



Cyfrowa infrastruktura badawcza dla humanistyki i nauk o sztuce DARIAH-PL

Katarzyna Klessa

Prezentacja projektu na Radę Naukową Dyscyplin Szkoły Nauk o Języku i Literaturze UAM, dn.
18.02.2021



Fundusze
Europejskie
Inteligentny Rozwój



Rzeczpospolita
Polska

Unia Europejska
Europejski Fundusz
Rozwoju Regionalnego





Konsorcjum DARIAH



CZYM JEST DARIAH

O DARIAH PL

O DARIAH ERIC

CELE I MISJA

Początki DARIAH sięgają 2008 roku, a od 2010 trwały starania przekształcenia konsorcjum w ERIC, co znalazło swój pozytywny finał



Więcej: dariah.pl, dariah.pl/grupy-robocze/ oraz dariah.eu



Projekt POIR DARIAH-PL: konsorcjum

- Uniwersytet Warszawski
 - Akademia Sztuk Pięknych w Warszawie
 - Instytut Badań Literackich PAN
 - **Instytut Chemii Bioorganicznej Polskiej Akademii Nauk, PCSS (Lider)**
 - Instytut Historii PAN
 - Instytut Języka Polskiego PAN
 - Instytut Podstaw Informatyki PAN
 - Instytut Slawistyki PAN
 - Instytut Sztuki PAN
 - Politechnika Wrocławska
 - **Uniwersytet im. Adama Mickiewicza w Poznaniu**
 - Uniwersytet Jagielloński
 - Uniwersytet Marii Curie-Skłodowskiej
 - Uniwersytet Mikołaja Kopernika
 - Uniwersytet Wrocłowski
 - Politechnika Poznańska
 - Partner biznesowy: EMC Computer Systems Austria GmbH (Dell)
-



Wyniki konkursu 4/4.2/2020

Projekty wybrane do dofinansowania								
Lp.	Numer wniosku	Nazwa wnioskodawcy	Tytuł projektu	Koszt kwalifikowalny projektu [zł]	Wnioskowana kwota dofinansowania [zł]	Rekomendowana kwota dofinansowania [zł]*	Końcowa liczba przyznanych punktów	Uwagi
1	POIR.04.02.00-00-D013/20	Instytut Nafty i Gazu - Państwowy Instytut Badawczy	Stanowiska badawcze wodorowe wysokociśnieniowe	7 190 000,00	5 033 000,00	5 033 000,00	24	
2	POIR.04.02.00-00-D006/20	Instytut Chemii Bioorganicznej Polskiej Akademii Nauk Poznańskie Centrum Superkomputerowo-Sieciowe (Lider)	Cyfrowa infrastruktura badawcza dla humanistyki i nauk o sztuce DARIAH-PL	117 411 774,00	99 800 000,00	99 800 000,00	23	
3	POIR.04.02.00-00-D011/20	Politechnika Krakowska im. Tadeusza Kościuszki	CePTE Centrum Proekologicznych Technologii Energetycznych	62 807 015,62	50 010 527,79	50 010 527,79	21,6	
4	POIR.04.02.00-00-D008/20	Politechnika Warszawska (Lider)	Krajowe laboratorium sieci i usług 5G wraz z otoczeniem	195 999 999,79	166 599 999,79	166 599 999,79	21,5	
5	POIR.04.02.00-00-D001/20	Uniwersytet Jagielloński	ATOMIN 2.0 - Centrum badań materiałowych w skali ATOMowej dla Innowacyjnej gospodarki	79 981 297,17	67 850 871,97	67 850 871,97	20,5	

<http://www.poir.opi.org.pl/Zakonczenie-oceny-wnioskow-zlozonych-w-ramach-konkursu-44.22020.html>



W projekcie przewidziano aspekty gospodarcze i niegospodarcze działalności

- cele szczegółowe to m.in.: zwiększenie potencjału badawczego polskich jednostek naukowych, zachowanie polskiego dziedzictwa kulturowego oraz utrwalenie tożsamości narodowej i europejskiej obywateli, zwiększenie zrozumienia różnic społecznych oraz kulturowych, wsparcie edukacji
- W wymiarze gospodarczym infrastruktura wspierać będzie rozwój turystyki regionalnej, wizualizacje w obszarze edukacji zwłaszcza z przeznaczeniem dla słabiej rozwiniętych regionów czy zastosowania w sektorze gier, a także zaawansowane zarządzanie multimedialnymi zasobami cyfrowymi



Laboratorium
inteligentnej analizy i
interpretacji

Laboratorium
zautomatyzowanego
wzbogacania

Laboratorium
nadzorowanego
semantycznego
odkrywania

Laboratorium
zaawansowanej
wizualizacji

Laboratorium źródeł



Dariah.lab w UAM i SNJL

- **Korpus multimodalny** (koordynator: prof. UAM dr hab. Maciej Karpiński, lokalizacja: Wydział Neofilologii UAM)
- **Węzeł filologiczny** (koordynator: prof. UAM dr hab. Mirosław Wobalis, lokalizacja infrastruktury: Wydział Nauk o Sztuce UAM)
- **Korpus notatek tłumaczy HANOI** (koordynator: dr Anna Jelec, lokalizacja: Wydział Anglistyki UAM)
- **Baza i narzędzia do badań percepcji muzyki, mowy i gestu** (koordynator: prof. UAM dr hab. Piotr Podlipniak, lokalizacja: Wydział Nauk o Sztuce UAM (Morasko))
- **Oprogramowanie do automatycznego datowania tekstów historycznych, normalizacji tekstów i optymalnego wyszukiwania informacji** (koordynator: prof. dr hab. Krzysztof Jassem, lokalizacja: Wydział Matematyki i Informatyki UAM)
- **Archiwum Audiowizualne** (koordynator: dr Kamila Kłudkiewicz, lokalizacja: Wydział Nauk o Sztuce UAM)





Struktura projektu w UAM

- **Komitet sterujący** - Prof. dr hab. Katarzyna Dziubalska-Kołaczyk
 - **Kierownik projektu** – prof. UAM dr hab. Katarzyna Klessa
 - **Koordynatorzy merytoryczni podzespołów** – prof. UAM dr hab. Katarzyna Klessa, prof. UAM dr hab. Krzysztof Jassem, prof. UAM dr hab. Maciej Karpiński, prof. UAM dr hab. Mirosław Wobalis, prof. UAM dr hab. Piotr Podlipniak, dr Kamila Kłudkiewicz, dr Anna Jelec
 - **Administracja** (CWP + wydziałowa)
-



Korpus multimodalny i studio (WN)

Koordynuje: Prof. UAM dr hab. Maciej Karpiński



Główne wyniki działań zespołu będą następujące:

- Korpus audiowizualny (wizja i fonia) komunikacji multimodalnej z wielowarstwową anotacją.
 - Techniki i narzędzia, które posłużą tworzeniu i analizowaniu podobnych korpusów na zamówienie, w tym:
 - Techniki rejestracji, obróbki i archiwizacji różnego typu danych audiowizualnych oraz parametrów ruchu
 - Techniki wspomagania oraz automatyzacji segmentacji i transkrypcji sygnału mowy
 - Zaplecze techniczne niezbędne do stworzenie korpusu, i umożliwiające tworzenie kolejnych tego rodzaju korpusów na zamówienie lub udostępniana zainteresowanym podmiotom.
-



Węzeł filologiczny (WFPiK)

Koordynuje: prof. UAM dr hab. Mirosław Wobalis



- **Pozyskiwanie źródeł** do badań filologicznych (m.in. zdigitalizowanie ok. 10 tysięcy stron ukazującego się w XIX wieku w zaborze rosyjskim „Kaliszanina”, ok. 75 tysięcy jednostek pochodzących z archiwum prof. Stanisława Bąby, 990 afiszy teatralnych z lat 1887-1935).
- Narzędzia informatyczne do **analizy źródeł**
- **Portal sieciowy wraz z zestawem zaawansowanych narzędzi** informatycznych służących m.in. do przeszukiwania, grupowania, oznaczania, statystyki, wizualizacji danych Źródłowych.
- Narzędzie informatycznego służące do tworzenia/ przetwarzania/analizy elektronicznych form **twórczości literackiej**
- Narzędzie informatyczne służące do tworzenia, przetwarzania i analizowania narracji interaktywnych i innych elektronicznych form twórczości literackiej.

(<http://wfpik.amu.edu.pl/strona-glowna/dzialalnosc-naukowa/dariah.pl-wezel-filologiczny>)



Narzędzia i korpus HANOI (WA)

Koordynuje: Dr Anna Jelec



1. Zestaw narzędzi HANOI do **automatycznej analizy własności i parametrów tekstu** (gęstość i struktura tekstu, procent zapisanej kartki, wykorzystane języki, symbole etc.). Aplikacja opiera się na algorytmie mierzącym m.in. gęstość zapisu i procent pokrycia kartki tekstem, pozwalając na ocenę notatek stworzonych do tłumaczenia konsekwentnego.
 2. **Skalowalny korpus notatek tłumaczy**, dający możliwość sortowania tekstów pod względem wykorzystanych języków, tematu wypowiedzi i stopnia zaawansowania zawodowego autora.
 3. **Narzędzie dydaktyczne** pozwalające użytkownikowi na załadowanie własnych notatek do korpusu i uzyskanie odczytu parametrów badanych przez HANOI
-



Narzędzia do badań muzyki, mowy, gestu (WNoS)

Koordynuje: Prof. UAM dr hab. Piotr Podlipniak

- **Zestaw aparatury** badawczej, pozwalający na przygotowanie, przeprowadzenie i analizę wyników badań percepcyjnych, z możliwością rejestracji parametrów fizjologicznych uczestników
 - Zestaw **oprogramowania** do automatycznej analizy symbolicznych reprezentacji różnych własności melodii, wypowiedzi ustnej oraz gestykulacji . Narzędzia powstaną w formie programowych wtyczek do dostępnych systemów operujących na materiale tagowanym. Pozwolą na automatyczną analizę ilościową nie tylko danego parametru anotowanego na jednej warstwie anotacji, lecz również analizy obejmujące wiele warstw anotacji.
 - **Baza przebiegów muzycznych**, która posłuży zarówno jako Źródło potencjalnych bodźców do badań percepcyjnych, jak i jako materiał do analiz i testowania powstających w ramach projektu narzędzi analitycznych.
-



Narzędzia datowania, normalizacji, wyszukiwania w tekstach (WMI)

Koordynuje: Prof. dr hab. Krzysztof Jassem

- narzędzie do automatycznej diachronicznej normalizacji tekstów polskich
- narzędzia do automatycznego datowania tekstów w obrębie ostatnich 20 dekad na podstawie wyłącznie ich treści, tzn. bez potrzeby użycia jawnych metadanych
- oprogramowanie, którego zadaniem jest optymalne wyszukiwanie informacji w tekstach historycznych
 - pozyskiwanie i normalizacja danych z polskich bibliotek cyfrowych oraz z zagranicznych bibliotek cyfrowych posiadających w swoich zasobach publikacje w języku polskim dla stworzenia historycznych słowników synonimów;
 - normalizacja i czyszczenie metadanych, identyfikacja anomalii w metadanych (np. błędnie datowanych tekstów);
 - ekstrakcja i normalizacja czystego tekstu z różnych formatów
 - Identyfikacje anomalii w metadanych

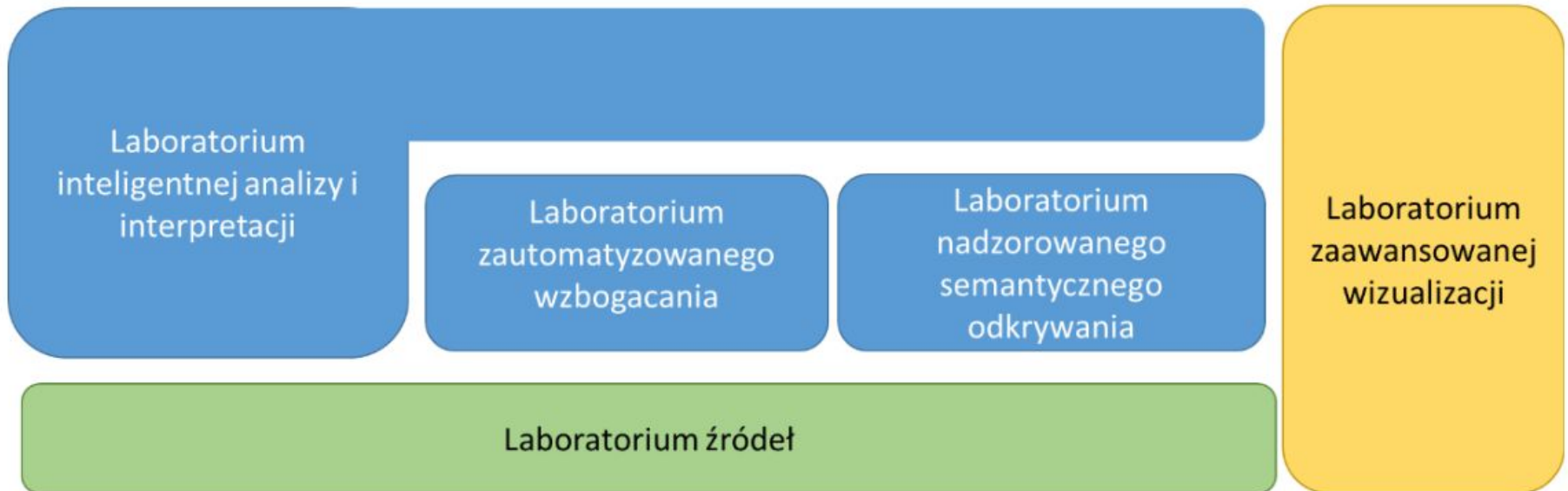


Archiwum Audiowizualne (WNoS)

Koordynuje: Dr Kamila Kłudkiewicz

Baza danych zapewniająca szeroki dostęp do trzech zróżnicowanych zespołów źródeł :

- reprodukcji z najstarszego zespołu Fototeki IHS UAM (10 tysięcy wizerunków szklanych diapozytywów i odbitek, w przyszłości: do 40 tysięcy) Fototeka Instytutu Historii Sztuki UAM jest kolekcją o ponad stuletniej historii. Grafiki, ryciny, z czasem przede wszystkim fotografie (diapozytywy, odbitki fotograficzne, slajdy), gromadzone na uniwersytetach, przyjęto nazywać fototekami. W Polsce przed 1939 r. powstały tylko trzy takie zbiory: w Poznaniu, Krakowie i Lwowie.
- dokumentów z archiwum IHS UAM - Na bogate archiwum Instytutu Historii Sztuki UAM składają się wyjątkowe zespoły, warte szerokiego udostępnienia ze względu na aktualne zainteresowania i trendy badawcze panujące w humanistyce (historia dyscyplin naukowych), bądź sławne i powszechnie znane w światowej historii sztuki postaci poznańskich badaczy.
- nagrań fonograficznych i dokumentacji z objazdów etnomuzykologicznych Instytutu Muzykologii UAM



- Laboratoria - to płaszczyzna współpracy partnerów konsorcjum
- Koordynatorem Lab. IAiI został Prof. UAM dr hab. Mirosław Wobalis



Dostęp do infrastruktury

- Budowana infrastruktura ma być dostępna dla **szerokiego grona odbiorców**, nie jest to infrastruktura budowana na potrzeby tylko jednej jednostki;
- Zasady dostępu do budowanej infrastruktury, wśród kluczowych założeń: warunki niedyskryminujące żadnego podmiotu (komercyjnego i publicznego).



Agenda badawcza

- Projekt DARIAH-PL POiR nie jest projektem badawczym, jest ukierunkowany na tworzenie infrastruktury
- Realizacji projektu towarzyszy jednak **agenda badawcza** obejmująca scenariusze badawcze, których realizacja będzie możliwa dzięki wykorzystaniu wytwarzanej infrastruktury - w ramach laboratoriów projektowych
- W toku realizacji projektu zaplanowano **seminaria DARIAH-PL**, otwarte dla wszystkich zainteresowanych osób; podczas seminarium będą prezentowane osiągnięte rezultaty - w kontekście agendy badawczej. **Zapraszamy!**





Dziękuję za uwagę

dariah@amu.edu.pl



Fundusze Europejskie
Inteligentny Rozwój



**Rzeczpospolita
Polska**

Unia Europejska
Europejski Fundusz
Rozwoju Regionalnego

