



Polish Open Science Conference 2024

**Data in the Service
of Science and Society**

📅 10-12.04.2024 📍 Cracow, Poland



Wiktor Florian

ACK Cyfronet AGH

Jakość Metadanych na Progu Ery Open Science: Wyzwania i Perspektywy

Metadata quality at the threshold of Open Science era: challenges and prospects

Próg Ery **Open Science**

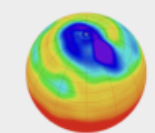
Nie tylko globalny dostęp do danych naukowych, ale i możliwość interaktywnej i dynamicznej wymiany, możliwość wykorzystania w czasie realnym.

Obecnie zaczynamy udostępniać informacje w momencie publikacji, śledzimy wszystko od tego momentu. Jednak nie śledzimy ważnych etapów w procesie badawczym, które wpływają na odtwarzalność. (ROSiE project)



Jakość i Braki **Metadanych**

- Data publikacji i jej wyświetlanie.
- Blisko 20% projektów posiadało 5 lub więcej braków metadanych. W pierwszej iteracji projekty składały się z 13 metadanych.



Centre for Environmental Data Analysis (CEDA) Document Repository

Published January 1, 9999 | Version v1

Technical note

Open

Meteosat - The operational programme

Unknown

Other: European Space Agency (ESA)

European Space Agency (ESA) brochure about the Meteosat missions from approximately 1980.

Year range

Start year

→ End year



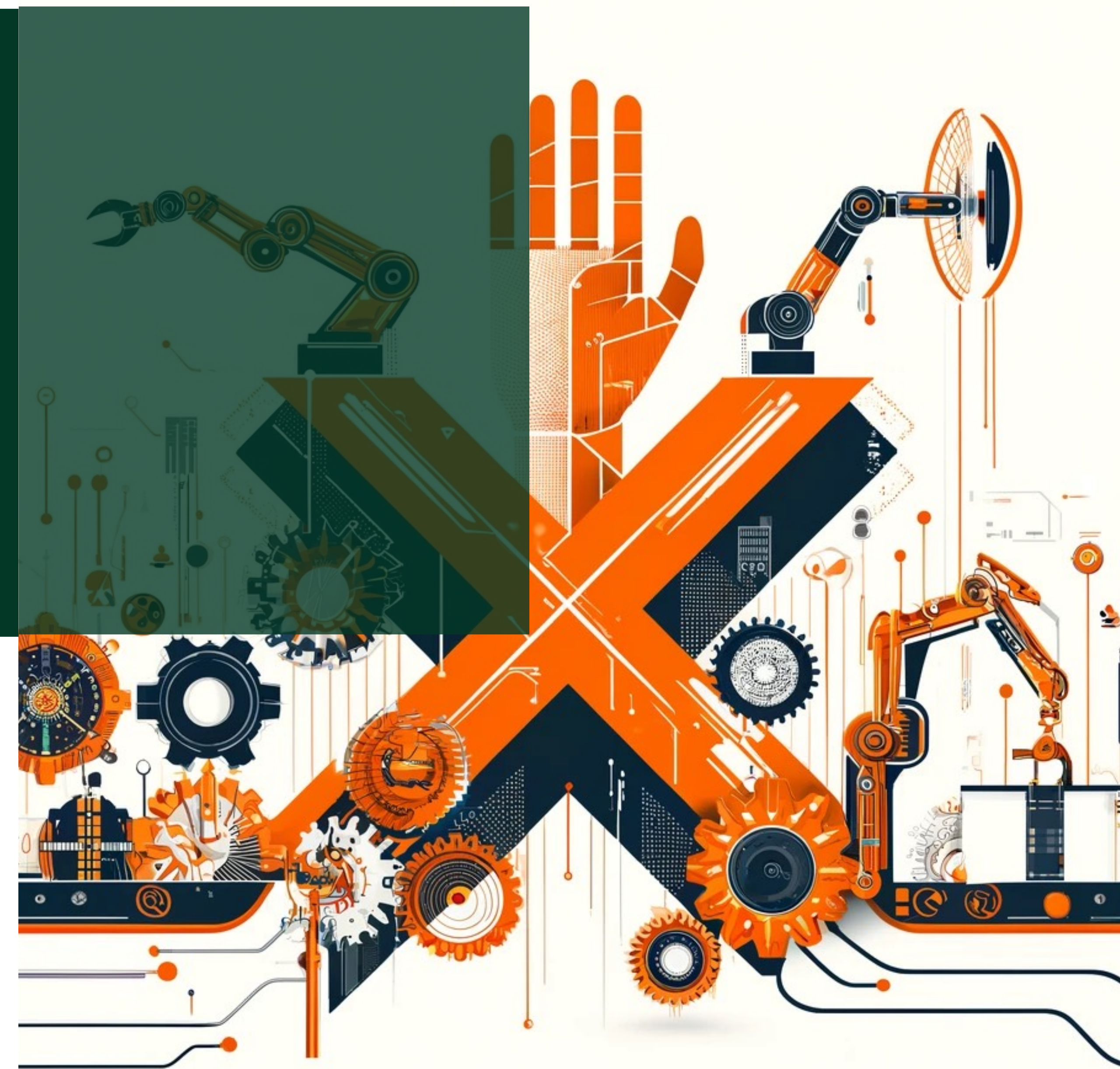
Ograniczenia **Automatyzacji**

Brak aktualnych produktów badawczych

Cytowanie - dane, które dostawaliśmy generowały niepoprawne cytowania. Skorzystaliśmy z DOI i Handle w celu odpytania API dostawcy PID o poprawne metadane.


Dane otrzymywane poprzez harvesting, zawierają dużo rekordów testowych np. Software o tytule "delete me", posiadający DOI.

Stosowanie agresywniejszych strategii deduplikacji, Pozwoliło na zredukowanie ilości rekordów "śmieciowych", jednak nadal są on obecnym zjawiskiem.



 EASY ▾ |  View all 2 versions |  Link to |  Share |  Cite |  Claim

Natuurbegraafplaats Prattenburg Rhenen

 **Publication** » *Other literature type* • 01 Jan 2021 • Embargo end date: 09 Apr 2021 • Dutch; Flemish • Publisher: Data Archiving and Networked Services

Authors: *Schorn, E.A.*;

DOI: [10.17026/dans-ztd-2qt4](https://doi.org/10.17026/dans-ztd-2qt4) 

Select a citation style

chicago ▾

Schorn, E.A. "Natuurbegraafplaats Prattenburg Rhenen." Data Archiving and Networked Services (DANS), 2021. doi:10.17026/dans-ztd-2qt4.

[COPY CITATION](#)

Select file format to download

No option selected ▾

Natuurbegraafplaats Prattenburg Rhenen

Version 1.0



E.A. Schorn, 2020, "Natuurbegraafplaats Prattenburg Rhenen", <https://doi.org/10.17026/dans-ztd-2qt4>, DANS Data Station Archaeology, V1

Cite Dataset ▾

Learn about [Data Citation Standards](#).

Description

KSP Archeologie heeft een archeologisch bureauonderzoek uitgevoerd voor de locatie aan de

```
@data{dans-ztd-2qt4_2020,
author = {E.A. Schorn},
publisher = {DANS Data Station Archaeology},
title = {{Natuurbegraafplaats Prattenburg Rhenen}},
year = {2020},
version = {V1},
doi = {10.17026/dans-ztd-2qt4},
url = {https://doi.org/10.17026/dans-ztd-2qt4}
}
```



ALL TOPICS ▾

18,462 result(s) found

Sort by Best match ▾

29760 datasets found on search for »test«

< 1 2 3 4 5 6 7 8 9 10 >

- Gerullis, L: Test**
Size: 8118 data points
<https://doi.pangaea.de/10.1594/PANGAEA.946783> – Score: 40.2
- Baranowski, UK (2019): Test-size diameters of 36 extant planktonic foraminiferal species, global oceans**
Supplement to: **Baranowski, UK (2013):** Phylogenetic significance of test morphology and ultrastructure in recent planktonic Foraminifera inferred from cladistic analyses. *Eberhard Karls University Tübingen*
Related to: **Rillo, MC; Sugawara, MTC; Cabella, B et al. (2019):** On the mismatch in the strength of competition among fossil and modern species of planktonic Foraminifera. *Global Ecology and Biogeography*
Size: 1038 data points
<https://doi.org/10.1594/PANGAEA.901825> – Score: 35.95
- Koppe, R: TEST DATASET**
Size: 1 data points
<https://doi.pangaea.de/10.1594/PANGAEA.947598> – Score: 35.07
- Koppe, R: TEST DATASET - do not touch**
Size: 7 data points
<https://doi.pangaea.de/10.1594/PANGAEA.920070> – Score: 34.56

April 10, 2015 (v2) Software Restricted

delete me please

delete, me

delete me please

Part of [Continuously Moving Table Magnetic Resonance Imaging](#)

Uploaded on June 25, 2015

1 more versions exist for this record

846 10

October 11, 2021 (v0.1.0) Other Restricted

delete me

jeanmichelscherer

delete me

Uploaded on October 11, 2021

24 1

June 5, 2018 (1.0.0) Dataset Open

DELETE ME PLEASE

Decan, Alexandre ; Mens, Tom ; Constantinou, Eleni

This is the replication package for paper "On the evolution of technical lag in the npm package dependency network" accepted for publication in ICSME 2018.

Niespójność oraz Problemy Standaryzacji Metadanych

Wprowadzając organizacje do EOSC, natknęliśmy się na specyficzny problem, otrzymaliśmy dane w których część organizacji się powtarza.



Politechnika Warszawska



RESEARCH PRODUCTS (5K)

PROJECTS (47)

DATA SOURCES (0)

ORGANIZATIONS (3)**Warsaw University of Technology****Organization** • Funded by: ANR | Wood-waste Panel, ANR | SmartBREATH, ANR | NITICS **+3 projects**[← Share](#)**POLITECHNIKA WARSZAWSKA****Organization** • Poland • Funded by: EC | Computer-based Visualisat..., EC | Common Learning Outcomes ..., EC | Management of a modern, m... **+44 projects**[← Share](#)**WUT****Organization** • Poland • Funded by: EC | INNUMAT, EC | PICs4All, EC | EUCARD-2 **+162 projects**[← Share](#)

Politechnika Krakowska



RESEARCH PRODUCTS (1K)

PROJECTS (46)

DATA SOURCES (3)

ORGANIZATIONS (2)**POLITECHNIKA KRAKOWSKA****Organization** • Poland • Funded by: EC | Development of two cycle ..., EC | The Activation of the Pub..., EC | Back to the Future - Buil... **+7 projects**[← Share](#)**PK****Organization** • Poland • Funded by: EC | PRESTO, EC | LOLIPEM, EC | KAM2SouthPL2 **+33 projects**[← Share](#)

AGH

X



RESEARCH PRODUCTS (35K)

PROJECTS (149)

DATA SOURCES (13)

ORGANIZATIONS (12)

**Wydawnictwa AGH****Organization** • Poland[Share](#)**AGH University of Science and Technology Press****Organization** • Poland[Share](#)**AGH Univeristy of Science and Technology Press****Organization** • Poland[Share](#)**AGH University of Science and Technology, Department of Telecommunications****Organization** • Poland • Funded by: CHIST-ERA | AMIS, CHIST-ERA | STAR[Share](#)**AGH University of Science and Technology****Organization** • Funded by: ANR | AMIS, ANR | META[Share](#)**Department of Telecommunications, AGH University of Science and Technology****Organization** • Funded by: ANR | STAR[Share](#)**University of Science and Technology AGH - Faculty of Applied Mathematics****Organization** • Funded by: FWF | Topology of Lorenz-like a...[Share](#)

Persistent Identifiers

(PIDs)

Ogromne trudności w deduplikacji. Różne wersje zawierające różne PIDy.

Trzy takie same publikacje, jedna zawiera trzy różne DOI, dla trzech różnych wersji, pozostałe dwie po jednym DOI.

Organizację posiadają blisko 20 różnych Persistent Identifiers.

(ISIN, ROR, GRID, WIKIDATA)

Perspektywy Rozwoju Jakości Metadanych Cz. 1



Wspólnotowe Podejście do Standaryzacji

Międzynarodowa współpraca, dialog między naukowcami, instytucjami i organizacjami pozwoli na ustanowienie wspólnych standardów.



Edukacja i świadomość wśród naukowców

Informowanie i kształcenie badaczy w zakresie tworzenia i zarządzania metadanymi jest kluczowe dla ich długoterminowej jakości i efektywności.



CDIF: Discovery Module

Kompleksowy przewodnik dostarczający wzorców i najlepszych praktyk dla infrastruktur badawczych z różnych dziedzin.

CDIF skupia się na ułatwianiu odkrywania zasobów, podkreślając znaczenie szczegółowych i dostępnych metadanych.

Perspektywy Rozwoju Jakości Metadanych Cz. 2



Nowe technologie w generowaniu metadanych

Zastosowania nowych technologii, takich jak sztuczna inteligencja, uczenie maszynowe czy przetwarzanie języka naturalnego.

Automatyzacja procesu tworzenia metadanych czy implementacja sztucznej inteligencji lepszego przeszukiwania repozytoriów.



Wykorzystanie Technologii Blockchain

Większa transparentność, bezpieczeństwo i niezmienność metadanych. Ułatwienie śledzenia pochodzenia danych i ich historii zmian.

ROSiE project - Comparison of existing blockchain technologies to safeguard responsible Open Science

Dziękuję
Za Uwagę