

archivist: wprowadzenie do pakietu, analiza reprodukowalności

Warsztaty badawcze 1

Konrad Komisarczyk, Mariusz Słapek



Politechnika Warszawska

`k.komisarczyk@student.mini.pw.edu.pl`
`m.słapek@student.mini.pw.edu.pl`

17 marca 2020

Plan prezentacji

1 Wprowadzenie do pakietu

2 Analiza reprodukowalności

Podstawowe informacje

archivist: An R Package for Managing, Recording and Restoring Data Analysis Results

Przemysław Biecek
University of Warsaw

Marcin Kosiński
Warsaw University of Technology

Data wydania artykułu: **6 grudnia 2017 roku**



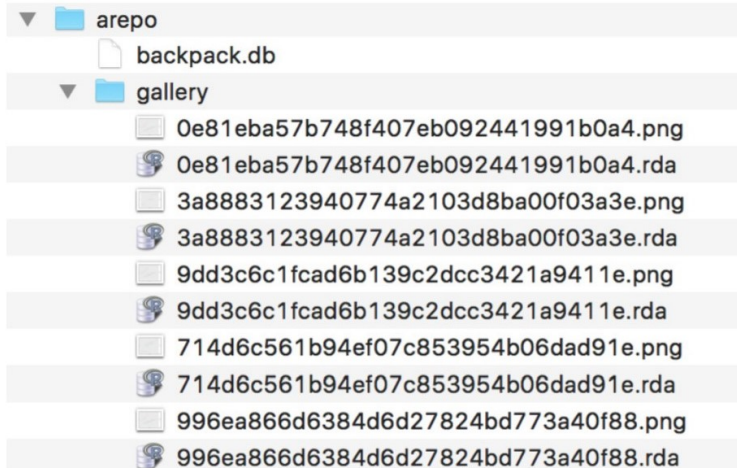
Wprowadzenie

archivist to pakiet pozwalający na łatwe i wygodne zarządzanie obiektami języka R, w tym: przechowywanie, wyszukiwanie i udostępnianie ich.

Struktura

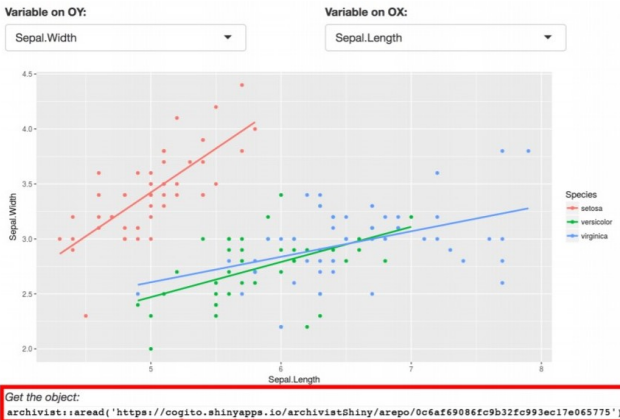
- Atrifact - obiekty R, które są zapisywane do repozytorium (identyfikowane przez hash MD5)
- Repository - kolekcja artefaktów
- MD5 hash - unikalny identyfikator artefaktu (32 literowy string)
- Tag - atrybut artefaktu, mają strukturę `key:value` (np. data utworzenia obiektu, typ - wszystkie na samym początku, typ modelu)

Struktura

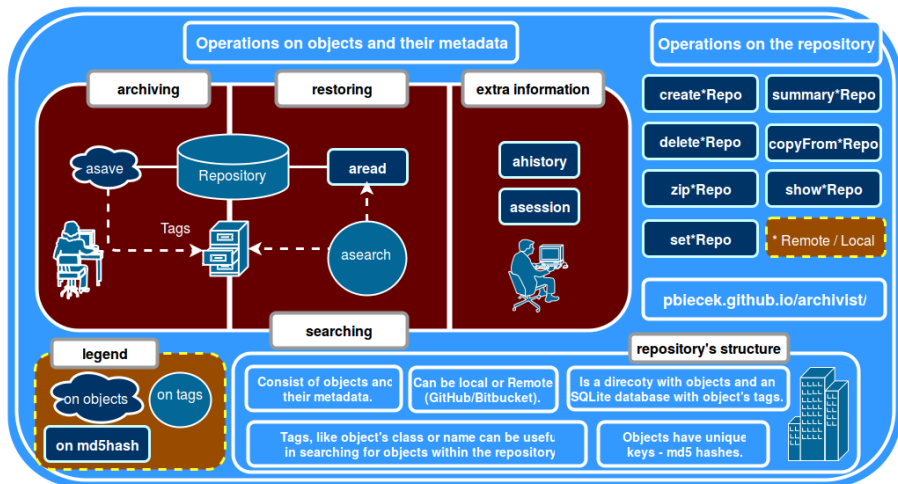


Motywacja

- Reprodukowalność
- Zarządzanie wieloma modelami
- Udostępnianie danych przez zewnętrzne aplikacje



Funkcjonalności



Prezentacja raportu *knitr*.