


README or not  
README  
that is the question





# Podstawowe pytania dotyczące dokumentacji projektu

Dla kogo?

Co?

Kiedy?

Jak?

Gdzie?

Dlaczego?

Kto?

# README w strukturze dokumentacji projektu

```
.
|-- CITATION
|-- README
|-- LICENSE
|-- requirements.txt
|-- data
|   |-- birds_count_table.csv
|-- doc
|   |-- notebook.md
|   |-- manuscript.md
|   |-- changelog.txt
|-- results
|   |-- summarized_results.csv
|-- src
|   |-- sightings_analysis.py
|   |-- runall.py
```

# Wybór repozytorium

Źródło:  
<https://guides.tricolib.brynmawr.edu/c.php?g=721140&p=5139491>

## Domain-specific repositories

[Finding domain-specific repositories](#)

[Astronomy repositories](#)

[Biology repositories](#)

[Chemistry repositories](#)

[Computer science repositories](#)

[Economics repositories](#)

[Education repositories](#)

[English repositories](#)

[Environmental sciences repositories](#)

[Fine arts repositories](#)

[Health studies repositories](#)

[History repositories](#)

[Linguistics repositories](#)

[Marine sciences repositories](#)

[Physics repositories](#)

[Political science repositories](#)

[Psychology repositories](#)

[Religion repositories](#)

[Sociology repositories](#)

## Astronomy repositories

- **[NASA Space Science Data Coordinated Archive \(NSSDCA\)](#)**

NSSDCA serves as the permanent archive for NASA space science mission data. "Space science" means astronomy and astrophysics, solar and space plasma physics, and planetary and lunar science. As permanent archive, NSSDCA teams with NASA's discipline-specific space science "active archives" which provide access to data to researchers and, in some cases, to the general public. [Information on submitting data.](#)

## Biology repositories

- **[Dryad](#)**

Dryad hosts research data underlying scientific and medical publications. Historically, the repository has been strongest in the life sciences. Most data in the repository is associated with peer-reviewed journal articles, but data associated with non-peer reviewed publications from other reputable sources (such as dissertations and books) is also accepted. [Information on submitting data.](#)

- **[GenBank](#)**

GenBank is the NIH genetic sequence database, an annotated collection of all publicly available DNA sequences. GenBank is part of the International Nucleotide Sequence Database Collaboration, which comprises the DNA DataBank of Japan (DDBJ), the European Nucleotide Archive (ENA), and GenBank at NCBI. These three organizations exchange data on a daily basis. [Information on submitting data.](#)

- **[Gene Expression Omnibus \(GEO\)](#)**

Supported by the NIH, the Gene Expression Omnibus (GEO) is a public functional genomics data repository supporting MIAME-compliant data submissions. Array- and sequence-based data are accepted. Tools are provided to help users query and download experiments and curated gene expression profiles. [Information on submitting data.](#)

- **[Protein Data Bank \(PDB\)](#)**

The PDB archive is the single worldwide repository of information about the 3D structures of large biological molecules, including proteins and nucleic acids. [Information on submitting data.](#)

## Chemistry repositories

- **[Cambridge Structural Database \(CSD\)](#)**

# Korzystanie z wzorców README,

np. z Cornell University

## Methodological information

1. **Description of methods for data collection or generation** (include links or references to publications or other documentation containing experimental design or protocols used)
2. **Description of methods used for data processing (describe how the data were generated from the raw or collected data)**
3. Any software or instrument-specific information needed to understand or interpret the data, including software and hardware version numbers
4. Standards and calibration information, if appropriate
5. Describe any quality-assurance procedures performed on the data
6. Definitions of codes or symbols used to note or characterize low quality/questionable/outliers that people should be aware of
7. People involved with sample collection, processing, analysis and/or submission

## Data-specific information

\*Repeat this section as needed for each dataset (or file, as appropriate)\*

1. Count of number of variables, and number of cases or rows
2. **Variable list, including full names and definitions (spell out abbreviated words) of column headings for tabular data**
3. **Units of measurement**
4. **Definitions for codes or symbols used to record missing data**
5. Specialized formats or other abbreviations used



# Dobre praktyki dla pliku README

- <https://data.research.cornell.edu/data-management/sharing/readme/>
- [https://data.4tu.nl/s/documents/Guidelines\\_for\\_creating\\_a\\_README\\_file.pdf](https://data.4tu.nl/s/documents/Guidelines_for_creating_a_README_file.pdf)
- [https://uwyo.libguides.com/ld.php?content\\_id=61896490](https://uwyo.libguides.com/ld.php?content_id=61896490)
- Format README: .TXT lub Markdown (.md) i zawartość „plain text”
- README jako dokument nawigacyjny datasetu (m.in. opis struktury plików)
- Wybrane wskazówki z wymagań dla repozytoriów CoreTrustSeal
  - zapewniona jest **kompletność i zrozumiałość danych i metadanych**
  - wszystkie przepływy danych są udokumentowane
  - metadane i dokumentacja wspierają zrozumienie i ponowne użycie dla określonej społeczności
- Zawartość zależy od specyfiki danej dyscypliny badawczej
- Strukturyzowana forma
- Standaryzowane nagłówki (patrz projekt SoMEF)
- Pole metadanych Opis (Description) jako kopia pliku README



# Informacje kontaktowe do autora

- [zbigniew.ruszczyk@ug.edu.pl](mailto:zbigniew.ruszczyk@ug.edu.pl)
- Tel. 58 523 32 67
- Uniwersytet Gdański
- Sekcja Zarządzania Danymi Badawczymi i Otwartej Nauki