID
- 0020,000E,- Series Instance UID
- 0020,0010 - Study ID
- 0020,0012 - Acquisition Number
- 0020,1070 - Other Study Numbers
  
- 0032,xxxx - Related to Study Status and Verification information
  
- 0038,xxxx - Related to Admission Information
  
- 0040,xxxx - Related to Procedure Information
- 0040,0002 - Scheduled Start date
- 0040,0003 - Scheduled Start Time
- 0040,0004 - Scheduled End Date
- 0040,0005 - Scheduled End Time
- 0040,0006 - Physician's Name
  
- 0040,0009 - Scheduled Step ID
- 0040,000B - Performing Physician ID Sequence
- 0040,0244 - Performed Start date
- 0040,0245 - Performed Start Time
- 0040,0250 - Performed End Date
- 0040,0251 - Performed End Time
- 0040,0253 - Performed Step ID
- 0040,050A - Specimen Accession Number
- 0040,0550 - Specimen Sequence
- 0040,0556 - Specimen identifier
- 0040,06FA - Slide identifier
- 0040,1001 - Requested Procedure ID
- 0040,1010 -Names of intended Recipient
- 0040,1011 - Intended Recipients of Results ID Sequence
- 0040,1101 - Person ID Code Sequence
- 0040,1102 - Person's Address
- 0040,1103 - Person's Telephone Number
- 0040,2004 - Issue Date of Imaging Service Request
- 0040,2005 - Issue Time of Imaging Service Request
- 0040,A030 - Verification Date/Time
- 0040,A032 - Observation Date/Time
- 0040,A121 - Date
- 0040,A122 - Time
- 0040,A123 - Person Name
- 0040,A124 - UID

- 0054,0400 - Image ID
- 2100,0040 - Creation Date
- 2100,0050 - Creation Time
- 2100,0160 - Owner ID
- 3002,0003 - RT Image Name
- 3005,0004 - Structure Set Name
- 3005,0008 - Structure Set Date
- 3005,0009 - Structure Set Time
- 3008,0024 - Treatment Control Point Date
- 3008,0025 - Treatment Control Point Time
- 3008,003A - Specified Treatment Time
- 3008,003B - Delivered Treatment Time
- 3008,0054 - First Treatment Date
- 3008,005A - Most Recent Treatment Date
- 3008,0162 - Safe Position Exit Date
- 3008,0164 - Safe Position Exit Time
- 3008,0166 - Safe Position Return Date
- 3008,0168 - Safe Position Return Time
- 3008,0250 - Treatment Date
- 3008,0251 - Treatment Time
- 300A,0003 - RT Plan Name
- 300A,0006 - RT Plan Date
- 300A,0007 - RT Plan Time
- 300E,xxxx - Related to Review Status and Information
- 4008,xxxx - Related to Interpretation Status and Information
-

- Osobny problem – dane obrazowe
  - Usuwanie identyfikujących informacji tekstowych z danych obrazowych
  - Zaburzanie twarzy
- Istniejące oprogramowanie
  - *Eigenstool*, Radiology Research at Henry Ford Health System, Detroit, USA
  - *Showcase*, Trillium Technology Inc., Ann Arbor, USA
  - *ImageJ*, Research Services Branch, National Institutes of Health, Bethesda, USA
  - *exDicom*, University of Nottingham School of Psychology, Nottingham, Wielka Brytania
  - *DicomWorks*, Puech and Boussel, Lyon, Francja
  - *DICOM Anonymizer by NeoLogica*, NeoLogica, Cairo Montenotte, Włochy
  - *DICOM Anonymizer by Universal PACS*, Universal PACS, New Orleans, USA
  - *Medical Image Resource Center - DicomEditor*, RSNA Radiology Informatics Committee, Oak Brook, USA
  - *Adobe Photoshop DICOM Plugin*, Adobe Systems Inc., San Jose, USA
  - *Matlab Image Processing Toolbox*, Mathworks Inc., USA

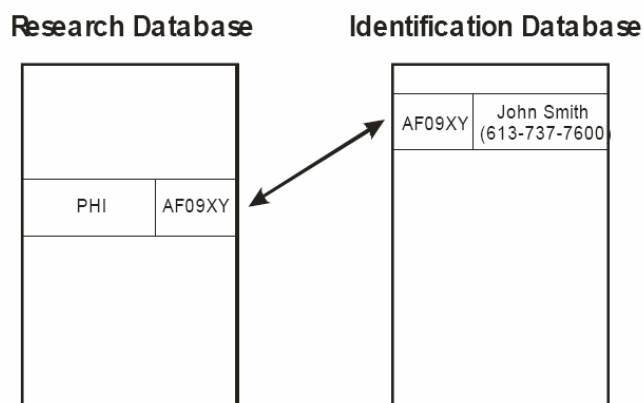
# Plan prezentacji

- Dane medyczne a dane osobowe
- Konieczność ochrony
- Prawo
- Dane identyfikacyjne
- Anonimizacja
- DICOM
- **Systemy bezpieczeństwa**

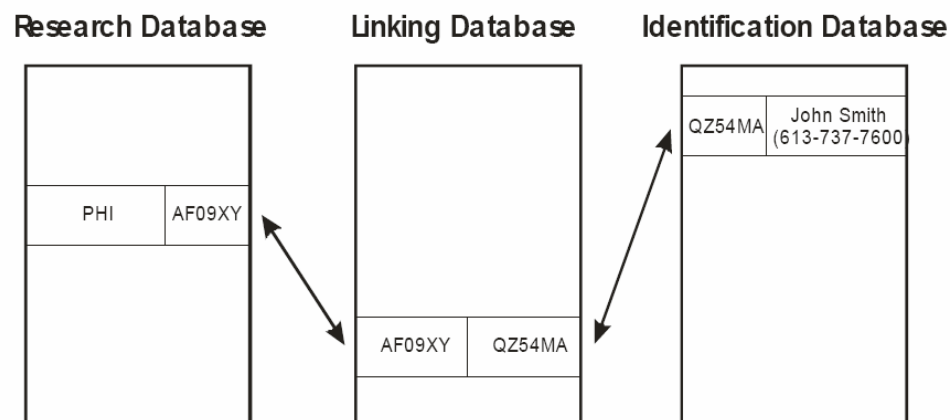
- Poszczególne metody anonimizacji stosuje się głównie w celu udostępniania danych
- W systemach medycznych (przechowujących i przetwarzających dane medyczne) bazuje się na połączeniu wielu metod w celu zwiększenia poziomu bezpieczeństwa
- Model separacji
  - Najczęściej stosowanym modelem jest model separacji danych
  - Dane są poddawane procesowi depersonalizacji – dane identyfikujące są oddzielane od pozostałych danych
  - Oba zestawy danych przechowywane są osobno (również fizycznie)
  - Skojarzenie danych identyfikujących z resztą danych realizowane za pośrednictwem pseudonimizacji
  - Część identyfikacyjna danych jest rzadziej potrzebna i może być obwarowana silniejszymi zabezpieczeniami
  - Część anonimowych danych może być udostępniana

# Systemy bezpieczeństwa

- Przykład zastosowania separacji i pseudonimizacji
  - Jednowarstwowy



- Dwuwarstwowy



# Systemy bezpieczeństwa

- Systemy złożone
  - Korzystają jednocześnie z wielu metod (deidentyfikacja + pseudonimizacja + uniejednoznacznianie)

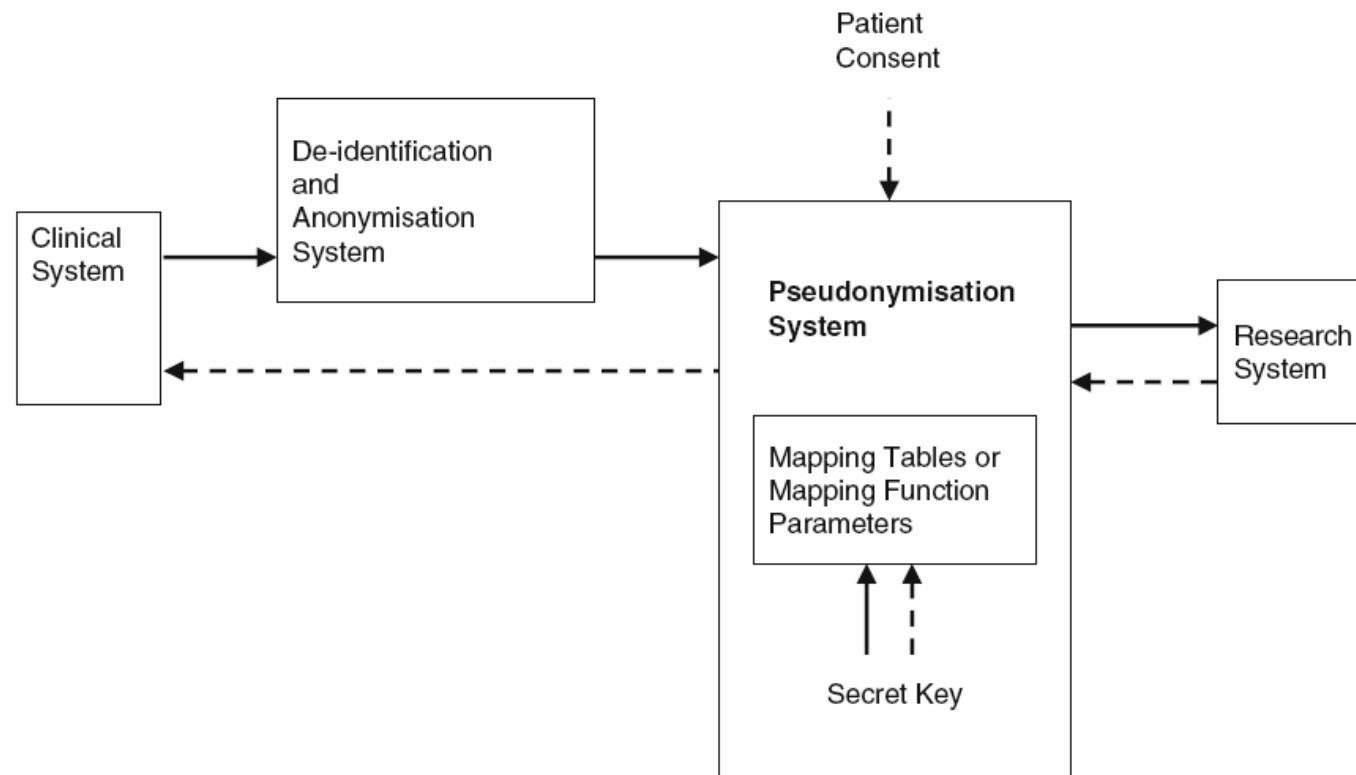


Fig 2. Reversible pseudonymization system and its interaction with other systems.

R.Noumeir, A.Lemay, „Pseudonymization of Radiology Data for Research Purposes”, J. of Digital Imaging, Vol.20, No 3, 2007

# Systemy bezpieczeństwa

- Systemy złożone
  - Korzystają jednocześnie z wielu metod (deidentyfikacja + pseudonimizacja + uniejednoznacznianie)

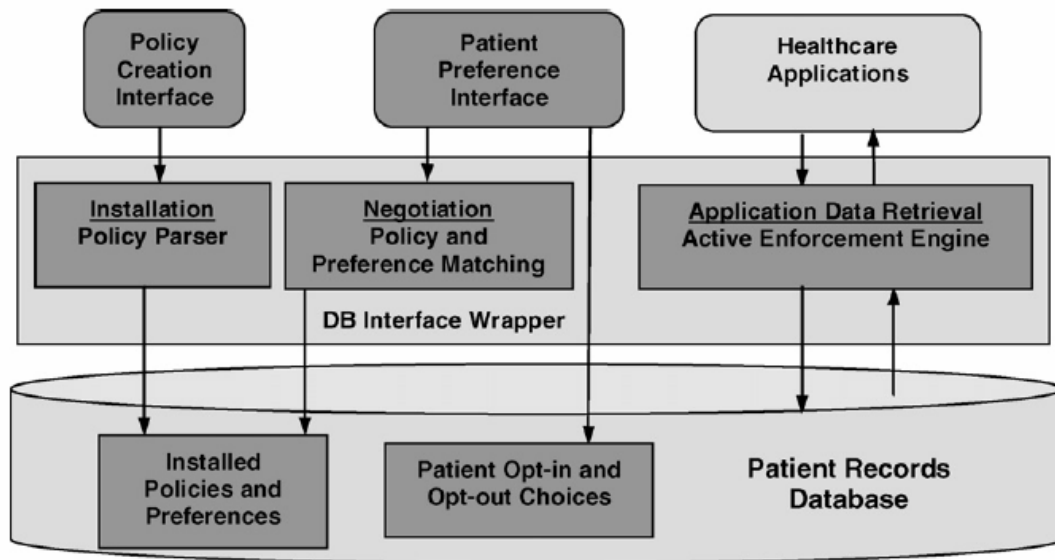


Fig. 1 – HDB Active Enforcement architecture.

R.Agrawal, C.Johnson, „Securing electronic health records without impeding the flow of information”, Int. J. of Medical Informatics 76, 2007

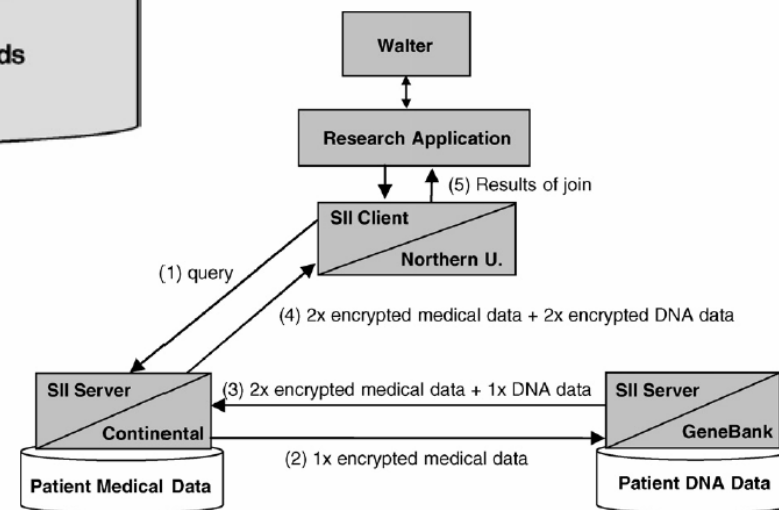


Fig. 5 – HDB Sovereign Information Integration.

# Podsumowanie - zalecenia

- Polskie prawo nie zobowiązuje do anonimizacji osobowych danych medycznych przetwarzanych w celach naukowych
- Nie jest wymagana zgoda pacjenta na przetwarzanie w celach naukowych
- Zaleca się jednak stosowanie metod anonimizacji tak często jak jest to możliwe – gdy dane identyfikacyjne nie są niezbędne
- Dobra praktyką jest również stosowanie tych metod w celu zwiększenia bezpieczeństwa danych – model separacji danych

;-)

- Ochrona osobowych danych medycznych jest jak seks nastolatków:
  - Wszyscy o tym myślą przez cały czas
  - Każdy o tym rozmawia przez cały czas
  - Każdy myśli, że inni to robią
  - Prawie nikt naprawdę tego nie robi
  - Tych kilku którzy naprawdę to robią:
    - robią to kiepsko
    - są pewni, że następnym razem zrobią to lepiej
    - nie robią tego bezpiecznie
  - Każdy przechwala się cały czas swoimi sukcesami, podczas gdy tak naprawdę niewielu ma jakiegokolwiek sukcesy

***Dziękuję za uwagę!***

***b.borucki@icm.edu.pl***

