

# Atviro mokslo gairių projektas

Saulius Gražulis

Vilnius, 2022

Vilniaus Universiteto Gyvybės mokslų centro Biotechnologijos institutas



Id: slides.tex 1524 2020-06-04 11:58:32Z saulius January 20, 2022



# Mokslo rezultatai yra bendras turtas

- Prieiga prie mokslo žinių yra *universali žmogaus teisė*<sup>1</sup>
- Visuomenė *turi teisę* pamatyti mokslinių tyrimų, padarytų už mokesčių mokėtojų pinigus, rezultatus;
- Mokslas yra bendra vertybė (“public good”) [Boulton(2021)];

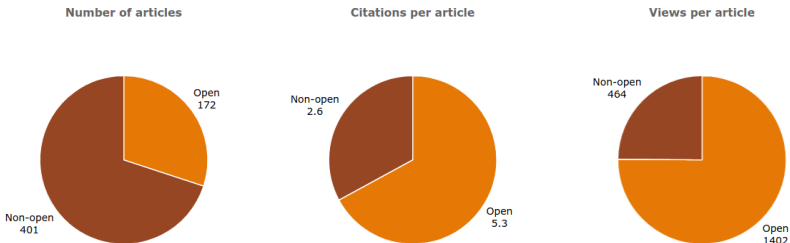
⇒ Mokslo rezultatai turėtų būti publikuojami atvirai.

---

<sup>1</sup><https://en.unesco.org/udhr>, žiūrėta 2022-01-13

## Atviri straipsniai cituojami daugiau!

Citations of open-access articles published since 2019



*Journal of Applied Crystallography, IUCr*

<https://journals.iucr.org/j/services/openaccess.html>

# Atviros duomenų bazės pasaulyje: PDB

<https://www.ebi.ac.uk/pdbe/>

The screenshot shows the Protein Data Bank in Europe (PDBe) homepage. At the top, there is a navigation bar with 'EMBL-EBI' logo and links for 'Services', 'Research', 'Training', and 'About us'. Below this is the main header 'Protein Data Bank in Europe' with the tagline 'Bringing Structure to Biology'. A search bar is located on the right side of the header. The main content area is divided into several sections: 'New PDBe-KB COVID-19 Data Portal' featuring a tree-like structure and text about SARS-CoV-2 structures; 'Featured structure' with a photo of a cat and text about microbial world; 'News' section with three articles: '3D-Beacons Network: protein structure data, all in one place' (16 September, 2021), 'AlphaFold's protein structure predictions now available to explore' (23 July, 2021), and 'A Tribute to Prof. Richard R. Ernst'; 'Events' section with a message that there are no upcoming events; 'Popular' section with a list of links like 'PDBe-KB', 'EMResearch', 'PDBeFold', etc.; 'Latest archive statistics' showing 184202 entries as of 17 November 2021; and a 'Tweets' section featuring a tweet from @PDBeurope about Alcohol dehydrogenases (ADH) enzymes. At the bottom right of the page, there is a video player showing a 3D protein structure.

(žiūrėta 2021-11-23)

# Atviros duomenų bazės pasaulyje: AlphaFold

<https://deepmind.com/research/open-source/alphafold-protein-structure-database><sup>2</sup>

DeepMind > Research > AlphaFold Protein Structure Database

OPENSOURCE

22 JUL 2021

SHARE

OPEN SOURCE LINKS

VIEW SOURCE

VIEW BLOG POST

VIEW PUBLICATION

FURTHER READING

Sciences

## AlphaFold Protein Structure Database

AlphaFold is our AI system that predicts a protein's 3D structure from its amino acid sequence. In CASP14, AlphaFold was the top-ranked protein structure prediction method by a large margin, producing predictions with high accuracy, many of which are competitive with experimentally-determined measurements.

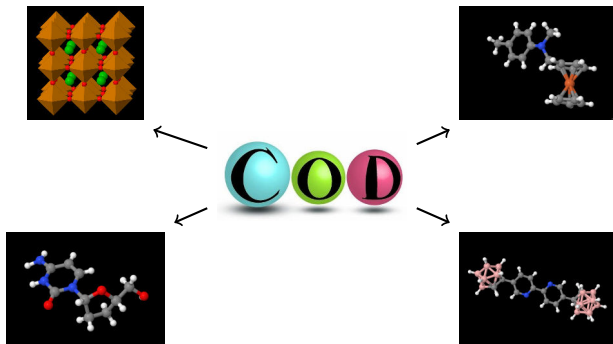
We've partnered with Europe's flagship laboratory for life sciences - EMBL's European Bioinformatics Institute (EMBL-EBI) - to create the AlphaFold Protein Structure Database to make these predictions freely available to the scientific community.

The initial release of the database covers all of the 20,000 proteins in the human proteome, along with the proteomes of several other biologically significant organisms, from E.coli to yeast, and from the fruit fly to the mouse. In the coming months we plan to expand the database to cover a large proportion of all the 100 million proteins catalogued in the [UniRef30 database](#).

“Our models are trained on structures extracted from the PDB”  
[Senior et al. (2020)].

<sup>2</sup>(žiūrėta 2021-11-23)

The Crystallography Open Database (COD)  
<https://www.crystallography.net>



483350 įrašų 2022-01-13, platinami pagal CC0 Licenciją

# COD duomenų bazės populiarumas

## Atviri duomenys gerai cituojami!<sup>3</sup>



Saulius Gražulis

Senior research fellow, [Vilnius University](#) Institute of Biotechnology  
Verified email at ibt.lt

[X-ray crystallography](#) [scientific databases](#) [software engineering](#) [computer languages](#)  
[bioinformatics](#)

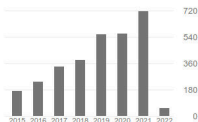
FOLLOW

TITLE	CITED BY	YEAR
<input type="checkbox"/> <a href="#">Crystallography Open Database—an open-access collection of crystal structures</a> S Gražulis, D Chateigner, RT Downs, AFT Yokochi, M Quirós, L Lutterotti, ... Journal of applied crystallography 42 (4), 726-729	1053	2009
<input type="checkbox"/> <a href="#">Crystallography Open Database (COD): an open-access collection of crystal structures and platform for world-wide collaboration</a> S Gražulis, A Daškevič, A Merkys, D Chateigner, L Lutterotti, M Quirós, ... Nucleic Acids Res 40, D420-D427	755	2012
<input type="checkbox"/> <a href="#">Structure of the tetrameric restriction endonuclease Ngo MIV in complex with cleaved DNA</a> M Deibert, S Gražulis, G Sasnauskas, V Siksnys, R Huber Nature structural biology 7 (9), 792-799	191	2000
<input type="checkbox"/> <a href="#">AceDRG: a stereochemical description generator for ligands</a> F Long, RA Nicholls, P Emsley, S Gražulis, A Merkys, A Vaitkus, ... Acta Crystallographica Section D: Structural Biology 73 (2), 112-122	161	2017
<input type="checkbox"/> <a href="#">Crystal Structure of Citrobacter freundii Restriction Endonuclease Cfr10I at 2.15 Å Resolution</a> D Bozic, S Gražulis, V Siksnys, R Huber Journal of molecular biology 255 (1), 176-186	130	1996
<input type="checkbox"/> <a href="#">COD::CIF::Parser: an error-correcting CIF parser for the Perl language</a> A Merkys, A Vaitkus, J Butkus, M Okulic-Kazanas, V Kairys, S Gražulis Journal of applied crystallography 49 (1), 292-301	129	2016

Cited by

VIEW ALL

	All	Since 2017
Citations	4360	2619
h-index	30	20
h10-index	47	37



Public access

VIEW ALL

6 articles	22 articles
not available	available

Based on funding mandates

Co-authors

EDIT

No co-authors

<sup>3</sup><https://scholar.google.com/citations?hl=en&user=BswX10sAAAAJ>, žiūrėta  
2022-01-13

TABLE 31  
OUTCOMES OF TESTS OF SIGNIFICANCE FOR FOUR  
PSYCHOLOGY RESEARCH JOURNALS

Journals: All Issues From January To December	Total Number of Research Reports (1)	Number of Research Re- ports Using Tests of Significance (2)	Number of Research Re- ports that Reject $H_0$ with $\Pr(E H_0) \leq .05$ (3)	Number of Research Re- ports that Fail to Reject $H_0$ (4)	Number of Research Reports That are Rep- lication of Previously Published Experiments (5)
Experimental Psychology (1955) Comparative and Physiological Psychology (1956)	124	106	105	1	0
Clinical Psychology (1955)	81	62	59	3	0
Social Psychology (1955)	39	32	31	1	0
Total	362	294	286	8	0

[Sterling(1959)]



<https://clinicaltrials.gov/>

 U.S. National Library of Medicine

**ClinicalTrials.gov**

Find Studies ▾

About Studies ▾

Submit Studies ▾

Resources ▾

ClinicalTrials.gov is a database of privately and publicly funded clinical studies conducted around the world.

Explore **400,873** research studies in all 50 states and in 220 countries.

See [listed clinical studies](#) related to the coronavirus disease (COVID-19)

Find a study (all fields optional)

Status ⓘ

- Recruiting and not yet recruiting studies
- All studies

<https://clinicaltrials.gov/ct2/results?cond=COVID-19>

Row	Saved	Status	Study Title	Conditions	Interventions
1	<input type="checkbox"/>	Recruiting	<a href="#">Duvelisib to Combat COVID-19</a>	• COVID-19	• Drug: Duvelisib • Procedure: Peripheral blood draw • Drug: Placebo
2	<input type="checkbox"/>	Recruiting	<a href="#">Relationship Between CT- Value With Prognosis in COVID-19 Patients</a>	• COVID-19	• Diagnostic Test: Cycle threshold
3	<input type="checkbox"/>	Recruiting	<a href="#">Observational Cohort of COVID-19 Patients at Raymond-Poincare</a>	• COVID-19	
4	<input type="checkbox"/>	Not yet recruiting	<a href="#">COVID-19 Surveillance Based on Smart Wearable Device</a>	• COVID-19	
5	<input type="checkbox"/>	Completed <a href="#">Has Results</a>	<a href="#">A Phase 2 Trial of Infliximab in Coronavirus Disease 2019 (COVID-19).</a>	• COVID-19	• Drug: Infliximab

Kas minutę – daugiau kaip po straipsnį!

[https://www.nlm.nih.gov/bsd/medline\\_cit\\_counts\\_yr\\_pub.html](https://www.nlm.nih.gov/bsd/medline_cit_counts_yr_pub.html)

Home > MEDLINE/PubMed Resources

## MEDLINE® Citation Counts by Year of Publication (as of January 2021)\*

MEDLINE consists of completed citations indexed with MeSH® (Medical Subject Headings®).

Year of Publication	Total # Citations	# Citations Published in US	%s Citations Published in US
2020*	362,528	138,112	38%
2019	898,145	345,923	39%
2018	866,977	343,605	40%
2017	848,776	343,947	41%
2016	862,829	351,138	41%
2015	878,403	367,373	43%



$$898145 / (365 \times 24 \times 60) = 1.7$$

# Pasiūlymas: Recenzuojamas duomenų bazės įrašas

- Duomenys šiuo metu recenzuojami prastai;
- Duomenys kartais vertingesni už patį straipsnį;
- Kokybiškas, praėjęs kolegų peržiūrą duomenų bazės įrašas turėtų būti užskaitomas kaip mokslo rezultatas („produkcija“);

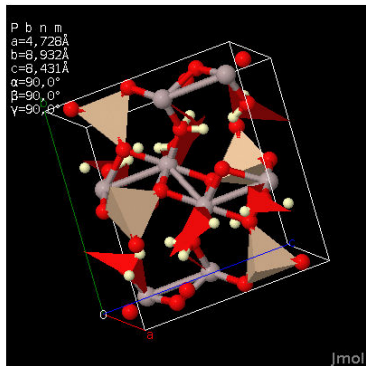
- Skaitmeninių duomenų kiekis auga (eksponentiškai);
- Duomenų tvarkymas tampa iššūkiu;
- Reikalinga „duomenų tvarkytojų“ (angl. “data stewards”) specialybė ir pareigybės (rolės);

- Mokslas turi būti atviras!
- Atviras mokslas naudingas mums visiems!
- XXIa. iššūkiai be atviro mokslo neišsprendžiami!
- Siūlomas gairių projektas atsižvelgia į šiuos aspektus:
  - kiek galima atvira, kiek reikia uždara;
  - infrastruktūros svarba;
  - duomenų specialistų poreikis;
- **Bet svarbiausia – Jūsų visų palaikymas!**

# Ačiū už dėmesį!



<http://en.wikipedia.org/wiki/Topaz>



**Coordinates**

[2207377.cif](#)

**Original IUCr paper**

[HTML](#)

<http://www.crystallography.net/2207377.html>



Boulton GS (2021) Science as a global public good. DOI 10.24948/2021.09, URL [https://council.science/wp-content/uploads/2020/06/Science-as-a-global-public-good\\_v041021.pdf](https://council.science/wp-content/uploads/2020/06/Science-as-a-global-public-good_v041021.pdf)



Senior AW, Evans R, Jumper J, Kirkpatrick J, Sifre L, Green T, et al. (2020) Improved protein structure prediction using potentials from deep learning. Nature 577(7792):706–710, DOI 10.1038/s41586-019-1923-7, <https://doi.org/10.1038/s41586-019-1923-7>



Sterling TD (1959) Publication decisions and their possible effects on inferences drawn from tests of significance — or vice versa. Journal of the American Statistical Association 54:30–34, DOI 10.1080/01621459.1959.10501497, <http://dx.doi.org/10.1080/01621459.1959.10501497>