

Architecture of large projects in bioinformatics (ADP)

Lecture 03

Łukasz P. Kozłowski

Warsaw, 2025

Architecture of large projects in bioinformatics (ADP)

Lecture 03

















<https://www.mimuw.edu.pl/~lukaskoz/teaching/adp/>

















Łukasz P. Kozłowski

Warsaw, 2025

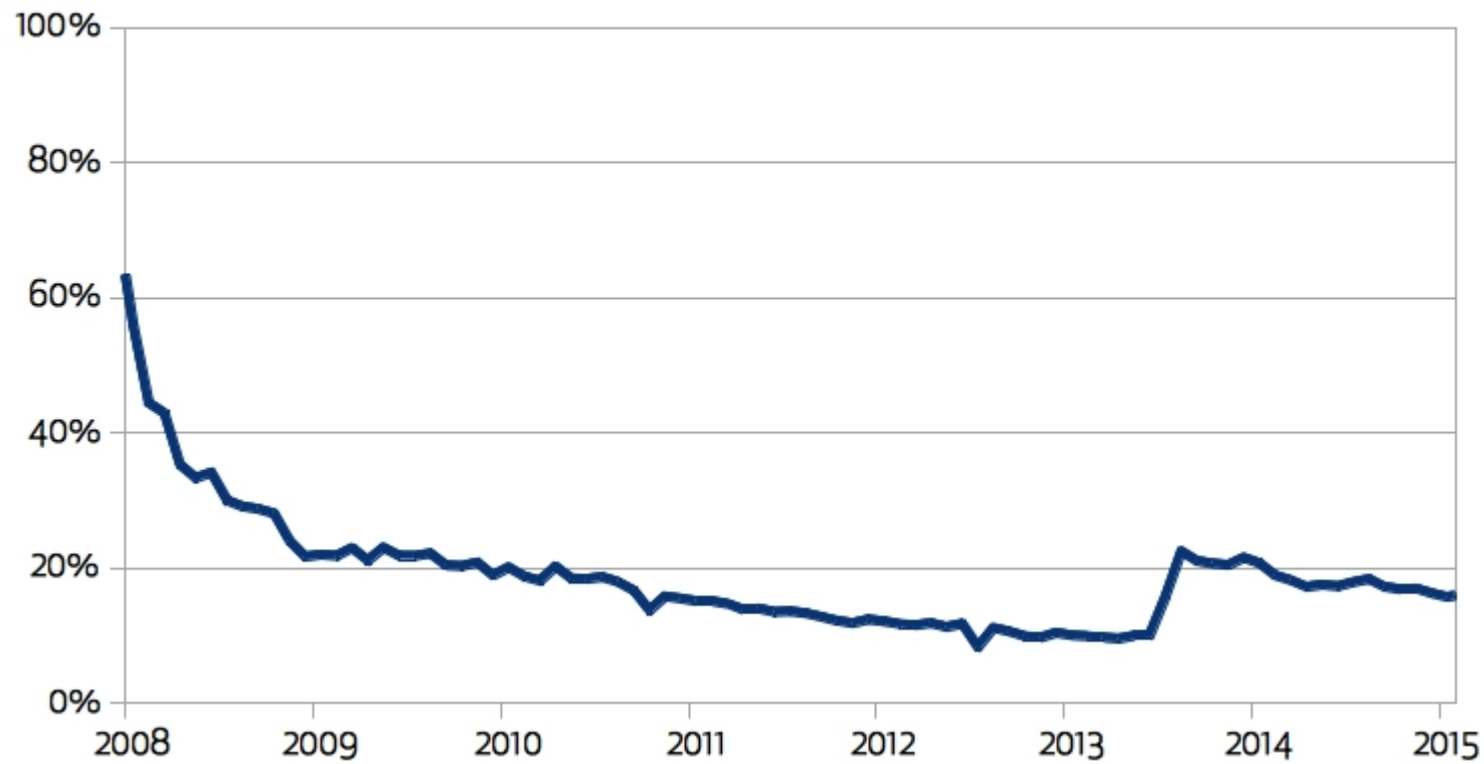
- 1. Data formats in bioinformatics**
- 2. Popular software libraries (BioPerl, BioPython)**
- 3. Most important bioinformatics databases (UniProt, PDB, RefSeq, GenBank, ENA, InterPro, etc.)**
- 4. Software licensing for scientific purposes. Free-software licensing. Patents.**
- 5. Generic model Organism database (GMOD) project - assumptions, history and usage**
- 6. Genome browsers, problem description and state of the solutions**

1. Data formats in bioinformatics
2. Popular software libraries (BioPerl, BioPython)
3. Most important bioinformatics databases (UniProt, PDB, RefSeq, GenBank, ENA, InterPro, etc.)
- 4. Software licensing for scientific purposes. Free-software licensing. Patents.**
5. Generic model Organism database (GMOD) project - assumptions, history and usage
6. Genome browsers, problem description and state of the solutions

 doc	REL: bumped version to 2020.12.21a	3 months ago
 src/cogent3	BUG: evo.hypothesis now corectly supports user value for init_alt	6 hours ago
 tests	BUG: evo.hypothesis now corectly supports user value for init_alt	6 hours ago
 .gitignore	DEV: wild card ignore coverage files	16 days ago
 .hgignore	DEV: wild card ignore coverage files	16 days ago
 .hgtags	Added tag 2020.12.21a for changeset 31dd23a514ab	3 months ago
 .readthedocs.yml	DOC: added version to readthedocs.yml	2 years ago
 ChangeLog	DOC: updated developer docs to point to github, git	2 years ago
 LICENSE	MAINT: updated license copyright info	14 months ago
 MANIFEST.in	MAINT: included the requirements.txt file in MANIFEST.in	4 months ago
 README.md	DOC: updated README	4 months ago
 c3dev-environment.yml	MAINT: added macos and ubuntu os tests in GitHub CI Actions, fixed #650	10 months ago
 pyproject.toml	TST: support test execution from root directory	20 days ago
 requirements.txt	Bump pillow from 8.1.1 to 8.1.2	23 days ago
 setup.py	MAINT: updated numba dependency to 0.53	yesterday
 tox.ini	MAINT: updated numba dependency to 0.53	yesterday

 doc	REL: bumped version to 2020.12.21a	3 months ago
 src/cogent3	BUG: evo.hypothesis now corectly supports user value for init_alt	6 hours ago
 tests	BUG: evo.hypothesis now corectly supports user value for init_alt	6 hours ago
 .gitignore	DEV: wild card ignore coverage files	16 days ago
 .hgignore	DEV: wild card ignore coverage files	16 days ago
 .hgtags	Added tag 2020.12.21a for changeset 31dd23a514ab	3 months ago
 .readthedocs.yml	DOC: added version to readthedocs.yml	2 years ago
 ChangeLog	DOC: updated developer docs to point to github, git	2 years ago
 LICENSE	MAINT: updated license copyright info	14 months ago
 MANIFEST.in	MAINT: included the requirements.txt file in MANIFEST.in	4 months ago
 README.md	DOC: updated README	4 months ago
 c3dev-environment.yml	MAINT: added macos and ubuntu os tests in GitHub CI Actions, fixed #650	10 months ago
 pyproject.toml	TST: support test execution from root directory	20 days ago
 requirements.txt	Bump pillow from 8.1.1 to 8.1.2	23 days ago
 setup.py	MAINT: updated numba dependency to 0.53	yesterday
 tox.ini	MAINT: updated numba dependency to 0.53	yesterday

Percentage of repositories licensed



GitHub

No license

“Public domain”

Public Domain

Open Source License

EULA

Patents



No license

When you make a creative work (which includes **code**)

the work is under exclusive copyright by default

No license

When you make a creative work (which includes **code**)

the work is under exclusive copyright by default

Unless you include a license that specifies otherwise,
nobody can copy, distribute, or modify your work
without being at risk of take-downs, shake-downs, or litigation.

No license

Disallowing use of your code might not be what you intend by
“no license.”

If your goal is to retain copyright, try an **open source license**

If your goal is to completely opt-out of copyright restrictions,
try a **public domain dedication** instead

No license

No license, means you have no permission from the creators of the software to use, modify, or share the software

No license

No license, means you have no permission from the creators of the software to use, modify, or share the software

Your options:

Ask the maintainers nicely to add a license. Unless the software includes strong indications to the contrary, lack of a license is probably an oversight. If the software is hosted on a site like GitHub, open an issue requesting a license and include a link to this site. If you're bold and it's fairly obvious what license is most appropriate, open a pull request to add a license – see “suggest this license” in the sidebar of the page for each license on this site (e.g., MIT).

Don't use the software. Find or create an alternative that is under an open source license.

Negotiate a private license. Bring your lawyer.

No license

“Public domain”

Public Domain

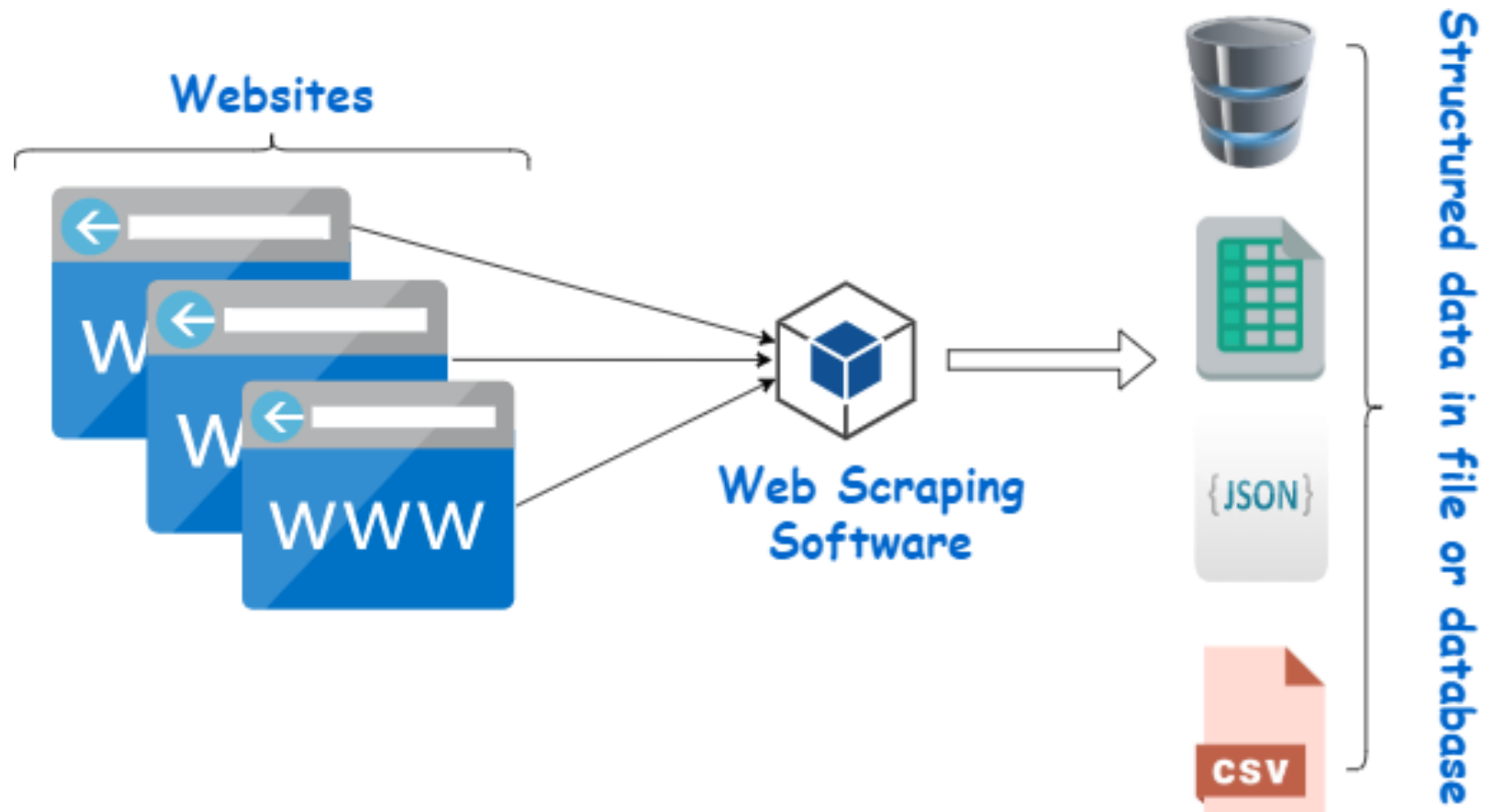
Open Source License

EULA

Patents

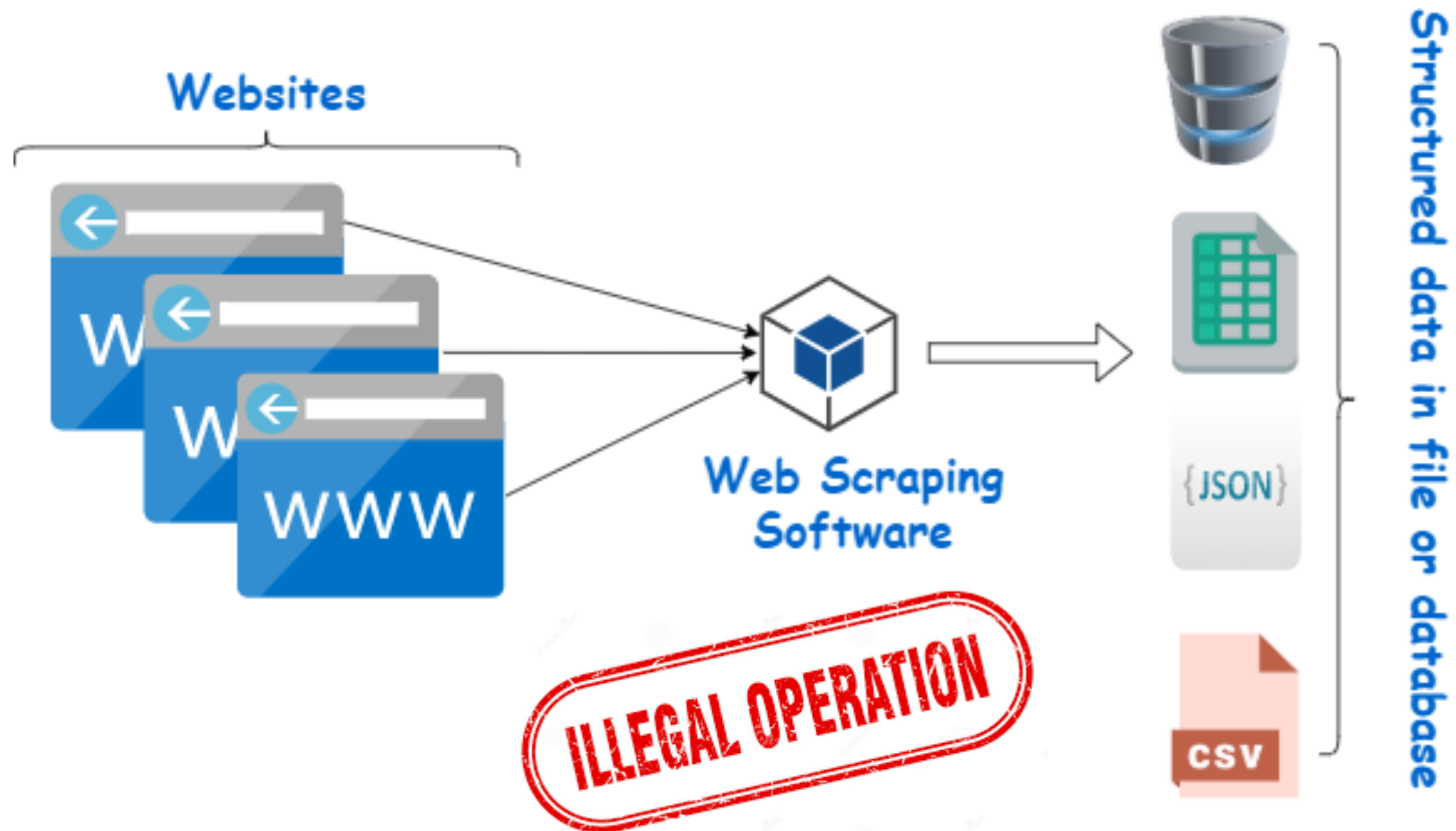
No license

“Public domain”



No license

“Public domain”







A license with no conditions whatsoever which dedicates works to the public domain. Unlicensed works, modifications, and larger works may be distributed under different terms and without source code.

Permissions

- Commercial use
- Distribution
- Modification
- Private use

Conditions

Limitations

- Liability
- Warranty



IPC 2.0 (Isoelectric Point Calculator 2.0) is public domain software.

=====

The IPC 2.0 software, including the web service and documentation, is donated to the public domain. You may therefore freely use it for any legal purpose you wish.

=====

Acknowledgement of authorship and citation in publications is appreciated:

Kozłowski LP (2016) IPC - Isoelectric Point Calculator. Biology Direct 2016;11:55.
PMID: 27769290, doi: 10.1186/s13062-016-0159-9

=====

Please, report evidence of IPC bugs to the Author:
lukasz[dot]kozłowski[dot]lpk[at]gmail[dot]com

=====

Disclaimer of warranty

THIS SOFTWARE IS PROVIDED "AS IS" WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESS OR IMPLIED, INCLUDING WITHOUT LIMITATION IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE.



When you should use it and when not



When you should use it and when not

You are sole author or all authors agreed to release the code under CC 0



When you should use it and when not

You are sole author or all authors agreed to release the code under CC 0

You do not want get in future any royalties, fees or patents related to the code



When you should use it and when not

You are sole author or all authors agreed to release the code under CC 0

You do not want get in future any royalties, fees or patents related to the code

Good choice for scientific data, software



When you should use it and when not

You are sole author or all authors agreed to release the code under CC 0

You do not want get in future any royalties, fees or patents related to the code

Good choice for scientific data, software

Not very good choice if you work in private company



IPC 2.0 (Isoelectric Point Calculator 2.0) is public domain software.

=====

The IPC 2.0 software, including the web service and documentation, is donated to the public domain. You may therefore freely use it for any legal purpose you wish.

=====

Acknowledgement of authorship and citation in publications is appreciated:

Kozłowski LP (2016) IPC - Isoelectric Point Calculator. Biology Direct 2016;11:55.
PMID: 27769290, doi: 10.1186/s13062-016-0159-9

=====

Please, report evidence of IPC bugs to the Author:
lukasz[dot]kozłowski[dot]lpk[at]gmail[dot]com

=====

Disclaimer of warranty

THIS SOFTWARE IS PROVIDED "AS IS" WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESS OR IMPLIED, INCLUDING WITHOUT LIMITATION IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE.



IPC 2.0 (Isoelectric Point Calculator 2.0) is public domain software.

IPC 2.0: prediction of **isoelectric point** and pK_a dissociation constants

[LP Kozlowski](#) - Nucleic Acids Research, 2021 - [academic.oup.com](#)

... the in silico prediction of **isoelectric point** and pK_a values. In this paper, I present **Isoelectric Point Calculator 2.0 (IPC 2.0)**, a web server for the prediction of **isoelectric points** and pK_a ...

☆ Zapisz Cytuj **Cytowane przez 81** [Powiązane artykuły](#) [Wszystkie wersje 9](#)

Please, report evidence of IPC bugs to the Author:
lukasz[dot]kozlowski[dot]lpk[at]gmail[dot]com

Disclaimer of warranty

THIS SOFTWARE IS PROVIDED "AS IS" WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESS OR IMPLIED, INCLUDING WITHOUT LIMITATION IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE.



IPC 2.0 (Isoelectric Point Calculator 2.0) is public domain software.

IPC 2.0: prediction of **isoelectric point** and pK_a dissociation constants

[LP Kozlowski](#) - Nucleic Acids Research, 2021 - [academic.oup.com](#)

... the in silico prediction of **isoelectric point** and pK_a values. In this paper, I present **Isoelectric Point Calculator 2.0 (IPC 2.0)**, a web server for the prediction of **isoelectric points** and pK_a ...

☆ Zapisz Cytuj **Cytowane przez 81** Powiązane artykuły Wszystkie wersje 9

IPC–isoelectric point calculator

[LP Kozlowski](#) - Biology direct, 2016 - Springer

... changing the pH, we can easily **calculate** the **isoelectric point**. The result will be almost certainly different than the real **isoelectric point** because many proteins are chemically modified (...)

☆ Zapisz Cytuj **Cytowane przez 366** Powiązane artykuły Wszystkie wersje 22



IPC 2.0 (Isoelectric Point Calculator 2.0) is public domain software.

IPC 2.0: prediction of **isoelectric point** and pK_a dissociation constants

[LP Kozlowski](#) - Nucleic Acids Research, 2021 - [academic.oup.com](#)

... the in silico prediction of **isoelectric point** and pK_a values. In this paper, I present **Isoelectric Point Calculator 2.0 (IPC 2.0)**, a web server for the prediction of **isoelectric points** and pK_a ...

☆ Zapisz Cytuj **Cytowane przez 81** [Powiązane artykuły](#) [Wszystkie wersje 9](#)

IPC–isoelectric point calculator

[LP Kozlowski](#) - Biology direct, 2016 - Springer

... changing the pH, we can easily **calculate** the **isoelectric point**. The result will be almost certainly different than the real **isoelectric point** because many proteins are chemically modified (...)

☆ Zapisz Cytuj **Cytowane przez 366** [Powiązane artykuły](#) [Wszystkie wersje 22](#)



IPC 2.0 (Isoelectric Point Calculator 2.0) is public domain software.

=====

IPC 2.0: prediction of isoelectric point and pK_a dissociation constants

[LP Kozlowski](#) - Nucleic Acids Research, 2021 - [academic.oup.com](#)

... the in silico prediction of **isoelectric point** and pK_a values. In this paper, I present **Isoelectric Point Calculator 2.0 (IPC 2.0)**, a web server for the prediction of **isoelectric points** and pK_a ...

☆ Zapisz  Cytuj **Cytowane przez 108** Powiązane artykuły Wszystkie wersje 10 Web of S

IPC–isoelectric point calculator

[LP Kozlowski](#) - Biology direct, 2016 - Springer

... Here, I present the **Isoelectric Point Calculator (IPC)**, a web ... According to the presented benchmarks, the newly developed **IPC** ... Moreover, the prediction of pI using the **IPC** pK_a 's leads to

☆ Zapisz  Cytuj **Cytowane przez 423** Powiązane artykuły Wszystkie wersje 24 Web of Sc

16

Citing Patents



IPC 2.0 (Isoelectric Point Calculator 2.0) is public domain software.

IPC 2.0: prediction of **isoelectric point** and pK_a dissociation constants

[LP Kozlowski](#) - Nucleic Acids Research, 2021 - [academic.oup.com](#)

... the in silico prediction of **isoelectric point** and pK_a values. In this paper, I present **Isoelectric Point Calculator 2.0 (IPC 2.0)**, a web server for the prediction of **isoelectric points** and pK_a ...

☆ Zapisz Cytuj **Cytowane przez 81** [Powiązane artykuły](#) [Wszystkie wersje 9](#)


IPC–isoelectric point calculator

[LP Kozlowski](#) - Biology direct, 2016 - Springer


... changing the pH, we can easily **calculate** the **isoelectric point**. The result will be almost certainly different than the real **isoelectric point** because many proteins are chemically modified (...)

☆ Zapisz Cytuj **Cytowane przez 366** [Powiązane artykuły](#) [Wszystkie wersje 22](#)

BMC Ecology and Evolution

[Home](#) [About](#) [Articles](#) [Submission Guidelines](#) [Collections](#)[Submit manuscript](#) [Software](#) | [Open access](#) | [Published: 28 June 2004](#)

RETRACTED ARTICLE: TREEFINDER: a powerful graphical analysis environment for molecular phylogenetics

[Gangolf Jobb](#) , [Arndt von Haeseler](#) & [Korbinian Strimmer](#)[BMC Evolutionary Biology](#) **4**, Article number: 18 (2004) | [Cite this article](#)**50k** Accesses | **1008** Citations | **21** Altmetric | [Metrics](#) This article was [retracted](#) on 05 November 2015

Retraction Note: TREEFINDER: a powerful graphical analysis environment for molecular phylogenetics

[Gangolf Jobb](#) , [Arndt von Haeseler](#) & [Korbinian Strimmer](#)

BMC Evolutionary Biology **15**, Article number: 243 (2015) | [Cite this article](#)

11k Accesses | **5** Citations | **105** Altmetric | [Metrics](#)

 The [Original Article](#) was published on 28 June 2004

Retraction

The editors of *BMC Evolutionary Biology* retract this article [1] due to the decision by the corresponding author, Gangolf Jobb, to change the license to the software described in the article. The software is no longer available to all scientists wishing to use it in certain territories. This breaches the journal's editorial policy on software availability [2] which has been in effect since the time of publication. The other authors of the article, Arndt von Haeseler and Korbinian Strimmer, have no control over the licensing of the software and support the retraction of this article.

Retraction Note: TREEFINDER: a powerful graphical analysis environment for molecular phylogenetics

[Gangolf Jobb](#) , [Arndt von Haeseler](#) & [Korbinian Strimmer](#)

[BMC Evolutionary Biology](#) **15**, Article number: 243 (2015) | [Cite this article](#)

11k Accesses | **5** Citations | **105** Altmetric | [Metrics](#)

 The [Original Article](#) was published on 28 June 2004

Retraction

The editors of *BMC Evolutionary Biology* retract this article [1] due to the decision by the corresponding author, Gangolf Jobb, to change the license to the software described in the article. The software is no longer available to all scientists wishing to use it in certain territories. This breaches the journal's editorial policy on software availability [2] which has been in effect since the time of publication. The other authors of the article, Arndt von Haeseler and Korbinian Strimmer, have no control over the licensing of the software and support the retraction of this article.

Retraction Note: TREEFINDER: a powerful graphical analysis environment for molecular phylogenetics

[Gangolf Jobb](#) , [Arndt von Haeseler](#) & [Korbinian Strimmer](#)

BMC Evolutionary Biology **15**, Article number: 243 (2015) | [Cite this article](#)

11k Accesses | **5** Citations | **105** Altmetric | [Metrics](#)

 The [Original Article](#) was published on 28 June 2004

Retraction

The editors of *BMC Evolutionary Biology* retract this article [1] due to the decision by the corresponding author, Gangolf Jobb, to change the license to the software described in the article. The software is no longer available to all scientists wishing to use it in certain territories. This breaches the journal's editorial policy on software availability [2] which has been in effect since the time of publication. The other authors of the article, Arndt von Haeseler and Korbinian Strimmer, have no control over the licensing of the software and support the retraction of this article.

No license

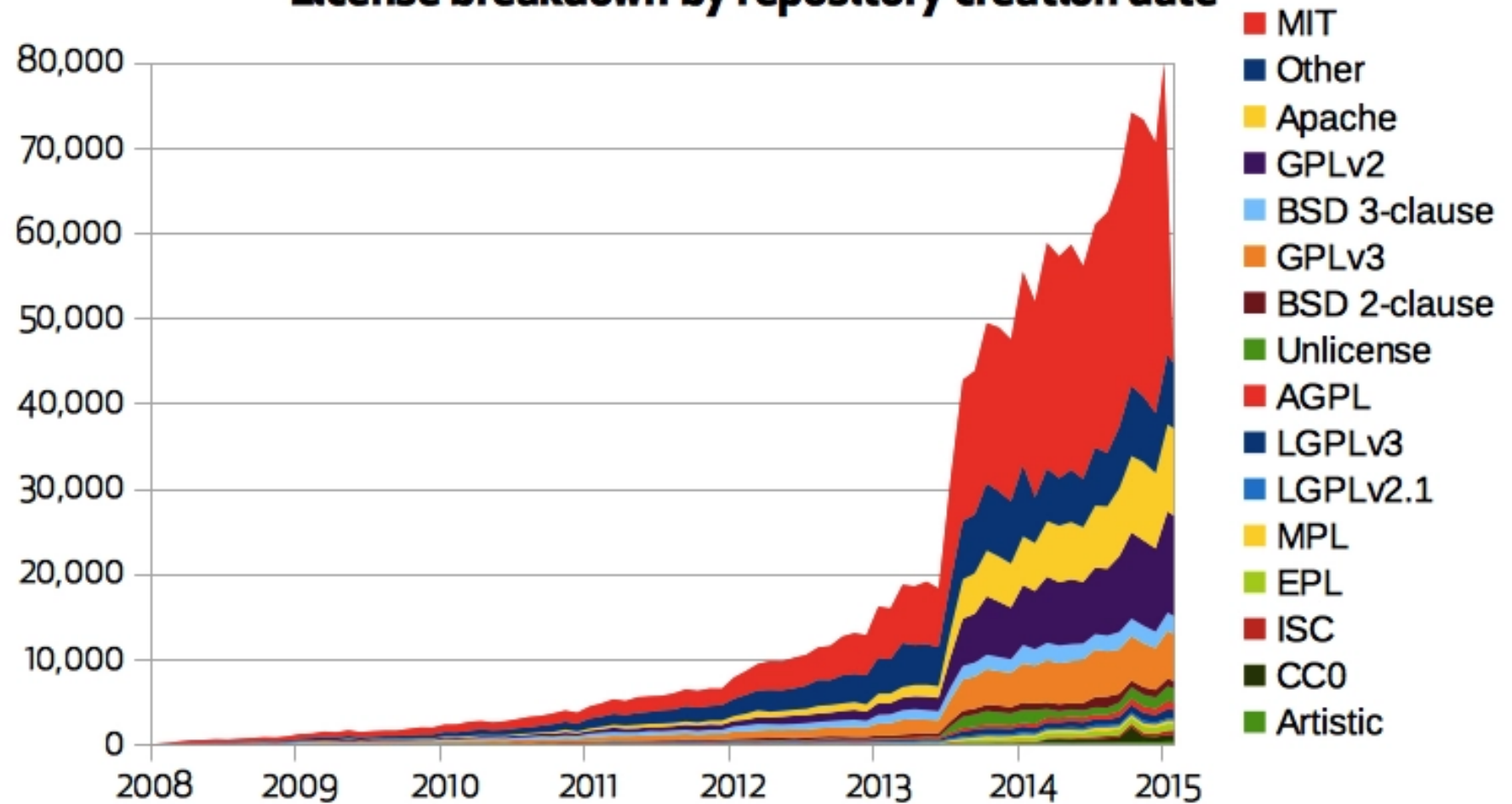
“Public domain”

Public Domain

Open Source License



License breakdown by repository creation date



GitHub



MIT License

A short and simple permissive license with conditions only requiring preservation of copyright and license notices. Licensed works, modifications, and larger works may be distributed under different terms and without source code.

Permissions

- Commercial use
- Distribution
- Modification
- Private use

Conditions

- License and copyright notice

Limitations

- Liability
- Warranty



Apache License 2.0

A permissive license whose main conditions require preservation of copyright and license notices. Contributors provide an express grant of patent rights. Licensed works, modifications, and larger works may be distributed under different terms and without source code.

Permissions

- Commercial use
- Distribution
- Modification
- Patent use
- Private use

Conditions

- License and copyright notice
- State changes

Limitations

- Liability
- Trademark use
- Warranty

GNU General Public License v3.0

GNU GPLv3

Permissions of this strong copyleft license are conditioned on making available complete source code of licensed works and modifications, which include larger works using a licensed work, under the same license. Copyright and license notices must be preserved. Contributors provide an express grant of patent rights.

Permissions

- Commercial use
- Distribution
- Modification
- Patent use
- Private use

Conditions

- Disclose source
- License and copyright notice
- Same license
- State changes

Limitations

- Liability
- Warranty



GNU AGPLv3

Permissions

- Commercial use
- Distribution
- Modification
- Patent use
- Private use

Conditions

- Disclose source
- License and copyright notice
- Network use is distribution
- Same license
- State changes

Limitations

- Liability
- Warranty

GNU GPLv3

Permissions

- Commercial use
- Distribution
- Modification
- Patent use
- Private use

Conditions

- Disclose source
- License and copyright notice
- Same license
- State changes

Limitations

- Liability
- Warranty

GNU LGPLv3

Permissions

- Commercial use
- Distribution
- Modification
- Patent use
- Private use

Conditions

- Disclose source
- License and copyright notice
- Same license (library)
- State changes

Limitations

- Liability
- Warranty



Mozilla Public License 2.0

Permissions of this weak copyleft license are conditioned on making available source code of licensed files and modifications of those files under the same license (or in certain cases, one of the GNU licenses). Copyright and license notices must be preserved. Contributors provide an express grant of patent rights. However, a larger work using the licensed work may be distributed under different terms and without source code for files added in the larger work.

Permissions

- Commercial use
- Distribution
- Modification
- Patent use
- Private use

Conditions

- Disclose source
- License and copyright notice
- Same license (file)

Limitations

- Liability
- Trademark use
- Warranty

No license

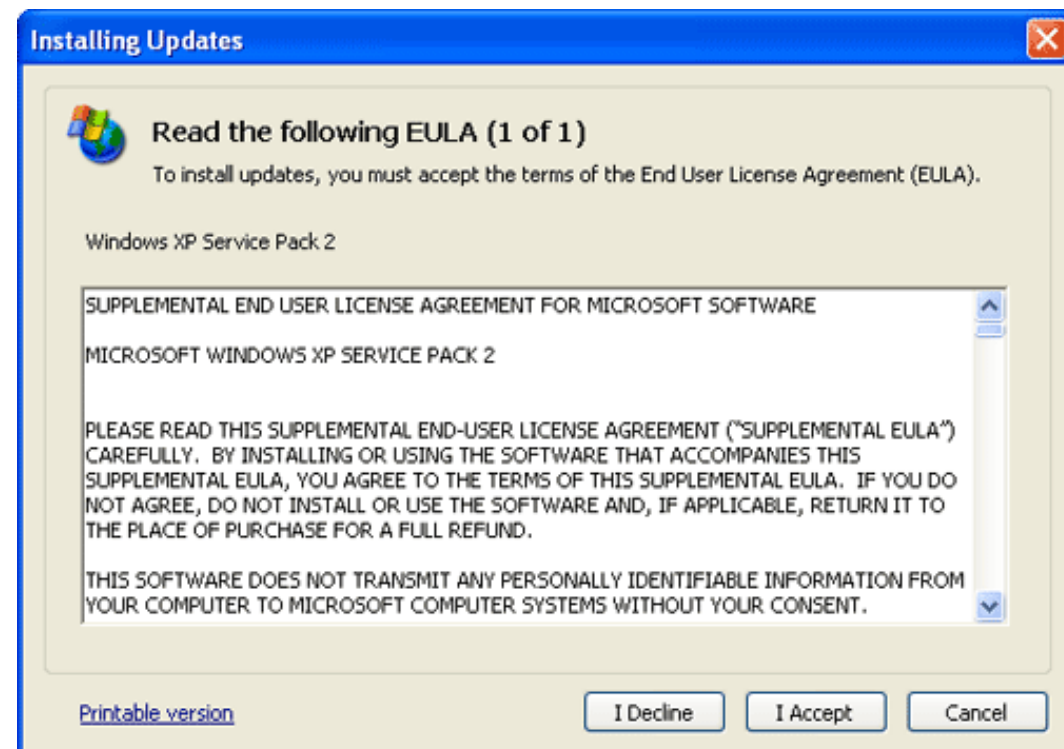
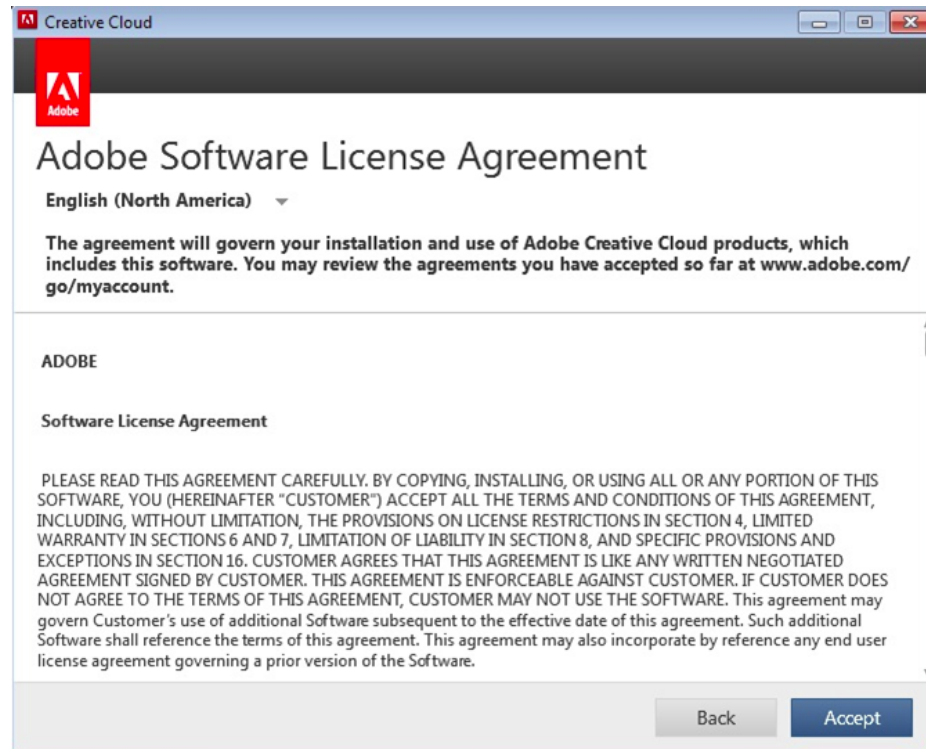
“Public domain”

Public Domain

Open Source License

EULA

Patents

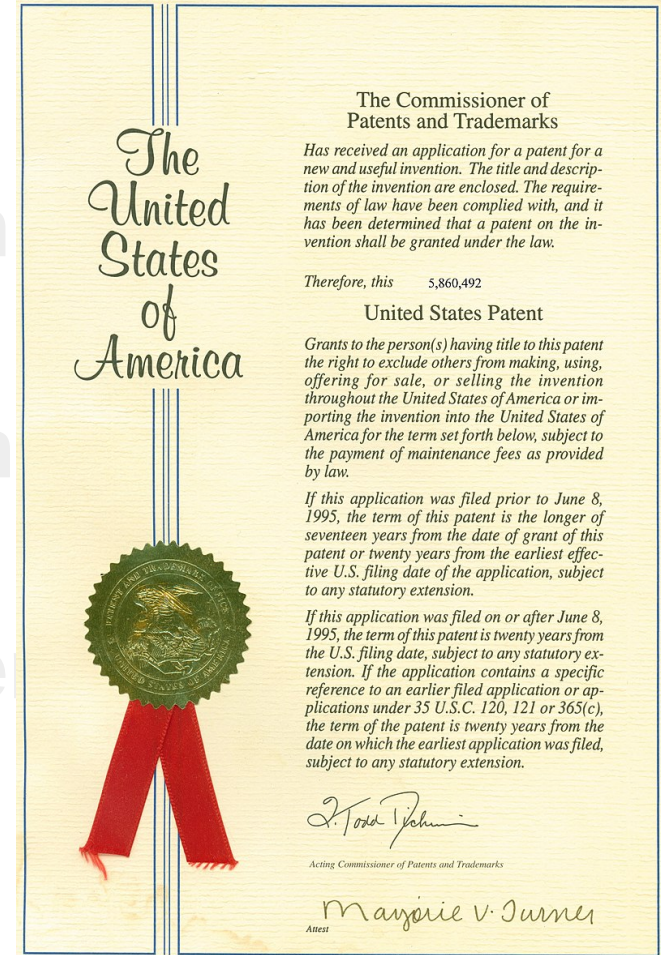
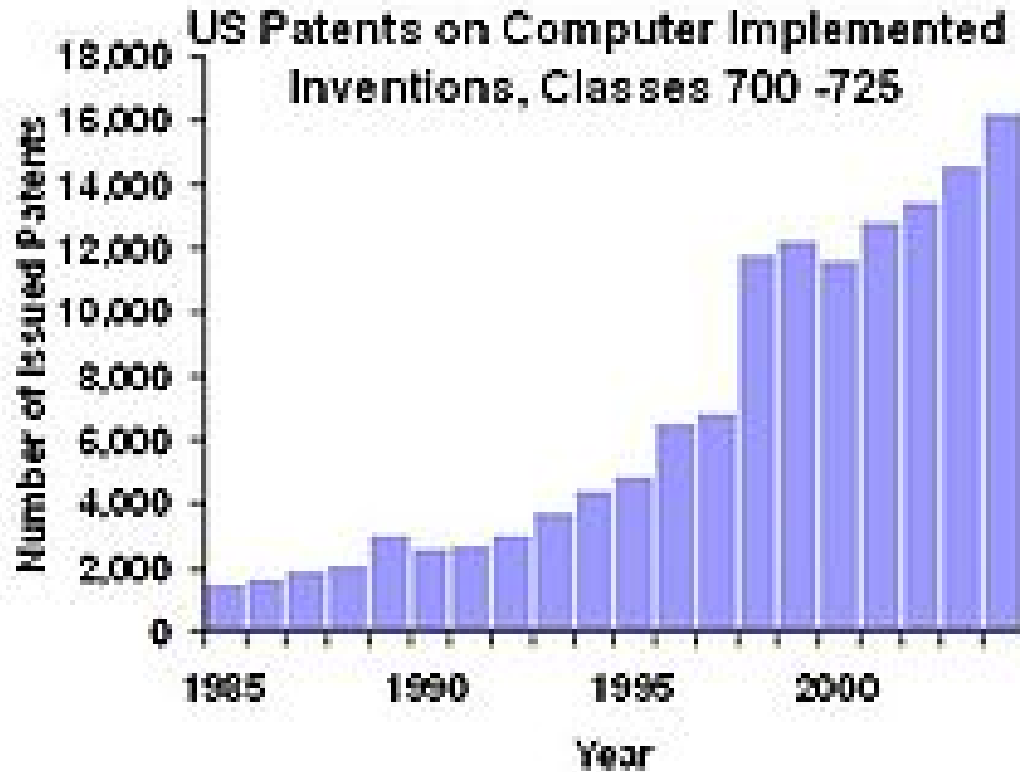




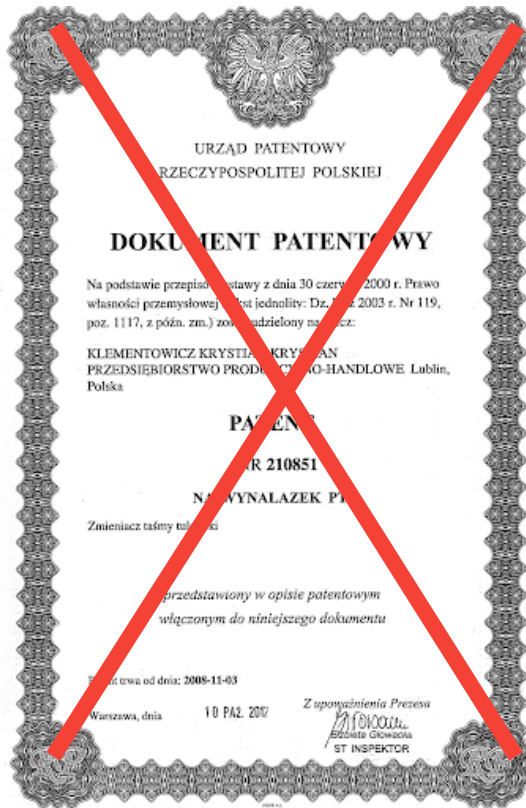
Example license

<https://choosealicense.com/terms-of-service/>

No license



Patents



(19)



(11)

EP 1 646 287 B1

(12)

EUROPEAN PATENT SPECIFICATION

(45) Date of publication and mention
of the grant of the patent:
10.01.2007 Bulletin 2007/02

(51) Int Cl.:

A21D 13/04 (2006.01)**A21D 2/36** (2006.01)**A23L 1/164** (2006.01)**A23L 1/00** (2006.01)**A23L 1/0522** (2006.01)

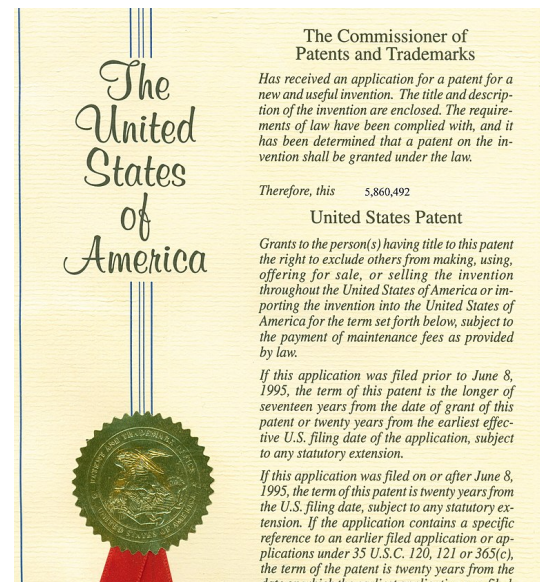
(21) Application number: **04774832.2**

(86) International application number:

PCT/NL2004/000524

(22) Date of filing: **22.07.2004**

(87) International publication number:

WO 2005/025319 (24.03.2005 Gazette 2005/12)(54) **PROCESSING OF TEFF FLOUR****VERARBEITUNG VON TEFF-MEHL****TRAITEMENT DE FARINE TEFF**

Patent is a **title** that gives its owner the legal right to exclude others from making, using, or selling an invention for a limited period of years (usually 20) in exchange for publishing an enabling public disclosure of the invention.

In most countries, patent rights fall under private law and the patent holder must sue someone infringing the patent in order to enforce his or her rights.







☐ Include non-patent literature (Google Scholar)

Search and read the full text of patents from around the world.

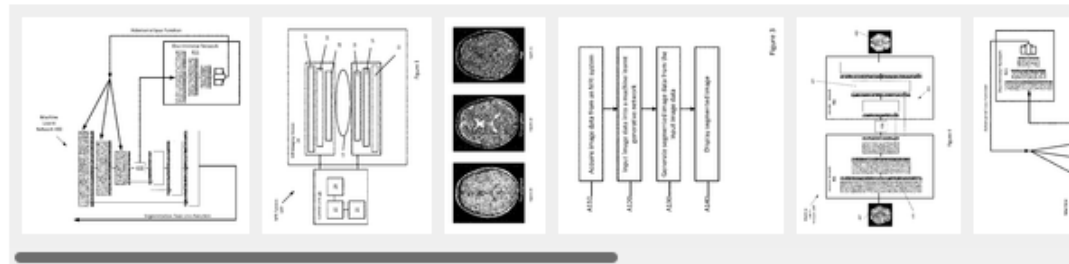
[← Back to results](#)  [deep learning image mri;](#)

Protocol independent image processing with adversarial networks

Abstract

Systems and methods are provided for generating a protocol independent **image**. A **deep learning** generative framework learns to recognize the boundaries and classification of tissues in an **MRI image**. The **deep learning** generative framework includes an encoder, a decoder, and a discriminator network. The encoder is trained using the discriminator network to generate a latent space that is invariant to protocol and the decoder is trained to generate the best output possible for brain and/or tissue extraction.

Images (9)



Classifications

• **A61B5/055** Detecting, measuring or recording for diagnosis by means of electric currents or magnetic fields; Measuring using microwaves or radio waves involving electronic [EMR] or nuclear

US20190046068A1

United States



Download PDF



Find Prior Art



Similar

Inventor: [Pascal Ceccaldi](#), [Benjamin L. Odry](#), [Boris Mailhe](#), [Mariappan S. Nadar](#)

Current Assignee : [Siemens Healthcare GmbH](#)

Worldwide applications

2018 • [US](#)

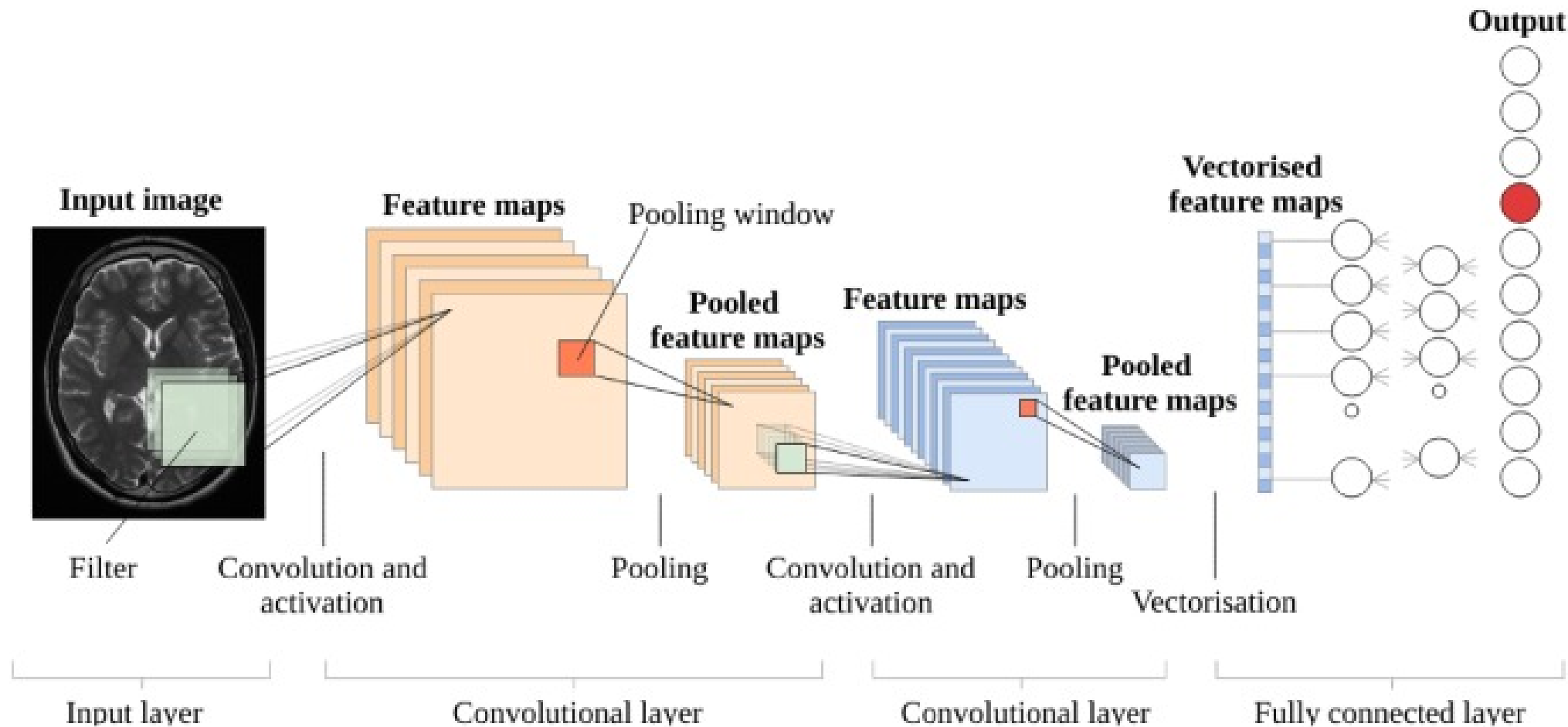
Application [US16/055,546](#) events 

2017-08-10 • Priority to [US201762543600P](#)

2018-08-06 • Application filed by [Siemens Healthcare GmbH](#)

2019-02-14 • Publication of [US20190046068A1](#)

2020-04-21 • Publication of [US10624558B2](#)



ResNet and DenseNet for classification

U-nets for segmentation

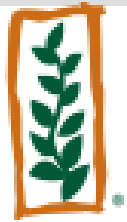
Before you start your project:

- check if other similar projects exists
- check if there are patents in similar area

Bioinformatics related areas where patents should be considered

- biotechnology**
- drug design**
- molecular biology**

MONSANTO



TOP 10 GENETICALLY Modified Foods

RawForBeauty.com



Canola



Soy



Corn



Rice



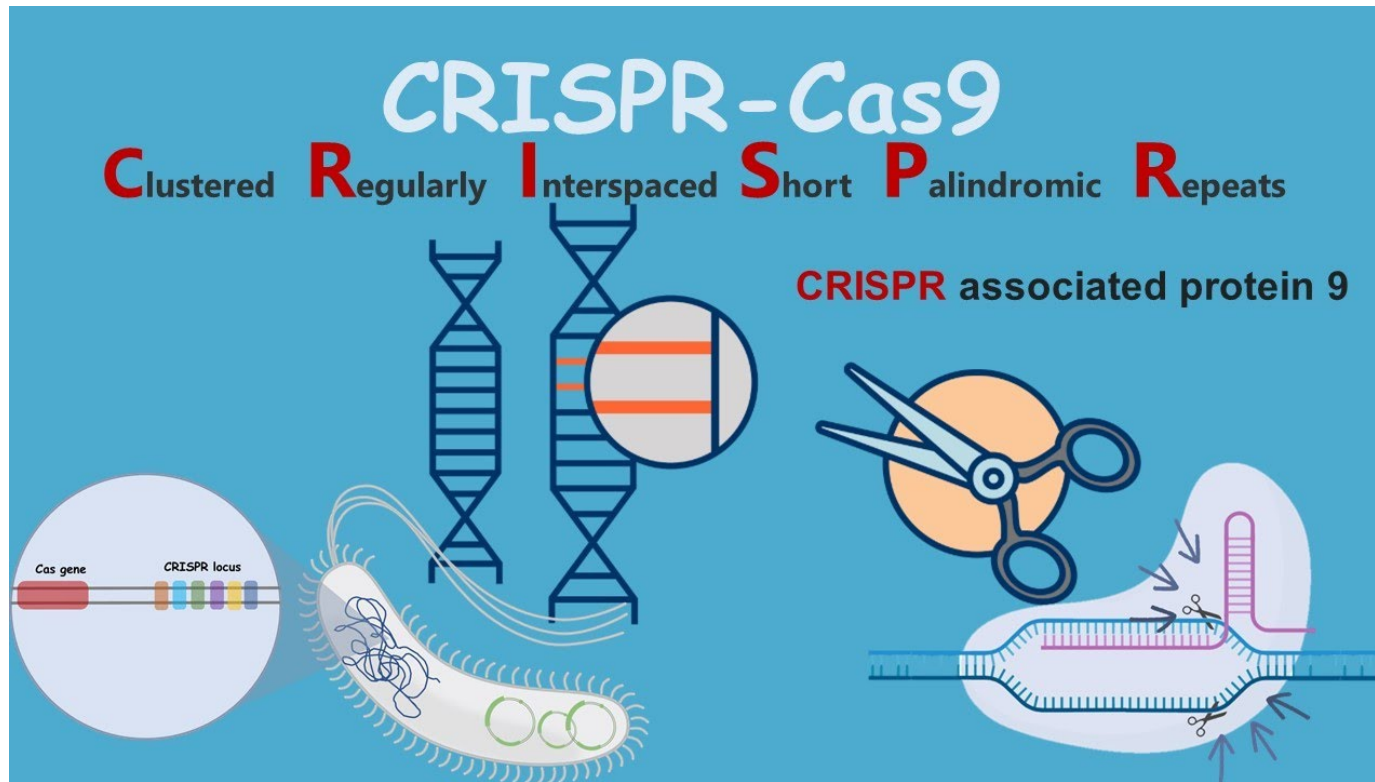
Tomatoes



Potatoes



Vegetables



The genome editor **CRISPR**, whose invention is at the heart of a fierce patent battle, typically uses an RNA molecule to guide a DNA-cutting enzyme such as Cas9 to a DNA sequence targeted for cutting

University of California vs. Broad Institute of MIT and Harvard

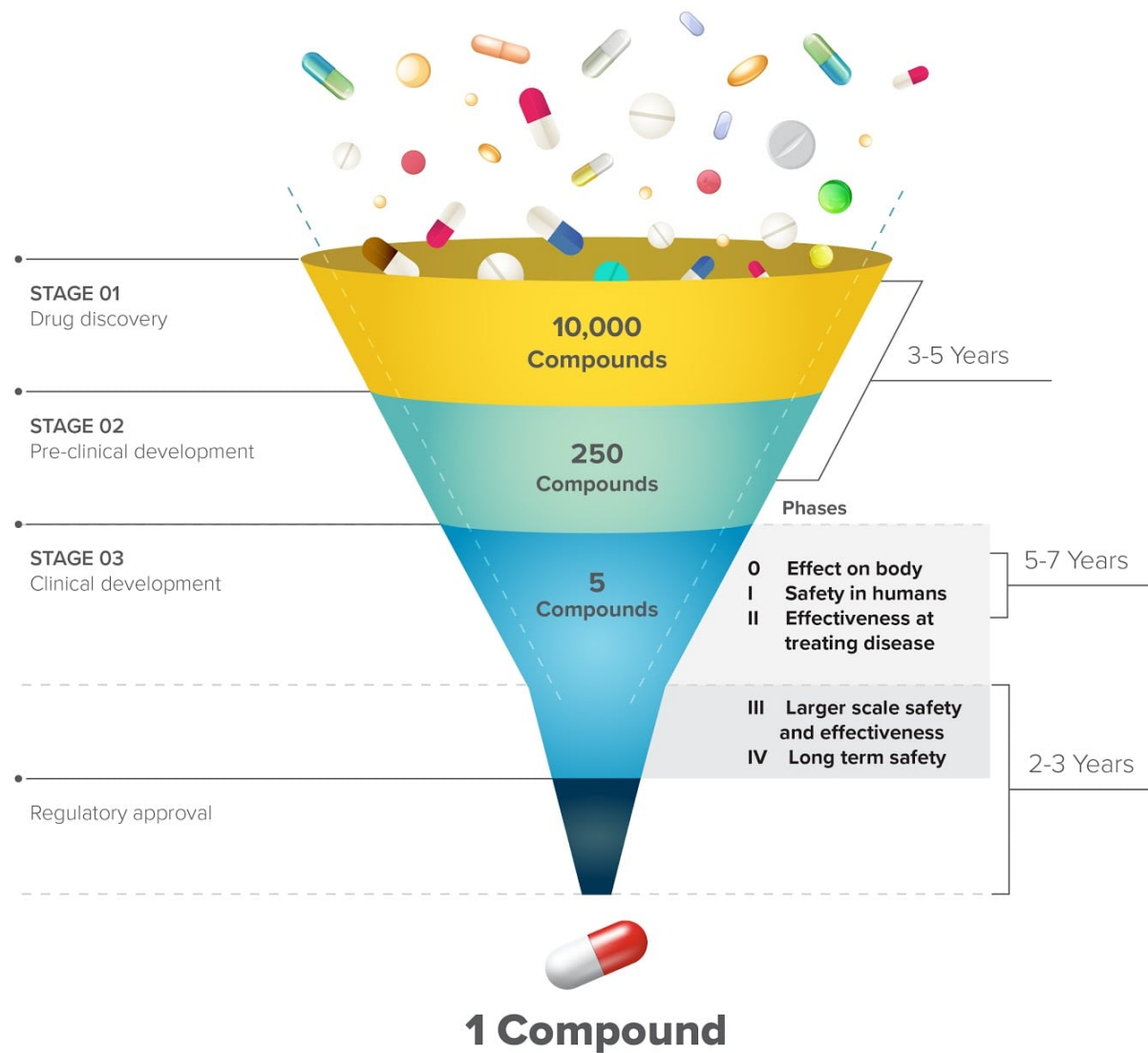


May 29, 2012

["Polymorphisms Associated With Parkinson's Disease"](#) patent

a variant in the SGK1 gene that may be protective against Parkinson's disease in individuals who carry the rare risk-associated LRRK2 G2019S mutation

Drug discovery and development timeline



16 March 2021



Adamed has signed an exclusive global licence agreement with Acadia Pharmaceuticals Inc., US development of a novel molecule which has been synthesized in Adamed laboratories for the potential treatment of psychiatric disorders

122 mln USD

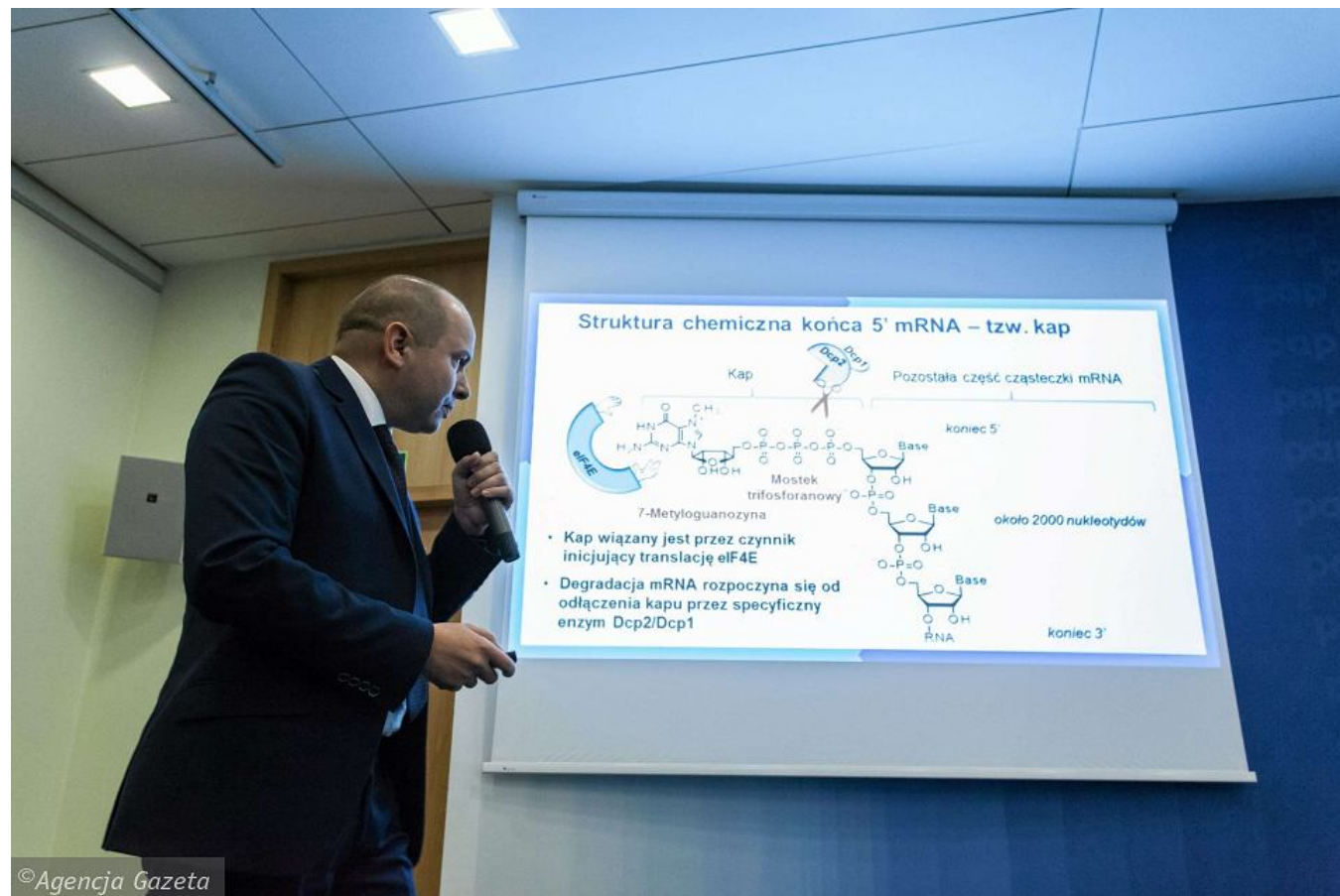
1 Nov 2020



OncoArendi Therapeutics S.A. has sign an exclusive global licence agreement with Galapagos NV

For OATD-01 molecule that is a Phase 2-ready chitotriosidase/acidic mammalian chitinase (CHIT1/AMCase) inhibitor for the treatment of idiopathic pulmonary fibrosis (IPF) and other diseases with a fibrotic component

€320 million



COMMERCIALISATION OF mRNA 5' CAP - TIMELINE

Before 2007	Studies on the properties of mRNA and 5' cap analogues
2007	Filing the patent application in Poland and in the USA (S-ARCA); the beginning of work on analogous solutions to protect against competition; publication in a renowned scientific journal
2008	Signing the contract on the co-ownership of the invention between UW and LSUHSC-S (S-ARCA)
2008	Filing the patent application in Poland and in the USA (B-ARCA); gaining the interest from research teams seeking potential cooperation; establishing the relation with University in Mainz and BioNTech company - the future investor
2010	Signing the contract on the co-ownership of the invention between UW and LSUHSC-S (B-ARCA)
2010	Agreement between UW and LSUHSC-S on the distribution of the revenue from selling the licence; negotiations with the investor (BioNTech); supporting the invention's credibility by the production and delivery of 5' cap analogue compound in the amount sufficient to start clinical trials
2010	Signing the licence contract between UW and LSUHSC-S with BioNTech company; first clinical trials set off
2013	Negotiations with BioNTech on the revision of the licence contract due to the global pharmaceutical company interested in the invention which was attracted by BioNTech
2015	Granting by BioNTech the sub-licence to pharmaceutical company Sanofi which conducts clinical trials on a broader scale – a contract worth 360 mln dollars
2016	Selling by BioNTech the sub-licence to Genentech company from Roche group – a contract worth 310 mln dollars

In general, in Poland and EU if you want to apply for patent related to the software you need to add the physical component

**e.g. dedicated device for obtaining the data,
for instance specific MRI scanner**

Software

vs

Databases

```
31 def __init__(self):
32     self.file = None
33     self.fingerprints = set()
34     self.logdups = True
35     self.debug = debug
36     self.logger = logging.getLogger(__name__)
37     if path:
38         self.file = open(os.path.join(path, 'requests.log'),
39                         'a')
40         self.file.seek(0)
41         self.fingerprints.update([x.request() for x in self.files])
42
43 @classmethod
44 def from_settings(cls, settings):
45     debug = settings.getbool('DEBUG', False)
46     return cls(job_dir(settings), debug)
47
48 def request_seen(self, request):
49     fp = self.request_fingerprint(request)
50     if fp in self.fingerprints:
51         return True
52     self.fingerprints.add(fp)
53     if self.file:
54         self.file.write(fp + os.linesep)
55
56 def request_fingerprint(self, request):
57     return request_fingerprint(request)
```



Software

vs

Databases

```
31 def __init__(self, settings):
32     self.file = None
33     self.fingerprints = set()
34     self.logdups = True
35     self.debug = debug
36     self.logger = logging.getLogger(__name__)
37     if path:
38         self.file = open(os.path.join(path, 'requests.log'),
39                         'a')
40         self.file.seek(0)
41         self.fingerprints.update([x.request() for x in self.requests])
42
43 @classmethod
44 def from_settings(cls, settings):
45     debug = settings.getbool('DEBUG', False)
46     return cls(job_dir(settings), debug)
47
48 def request_seen(self, request):
49     fp = self.request_fingerprint(request)
50     if fp in self.fingerprints:
51         return True
52     self.fingerprints.add(fp)
53     if self.file:
54         self.file.write(fp + os.linesep)
55
56 def request_fingerprint(self, request):
57     return request_fingerprint(request)
```



Data, media, etc.

Use: CC0-1.0, CC-BY-4.0, and CC-BY-SA-4.0 are open licenses used for non-software material ranging from datasets to videos.

Note that CC-BY-4.0 and CC-BY-SA-4.0 should not be used for software.





CC0 1.0 Universal (CC0 1.0) Public Domain Dedication

Creative Commons Zero v1.0 Universal

The [Creative Commons CC0 Public Domain Dedication](#) waives copyright interest in a work you've created and dedicates it to the world-wide public domain. Use CC0 to opt out of copyright entirely and ensure your work has the widest reach. As with the Unlicense and typical software licenses, CC0 disclaims warranties. CC0 is very similar to the Unlicense.

Permissions

- Commercial use
- Distribution
- Modification
- Private use

Conditions

Limitations

- Liability
- Patent use
- Trademark use
- Warranty

Creative Commons Attribution 4.0 International

Permits almost any use subject to providing credit and license notice. Frequently used for media assets and educational materials. The most common license for Open Access scientific publications. Not recommended for software.

Permissions

- Commercial use
- Distribution
- Modification
- Private use

Conditions

- License and copyright notice
- State changes

Limitations

- Liability
- Patent use
- Trademark use
- Warranty

Creative Commons Attribution Share Alike 4.0 International

Similar to [CC-BY-4.0](#) but requires derivatives be distributed under the same or a similar, [compatible](#) license. Frequently used for media assets and educational materials. A previous version is the default license for Wikipedia and other Wikimedia projects. Not recommended for software.

Permissions

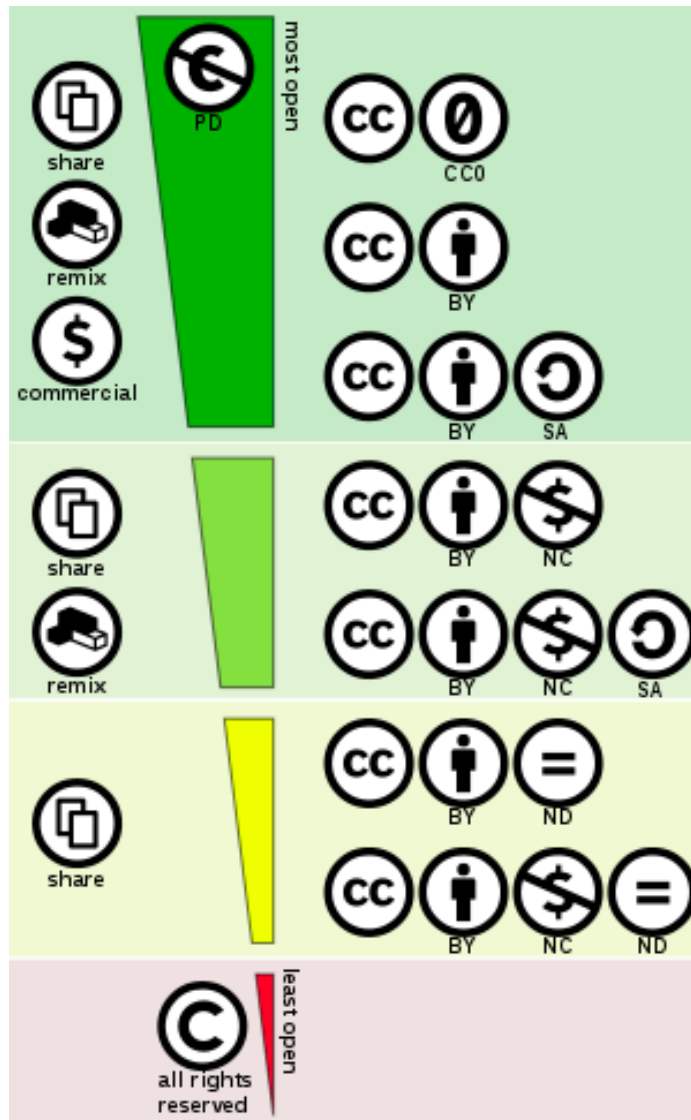
- Commercial use
- Distribution
- Modification
- Private use

Conditions

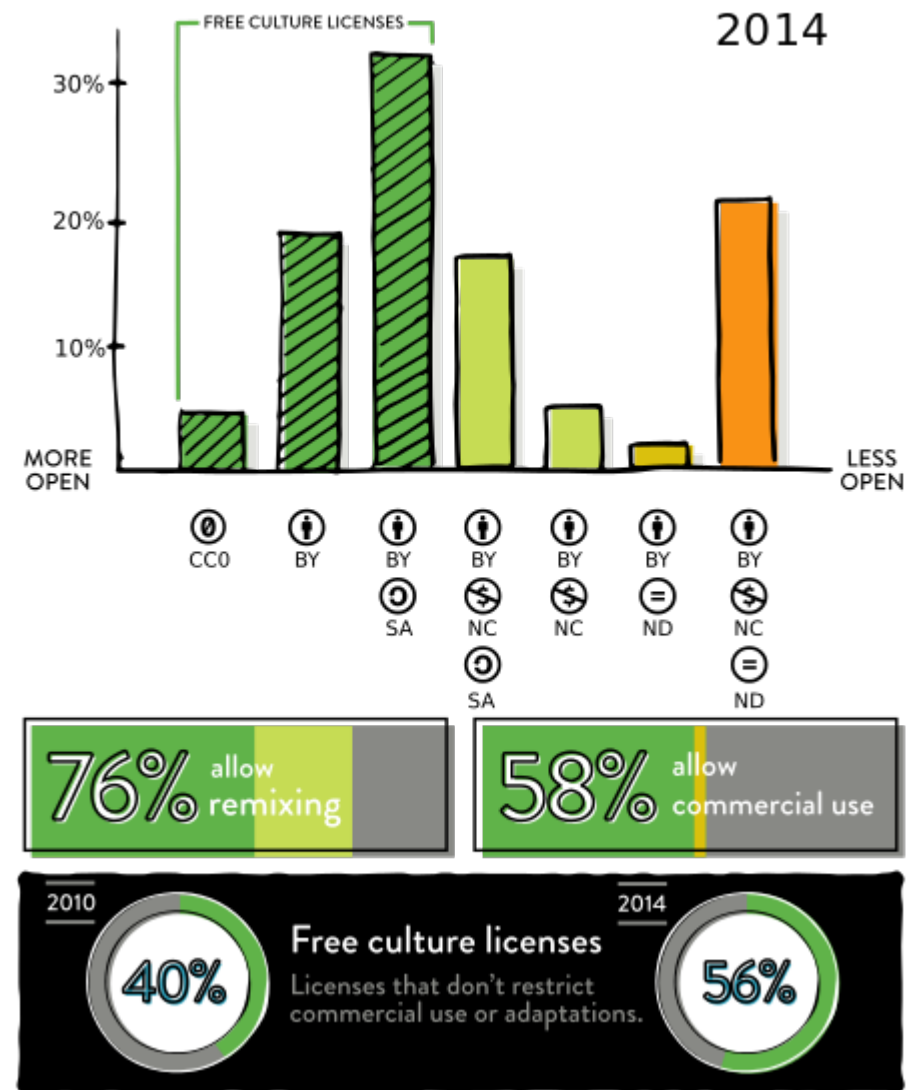
- License and copyright notice
- Same license
- State changes

Limitations

- Liability
- Patent use
- Trademark use
- Warranty



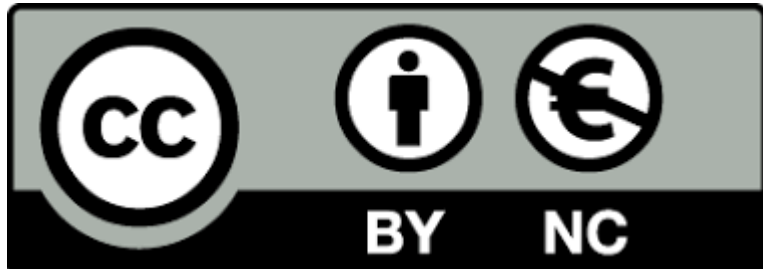
Creative Commons license use



READ: https://en.wikipedia.org/wiki/Creative_Commons_license







Creative Commons Attribution (CC BY 4.0)



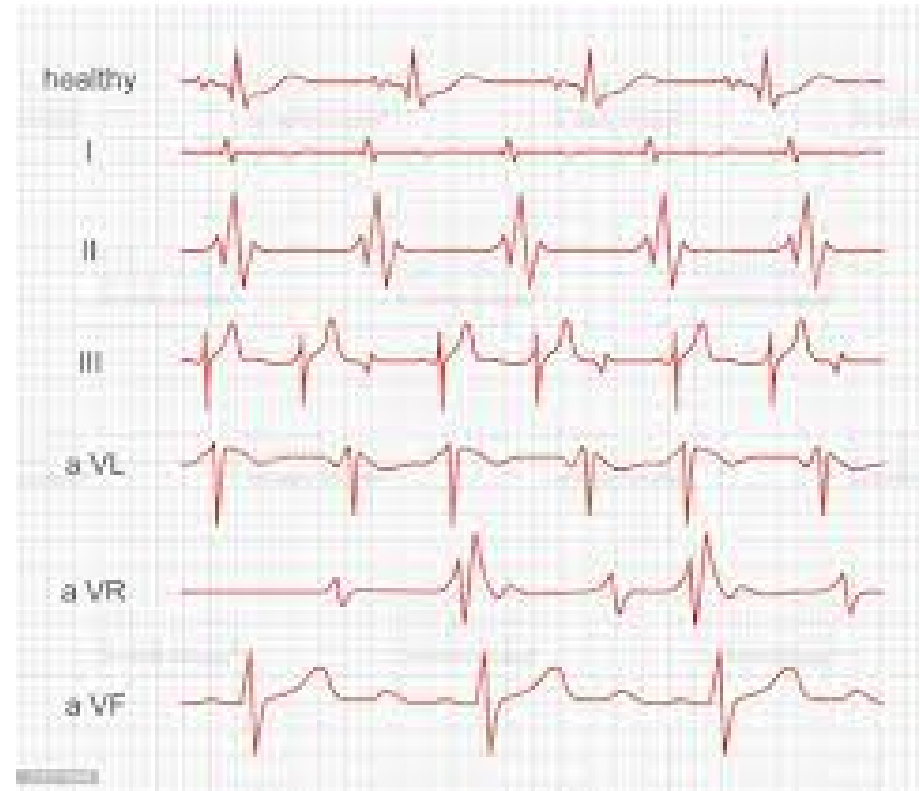
Medical data



Medical data



Medical data



HISTORIA CHOROBY

Pieczętka Zakładu
Dział A

PORADNI

Nr karty

Data zarejestr.

Nr ks. zdrowia

Nazwisko Imię Płeć: M Ż*

Data urodzenia PESEL

--	--	--	--	--	--	--	--	--	--

 NIP

Adres zawód

Miejsce pracy NIP pracodaw.

Rodzaj ubezpiecz. Oddział NFZ nr ubezpiecz.

Ubezpieczony		Nieubezpieczony	
czynny	bierny	samo- płatny	leczone bezpłatnie

Symbol grupy
produkcji i usług
(wpisać odpowiednią cyfrę)

--

Grupa krwi		Rh
podpis lekarza		

Data	Wywiad: objawy, rozpoznanie, leczenie uwagi i podpis lekarza	Nr statystyczny choroby	Czy pierwsze zachorowanie wpisać „tak” lub „nie”	Niezdolność do pracy od - do

Mz/Og-2

Uwaga! Wypełniać długopisem lub maszynowo

HISTORIA CHOROBY

Pieczętka Zakładu
Dział A

PORADNI

Nr karty
Data zarejestr.
Nr ks. zdrowia

Nazwisko Imię Płeć: M Ż*

Data urodzenia PESEL

--	--	--	--	--	--	--	--	--	--

 NIP

Adres zawód

Miejsce pracy

Rodzaj ubezpiecz.

Ubezpieczony			
czynny	bierny		

Data			

krwi		Rh	
podpis lekarza			

pierwsze porowanie sać „tak” b „nie”		Niezdolność do pracy od - do	

Mz/Og-2 Uwaga! Wypełniać długopisem lub maszynowo

HISTORIA CHOROBY

Pieczętka Zakładu
Dział A

PORADNI

Nr karty
Data zarejestr.
Nr ks. zdrowia

Nazwisko Imię Płeć: M Ż*

Data urodzenia PESEL

--	--	--	--	--	--	--	--	--	--

 NIP

Adres zawód

Miejsce pracy

Rodzaj ubezpiecz.

Ubezpieczony		Krew		Rh	
czynny	bierny				

GDPR (RODO)

Data	Niezdolność do pracy	
	od	do

Mz/Og-2 Uwaga! Wypełniać długopisem lub maszynowo

READ: https://en.wikipedia.org/wiki/General_Data_Protection_Regulation

Podmioty uprawnione do dostępu do dokumentacji medycznej:

- pacjent /przedst. ustawowy/ osoba upoważniona przez niego za życia
- osoby prowadzące kontrole
- organ rentowy, Woj. Komisja ds. orzekania o zdarzeniach medycznych
- sąd/prokurator
- okręgowy rzecznik odpowiedzialności zawodowej
- organy władzy publicznej, NFZ, konsultant krajowy i wojewódzki
- inny zakład opieki zdrowotnej zakład ubezpieczeń –za zgodą pacjenta

Ustawa o prawach pacjenta z dnia 06 listopada 2008 r. art. 26

http://www.forensic.umed.wroc.pl/files/archiwum/Prawo_medyczne_Dokumentacja_medyczna.pdf

Useful links:

<https://opensource.org>

<https://choosealicense.com>

[http://www.forensic.umed.wroc.pl/files/archiwum/
Prawo_medyczne_Dokumentacja_medyczna.pdf](http://www.forensic.umed.wroc.pl/files/archiwum/Prawo_medyczne_Dokumentacja_medyczna.pdf)

**In two weeks (25.03.2025)
we will present essays**

**In two weeks (25.03.2025)
we will present essays**

It should take ~10-15 min plus 2-5 min for questions

**In two weeks (25.03.2025)
we will present essays**

It should take ~10-15 min plus 2-5 min for questions

Presenting via the projector in class (thus send the presentation at least one day before)

**In two weeks (27.03.2025)
we will present essays**

It should take ~10-15 min plus 2-5 min for questions

Presenting via the projector in class (thus send the presentation at least one day before)

You present it in the form of power point or html (as you wish)

In two weeks (25.03.2025) we will present essays

It should take ~10-15 min plus 2-5 min for questions

Presenting via the projector in class (thus send the presentation at least one day before)

You present it in the form of power point or html (as you wish)

Need to cover the topic (introduction, definitions, examples, more detailed description of interesting cases)

Katsiaryna Dubrouskaya "Hidden Markov Models in Bioinformatics"
Ryszard Kobiera "Stress induced DNA duplex destabilization"
Zuzanna Milczarska "Pharmacoprint: A Combination of a Pharmacophore Fingerprint and Artificial Intelligence as a Tool for Drug Design"
Nikola Wiejak "Higher-order organization of complex networks"
Ignacy Makowski "CellBox: Interpretable Machine Learning for Perturbation Biology with Application to the Design of Cancer Combination Therapy"
Joanna Dąbrowska "Integration of single cell data"
Julia Szkóp "Folding and Stability of Ancient Peptides"
Kacper Pietrzyk "Multiplex spatial imaging analysis CellCharter"
Agata Paluch "Deep learning for epigenomic prediction"
Oliwia Kozłowska "Using PyMOL as a platform for computational drug design"
Justyna Kowalska "A deep learning model for DNA enhancer prediction based on nucleotide position aware feature encoding"
Veranika Kananovich "Biologically informed variational autoencoders (with gene ontology integration)"
Michał Zgieb "Calculating Biodiversity Based on Phylogenetic Trees"
Anna Szymik "ATAC-STARR-seq"
Michał Stanowski "Algorithms of genomes assembly"
Stanisław Gołębiowski "MMseqs2"
Julia Świątkowska "Integrating AlphaFold with Molecular Docking: A New Era in Drug Design"
Mikołaj Dziewiatowski "Histopathology and Mutational Signatures for Cancer Classification"
Barbara Pawłowska "Cheminformatics Microservice: Unifying access to open cheminformatics toolkits"

Must see:

EWSC: Protein design using deep learning, David Baker

<https://www.youtube.com/watch?v=-H27Kv5duYA>

Also:

<https://www.youtube.com/watch?v=HnT1VWzdFWc>



Thank you for your time
and
See you at the next lecture

Any other
questions & comments

lukaskoz@mimuw.edu.pl