Vilniaus universitetas

# Atvirojo mokslo perspektyvos

## Publikacijos, programos, duomenys

Saulius Gražulis  Algirdas Grybauskas
Andrius Merkys  Antanas Vaitkus

Vilniaus universitetas
Gyvybės mokslų centras
Biotechnologijos institutas

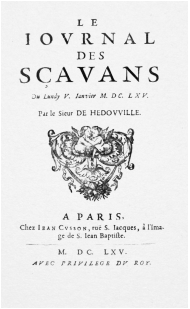
https://www.crystallography.net/cod/archives/2024/slides/atviram-mokslui/slides.pdf

Vilnius, 2024

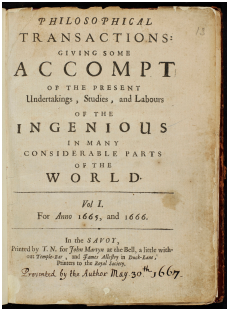
Id: slides.tex 2935 2024-10-21 07:59:09Z vantanas  2024 m. spalio 21 d.

# Mokslas gyvas žinių mainais

Prieš maždaug 360 metų atsirado pirmieji žurnalai *mokslo žinioms* publikuoti:

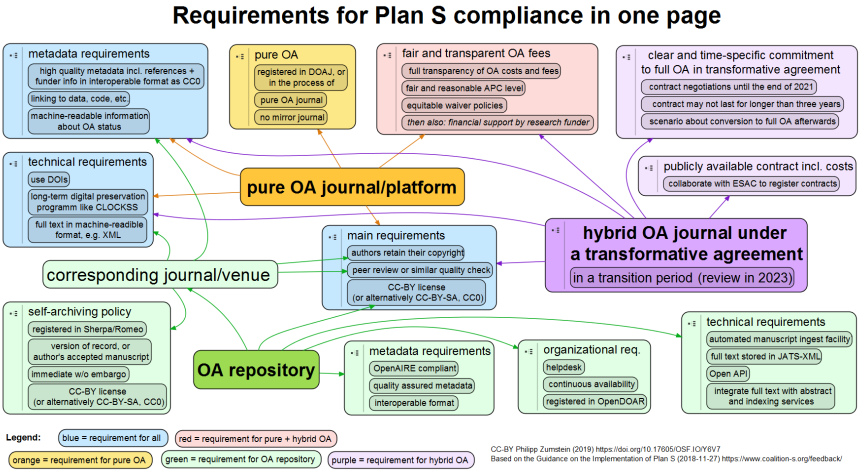

LE
IOVRNAL
DES
SÇAVANS
Du Lundy 5. Ianvier M. DC. LXV.
Par le Sieur DE HEDOVVILLE.
A PARIS.
Chez IEAN CVSSON, rue S. Iacques, à l'Image de S. Iean Baptiste.
M. DC. LXV.
AVEC PRIVILEGE DV ROY.

Journal des savants
(1665 sausis)


PHILOSOPHICAL
TRANSACTIONS:
GIVING SOME
ACCOMPT
OF THE PRESENT
Undertakings, Studies, and Labours
OF THE
INGENIOUS
IN MANY
CONSIDERABLE PARTS
OF THE
WORLD.
Vol. I.
For Anno 1665, and 1666.
In the SAVOY,
Printed by T. N. for John Martyn at the Bell, a little without Temple-Bar, and James Allestry in Duck-Lane, Printers to the Royal Society.
Perused, by the Author May 30th 1667.

Philosophical Transactions of the
Royal Society (1665 kovas)

# Atvira prieiga prie straipsnių




https://osf.io/dy6v7/

# Atviros ir laisvos programos

Programinė įranga turi būti laisva ir visiems prieinama!

FREE SOFTWARE FOUNDATION

https://www.gnu.org/

"The users have the freedom to **run**, **copy**, distribute, **study**, change and **improve** the software"

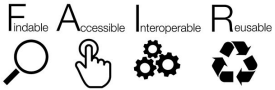
open source initiative®

https://opensource.org/

https://opensource.org/license

# Atviri ir FAIR duomenys

US National Weather Records Center: 400 million punched cards (1960)




FAIR (Wilkinson ir kt. 2016), FAIR/O

NIH National Library of Medicine
National Center for Biotechnology Information

PubMed.gov

PROTEIN DATA BANK

Years

# Atviri duomenys kristalografijoje

Vilniaus universitetas

Atviros prieigos:

# Atviri duomenys kristalografijoje

Atviros prieigos:

- Protein Data Bank (Berman ir kt. 2007);

# Atviri duomenys kristalografijoje

Atviros prieigos:

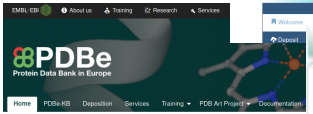
- Protein Data Bank (Berman ir kt. 2007);

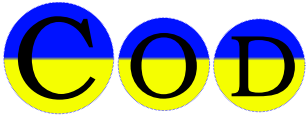

# Atviri duomenys kristalografijoje

Atviros prieigos:

- Protein Data Bank (Berman ir kt. 2007);
- Crystallography Open Database (and its “sisters”);





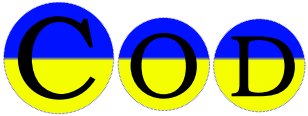



# Atviri duomenys kristalografijoje

Atviros prieigos:

- Protein Data Bank (Berman ir kt. 2007);
- Crystallography Open Database (and its “sisters”);
- Bilbao Cryst. Server (Aroyo ir kt. 2006)



# Atviri duomenys kristalografijoje

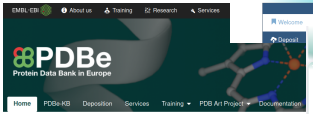
Atviros prieigos:

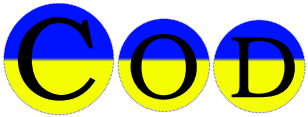
- Protein Data Bank (Berman ir kt. 2007);
- Crystallography Open Database (and its “sisters”);
- Bilbao Cryst. Server (Aroyo ir kt. 2006)

Uždaros:

- CCDC
- ICSD
- PDF
- Pauling File
- ...





# Atviri duomenys kristalografijoje

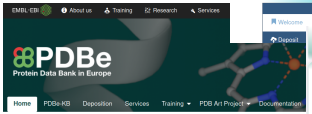
Atviros prieigos:

- Protein Data Bank (Berman ir kt. 2007);
- Crystallography Open Database (and its “sisters”);
- Bilbao Cryst. Server (Aroyo ir kt. 2006)

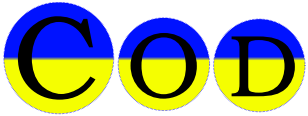
Uždaros:

- CCDC
- ICSD
- PDF
- Pauling File
- ...










Šiuo metu žinoma apie $\mathbf{10^6 - 10^7}$ kristalinių struktūrų.

# COD atsiradimas

But what if crystallographers work together to establish a public domain database with all relevant crystallographic data? This would not only overcome the current situation with 'fragmented' databases, it would also prevent for becoming dependent from monopolists.

What would be needed?

1. A small team of engaged scientists with some experience in database and software design to coordinate the project.

2. The authors (i.e. the scientific community = YOU) who provides the project with database entries (note, that if you have'nt sold your experimental results exclusively, you are free to distribute the data to such a database, even if they have already been part of a publication - and a lot of good data have never been published).

3. Free software a) for maintaining the database, b) for data evaluation and calculation of derived data (e.g. calculated powder pattern from crystal structures for search-match purposes), c) for browsing and retrieval.

gemstonede (Dr. Michael BERNDT) Fri Feb 14, 2003 1:26 pm

# COD atsiradimas

But what if crystallographers work together to establish a public domain database with all relevant crystallographic data? This would not only overcome the current situation with 'fragmented' databases, it would also prevent for becoming dependent from monopolists.

What would be needed?

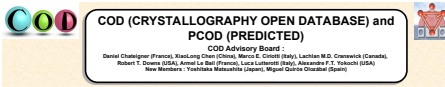
1. A small team of engaged scientists with some experience in database and software design to coordinate the project.

2. The authors (i.e. the scientific community = YOU) who provides the project with database entries (note, that if you have'nt sold your experimental results exclusively, you are free to distribute the data to such a database, even if they have already been part of a publication - and a lot of good data have never been published).

3. Free software a) for maintaining the database, b) for data evaluation and calculation of derived data (e.g. calculated powder pattern from crystal structures for search-match purposes), c) for browsing and retrieval.

gemstonede (Dr. Michael BERNDT) Fri Feb 14, 2003 1:26 pm

COD įkūrimas (2003), IUCr pranešimas (2005):




(Chateigner et al., 2005)

# COD atsiradimas

But what if crystallographers work together to establish a public domain database with all relevant crystallographic data? This would not only overcome the current situation with 'fragmented' databases, it would also prevent for becoming dependent from monopolists.

What would be needed?

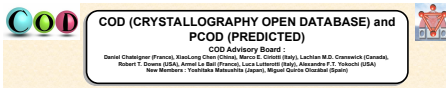
1. A small team of engaged scientists with some experience in database and software design to coordinate the project.

2. The authors (i.e. the scientific community = YOU) who provides the project with database entries (note, that if you have'nt sold your experimental results exclusively, you are free to distribute the data to such a database, even if they have already been part of a publication - and a lot of good data have never been published).

3. Free software a) for maintaining the database, b) for data evaluation and calculation of derived data (e.g. calculated powder pattern from crystal structures for search-match purposes), c) for browsing and retrieval.

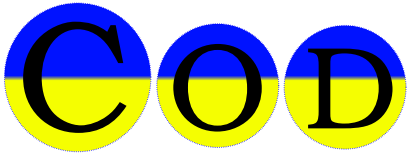
gemstonede (Dr. Michael BERNDT) Fri Feb 14, 2003 1:26 pm

COD įkūrimas (2003),
IUCr pranešimas (2005):




(Chateigner et al., 2005)

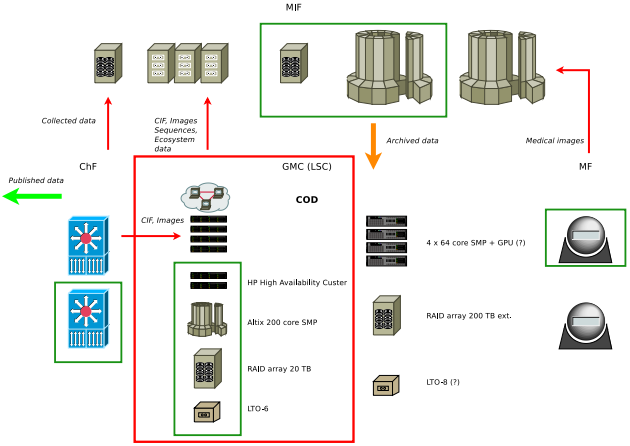
Po 20 metų...



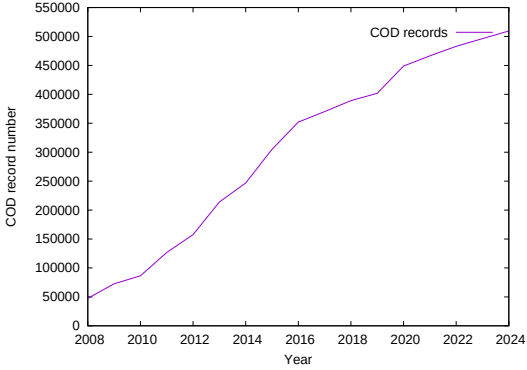

**there are 518746 entries in the COD** https://crystallography.net

# COD perkėlimas į Vilniaus universitetą

2007 m. COD serveriai perkelti į Biotechnologijos institutą, kuris 2016 m. prisijungė prie Vilniaus universiteto

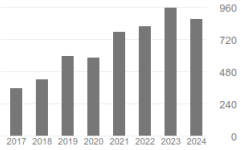


# COD augimas



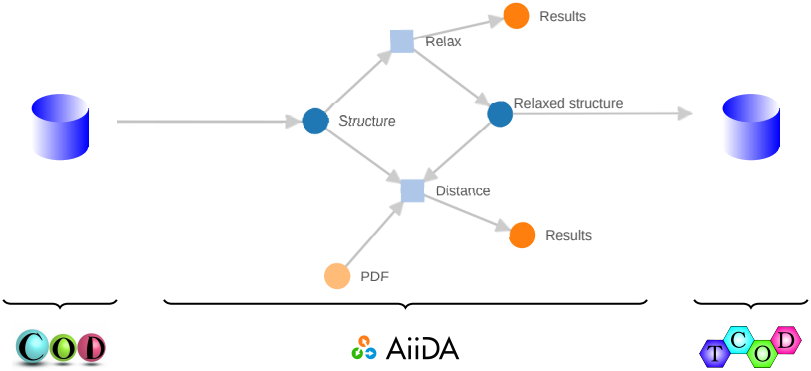

| TITLE | CITED BY | YEAR |
|---|---|---|
| Crystallography Open Database–an open-access collection of crystal structures<br>S Gražulis, D Chateigner, RT Downs, AFT Yokochi, M Quirós, L Lutterotti, ...<br>Journal of applied crystallography 42 (4), 726-729 | 1786 | 2009 |
| Crystallography Open Database (COD): an open-access collection of crystal structures and platform for world-wide collaboration<br>S Gražulis, A Daškevič, A Merkys, D Chateigner, L Lutterotti, M Quirós, ...<br>Nucleic Acids Res 40, D420-D427 | 1258 | 2012 |
| AceDRG: a stereochemical description generator for ligands<br>F Long, RA Nicholls, P Emsley, S Gražulis, A Merkys, A Vaitkus, ...<br>Acta Crystallographica Section D: Structural Biology 73 (2), 112-122 | 340 | 2017 |

| | | |
|---|---|---|
| Citations | 7155 | 4607 |
| h-index | 36 | 23 |
| i10-index | 61 | 46 |




(Šaltinis: Google Scholar, 2024-10-20)

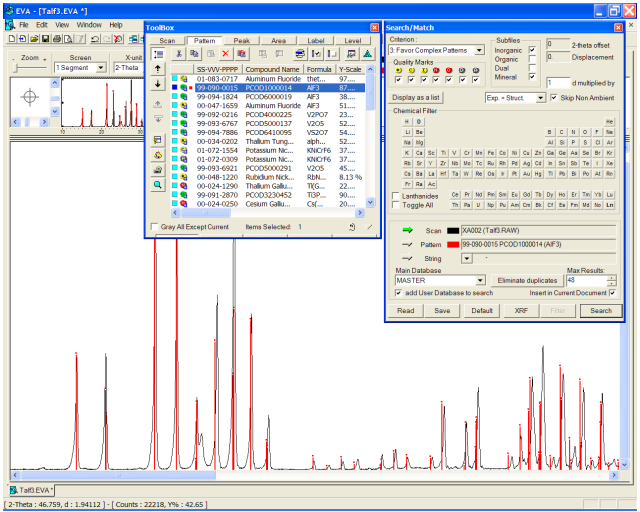
# COD panaudojimas




- COD duomenys panaudojami skaičiuojamojoje medžiagotyroje

(Merkys ir kt. 2017; Pizzi 2018)

# COD panaudojimas

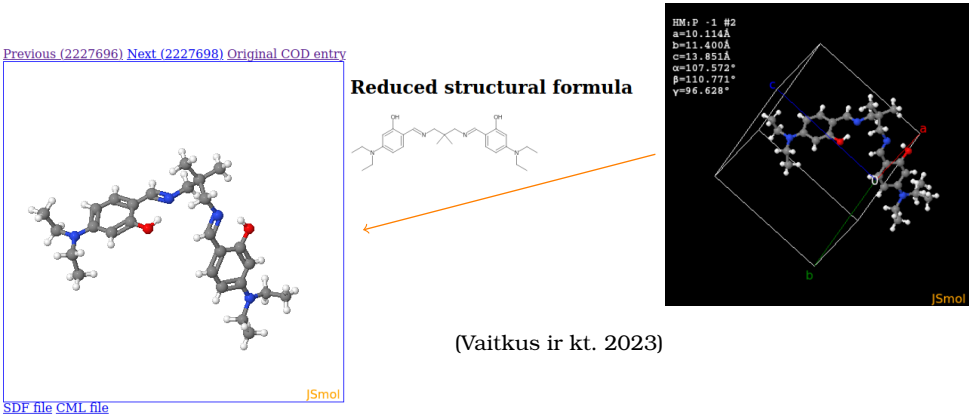



PCOD duomenų panaudojimas medžiagų identifikavimui, ir galimas analogiškas COD panaudojimas.

Paveikslėlis panaudotas leidus Armel Le Bail (Le Bail 2008)

# COD panaudojimas

https://molecules.crystallography.net/




(Vaitkus ir kt. 2023)

**Reduced canonical SMILES:**

CCN(c1ccc(c(c1)O)/C=N/CC(C/N=C/c1ccc(cc1O)N(CC)CC)(C)C)CC **(x1)** PubChem

# Nauji publikavimo būdai

Siūlomas naujas publikacijos tipas:

Recenzuojamas duomenų bazės įrašas!

COD duomenys gali būti **cituojami**:

- Nukreipiant į mašina skaitomus duomenis:

  Srivastava, R. C.; Klooster, W. T.; Koetzle, T. F. “Neutron Structures of Ammonium Fluoroberyllate” (1999) *The Crystallography Open Database*, rev. 176759, the COD Advisory Board (eds.), https://www.crystallography.net/cod/2002926.cif@176759. [Retrieved 2016-09-21 16:48 EEST]

- Nukreipiant į žmogui skaitomą Web puslapį:

  Srivastava, R. C.; Klooster, W. T.; Koetzle, T. F. “Neutron Structures of Ammonium Fluoroberyllate” (1999) *The Crystallography Open Database*, rev. 176759, the COD Advisory Board (eds.), https://www.crystallography.net/cod/2002926. [Retrieved 2016-09-21 16:48 EEST]

# COD perspektyvos

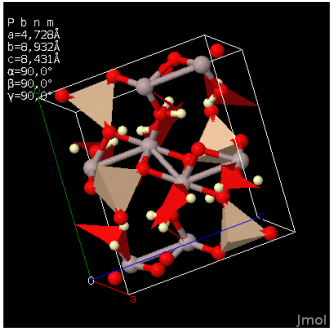
Ilgalaikiai COD tikslai:

- Surinkti galimai visas žinomas kristalų struktūras (nesančias PDB);
- Įgyvendinti duomenų recenzavimo mechanizmus, iš esmės pagerinant publikuojamų duomenų kokybę;
- Prisidėti prie Atviro mokslo, užtikrinant atvirą prieigą prie FAIR duomenų;
- Telkti mokslo bendruomenę ir populiarinti Atviro mokslo principus.

# KICIS komanda

Antanas Vaitkus  Algirdas Grybauskas  Saulius Gražulis  Andrius Merkys

# Ačiū už dėmesį!




| | |
|---|---|
| **Coordinates** | 2207377.cif |
| **Original IUCr paper** | HTML |

https://en.wikipedia.org/wiki/Topaz

https://www.crystallography.net/2207377.html

https://www.crystallography.net/cod/archives/2024/slides/atviram-mokslui/slides.pdf

# Nuorodos I



Aroyo, Mois Ilia et al. (Jan. 2006). “Bilbao crystallographic server: I. Databases and crystallographic computing programs.” In: *Zeitschrift für Kristallographie - Crystalline Materials* 221.1, pp. 15–27. ISSN: 2194-4946. DOI: 10.1524/zkri.2006.221.1.15.


Berman, Helen et al. (2007). “The worldwide Protein Data Bank (wwPDB): ensuring a single, uniform archive of PDB data.” In: *Nucleic Acids Research* 35, pp. D301–D303. DOI: 10.1093/nar/gkl971. eprint: http://nar.oxfordjournals.org/content/35/suppl_1/D301.full.pdf+html. URL: http://nar.oxfordjournals.org/content/35/suppl_1/D301.abstract.


Gražulis, Saulius, Daniel Chateigner, et al. (2009). “Crystallography Open Database – an open-access collection of crystal structures.” In: *Journal of Applied Crystallography* 42, pp. 726–729. DOI: 10.1107/S0021889809016690. URL: http://dx.doi.org/10.1107/S0021889809016690.


Gražulis, Saulius, Adriana Daškevič, et al. (2012). “Crystallography Open Database (COD): an open-access collection of crystal structures and platform for world-wide collaboration.” In: *Nucleic Acids Research* 40, pp. D420–D427. DOI: 10.1093/nar/gkr900. URL: http://nar.oxfordjournals.org/content/40/D1/D420.abstract.


Gražulis, Saulius, Andrius Merkys, et al. (2018). “Crystallography Open Database (COD).” In: *Handbook of Materials Modeling*. Springer International Publishing, pp. 1–19. DOI: 10.1007/978-3-319-42913-7_66-1.

# Nuorodos II


Le Bail, Armel (2008). “Frontiers Between Crystal-structure Prediction and Determination by Powder Diffractometry.” In: *Powder Diffraction Suppl.*, S5–S12. DOI: 10.1154/1.2903488.

Merkys, Andrius et al. (Nov. 15, 2017). “A posteriori metadata from automated provenance tracking: Integration of AiiDA and TCOD.” In: *Journal of Cheminformatics* 9.1, p. 56. DOI: 10.1186/s13321-017-0242-y. arXiv: 1706.08704v3 [cond-mat.mtrl-sci]. URL: https://jcheminf.springeropen.com/articles/10.1186/s13321-017-0242-y.

Pizzi, Giovanni (2018). “Open-Science Platform for Computational Materials Science: AiiDA and the Materials Cloud.” In: *Handbook of Materials Modeling*. Springer International Publishing, p. 1. DOI: 10.1007/978-3-319-42913-7_64-1.

Vaitkus, Antanas et al. (Dec. 2023). “A workflow for deriving chemical entities from crystallographic data and its application to the Crystallography Open Database.” In: *Journal of Cheminformatics* 15.1. ISSN: 1758-2946. DOI: 10.1186/s13321-023-00780-2.

Wilkinson, Mark D. et al. (Mar. 2016). “The FAIR guiding principles for scientific data management and stewardship.” In: *Scientific Data* 3.1. DOI: 10.1038/sdata.2016.18.