

# Działania projektu EuropeanaLocal w Polsce

Marcin Werla, PCSS <[mwerla@man.poznan.pl](mailto:mwerla@man.poznan.pl)>

# Co już zrobiliśmy?

- Spotkanie inaugurujące działalność projektu
  - 8 września 2009 r. w Poznaniu
  - 32 osoby z 22 instytucji wstępnie zainteresowanych współpracą
- Stworzenie listy kontaktowej przedstawicieli instytucji zainteresowanych udostępnieniem swoich obiektów cyfrowych w Europeanie w ramach działań projektu EuropeanaLocal
  - W skali projektu lista kontaktowa ma obecnie około 1000 osób, z czego około 50 osób jest z Polski

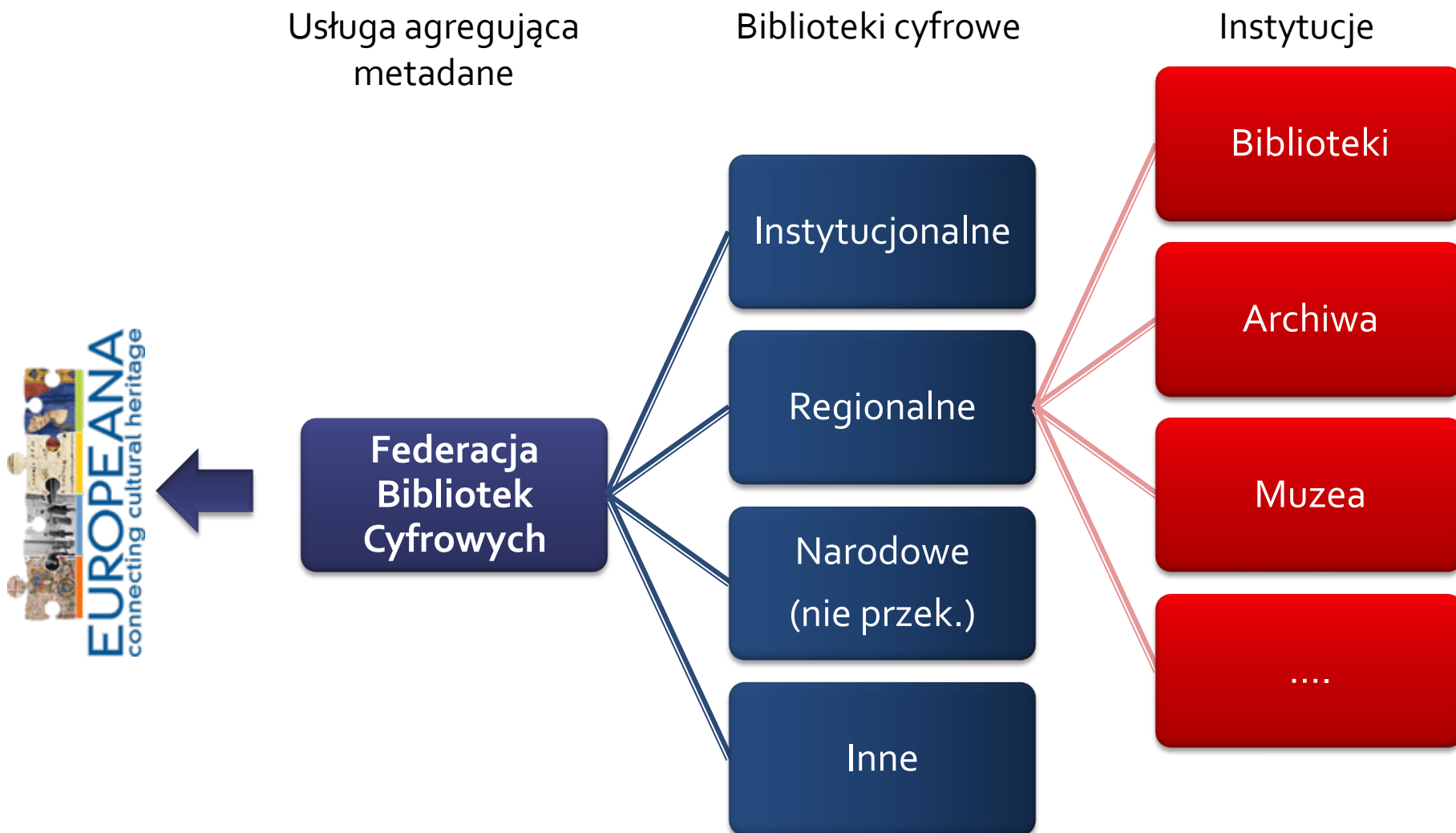
# Co już zrobiliśmy?

- Ankiety
  - Grudzień 2008 – luty 2009
  - Ankieta dotycząca kolekcji i metadanych
    - Zawiera opis kolekcji przeznaczonych do udostępnienia w Europeanie
    - Zawiera referencje do standardów metadanych wykorzystywanych do opisu obiektów w tych kolekcjach
    - Jej podstawą była oryginalna ankieta Europeany dla instytucji zainteresowanych współpracą
  - Ankieta techniczna
    - Dotyczy oprogramowania wykorzystywanego przez biblioteki cyfrowe zainteresowane współpracą z Europeaną
    - Robiona w celu rozpoznania ewentualnych obszarów pomocy technicznej świadczonej przez partnerów technicznych EuropeanaLocal
    - Wg wyników ankiety: dLibra najpopularniejszym w Europie oprogramowaniem do budowy bibliotek cyfrowych! 😊

# Co już zrobiliśmy?

- Udział w pracach Europeany związanych z opracowywaniem różnego rodzaju wytycznych dla dostawców treści
  - Wytyczne organizacyjne
  - Wytyczne odnośnie metadanych
- Wymiana doświadczeń z innymi partnerami projektu - agregatorami
- Poza działaniami związanymi wprost z zakresem działań projektu następuje również rozwój krajowego agregatora metadanych, czyli Federacji Bibliotek Cyfrowych

# Federacja Bibliotek Cyfrowych jako agregator metadanych



[▶ Strona główna](#)[▶ Kolekcje](#)[▶ Nowe konto](#)[▶ Logowanie](#)[▶ Kontakt](#)

obecnie czytających: 80



### Biblioteka

[▶ Publiczne Cyfrowe Archiwum Agnieszki Osieckiej](#)

- ▶ Archiwum Agnieszki Osieckiej
- Archiwum fotograficzne
- Dokumenty nabyte
- Wystawy

### Wyszukiwanie w indeksach

- ▶ Indeks tytułów
- ▶ Indeks twórców
- ▶ Indeks słów kluczowych

### Informacje

- ▶ Informacje o Archiwum
- ▶ Strona Fundacji
- ▶ Kontakt z nami

### Statystyki

Liczba publikacji: 9  
Obecnie czytających: 80  
Łączna liczba czytelników od dnia 2009-04-28: 2865

- ▶ Najczęściej oglądane
- ▶ Najlepiej oceniane
- ▶ Więcej statystyk...

### Kanały RSS

- ▶   25 ostatnich publikacji

Publiczne Cyfrowe Archiwum Agnieszki Osieckiej

Federacja Bibliotek Cyfrowych

Szukaj

Zakres: **Wszędzie** |  |  ▼

Wyszukiwanie zaawansowane...

Poprawne formułowanie zapytań

### Ostatnio dodane

1. Dziennik, tom XII - 14.05.1951-24.09.1951 r.
2. Dziennik, tom X - 17.04.1951-26.04.1951
3. Dziennik, zeszyt VII - 21.11.1950 - 30.11.1950 r.
4. Zdjęcia obiektów Agnieszki Osieckiej
5. Dziennik, tom IV - 24.05.1950 - 25.06.1950 r.
6. Dziennik, tom III - 10.04.1950 - 14.05.1950 r.
7. Dziennik, tom II - 6.03.1950 - 10.04.1950
8. Dziennik, tom I - 18.05.1949 - 17.02.1950
9. Pamiętnik - 27.12.1945 - 02.01.1946 r.




Więcej...



### Najczęściej czytane publikacje

1. Dziennik, tom XII - 14.05.1951-24.09.1951 r. [7]
2. Pamiętnik - 27.12.1945 - 02.01.1946 r. [7]
3. Zdjęcia obiektów Agnieszki Osieckiej [6]
4. Dziennik, tom I - 18.05.1949 - 17.02.1950 [6]
5. Dziennik, tom X - 17.04.1951-26.04.1951 [2]
6. Dziennik, tom IV - 24.05.1950 - 25.06.1950 r. [1]
7. Dziennik, tom III - 10.04.1950 - 14.05.1950 r. [1]

Więcej...

### Dodatki

- ▶  [Plany wprowadzania publikacji](#)
- ▶  [Statystyki](#)
- ▶  [Najlepiej oceniane publikacje](#)

- ▶  [Biblioteki cyfrowe dLibra](#)
- ▶  [Wtyczka umożliwiająca wyszukiwanie](#)

# Agregatory metadanych dla Europeany



- Wg aktualnej wersji specyfikacji funkcjonalnej Europeany (Europeana Outline Functional Specification, wer. 1.7) zadania agregatora to:
  1. Zebranie informacji na temat dostawców treści i ich systemów informatycznych
  2. Zebranie informacji o obiektach do udostępnienia w Europeanie
  3. Usunięcie duplikatów, uporządkowanie metadanych, normalizacja, wzbogacenie ich
  4. Potwierdzenie dostępności obiektów cyfrowych
  5. Udostępnienie danych Europeanie przy pomocy protokołu OAI-PMH

[http://dev.europeana.eu/public\\_documents/EDLnet%20D2.5\\_Outline\\_Functional\\_Specifications20090301\\_version%201.7\\_consWithoutHistory\\_lossless.pdf](http://dev.europeana.eu/public_documents/EDLnet%20D2.5_Outline_Functional_Specifications20090301_version%201.7_consWithoutHistory_lossless.pdf)

# FBC jako agregator metadanych dla Europeany



- Zebranie informacji na temat dostawców treści i ich systemów informatycznych
  - Obecnie – tylko nazwa i logo biblioteki cyfrowej, adres WWW, adres interfejsu OAI-PMH dla obiektów dostępnych i (opcjonalnie) planowanych
  - Wkrótce – Szczegółowy opis wraz z listą instytucji, przykładowymi polecanymi obiektami, statystykami itp.
    - <http://fbc.pionier.net.pl/test/>



# FBC jako agregator metadanych dla Europeany



- Zebranie informacji o obiektach do udostępnienia w Europeanie
  - Realizowane przy pomocy protokołu OAI-PMH
    - Wymagana ścisła zgodność ze specyfikacją protokołu
    - W wyjątkowych przypadkach odstępstwa (PBI)
  - Obecnie pobierane metadane to tylko Dublin Core
    - Do końca roku planowane rozszerzenie do ESE

# FBC jako agregator metadanych dla Europeany

- **Usunięcie duplikatów**, uporządkowanie metadanych, normalizacja, wzbogacenie ich
  - Ma sens głównie w kontekście bibliotek
    - W muzeach i archiwach każdy obiekt jest traktowany jako unikalny
  - Oparte na porównywaniu metadanych, z uwzględnieniem drobnych różnic w zapisie
  - Około 0.2% agregowanych obiektów trafia na listę potencjalnych duplikatów
  - Takie wykrywanie duplikatów ma sens w przypadku koordynowania digitalizacji
    - W przypadku Europeany bardziej chodzi o duplikaty powstałe w skutek agregowania tych samych obiektów przez różne agregatory
    - Wystarczy porównywanie adresów URL obiektów lub identyfikatorów URI

# FBC jako agregator metadanych dla Europeany

- Usunięcie duplikatów, **uporządkowanie metadanych, normalizacja**, wzbogacenie ich
  - Na poziomie FBC budowane są automatycznie słowniki agregowanych metadanych
    - Osobno dla każdego elementu opisu
    - Osobno dla każdego języka opisu
  - Różnice pomiędzy metadanymi pochodzącymi z różnych BC wpływają negatywnie na możliwości wyszukiwawcze użytkowników końcowych
  - Dlatego też istotne jest uporządkowanie i normalizacja metadanych w tak dużym stopniu jak to tylko możliwe
  - Podstawowa analiza pokazuje, które elementy są kluczowe
    - Analiza została przeprowadzona w kwietniu 2009 na danych 214 254 zagregowanych obiektów

# FBC jako agregator metadanych dla Europeany

Element DC	Liczba unikalnych wartości	Liczba wystąpień tego elementu DC w opisach obiektów	Średnia liczba wyst. na poj. wartość
<b>format</b>	39	209 789	5 379,2
<b>language</b>	195	210 529	1 079,6
<b>type</b>	822	211 816	257,7
<b>rights</b>	1 192	246 093	206,5
<b>coverage</b>	66	2 390	36,2
<b>publisher</b>	18 002	310 764	17,3
<b>contributor</b>	12 979	83 464	6,4
<b>subject</b>	78 440	438 871	5,6
<b>relation</b>	9 292	48 319	5,2
<b>date</b>	47 581	209 589	4,4
<b>identifier</b>	6 426	27 666	4,3
<b>description</b>	43 657	180 391	4,1
<b>source</b>	16 996	52 506	3,1
<b>creator</b>	21 908	67 503	3,1
<b>title</b>	210 745	227 039	1,1

# FBC jako agregator metadanych dla Europeany

- Format
  - W 99% opisów: typ MIME (np. text/html, image/x.djvu)
- Język
  - Większość przypadków: ISO 639-2 (pol, ger, lat, fre etc.)
  - Czasami jedna wartość „pol, ger” zamiast odrębnych „pol”, „ger”
- Prawa
  - Nazwa instytucji, która przechowuje oryginalny obiekt
- Typ
  - ...

# FBC jako agregator metadanych dla Europeany

Wartości (top 20)	Liczba obiektów opisanych tą wartością	% zagregowanych obiektów	% zagreg. obiektów (po czyszczeniu)
czasopismo	44 709	20,9%	33,8%
gazeta	32 921	15,4%	31,3%
gazety	23 119	10,8%	
Czasopismo	20 965	9,8%	
książka	12 503	5,8%	
Gazeta	11 098	5,2%	
pocztówka	5 768	2,7%	
czasopisma	4 962	2,3%	
text	4 452	2,1%	
grafika	3 863	1,8%	
fotografia	3 596	1,7%	
artykuł z czasopisma	3 164	1,5%	2,6%
artykuł	2 455	1,1%	
Czasopisma	1 710	0,8%	
dzienniki urzędowe	1 516	0,7%	
stary druk	1 222	0,6%	1,1%
starodruk	1 221	0,6%	
rysunek	1 094	0,5%	
rękopis	1 062	0,5%	
mapa	1 028	0,5%	
<b>Suma</b>		<b>85,1%</b>	<b>68,9%</b>

# FBC jako agregator metadanych dla Europeany

- Usunięcie duplikatów, uporządkowanie metadanych, normalizacja, **wzbogacenie ich**
  - ESE ver. 3.1 składa się z:
    - A. 15 elementów Dublin Core
      - + 21 kwalifikatorów/zawężeń Dublin Core
    - B. 1 element Dublin Core terms
    - C. 11 elementów specyficznych dla Europeany
  - Większość elementów z grupy A i B powinna zostać pobrana z agregowanej biblioteki cyfrowej
  - Część elementów A i B może być określona na podstawie innych elementów
    - Taka możliwość zależy od standardów opisu przyjętych przez poszczególne biblioteki
  - Większość (całość?) elementów C może być określona na podstawie elementów A i B

# FBC jako agregator metadanych dla Europeany

- 11 elementów specyficznych dla Europeany
  - **isShownBy, isShownAt**
    - Odnośniki do obiektów wykorzystywane w interfejsie Europeany
  - **unstored**
    - Miejsce na wszystko (czego nie udało się przenieść do ESE)
  - **object** – Odnośnik do miniaturki/próbki obiektu
    - Stworzenie tego odnośnika może być czasem zautomatyzowane
      - <http://www.wbc.poznan.pl/dlibra/docmetadata?id=2752>
      - <http://www.wbc.poznan.pl/Content/2752>
      - <http://www.wbc.poznan.pl/image/edition/2752>
  - **hasObject**
    - *true* lub *false* – wskazuje, czy pole „object” jest dostępne



# FBC jako agregator metadanych dla Europeany

- 11 elementów specyficznych dla Europeany
  - **provider**
    - Nazwa dostawcy treści (tego, kto wysyła dane do Europeany, czyli agregatora)
  - **country**
    - Kraj dostawcy treści (ISO 3166)
  - **language**
    - Język obowiązujący w kraju dostawcy treści (ISO 639-1)
  - **uri**
    - Unikalny identyfikator obiektu

# FBC jako agregator metadanych dla Europeany

- 11 elementów specyficznych dla Europeany
  - **type**
    - Jedna z wartości TEXT, IMAGE, SOUND, VIDEO
    - Można w dużej mierze osiągnąć automatycznie
      - na podstawie dc:type i dc:format
  - **userTag**
    - Tagi stworzone przez użytkowników (Europeany??)
  - **year**
    - 4 cyfrowa (???) liczba w kalendarzu gregoriańskim, wykorzystywana na osi czasu
    - W wielu przypadkach może być automatycznie uzyskana z dc:date

# FBC jako agregator metadanych dla Europeany

- Prosty proces analizy daty
  - Podstawowy wskaźnik: długość wartości DC:date

Długość	Liczba wystąpień	%
4	92 606	44,03%
10	82 182	39,07%
9	12 833	6,10%
6	5 133	2,44%
11	4 772	2,27%
5	2 420	1,15%
13	2 038	0,97%
7	1 975	0,94%
8	1 484	0,71%
16	866	0,41%

- Wartości dc:date o 10 najczęściej występujących długościach opisują 98,09% wszystkich obiektów

# FBC jako agregator metadanych dla Europeany

- Prosty proces analizy daty
  - Poszukiwanie wzorców – krok 1

Wzorzec	Liczba wystąpień	%
DDDD	92 402	43,93%
DDDD!DD!DD	81 162	38,59%
DDDD!DDDD	9 029	4,29%
!DDDD!	4 350	2,07%
!ca DDDD!	3 219	1,53%
!DDDD!DDDD!	2 208	1,05%
DDDD!	1 783	0,85%
DDDD!DD	1 354	0,64%
!ante DDDD!	924	0,44%
DDDD!D!DDDD	836	0,40%

- Wartości dc:date pasujące do najpopularniejszych 10 wzorców opisują 93,79% wszystkich obiektów

# FBC jako agregator metadanych dla Europeany

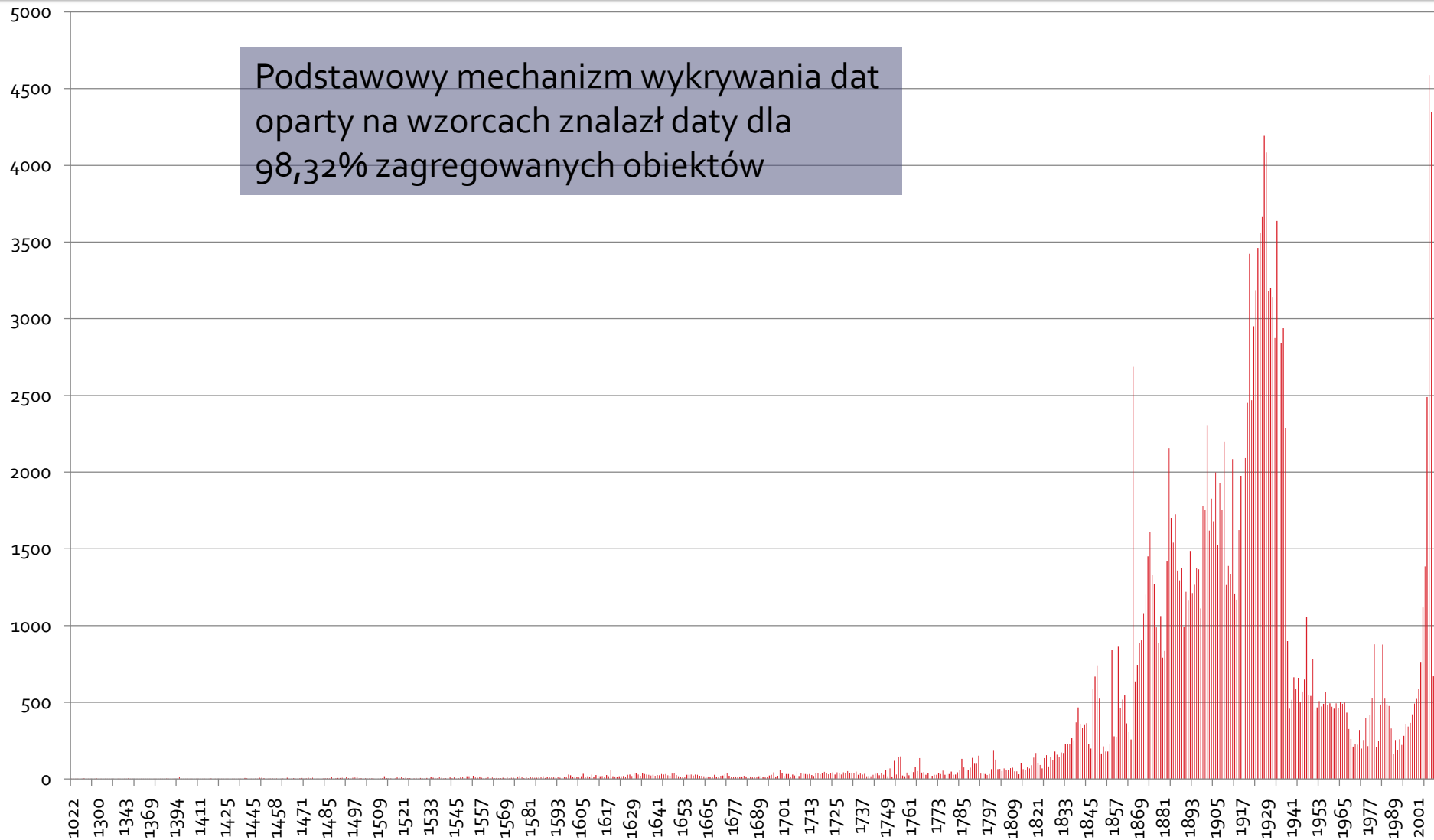
- Prosty proces analizy daty
  - Poszukiwanie wzorców – krok 2

Wzorzec	Liczba wystąpień	%
DDDD	92 402	43,93%
DDDD.DD.DD	62 710	29,82%
DDDD-DD-DD	18 287	8,69%
DDDD-DDDD	8 935	4,25%
[DDDD]	4 327	2,06%
[ca DDDD]	3 208	1,53%
[DDDD-DDDD]	2 202	1,05%
[ante DDDD]	924	0,44%
DDDD.	906	0,43%
DDDD.DD	840	0,40%

- Wartości dc:date pasujące do najpopularniejszych 10 wzorców opisują 92,59% wszystkich obiektów

# FBC jako agregator metadanych dla Europeany

Podstawowy mechanizm wykrywania dat oparty na wzorcach znalazł daty dla 98,32% zagregowanych obiektów



# FBC jako agregator metadanych dla Europeany



- Potwierdzenie dostępności obiektów cyfrowych
  - Zapewniane przez okresową synchronizację metadanych (obecnie raz na dobę)
- Udostępnienie danych Europeanie przy pomocy protokołu OAI-PMH
  - Interfejs OAI-PMH FBC jest gotowy od początku czerwca 2009
  - Polska to jeden z 4 pierwszych krajów, w których zgłoszono agregatory gotowe do testów wymaganych do przyłączenia
  - Efekty testowego pobrania metadanych będą prezentowane przedstawicielom Komisji Europejskiej 3 lipca w Luksemburgu
  - Biblioteka Narodowa jest podłączona do Europeany przez TEL, a więc mimo że jest widoczna na stronach WWW FBC, nie będzie widoczna w interfejsie OAI-PMH
    - NAC zapewne przez APEnet
    - Co z normalizacją metadanych?

# Co robimy?

- Dalszy rozwój FBC
  - Rozwój mapowania na ESE, normalizacja metadanych
    - Liczymy na uczestników wrześniowych warsztatów DC w Gnieźnie
  - Przygotowanie do nowego modelu metadanych Europeany opartego na protokole OAI-ORE
    - Pozwoli na przeszukiwanie treści publikacji na poziomie FBC
  - Stworzenie pakietu dystrybucyjnego oprogramowania FBC
  - Międzynarodowe szkolenie z zakresu budowania agregatorów metadanych
    - Przy tegorocznej edycji konferencji European Conference on Digital Libraries (ECDL 2009)
      - wrzesień/październik 2009 r., Grecja
      - <http://ecd12009.eu/>



# Co robimy?

- Projekt ACCESS IT
  - Finansowany z programu CULTURE
  - Od maja 2009 do kwietnia 2011
  - Polega na pomocy w zdobyciu wiedzy niezbędnej do budowy bibliotek cyfrowych i przyłączania ich do Europeany
    - Tworzenie centrów kompetencji
    - Opracowanie certyfikowanych szkoleń
  - Rozszerzenie działania EuropeanaLocal na Turcję i Serbię, „pogłębienie” działania w Grecji
  - Poza Turcją, Serbią i Grecją udział w projekcie biorą:
    - MDR Partners – zarządzanie projektem, wsparcie merytoryczne
    - IChB PAN PCSS – główny partner merytoryczny

# Co robimy?

- Współpraca z Muzeum Narodowym w Warszawie
  - Nawiązana na konferencji PBC 2008
  - Budowa Cyfrowego Muzeum Narodowego w Warszawie
  - Bliska współpraca z pracownikami muzeum, pozwalająca na opracowanie „muzealnej” dLibry zgodnej z systemem MONA
  - W przyszłości budowa Federacji Muzeów Cyfrowych? 😊

Dziękuję za uwagę!

Pytania?

# Działania projektu EuropeanaLocal w Polsce

Marcin Werla, PCSS <[mwerla@man.poznan.pl](mailto:mwerla@man.poznan.pl)>