

Zadanie

*„Model normatywny podnoszenia kompetencji cyfrowych w szkołach wyższych”
finansowane w ramach dotacji celowej pochodzącej ze środków Ministerstwa
Edukacji i Nauki na podstawie Umowy Nr MEiN/2023/CTC/2625*

Konferencja naukowa

**„Wzmacnianie odporności szkół wyższych
na cyberataki przez podnoszenie kompetencji cyfrowych”**

Usuwanie niespójności baz danych

dr inż. Kajetan Wojsyk



UKSW UNIwersytet Kardynała
STEFANA WYSZYŃSKIEGO
W WARSZAWIE

CLTC Centrum Liderów
Transformacji Cyfrowej



Uczelnia – to nie budynki, to ludzie – obiekty do zaatakowania!

• Skuteczny atak to:

- atak ze strony, z której atakowany ataku się nie spodziewa;
 - atak w miejscu, o którym nie pomyślano, że może być podatnością;
 - atak w sytuacjach niejasnych, w których nie wiadomo, kto odpowiada za ochronę;
 - atak z wykorzystaniem **podstępu, socjotechniki, bez śladów** (zob. Kevin Mitnick);
 - atak niewykryty lub wykryty ze znacznym opóźnieniem, gdy jego skutki są już kosztowne, widoczne - i bolesne dla zaatakowanego...
- Im większa **iluzja bezpieczeństwa**, tym słabszy system ochrony, o ile w ogóle istnieje...
 - Im niższa wiedza o **podatnościach** obiektu będącego przedmiotem ochrony, tym większe szanse na skuteczny atak ma atakujący.
 - Istnieje wiele książek napisanych przez wybitnych znawców przedmiotu, którzy – albo zawodowo pracując „po jasnej stronie mocy”, albo na nią przechodząc, tłumaczą zawile problemy w zrozumiałym dla niespecjalistów języku, wzbogacając przekaz opisami rzeczywistych, wziętych z życia sytuacji.
 - Problemem jest powszechne przekonanie, że „**ktoś stoi na straży bezpieczeństwa**”, ktoś to całe „cyber...” ogarnia; my jesteśmy urzędnikami, pracownikami nauki, studentami, administracją taką czy inną, to nie nasza sprawa, my tu tylko sprzątamy...
 - I to właśnie jest niebezpieczne! Człowiek jest najszabszym, najbardziej podatnym elementem systemu, obiektem ataku!



- W roku 2017, w trakcie prac związanych z przekształcaniem i tworzeniem **rejestrów medycznych** dostrzeżono istotną **słabość tych rejestrów**, polegającą na ich silosowości, na **braku automatycznej aktualizacji danych** wtórnych, tworzonych w innych systemach pierwotnych. Od działania systemu informacji w ochronie zdrowia zależy czas reakcji ekipy ratunkowej na zagrożenie życia człowieka i szybkość udzielenia mu pomocy!
- Poszukiwano **sposobu zapewnienia jakości danych**, które spełniałyby kilka warunków:
 - Źródłowość (by dane były dostarczane ze źródła ich tworzenia, bez pośrednictwa mogącego stanowić miejsce ich zafałszowania),
 - Aktualność (zgodność ze stanem faktycznym w dniu wykorzystywania),
 - Możliwość stworzenia mechanizmów **raportowania zmian i niezgodności**, które mogłyby, a nawet musiałyby się pojawić w rejestrach silosowych, budowanych odrębnie, w oderwaniu od rejestrów pierwotnych.

Rozważania nad zapewnieniem wspomnianych warunków doprowadziły do powstania w roku 2017 koncepcji tzw. „**adresu uniwersalnego**”.

Pojęcie „adresu uniwersalnego”

- Istotą koncepcji jest **równoczesne i mechanistyczne** wykorzystanie przepisów prawa, które wprowadzie odrębnie regulują pewne pozornie niezwiązane z sobą obszary, ale przypominają w działaniu **kompletne, lecz niepołączone ze sobą ogniwa łańcucha**, czyli jako zbiór praw nie dają efektu synergii. Analogia mechaniczna: ogniwa łańcucha niepołączone ze sobą nie pozwalają zrealizować **jedynego prostego zadania** – przeniesienia wzdłużnych sił rozciągających, jeżeli choć **jedno** ogniwo jest wadliwe.
- Przykład dwóch tylko przepisów wykonawczych do ustawy - *Prawa geodezyjnego i kartograficznego* (Dz.U.2023.1752 t.j. z dnia 2023.08.31):
 - *Rozporządzenie w sprawie ewidencji miejscowości, ulic i adresów* (Dz.U.2021.1368) oraz
 - *Rozporządzenie w sprawie ewidencji gruntów i budynków* (Dz.U.2021.1390),powołanych w załączniku nr 1 do *Rozporządzenia Rady Ministrów w sprawie Krajowych Ram Interoperacyjności, minimalnych wymagań dla rejestrów publicznych i wymiany informacji w postaci elektronicznej oraz minimalnych wymagań dla systemów teleinformatycznych* (Dz.U.2017.2247 t.j. z dnia 2017.12.05) łącznie jako podstawy utworzenia **trzeciego obiektu interoperacyjności – punktu adresowego lub działki** występujących w **architekturze rejestrów publicznych** (obok identyfikatorów PESEL i REGON). Dostrzeżenie związku tych przepisów z **bezpieczeństwem fizycznym człowieka** doprowadziło do utworzenia pojęcia „**adresu uniwersalnego**” - **zestawu kodu pocztowego, kodu terytorialnego, kodów miejscowości, kodów ulic, współrzędnych X i Y w układzie prostokątnym oraz numeru budynku - w ściśle zdefiniowanej automatycznie tworzonej strukturze. Ów automatyzm ma tu fundamentalne znaczenie!**

Co wnosi do bezpieczeństwa „adres uniwersalny”?

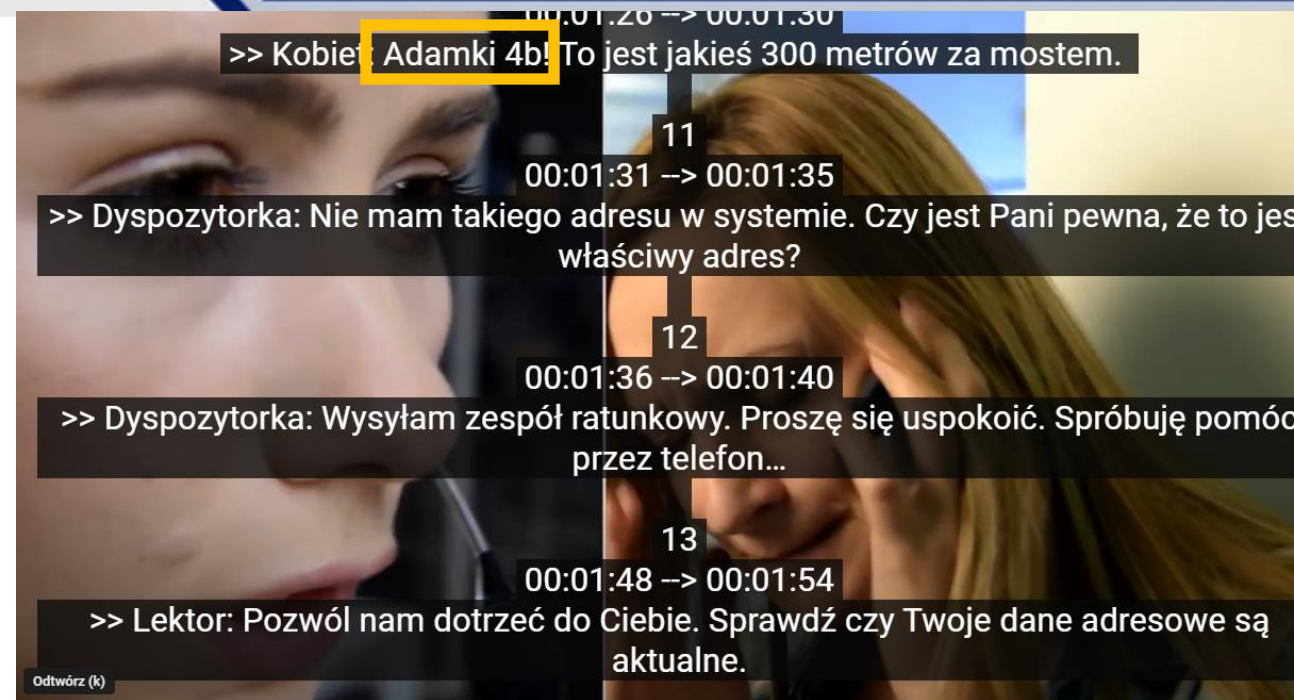
- Adres uniwersalny zalecany jest do stosowania w załączniku do Uchwały nr 28 Rady Ministrów z dnia 18 lutego 2021 r. w sprawie Programu **otwierania danych** na lata 2021-2027 (M.P.2021.290)
- **Adresy uniwersalne** tworzone są **automatycznie** z odpowiednich danych i **kodów** tych danych przesyłanych ze źródeł ich wytworzenia (gmin i powiatów) do Państwowego Rejestru Granic (GUGiK) – z którego przez odpowiednią ekstrakcję zgromadzonych w nim danych automatycznie tworzy się **otwarty zbiór – adruni.csv** (<https://danepubliczne.punktyadresowe.pl/>)
- **Nazwy obiektów** takich jak ulice (place, aleje, itp.) tworzone są przez Rady Gmin i ogłaszane są w Wojewódzkich Dziennikach Urzędowych, są więc nazwami **prawnie obowiązującymi**. Główny Urząd Statystyczny tworzy kody owych obiektów i również publikuje je jako **zbiory otwarte** (rejestr TERYT).
- Wystarczyło zbudować mechanizm **porównujący** odpowiednio oba wymienione wyżej zbiory danych – i uzyskać **możliwość wychwytywania niezgodności danych** oraz jednoznacznego ustalenia, **jakiego miejsca w terenie zapytanie o lokalizację zdarzenia dotyczy**. Owo miejsce w terenie jest powiązaniem opisów tekstowych z **rzeczywistością, w której my – ludzie – żyjemy**.
- **Adres** po uzupełnieniu go o współrzędne geodezyjne staje się **punktem adresowym**.

Bezpieczeństwo - czego, czy czyje?

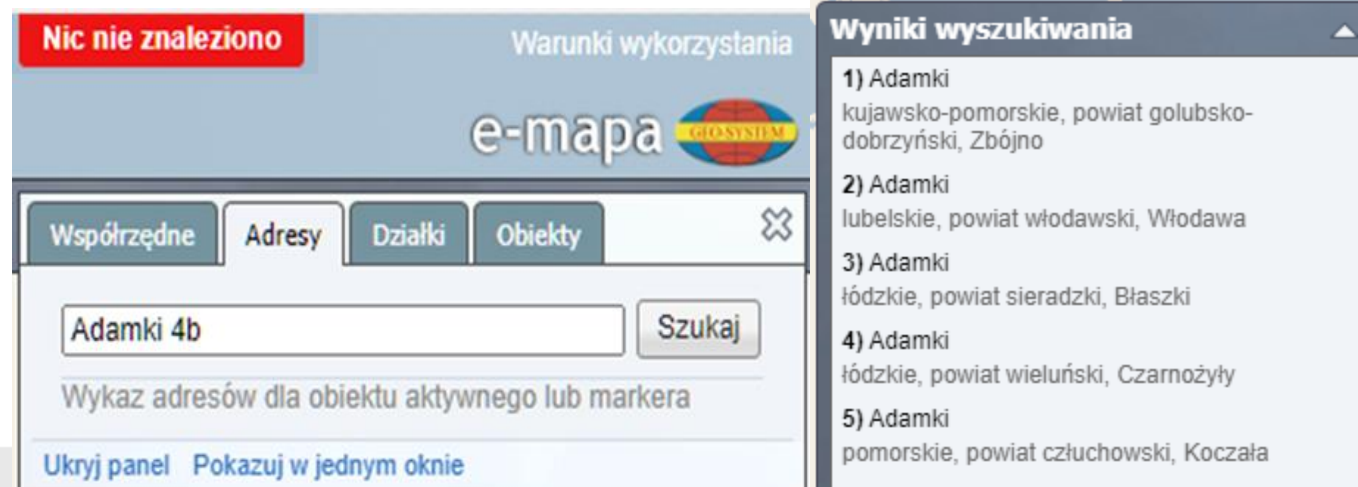
- Myśląc o bezpieczeństwie, zawsze myślimy o określonym:
 - obiekcie fizycznym,
 - działaniu,
 - czynności.

Praca w **ochronie zdrowia** determinuje postrzeganie bezpieczeństwa jako **fizycznego bezpieczeństwa człowieka**, zależnego od systemów ratownictwa – i **od danych** przetwarzanych w tych systemach, w szczególności systemach informacyjnych o **lokalizacji zdarzenia**...

<https://polska.e-mapa.net>



<https://www.youtube.com/watch?v=Vo44PCh5VUE> (akcja GUGiK)



Co ma informacja w ochronie zdrowia do geodezji?

Ustawa o systemie informacji w ochronie zdrowia (t.j. Dz.U. 2022 poz. 1555)

- Art. 4 ust 5. **Dane adresowe usługodawców oraz usługobiorców** wykorzystywane w systemie informacji są **systematycznie aktualizowane i weryfikowane w oparciu o referencyjne bazy danych**, o których mowa w art. 4 ust. 1a pkt 4–6 i 8 ustawy z dnia 17 maja 1989 r. – **Prawo geodezyjne i kartograficzne** (Dz. U. z 2020 r. poz. 2052), za pośrednictwem usług dostępnych w **geoportalu infrastruktury informacji przestrzennej**, o którym mowa w art. 13 ust. 1 ustawy z dnia 4 marca 2010 r. o **infrastrukturze informacji przestrzennej** (Dz. U. z 2021 r. poz. 214).

2023-11-30 Według informacji portalu wadowice24.pl policja we wtorek otrzymała zgłoszenie o zaginięciu 14-latki z Andrychowa. Dziewczynka miała jechać do pobliskich Kęt. **Rano zatelefonowała do ojca, że źle się czuje i nie wie, gdzie jest.** Rodzic zgłosił sprawę policji, która **wczesnym popołudniem** opublikowała komunikat o poszukiwaniu dziecka. **Krótko potem** nastolatka została odnaleziona. Za **późno**... A przecież wystarczyło, po zgłoszeniu zaginięcia dziecka przez ojca namierzyć lokalizację telefonu!
Więcej: <https://wiadomosci.radiozet.pl/polska/andrychow-14-letnia-natalia-zamarzla-przed-sklepem-mieszkanicy-sa-w-szoku>

Stworzenie mechanizmu porównywania rejestrów

- Poszukiwanie niezgodności danych adresowych (umożliwiających np. wysłanie drona z pomocą czy skierowanie ekip ratowniczych **w konkretne miejsce w przestrzeni**) doprowadziło do **utworzenia aplikacji** pozwalających na precyzyjne i szybkie (liczone w 3-4 minutach) lokalizowanie niezgodności danych adresowych w dowolnej gminie, a co za tym idzie, szybkie ich usuwanie czyli doprowadzanie danych do stanu zgodnego ze stanem faktycznym.
- Mechanizm porównywania danych korzysta **bezpośrednio** z istniejącego od lat systemu kodów przypisywanych do obiektów terytorialnych w rejestrze **TERYT** przez **Główny Urząd Statystyczny** oraz z danych źródłowych, pochodzących z urzędów prowadzących ewidencję gruntów i budynków oraz ewidencję miejscowości, ulic i adresów gromadzonych w **Państwowym Rejestrze Granic (PRG)** prowadzonym przez **Główny Urząd Geodezji i Kartografii**.
- Tworzenie i wykorzystywanie ww. danych ma oparcie w przepisach prawa. Każda gmina powinna z proponowanego mechanizmu lokalizowania niezgodności korzystać.
Kolejny slajd pokazuje wynik informacji o realizacji poszczególnych procedur porównywania **danych otwartych** – dostępnych dla każdego.

Sprawność generowania zbiorów danych i raportów

Sąsiedni zrzut ekranu pokazuje proces tworzenia plików różnych raportów dotyczących obszaru konkretnej gminy (Łask w powiecie łaskim) w dniu 2023-11-25.

Raporty zawierają bazy danych **wszystkich punktów adresowych**, obiektów, różnic między prawnie umocowanymi nazwami a nazwami wpisanymi w wydziale geodezji, niepoprawne numery budynków itd. – z dokładnością do 1 znaku, a nawet jego wielkości...

```
Wiersz polecenia
TERYT: 100302
Spis wszystkich ulic, alej, placów itd. w kraju został utworzony
Spis ulic dla kodu TERYT 100302 został utworzony
Słownik aktualnych nazw ulic został utworzony.
Pliki 2023-11-25_adruni_przestawiony.csv i 2023-11-25_adruni.csv zostały utworzone.
Pliki 2023-11-25_Adresy_uniwersalne.csv i 2023-11-25_Bledne_PNA.csv zostały utworzone.
Plik 2023-11-25_Niesłownikowe_nazwy_ulic.csv został utworzony.
Pliki 2023-11-25_poprawne_nr_bud.csv i 2023-11-25_niepoprawne_nr_bud.csv zostały utworzone.
Plik 2023-11-25_result.csv został utworzony.
Pliki 2023-11-25_Miejsc_ulic_unikat.csv, 2023-11-25_Miejsc_ulic_zeros.csv,
2023-11-25_Miejsc_ulic_zeros_not_empty.csv, 2023-11-25_Miejsc_ulic_all_zeros.csv
i 2023-11-25_Miejsc_zeros.csv zostały utworzone
Plik 2023-11-25_Bledne_znaki.csv został utworzony.
Pliki 2023-11-25_Niesłownikowe_nazwy_ulic_skroc.csv
i 2023-11-25_Słownikowe_nazwy_ulic_skroc.csv zostały utworzone.
Podaj datę pierwszego pliku adruni.csv (rrrr-mm-dd): 2020-01-03
Podaj datę drugiego pliku adruni.csv (rrrr-mm-dd): 2023-11-25
Plik 2023-11-25_100302_cechy.csv został utworzony.
Pliki 2023-11-25_sprawdz_kod_ulicy.csv i 2023-11-25_bez_punktow_adresowych.csv zostały utworzone.
Plik gminy_szeroko.txt został utworzony.
Raport zbiorczy został zapisany do pliku: 2023-11-25_100302_Raport_zbiorczy.csv
TERYT,Gmina,Województwo,Powiat,Data
100302,Łask,ŁÓDZKIE,Łaski,2023-11-25
Czas rozpoczęcia:
11:10
Czas zakończenia:
11:11
```

Tworzenie pliku zawierającego różnice między stanami określonymi datami – 2020-01-03 i 2023-11-25 – zawartość pokazana jest na nast. slajdzie...

Zawartość pliku 2023-11-25_2020-01-03_adruni_zmienione.csv (fragm.)

data	teryt	miescowosc	ulica	numer	adruni
2020-01-03	100302	Kolonia Bałucz		40	;98100 100302 0706898 0706898 00000 418868 503984 40
2023-11-25	100302	Bałucz		40	;98100 100302 0706881 0706881 00000 418868 503984 40
2020-01-03	100302	Młynisko		12A	;98100 100302 0706906 0706906 00000 417635 501877 12A
2023-11-25	100302	Borszewice Kościelne		12A	;98100 100302 0706964 0706964 00000 417635 501877 12A

2020-01-03	100302	Mikołajki		406	23
2023-11-25	100302	Mikołajki		406	23
2020-01-03	100302	Łopaty		731	117
2023-11-25	100302	Łopaty		731	6
2020-01-03	100302	Ostrów		715	60
2023-11-25	100302	Ostrów		715	7
2020-01-03	100302	Ostrów		621	17
2023-11-25	100302	Ostrów		621	39
2020-01-03	100302	Ostrów		351	37D
2023-11-25	100302	Ostrów		2351	1
2020-01-03	100302	Ostrów		697	46a
2023-11-25	100302	Ostrów		2697	9
2020-01-03	100302	Ostrów		746	26
2023-11-25	100302	Ostrów		2746	51
2020-01-03	100302	Ostrów		689	28
2023-11-25	100302	Ostrów		2689	55

Informacja o obiekcie

- Adresy i ulice

Punkt adresowy

Gmina Łask (100302)

Miejscowość Borszewice Kościelne (0706964)

Kod pocztowy 98-100

Ulica

Numer 12A

Współrzędne 1992 417635 501877

Adres uniwersalny 98100|100302|0706964|0706964|00000|417635|501877|12A|

Źródło PRG

Otwórz w Geoportal.gov.pl

Współrzędne

Adresy

Działki

Obiekty

stopnie, minuty, sekundy np. 52° 40' 25.6"

B 51° 37' 30.65" L 19° 01' 37.65"

zapis dziesiętny np. 52.67370428

B 51.62518128 L 19.02712364

układy współrzędnych prostokątnych płaskich

Układ 1992 (EPSG:2180)

X 417635 Y 501877

what3words np. skalne.kolumna.kredens [2]

jednakowy.napiwki.naliczać

adres uniwersalny

417635|501877

link z innego portalu mapowego

Pokaż Ukryj znacznik położenia

Ukryj panel Pokazuj w jednym oknie

Przykłady zawartości różnych raportów (CSV)

Nazwa_pliku;Liczba_rekordow

2023-11-25_100302_cechy.csv;4

2023-11-25_100302_Weryfikator_TERYT.csv;275

2023-11-25_2020-01-03_adruni_zmienione.csv;332

2023-11-25_Adresy_uniwersalne.csv;6593

2023-11-25_adruni.csv;6593

2023-11-25_adruni_przestawiony.csv;6593

2023-11-25_bez_punktow_adresowych.csv;23

2023-11-25_Bledne_PNA.csv;0

2023-11-25_Bledne_znaki.csv;0

2023-11-25_Miejsc_ulic_all_zeros.csv;0

2023-11-25_Miejsc_ulic_unikat.csv;249

2023-11-25_Miejsc_ulic_zeros.csv;46

2023-11-25_Miejsc_ulic_zeros_not_empty.csv;0

2023-11-25_Miejsc_zeros.csv;0

2023-11-25_niepoprawne_nr_bud.csv;696

2023-11-25_Nieslownikowe_nazwy_ulic.csv;3823

2023-11-25_Nieslownikowe_nazwy_ulic_skroc.csv;249

2023-11-25_poprawne_nr_bud.csv;5897

2023-11-25_result.csv;1

2023-11-25_Slownik.csv;48247

2023-11-25_Slownikowe_nazwy_ulic_skroc.csv;0

2023-11-25_sprawdz_kod_ulicy.csv;2770

2023-11-25_Weryfikator_TERYT.csv;296148

Cecha;Liczba
al.;1
pl.;6
rondo;1
ul.;267

WOJ_POW_GMI;SYM;ULIC;WOJEWODZTWO;POWIAT;GMINA;MIJSCOWOSC;ULICA
100302;0976296;11202;ŁÓDZKIE;łaski;łask;łask;pl. 11 Listopada
100302;0976296;20427;ŁÓDZKIE;łaski;łask;łask;ul. Jana Sobieskiego
100302;0976296;21925;ŁÓDZKIE;łaski;łask;łask;ul. Szeroka
100302;0976296;03686;ŁÓDZKIE;łaski;łask;łask;pl. Jarosława Dąbrowskiego
100302;0976296;16262;ŁÓDZKIE;łaski;łask;łask;pl. Józefa Piłsudskiego
100302;0976296;21786;ŁÓDZKIE;łaski;łask;łask;pl. Szarych Szeregów
100302;0976296;11710;ŁÓDZKIE;łaski;łask;łask;ul. Łódzka
100302;0976296;37611;ŁÓDZKIE;łaski;łask;łask;pl. Plac Lotników Łaskich
100302;0976296;06423;ŁÓDZKIE;łaski;łask;łask;pl. Gwiazdzisty
100302;0976296;14004;ŁÓDZKIE;łaski;łask;łask;ul. Narodowa
100302;0707113;27064;ŁÓDZKIE;łaski;łask;Okup Wielki;ul. Bałucka
100302;0976296;01528;ŁÓDZKIE;łaski;łask;łask;ul. Boczna
100302;0976296;43375;ŁÓDZKIE;łaski;łask;łask;ul. Mirosława Bogusa
100302;0707107;27064;ŁÓDZKIE;łaski;łask;Okup Mały;ul. Bałucka
100302;0707165;23129;ŁÓDZKIE;łaski;łask;Ostrów;ul. Tulipanowa
100302;0707165;13283;ŁÓDZKIE;łaski;łask;Ostrów;ul. Morelowa
100302;0707165;13175;ŁÓDZKIE;łaski;łask;Ostrów;ul. Modrakowa
100302;0976296;57276;ŁÓDZKIE;łaski;łask;łask;rondo NSZZ Solidarność
100302;0707165;23033;ŁÓDZKIE;łaski;łask;Ostrów;ul. Truskawkowa
100302;0707165;07722;ŁÓDZKIE;łaski;łask;Ostrów;ul. Kaczeńcowa
100302;0707165;37633;ŁÓDZKIE;łaski;łask;Ostrów;ul. Kwitnącej Łąki
100302;0707165;35588;ŁÓDZKIE;łaski;łask;Ostrów;ul. Jesiennych Liści
100302;0707165;35585;ŁÓDZKIE;łaski;łask;Ostrów;ul. Szumiących Traw

TERYT;Liczba_blednych;Procent;Z_dnia
100302;696;89,4433;2023-11-25

2023-22-25_100302_Raport_zbiorczy.csv

Podsumowanie



Obraz przedstawia człowieka usiłującego oddzielić prawdę od fałszu, w mroku rozświetlanym światłem wiedzy, w czasie ograniczonym długością życia - i w niedostatku...

- Metoda wyszukiwania niezgodności (niespójności) danych oparta o **porównanie** odrębnych zbiorów (z których jeden uznajemy za odniesienie, a drugi jest obiektem porównywanym), pozwala wyodrębnić wszystkie **różniące się** wartości tej samej - teoretycznie - danej.
- Zasadą jest stosowanie **słowników** zawierających kody skojarzone z wartościami danych.

Przykład:

Opisywaną metodę zastosowano do **Rejestru Aptek**. W dniu 2023-12-04 stan jakości danych adresowych (pozwalających na określenie **lokalizacji** aptek w sposób automatyczny) określany jako stosunek liczby automatycznie zlokalizowanych punktów adresowych do wszystkich rekordów wyniósł **52,31%**

Rozwój kompetencji cyfrowych w 3 dekadzie XXI wieku...

- Do podstawowych kompetencji cyfrowych należy zaliczyć:
 - Umiejętność lokalizowania **źródeł** danych;
 - Umiejętność **rozpoznawania i pomiaru jakości** źródeł danych;
 - Umiejętność tworzenia informacji w oparciu o zweryfikowane dane – dane niezsweryfikowane stanowią podatność na zagrożenie!
 - Umiejętność stosowania prawnych środków poprawy jakości danych;
 - Umiejętność korzystania z narzędzi sztucznej inteligencji (AI) w sposób umożliwiający kontrolę jej „produktów”;
 - Umiejętność systematycznego śledzenia stanu zaawansowania usług świadczonych drogą elektroniczną i korzystania z nich;
 - Umiejętność poszukiwania związków pomiędzy **pozornie** niezwiązanymi ze sobą zbiorami danych i wiele innych, nad których określeniem zapewne zostaną podjęte prace badawcze, a nawet zastanawianie się, co ma haker do pociągu:
[\(https://zaufanatrzeciastrona.pl/post/o-trzech-takich-co-zhakowali-prawdziwy-pociag-a-nawet-30-pociagow/\)](https://zaufanatrzeciastrona.pl/post/o-trzech-takich-co-zhakowali-prawdziwy-pociag-a-nawet-30-pociagow/)...

czyli ciągłe doskonalenie własnej wiedzy o zagrożeniach, jakie mogą zaistnieć dosłownie wszędzie: w domu, w szkole i w podróży...

Współrzędne | Adresy | Działki | Obiekty

stopnie, minuty, sekundy np. 52° 40' 25.6"

B 51 ° 37 ' 30.65 " L 19 ° 01 ' 37.65 "

zapis dziesiętny np. 52.67370428

B 51.62518128 L 19.02712364

układy współrzędnych prostokątnych płaskich

Układ 1992 (EPSG:2180)

X 417635 Y 501877

what3words np. skalne.kolumna.kredens [?]

jednakowy.napiwki.naliczać

adres uniwersalny

|||||417635|501877

link z innego portalu mapowego

Pokaż Ukryj znacznik położenia

Ukryj panel Pokazuj w jednym oknie

Dziękuję za uwagę