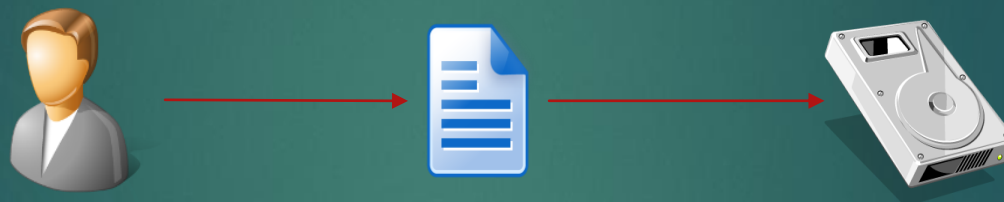




Składowanie i dostęp do danych w rozproszonym systemie ochrony własności intelektualnej

ANDRZEJ SOBECKI, POLITECHNIKA GDAŃSKA
INFOBAZY 2014

Podstawowy proces gromadzenia dokumentów



Trudności:

- Weryfikacja dokumentu
- Ochrona przed kopiowaniem treści
- Współdzielenie
- Dostęp do dokumentów podobnych

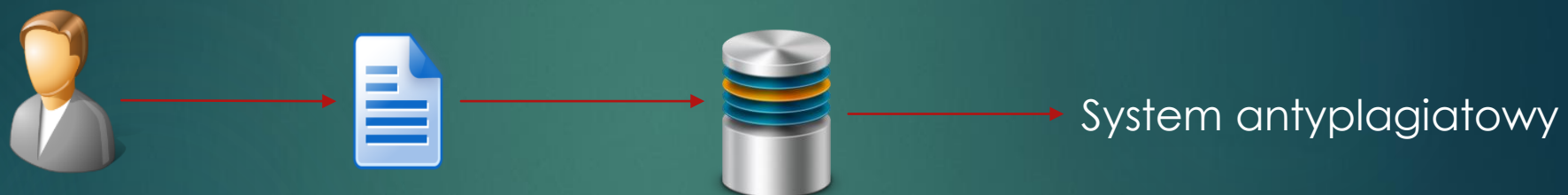
Podstawowy proces gromadzenia dokumentów



Trudności:

- Weryfikacja dokumentu
- Ochrona przed kopiowaniem treści
- Dostęp do dokumentów podobnych

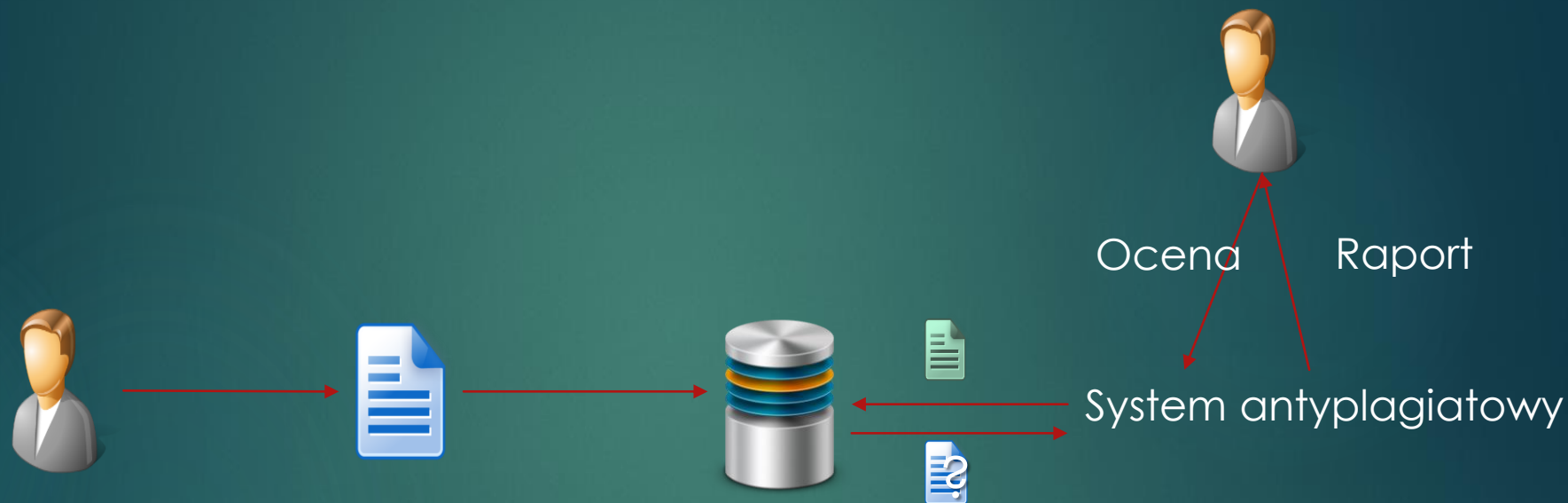
Proces gromadzenia dokumentów rozszerzony o analizę antyplagiatową



Trudności:

- Weryfikacja dokumentu (problem pozostaje dopóki ktoś nie zaakceptuje)
- Ochrona przed kopiowaniem treści
- Dostęp do dokumentów podobnych

Gromadzenie dokumentów zatwierdzonych przez operatora



Trudności:

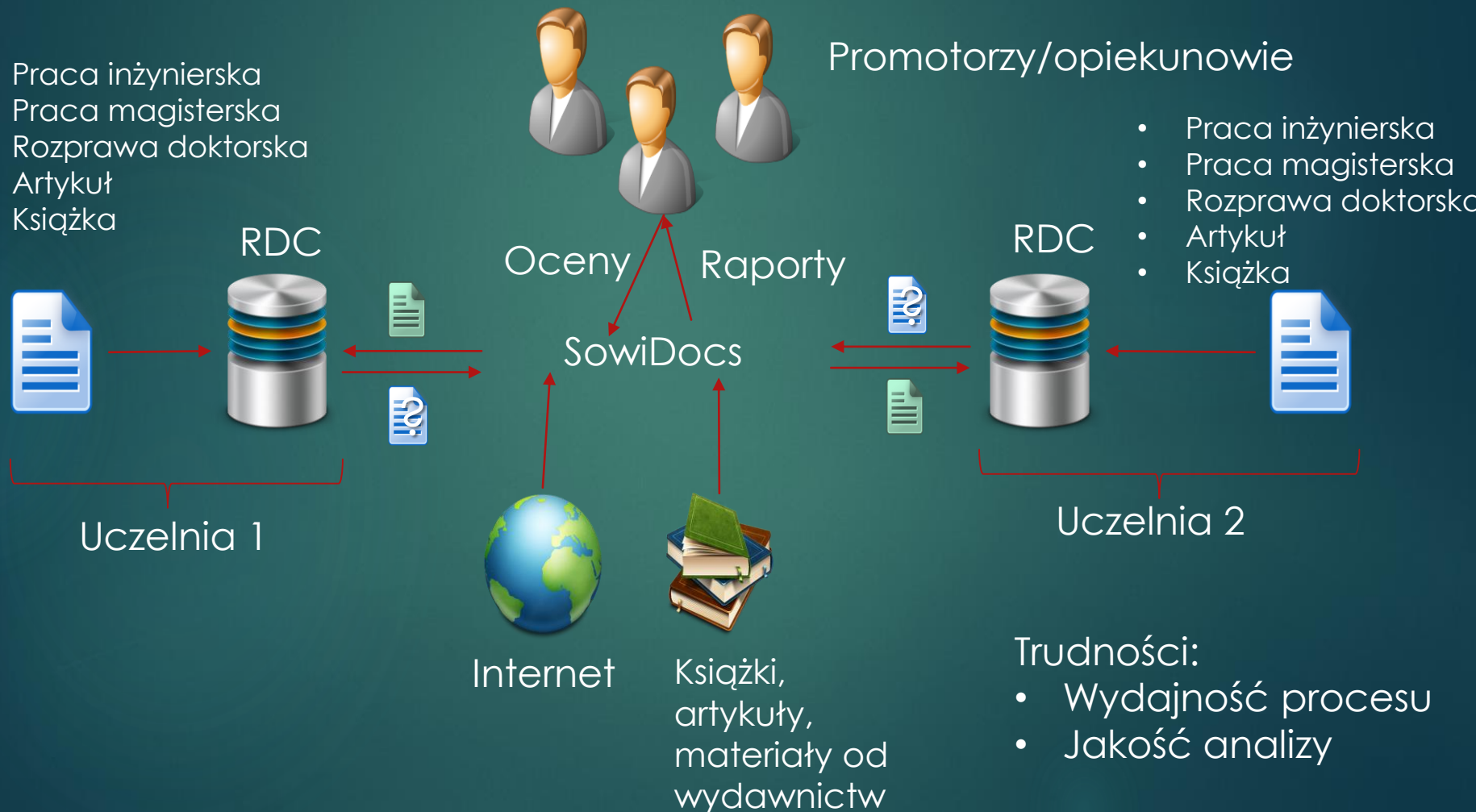
- Ochrona przed kopiowaniem treści
- Wydajność procesu gromadzenia danych
- Jakość analizy

Współpraca wielu organizacji w procesie gromadzenia dokumentów



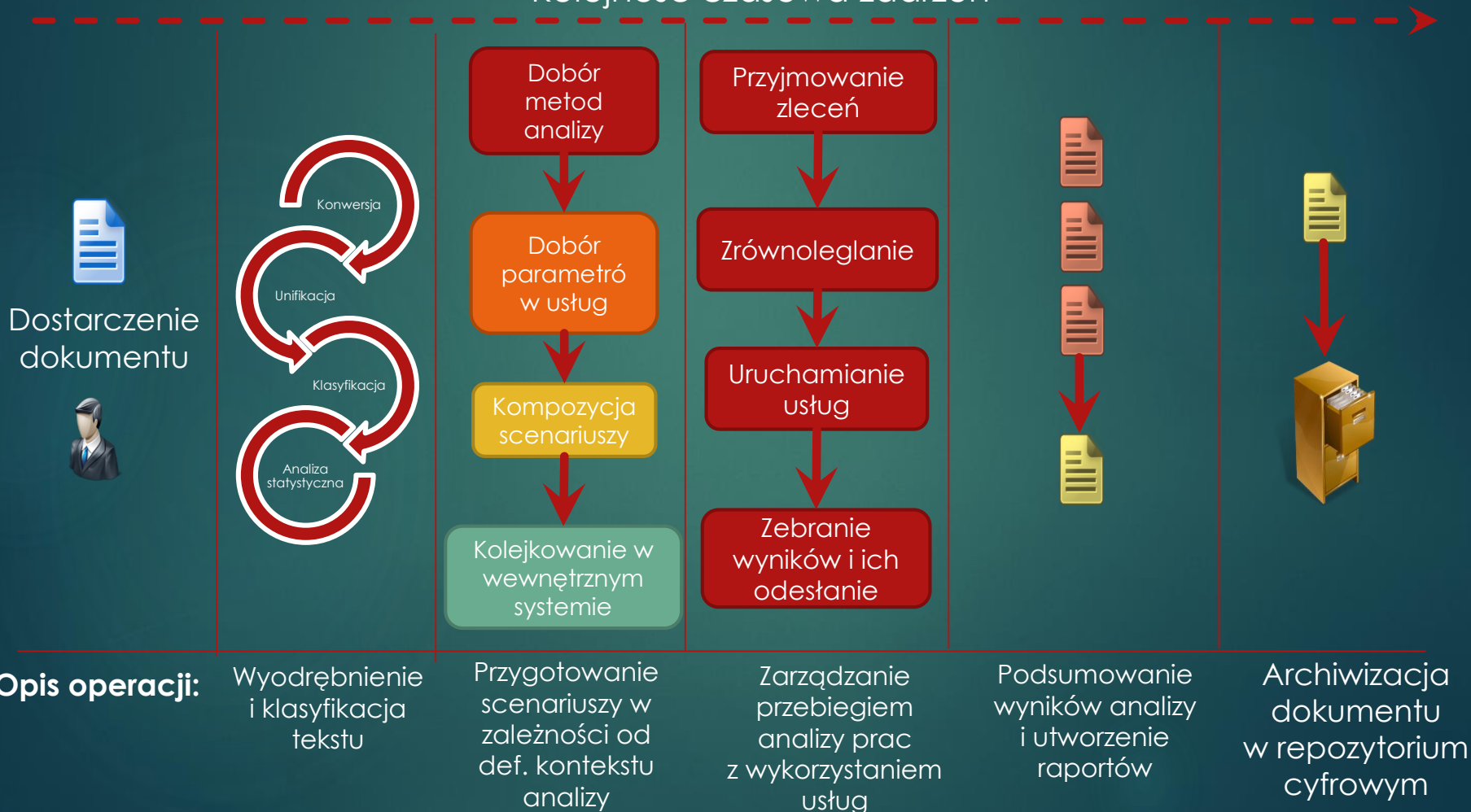
SOWI – Studium przypadku

- Praca inżynierska
- Praca magisterska
- Rozprawa doktorska
- Artykuł
- Książka



Jakość i wydajność procesu analizy antyplagiatowej

Kolejność czasowa zdarzeń



Jakość i wydajność procesu analizy antyplagiatowej – metody analizy

- ▶ Wyszukiwanie dokumentów podobnych
 - ▶ Generowanie metryk dokumentów
 - ▶ Wyszukiwanie metryk dokumentów podobnych w repozytoriach
 - ▶ Tworzenie słów kluczowych na podstawie treści
 - ▶ Wyszukiwanie w Internecie materiałów podobnych
- ▶ Analiza podobieństwa dokumentów
 - ▶ Analiza statystyczna - zgodności dokumentu z profilem użytkownika, wykrywanie prób oszustwa systemu antyplagiatowego
 - ▶ Wyszukiwanie fragmentów identycznych
 - ▶ Wyszukiwanie fragmentów podobnych – odmiana wyrazu, zamiana wyrazów, wstawienie dodatkowych fragmentów

Jakość i wydajność procesu analizy antyplagiatowej – architektura

SowiDocs

- ▶ Obsługa żądań użytkowników
- ▶ Analiza statystyczna
- ▶ Konwersja dokumentów
- ▶ Kompozycja scenariuszy analiz
- ▶ Obsługa kolejowania scenariuszy analiz
- ▶ Podsumowanie wyników analizy
- ▶ Raportowanie
- ▶ Zbieranie ocen archiwizacja dokumentu

Platforma KASKADA

- ▶ Wykorzystanie mocy obliczeniowej klastra superkomputerowego



Scenariusze analizy

- ▶ Realizacja przygotowanych kompozycji usług:
 - ▶ Wyszukiwanie dokumentów podobnych
 - ▶ Analiza podobieństwa dokumentów
 - ▶ Raportowanie

Repozytoria



Internet



SowiSync



SowiSync

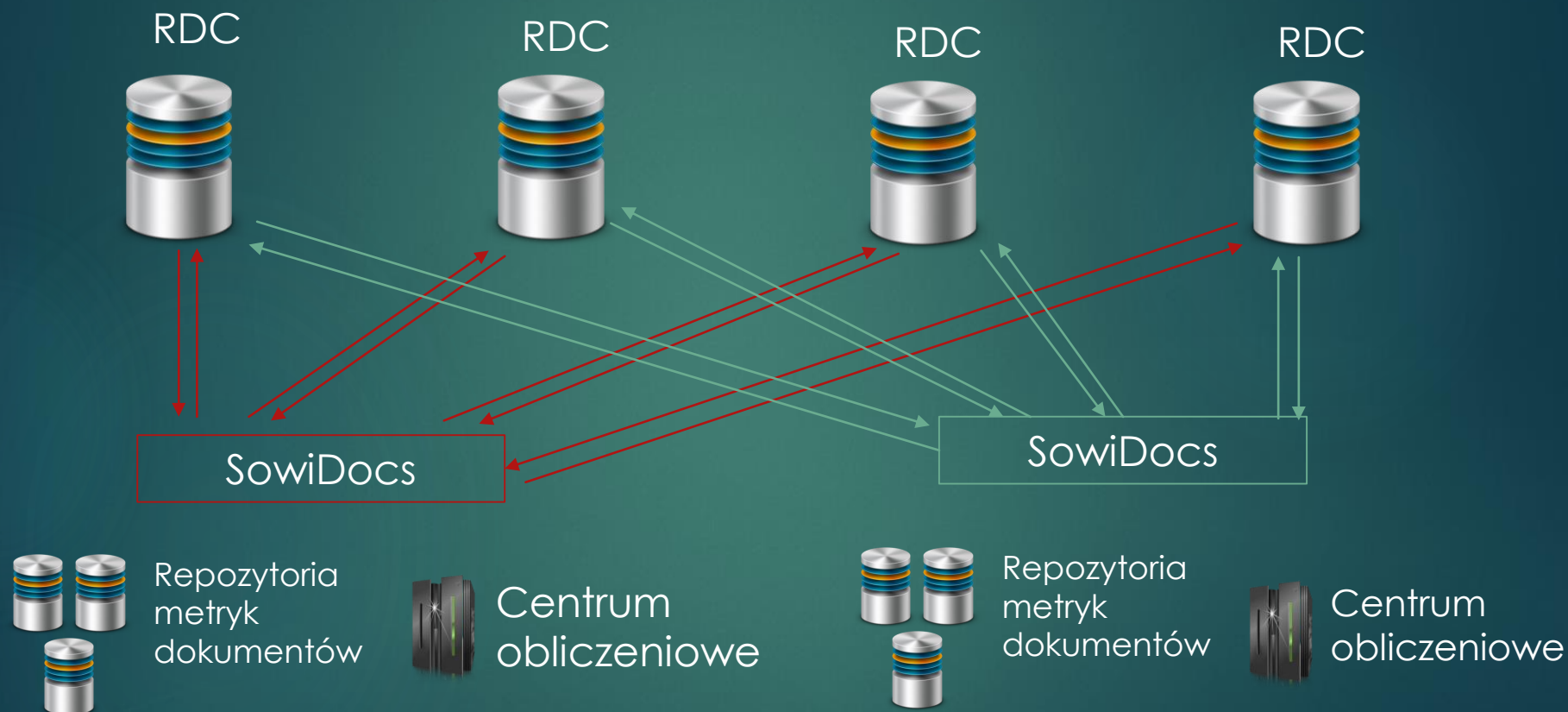


SowiSync



Materiały
wydawnictw

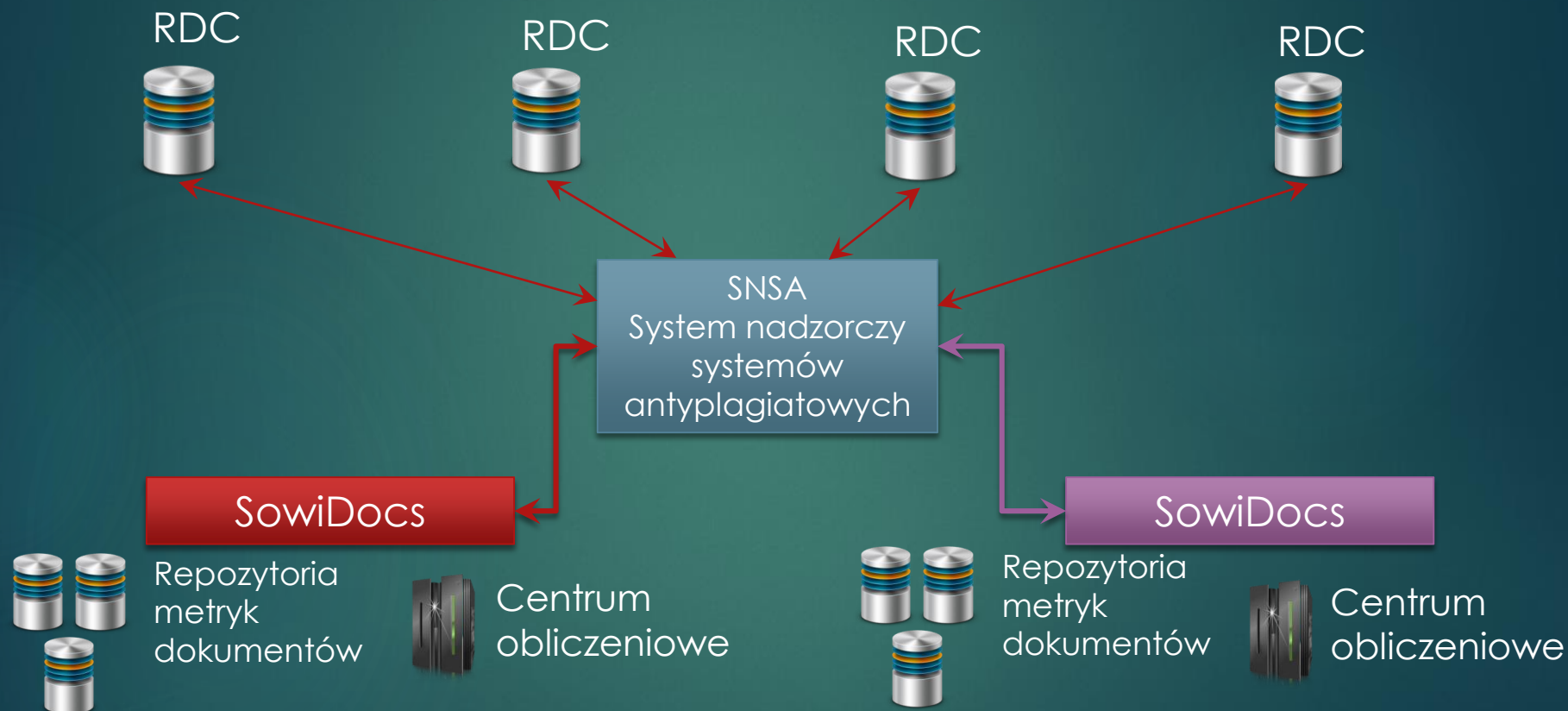
Architektura rozproszona



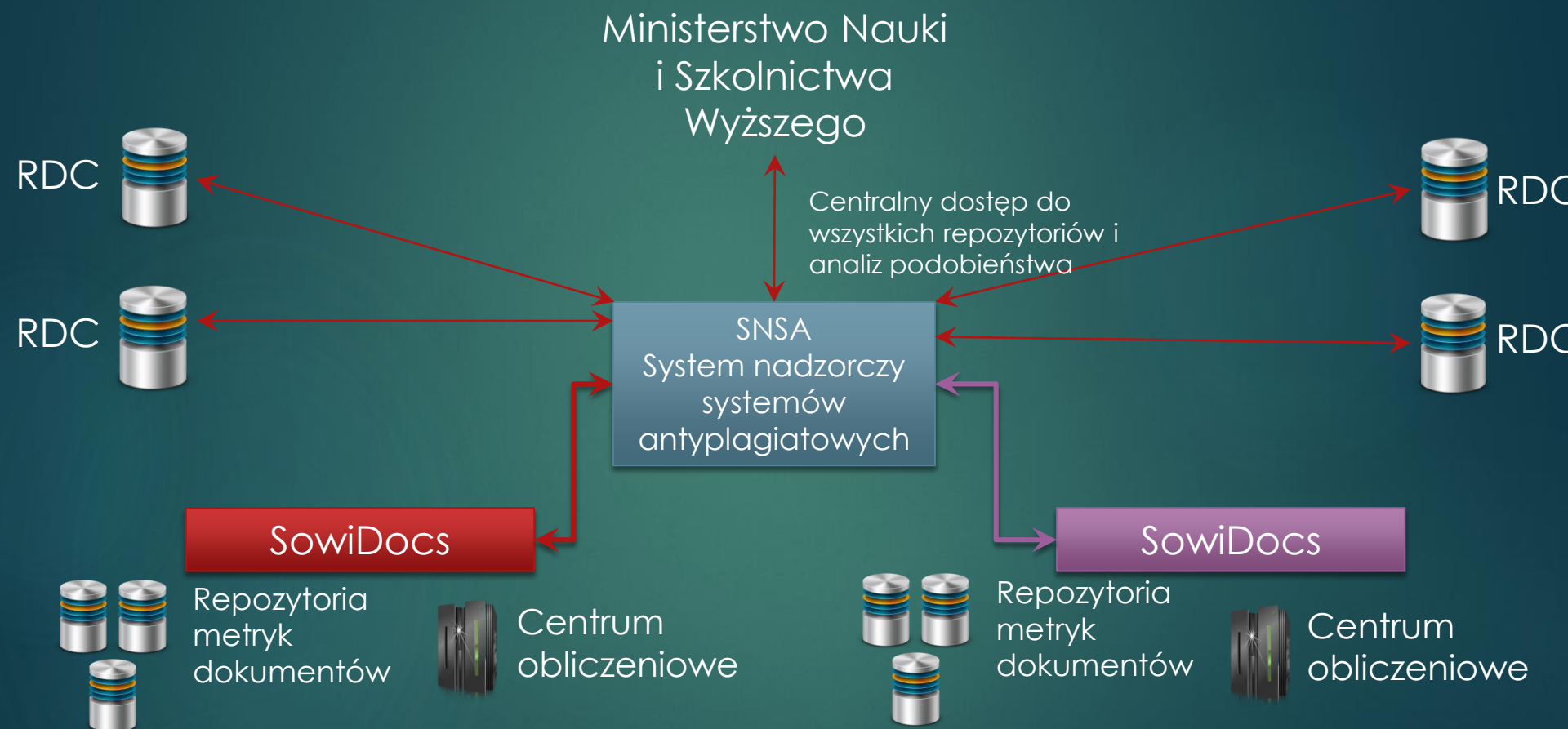
Trudności:

- Spójność danych
- Aktualność danych

Architektura rozproszona



Architektura rozproszona



Zakończenie

- ▶ Zalety wykorzystania systemu SOWI
 - ▶ Automatyczna obsługa analiz podobieństwa i archiwizacji dokumentów
 - ▶ Realizacja obliczeń na superkomputerze
 - ▶ Wykorzystanie dokładniejszych algorytmów (wyszukiwanie fragmentów podobnych)
 - ▶ Pozostawienie zasobów po stronie uczelni, możliwość utworzenia własnego repozytorium prac dyplomowych
 - ▶ Możliwość stworzenia węzła antyplagiatowego po udostępnieniu serwerów – obniżenie kosztów analizy antyplagiatowej / możliwy zysk dla uczelni