

# Využití veřejných dat pro výzkumné účely z perspektivy uživatele

Jak na data ve veřejné správě

Digitální a informační agentura, Praha, 2025-06-03

Radim Hladík

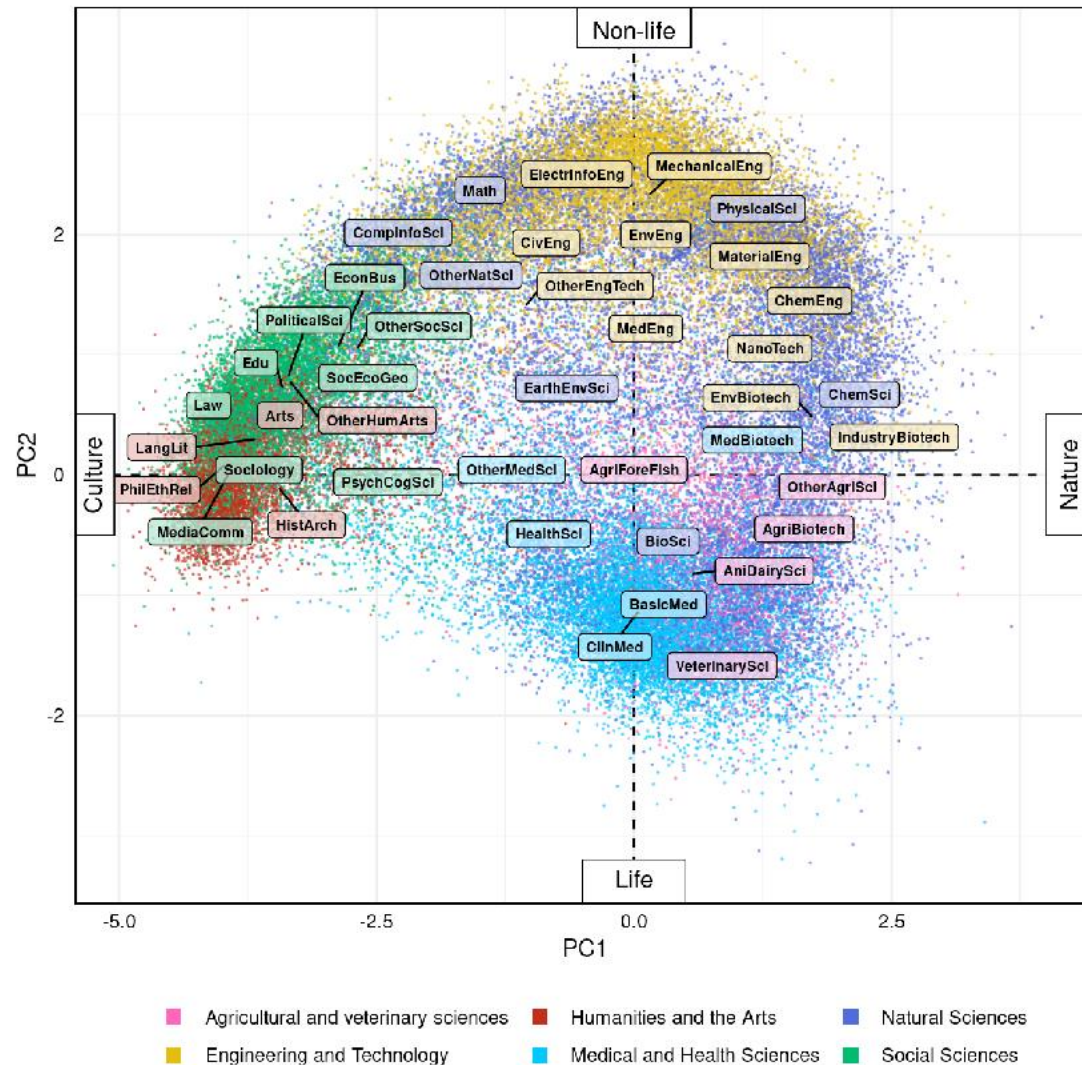
Filosofický ústav AVČR & Česká asociace pro digitální humanitní vědy

hladik@flu.cas.cz, @hlageek.bsky.social

# Agenda

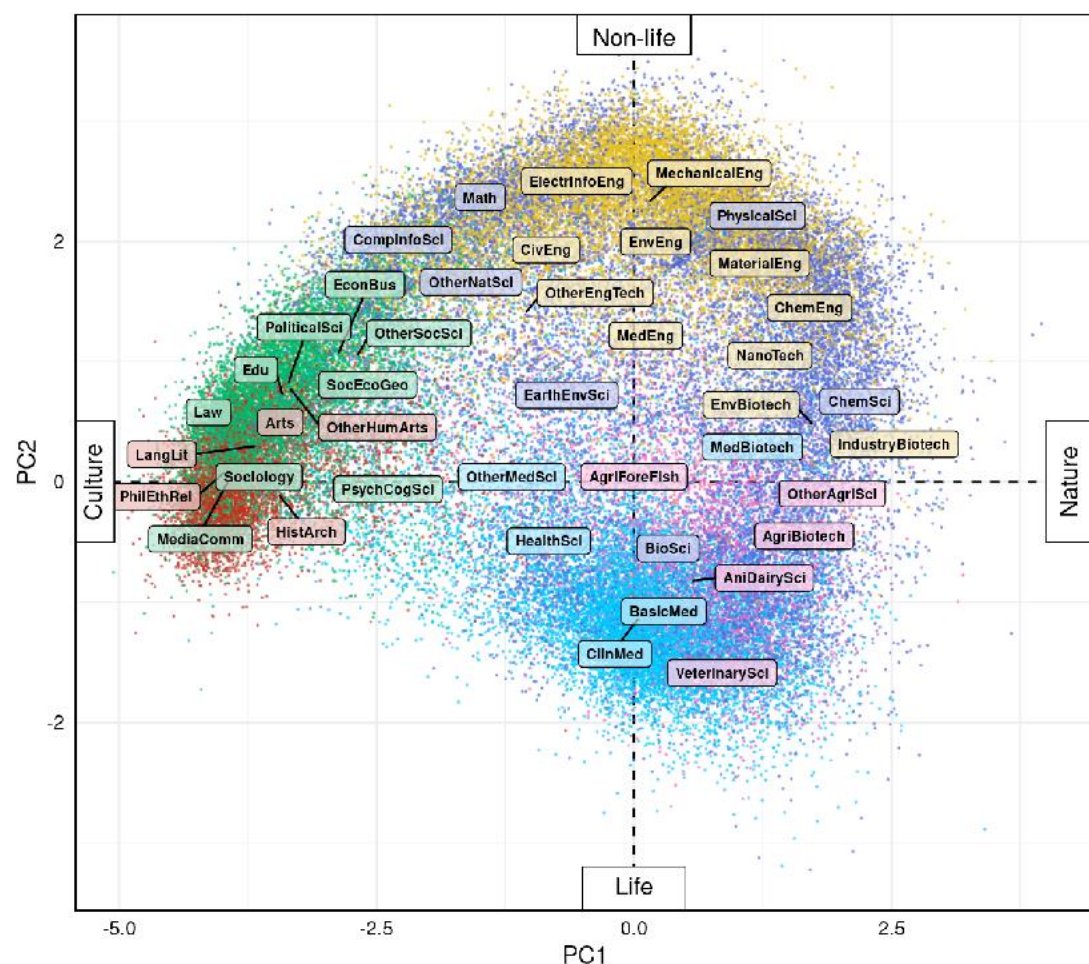
- Příklady využití administrativních dat z IS VaVal pro výzkumné účely
  - Mapa české vědy
  - Porovnání grantových agentur
- (Ne)získávání dat veřejné správy a práce s nimi - anekdotické zkušenosti
  - GAČR
  - ČSÚ
  - IS VaVal
  - RVVI
- Zamyšlení nad problémy a cestami k jejich překonávání

# Mapa české vědy

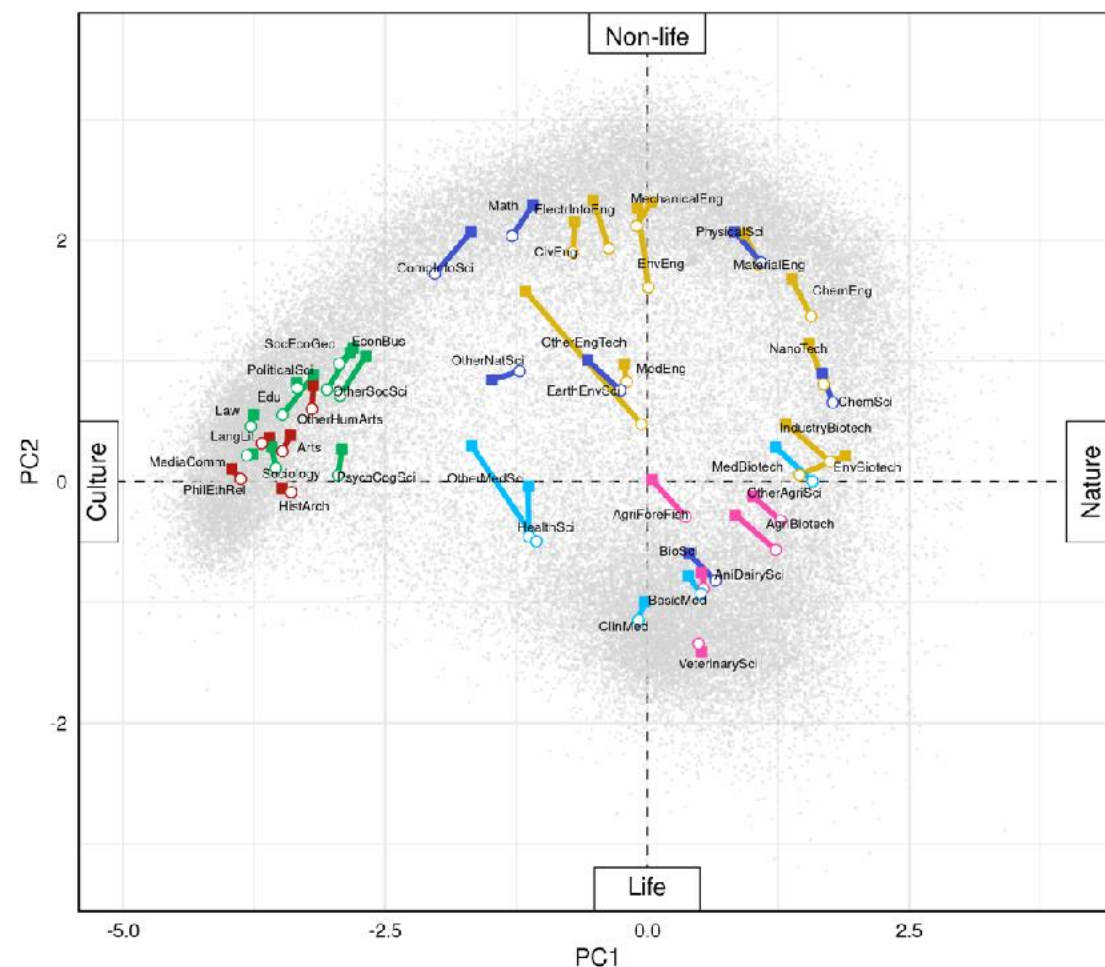


- pozice výzkumníků a výzkumnic na základě tematické struktury oborů
- data z Informačního systému výzkumu, vývoje a inovací (IS VaVal)

# Rozdíly v průměrné pozici žen a mužů



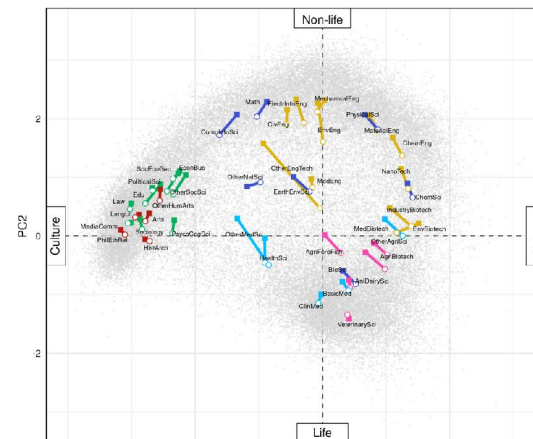
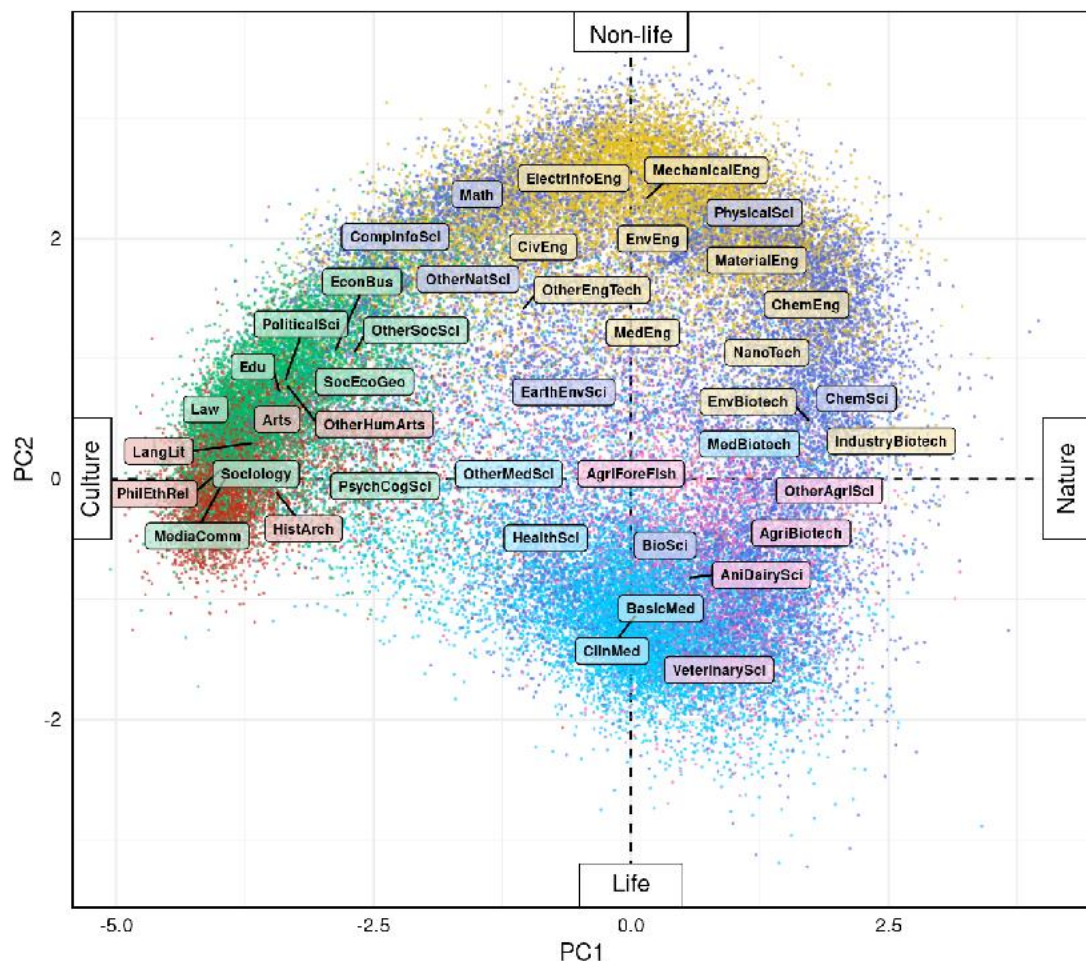
- Agricultural and veterinary sciences
- Humanities and the Arts
- Natural Sciences
- Engineering and Technology
- Medical and Health Sciences
- Social Sciences



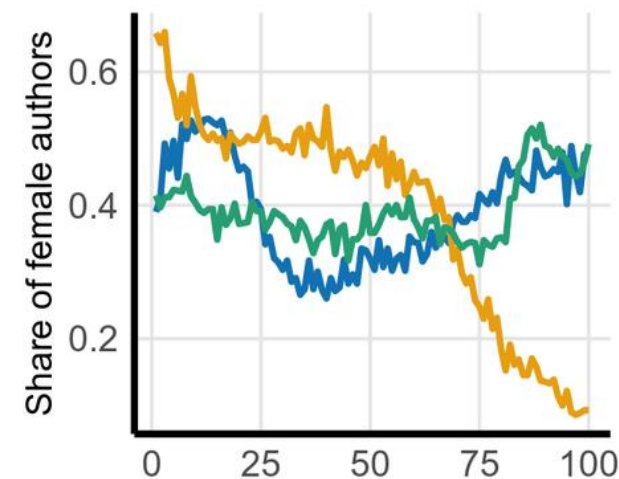
- Agricultural and veterinary sciences
- Humanities and the Arts
- Natural Sciences
- Engineering and Technology
- Medical and Health Sciences
- Social Sciences
- Female
- Male



# Rozdíly v průměrné pozici žen a mužů

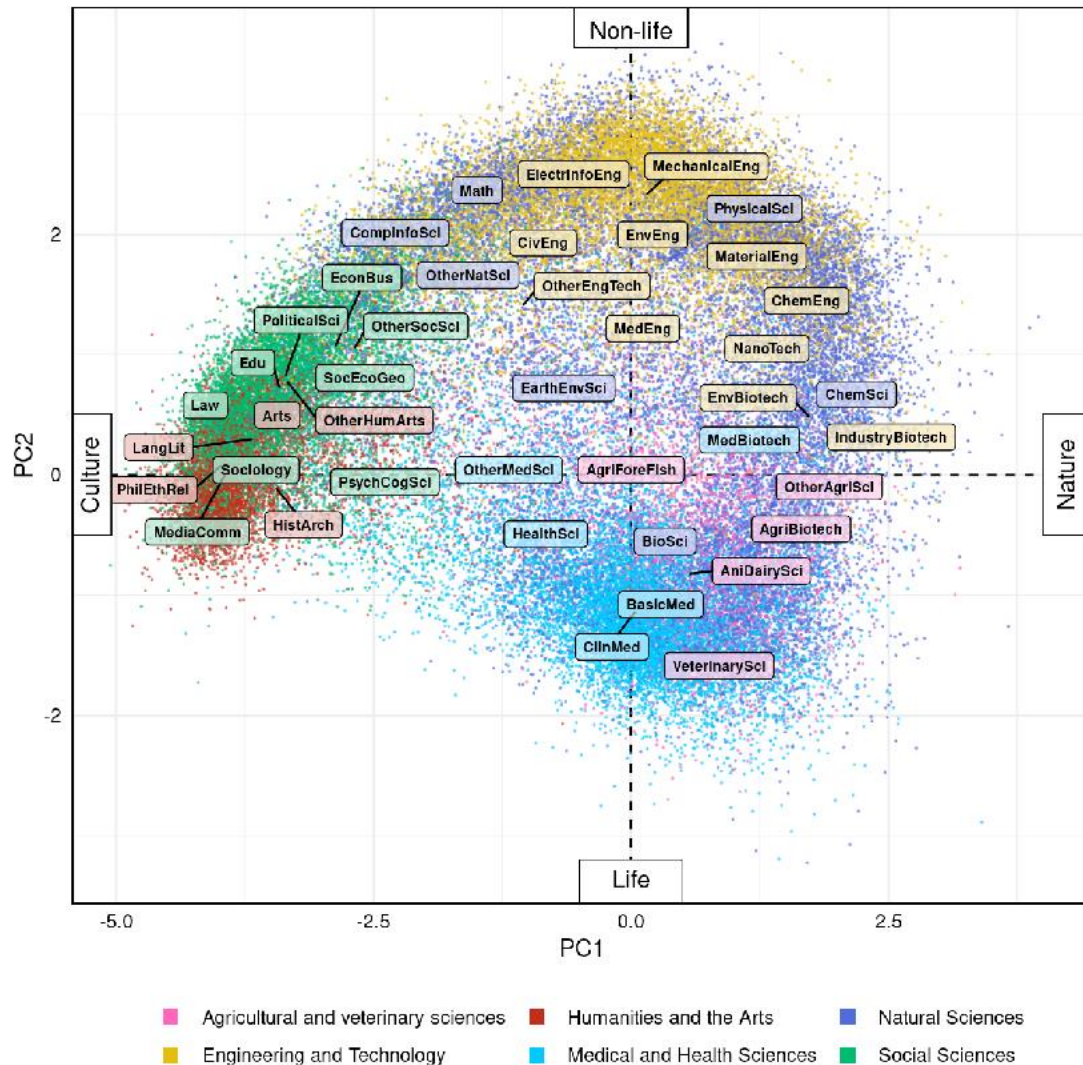


## Female authors



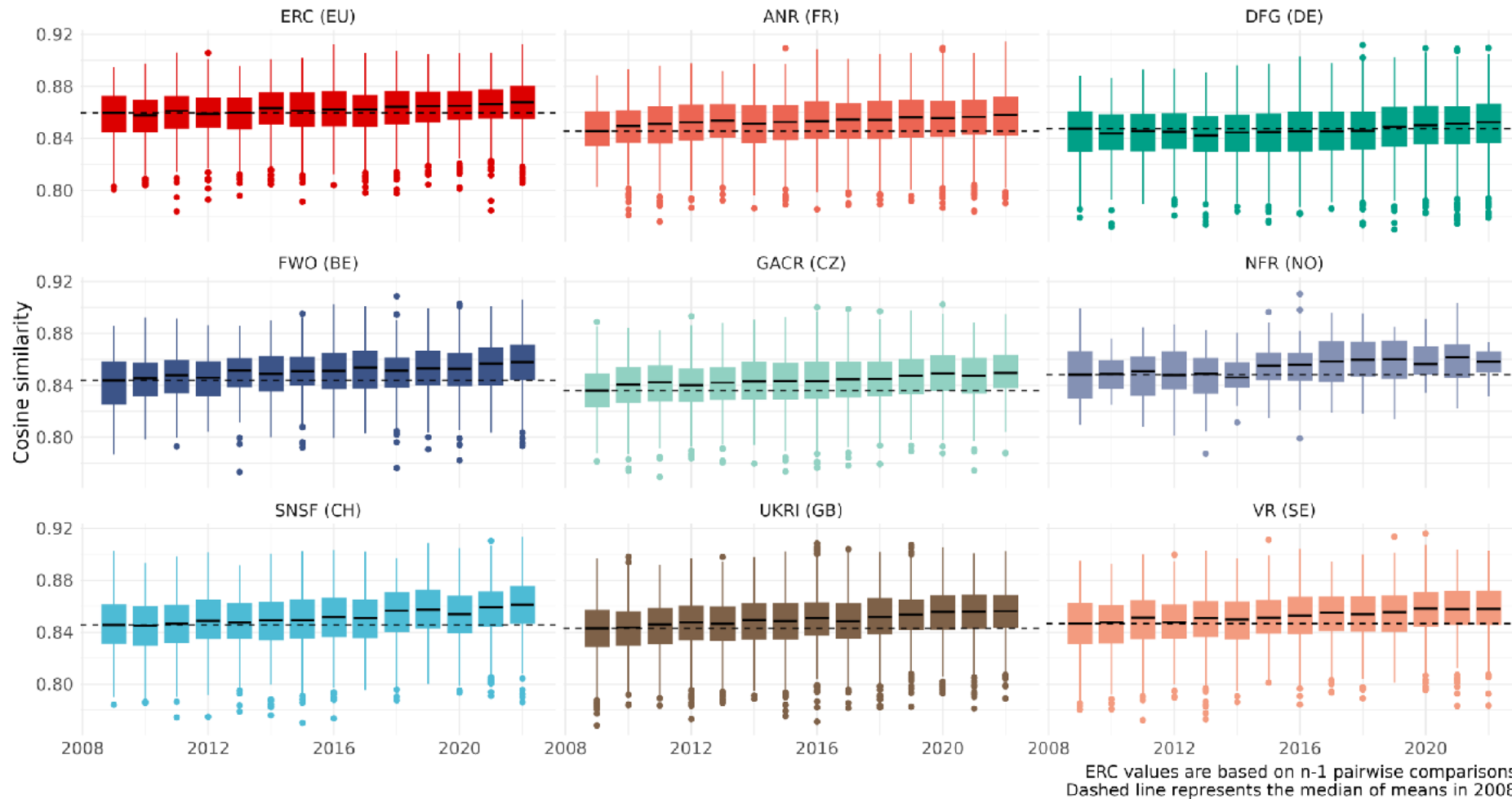
Axis — Culture (0) - Nature (100) — Life (0) - Non-life (100) — Materials (0) - Methods (100)

# Implikace pro praxi



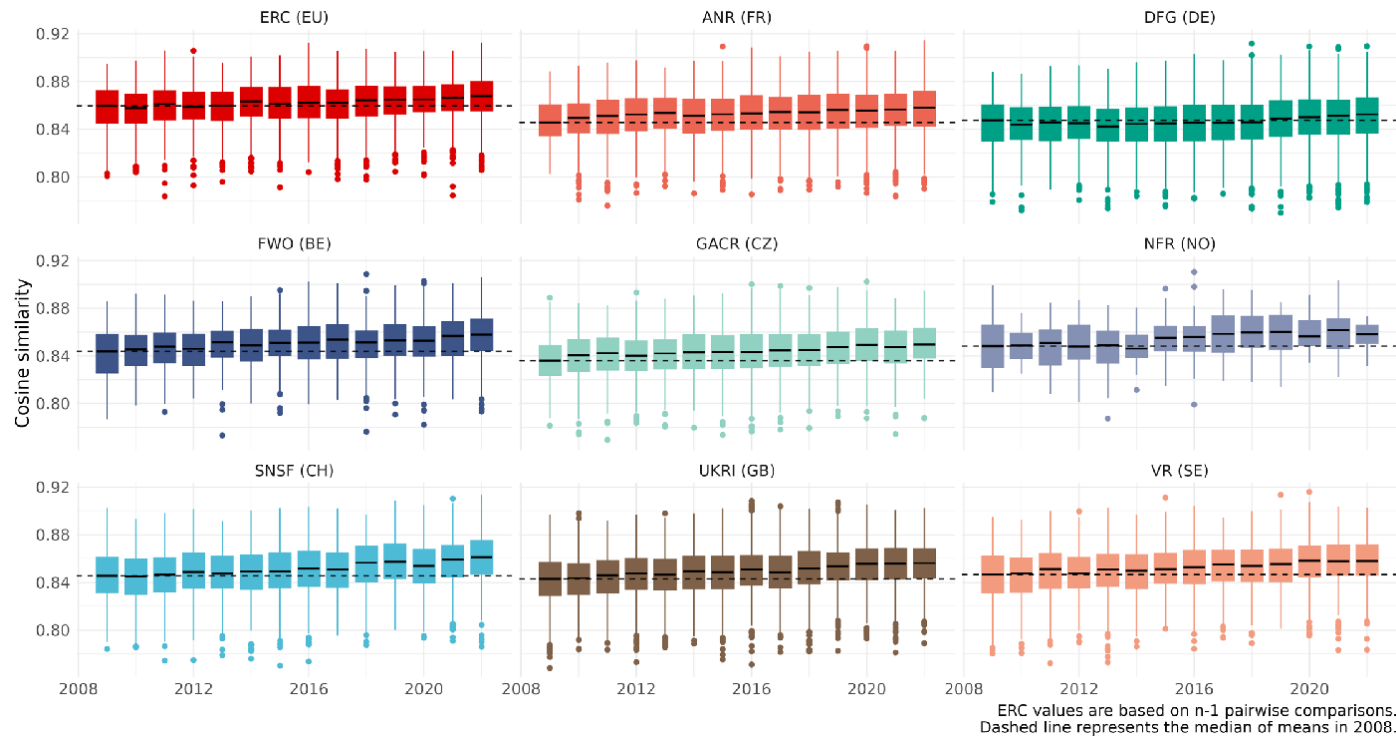
- genderová diverzita sleduje primárně rozdíly v tematickém zaměření (životné vs. neživotné objekty výzkumu) spíše než oborové rozlišení mezi SSH a STEM
- nedostatečné propojení oborů, zejména sociální a kulturní aspekty ve zdravotnictví
  - ale podporu interdisciplinarit nelze aplikovat mechanicky

# Porovnání grantových agentur



- trendy v podobnosti grantových projektů
- ERC benchmark
- data za GAČR z IS VaVal

# Implikace pro praxi



- GAČR sleduje mezinárodní trendy v prezentaci projektů
- sleduje také trendy v poskytování prostředků?
- je tento trend žádoucí pro diverzifikaci grantového portfolia (cf. autonomie DFG)?
- dobrá dostupnost dat za ČR



# Přístup k datům - GAČR

## 1. Přihlášky do veřejných soutěží ve VaVal GAČR

- Unikátní identifikátor přihlášky
- Jméno, příjmení a tituly hlavního navrhovatele a spolunavrhovatelů (nebo anonymizovaný identifikátor) pouze u úspěšných projektů
- Název projektu (ČJ a AJ)
- Rok podání přihlášky
- Obor(y)
- Typ soutěže
- Abstrakt žádosti (ČJ a AJ) pouze u úspěšných projektů
- Panel
- Navržená doba řešení projektu pouze u úspěšných projektů
- Požadované náklady dle let pouze u úspěšných projektů
- Přiznané náklady dle let u podpořených žádostí

## 2. Seznam členů oborových komisí

- Jméno, příjmení a tituly člena komise
- Typ komise dle soutěže a oborů
- Funkce v oborové komisi
- Rok působení v komisi

## 3. Seznam členů hodnotících panelů

- Jméno, příjmení a tituly člena hodnotícího panelu
- Typ panelu dle soutěže a oborů
- Rok působení v panelu

- žádost podána 2020-11-16
- rozhodnutí 2021-04-12
  - „*Předsednictvo Váš výzkum vítá a považuje jej za velmi přínosný. Bohužel však naše možnosti součinnosti jsou omezené, bližší informace naleznete v příloze.*“
- rozsah poskytnutých dat o nepodpořených žádostech neumožnil komparativní výzkum realizovat
- varianta poskytnutí údajů o slovních frekvencích (navržena v žádosti jako alternativa vyžadující předzpracování) nebyla zohledněna

# Přístup k datům - GAČR

## 1. Přihlášky do veřejných soutěží ve VaVaI GAČR

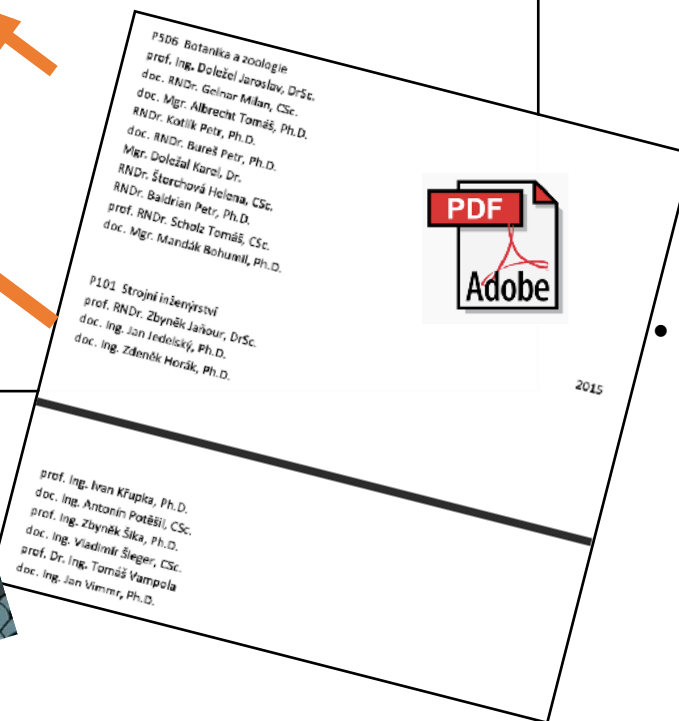
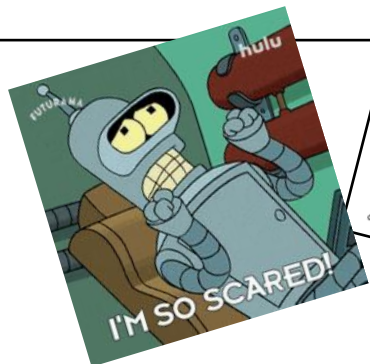
- Unikátní identifikátor přihlášky
- Jméno, příjmení a tituly hlavního navrhovatele a spolunavrhovatelů (nebo anonymizovaný identifikátor) pouze u úspěšných projektů
- Název projektu (ČJ a AJ)
- Rok podání přihlášky
- Obor(y)
- Typ soutěže
- Abstrakt žádosti (ČJ a AJ) pouze u úspěšných projektů
- Panel
- Navržená doba řešení projektu pouze u úspěšných projektů
- Požadované náklady dle let pouze u úspěšných projektů
- Přiznané náklady dle let u podpořených žádostí

## 2. Seznam členů oborových komisí

- Jméno, příjmení a tituly člena komise
- Typ komise dle soutěže a oborů
- Funkce v oborové komisi
- Rok působení v komisi

## 3. Seznam členů hodnotících panelů

- Jméno, příjmení a tituly člena hodnotícího panelu
- Typ panelu dle soutěže a oborů
- Rok působení v panelu



- žádost podána 2020-11-16
- rozhodnutí 2021-04-12
  - „Předsednictvo Váš výzkum vítá a považuje jej za velmi přínosný. Bohužel však naše možnosti součinnosti jsou omezené.“
- rozsah poskytnutých dat o nepodpořených žádostech neumožnil komparativní výzkum realizovat
- varianta poskytnutí údajů o slovních frekvencích (navržena v žádosti jako alternativa vyžadující přepracování) nebyla zohledněna
- ☒ na základě smlouvy o mlčenlivosti data o juniorské soutěži poskytnuta PhD. studentovi - 2024-01-24

# Přístup k datům - ČSÚ

- 2024-01-24
- missing value
- 2024-02-12
- 2024-02-13
- missing value
- 2024-09-13
- missing value

obracím se na Vás s žádostí o data ohledně přechodu mezi kategoriemi zaměstnání.

omlouvám se za další email, ale jelikož jsem dosud neobdržel žádnou odpověď, rád bych se aspoň ujistil, že Vám předchozí zpráva v pořádku dorazila a žádost je ve frontě na zpracování. Předem děkuji za potvrzení přijetí a jsem s pozdravem

je dobře, že se připomínáte. Omlouvám se, že jsem Vám neodpověděla.

Jelikož data nemáme ihned připravena tak, jak byste potřeboval a je nás na zpracování a přípravu dat velice málo, je Vaše žádost v pořadí.

dovoluji si ještě jednou se připomenout s níže zmiňovanou žádostí o data. Zpracování a výpočty si klidně provedu sám, stačí mi to v té nejsurovější podobě, jakou lze poskytnout. Tedy, předpokládám, stačí nahradit citlivé údaje konzistentními anonymizovanými identifikátory. Doufám v kladné vyřízení a zdravím

# O Jindřichovi, Jindře a ČSÚ

- křestní jména ve velkých datech mohou složit jako indikátor vnímaného genderu, nebo i etnicity (např. census USA)
- v českém kontextu i kvůli zákonné úpravě pojmenování velmi přesné (neplatí ve všech kulturně-jazykových kontextech)
- ČSÚ místo dat pro výzkum poskytuje infografiky pro média 🙋 -> nutno využívat jiné služby



## Dětská jména v infografice ČSÚ

13. 08. 2024

Oznámení

Jiná skupina

Pravděpodobně znáte nějakou Janu, Terezu či Elišku a také Jana, Petra, Jakuba...

## Pozvánka na tiskovou konferenci: Dětská jména

22. 05. 2025

Oznámení

Jiná skupina

Pozvánka na tiskovou konferenci: Dětská jména





# Zkušenosti s IS VaVal

- Tři úrovně (možný zdroj problémů)
  - Úřad vlády
  - Poskytovatelé (agentury, ministerstva, AVČR, ...)
  - Příjemci (univerzity, ústavy, ...)
- 👍 API přístup + dokumentace
- 👍 otevřené datové sady
- 👍 Jeden z nejlepších evropských systémů
- 🗨️ Algoritmy nejsou transparentní
- 🗨️ Nižší kvalita dat, přičemž opravy podkladů nemohou správci vymáhat

- 👍 responzivní a ochotný tým

I am a PhD student in Economics under the supervision of Professor Aghion at the Collège de France and the Paris School of Economics. Together with my colleague Raphaël Wargon (cc'd on this email), We are currently working on innovation dynamics in Eastern European countries. We recently read your article "Mapping Knowledge: Topic Analysis of Science Locates Researchers in Disciplinary Landscape" and found it very insightful.

We are particularly interested in obtaining data on publications in the Czech Republic and were intrigued by the potential of extracting information from [isvavai.cz](https://isvavai.cz) as mentioned in your work.

Thank you very much for your response!

We have started to look at the data and have also had an exchange with the people managing the ISVAVAI API, who gave us advice for our project. It seems that we need a bit more time to process all the documentation in Czech, but we would be very happy to discuss the data (and perhaps our project, if it has advanced) with you soon!

# Kvalita dat v IS VaVal

## ÚČASTNÍCI PROJEKTU

Počet příjemců	1
Počet dalších účastníků projektu	0
Příjemce	<a href="#">Ústav makromolekulární chemie AV ČR, v. v. i.</a>
Řešitel	<b>Mgr. Zuzana Morávková, Ph.D.</b> (státní příslušnost: CZ - Česká republika, vedídk: <a href="#">9014701</a> )

ID řešitele projektu se neshoduje s ID autora publikací vykázaných v projektu

Výčet všech uvedených jednotlivých tvůrců	<b>Patrycja Bober</b> (státní příslušnost: PL - Polská republika, domácí tvůrce: A, orcid: <a href="#">0000-0002-7117-8604</a> , scopusid: <a href="#">36141047100</a> , researchid: <a href="#">J-9778-2012</a> ) <b>Zuzana Morávková</b> (státní příslušnost: CZ - Česká republika, domácí tvůrce: A, vedídk: <a href="#">2677741</a> , researchid: <a href="#">D-6162-2012</a> )
---	--

# Kvalita dat v IS VaVal

## ÚČASTNÍCI PROJEKTU

Počet příjemců	1
Počet dalších účastníků projektu	0
Příjemce	<a href="#">Biologické centrum AV ČR, v. v. i.</a>
RIS ZED - ID Akce	<b>Z210501000051</b> (externí ID: GJ1922187Y_60077344, agregační ID: Z210501000000, rpd)
Řešitel	<b>Mgr Igor Koloniuk, Ph.D.</b> (státní příslušnost: CZ - Česká republika, vedídk: <a href="#">7757840</a> )

ID řešitele není u publikací uvedeno, národnost nekonzistentní

Výčet všech uvedených jednotlivých tvůrců	<p><b>Sára Brázdová</b> (státní příslušnost: CZ - Česká republika, domácí tvůrce: A, vedídk: <a href="#">8959240</a>)</p> <p><b>Jana Fránová</b> (státní příslušnost: CZ - Česká republika, domácí tvůrce: A, vedídk: <a href="#">3020630</a>, orcid: <a href="#">0000-0001-8451-1454</a>, researcherid: <a href="#">G-9470-2014</a>)</p> <p><b>Igor Koloniuk</b> (státní příslušnost: UA - Ukrajina, domácí tvůrce: A, orcid: <a href="#">0000-0002-5893-6683</a>, researcherid: <a href="#">G-9526-2014</a>) ???</p> <p><b>Alena Matyášová</b> (státní příslušnost: CZ - Česká republika, domácí tvůrce: A, vedídk: <a href="#">9475097</a>)</p> <p><b>Jaroslava Příbylová</b> (státní příslušnost: CZ - Česká republika, domácí tvůrce: A, vedídk: <a href="#">2845768</a>, orcid: <a href="#">0000-0003-1807-8044</a>, researcherid: <a href="#">G-9517-2014</a>)</p> <p><b>Jana Veselá</b> (státní příslušnost: CZ - Česká republika, domácí tvůrce: A, vedídk: <a href="#">6214215</a>)</p> <p><b>É. Várallyay</b> (státní příslušnost: HU - Maďarsko)</p>
---	--

# Kvalita dat v IS VaVal

Popis výsledku v anglickém jazyce	Článek se zabývá způsoby potlačení interferencí u systémů druhé a třetí generace (2G a 3G). Potlačení interferencí metodou SAIC je v současné době vysoce aktuální, protože interference jsou dominantním omezujícím faktorem kapacity systémů 2G a 3G. V příspěvku je popisován jeden a několika možných algoritmů a jsou uvedeny výsledky jeho simulace v prostředí MATLAB.
Popis výsledku v anglickém jazyce	Der Beitrag beschäftigt sich mit dem Echo der Brünner Vorlesungen von Karl Kraus im Brünner Tagesboten sowie mit Kraus Auseinandersetzung mit Karl Hans Strobl und Hans Müller, die er beide als typische Vertreter der Literatur der Lüge bekämpfte. Leo Reiss, der Herausgeber der Brünner expressionistischen Zeitschrift "Der Mensch", und Karl Wilhelm Fritsch, der Verfasser des Aufsatzes "Blutschuld der Phrase", waren Bewunderer von Kraus, die für ähnliche Werte warben wie er. Ihr Einfluss auf das literarische Leben in Mähren blieb allerdings hinter dem von Strobl bzw. Müller weit zurück.
Výčet všech uvedených jednotlivých tvůrců	<b>Lukáš Dřínovský</b> (státní příslušnost: CZ - Česká republika, domácí tvůrce: A, vedidk: <a href="#">3899993</a> ) <b>Zdeněk Šnirch</b> (státní příslušnost: CZ - Česká republika, domácí tvůrce: A, vedidk: <a href="#">4774396</a> ) <b>Studijní agenda - rozvrhář</b> (státní příslušnost: CZ - Česká republika) <b>Rostislav Drochytka</b> (státní příslušnost: CZ - Česká republika)
Výčet všech uvedených jednotlivých tvůrců	<b>Miroslav Zatloukal</b> (státní příslušnost: CZ - Česká republika, domácí tvůrce: A, vedidk: <a href="#">3417867</a> ) <b>Studijní agenda - rozvrhář</b> (státní příslušnost: CZ - Česká republika)



# Kvalita dat v IS VaVal

Kód UT WoS článku podle Web of Science	10.2495/EEIA100
Kód UT WoS článku podle Web of Science	BLS54
Kód UT WoS článku podle Web of Science	SCOPUS, Code874

## Property talk:P8372

[Property](#) [Discussion](#)

## Documentation

**Web of Science work ID** [Start a query](#) – [Current uses](#) -  
unique academic work identifier assigned in Web of Science bibliograph  
not include WOS: prefix)

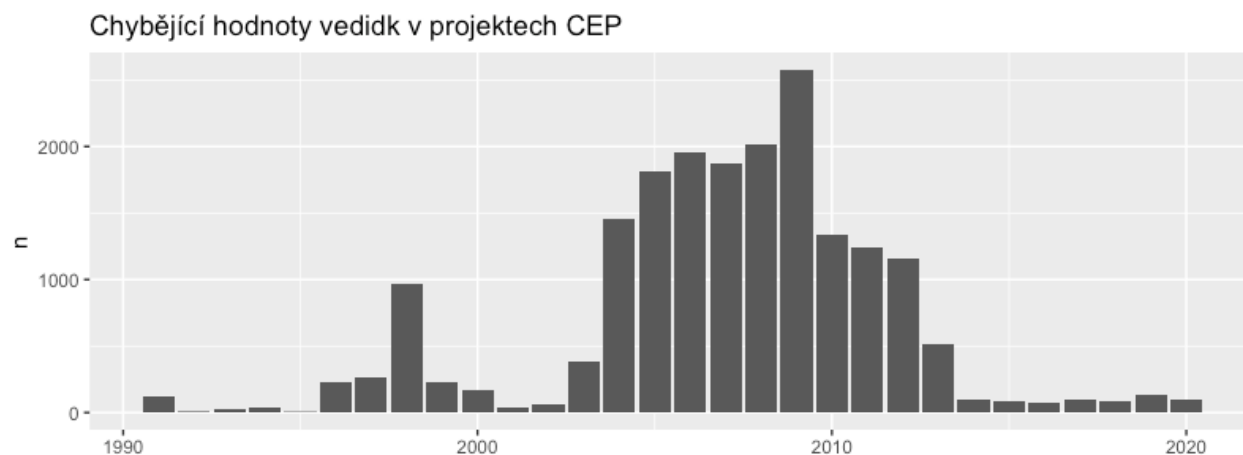
**Applicable** [Web of Science](#) (Q1201886)  
**"stated in"**  
**value**

**Data type** [External identifier](#)

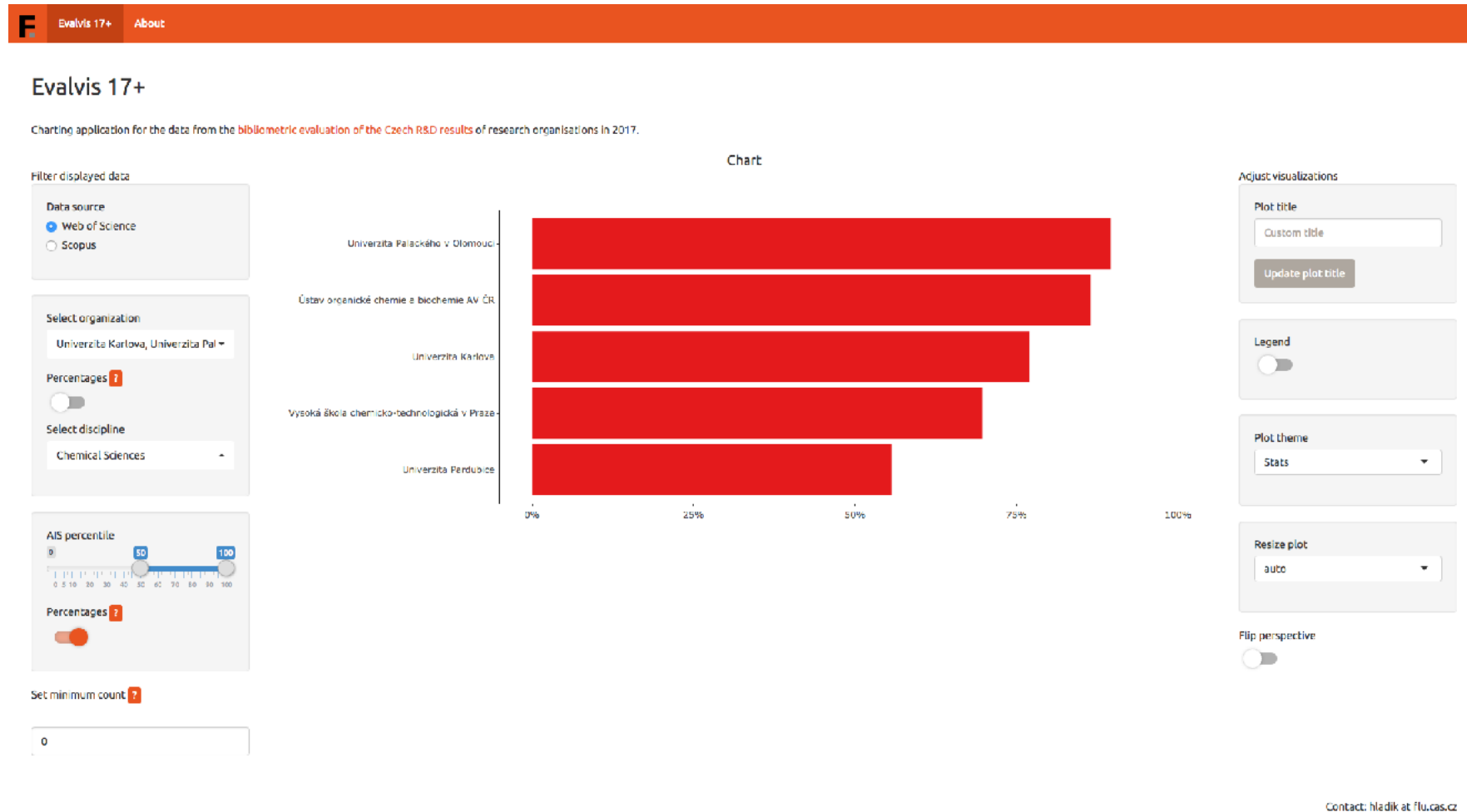
**Allowed values** [A-Z\d]{15}



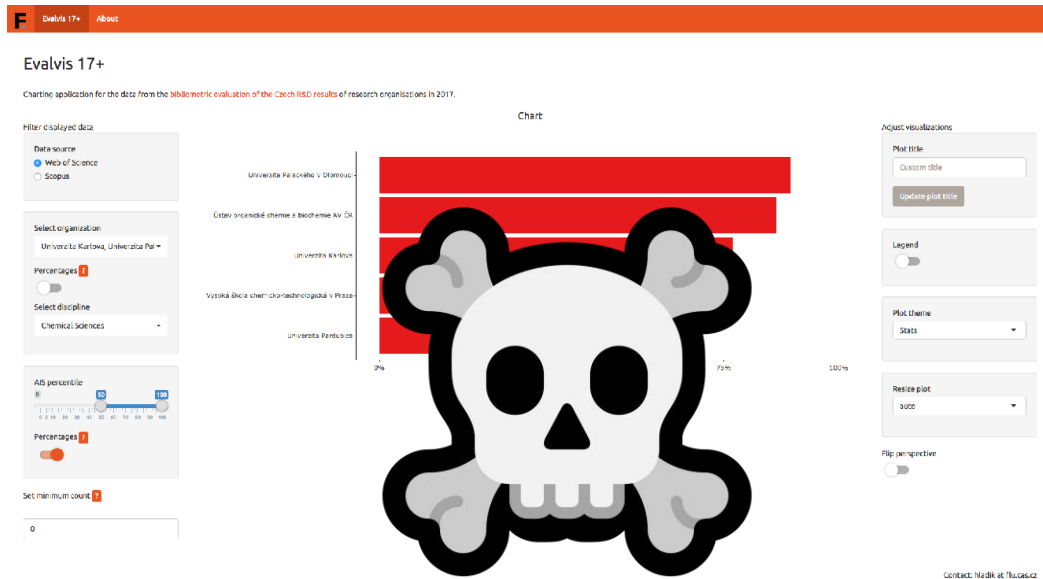
# Kvalita dat v IS VaVal



# Evalvis - zrod a pád aplikace



# Evalvis - zrod a pád aplikace



jsem zpracovatelka ASEP na Archeologickém ústavu AV ČR, Brno. Dávala jsem se dnes na Vaši aplikaci Evalvis 17+  
<http://evalvis.flu.cas.cz/>

....

(ARÚB AV ČR měl v této bibliometrické zprávě pouze 3 výsledky ve WoS s indikátorem AIS, další 4 výsledky z WoS byly neanalyzované. Ale rádi bychom do budoucna grafy poskytované Vaší aplikací využívali.)

Výsledek	časopis	ISSN	AJS
Epistemic logics for sceptical agents	journal of logic and computation	0955-792x	0,45
Frayfish in Czech cultural space: the longest docu	knowledge and management of	19619502	0,26
Čarnap on unfixed science	studies in history and philosophy	393681	0,25
The tempting girl, I know so well': Representation	early science and medicine	13837427	0,20
The Devaluation of the Subject in Popper's Theory	philosophy of the social sciences	483931	0,20
Logically Incorrect Arguments	argumentation: an international	0920-427x	0,18
Exploring Pan-Africanism's theories: from race-ba	journal of african cultural studie	13696815	0,05
Jeden vel'ký kolektiv': Folklór ako vystúpenie a o	sociológia	491225	0,03
Žďařilost, zvládání, teritorium: úvaha k ětosu hra	československá psychologie	0009-062x	0,02
...udický princip v kontextu otázky po žďařilém žive	československá psychologie	0009-062x	0,02
Žěda o well-being - několik (vice méně kritických)	československá psychologie	0009-062x	0,02
Metodologické srovnání práce s fenomenálními p	československá psychologie	0009-062x	0,02

A	B	C	D	E	F	G	H	I	J	K	L
k uplatněn	UT WoS	Výsledkek	rev časopis	print Aut	Aut	drodní spol	Přísn	FORD	číslní jednotka		
2017 WOS.000	The Plate	VIRUSES	Svooda,	Svooda,	NEPRAVDA	Q1	1.6	3 cloy	Ústav molekuliární genetiky AV ČR, v. i.		
2017 WOS.000	Academir	ASUB J IN	Vostal, F	Stockelov	NEPRAVDA	Q2	5.8	Mecir Sociologický ústav AV ČR, v. i.			
2017 WOS.000	Unitary	CONSC	O Marvan,	Polak, M	NEPRAVDA	Q2	5.1	Psych Západočeská univerzita v Plzni			
2015 WOS.000	"The tem"	EARLY SO	Stordhov	Storchov	NEPRAVDA	Q3	6.3	Philor Ficsofský ústav AV ČR, v. i.			
2015 WOS.000	Academir	ASUB J IN	Vostal, F	Stockelov	NEPRAVDA	Q3	1.2	Comp Sociologický ústav AV ČR, v. i.			
2017 WOS.000	Algebras	LOGIC	Punccho	Punccho	NFPAVDA	Q3	1.2	Comp Ficsofský ústav AV ČR, v. i.			
2017 WOS.000	Algebras	LOGIC	Punccho	Punccho	NEPRAVDA	Q3	1.1	Mathy Ficsofský ústav AV ČR, v. i.			
2015 WOS.000	Garnap	STUDIES	Niev, A	Riez, A	NEPRAVDA	Q3	6.3	Philor Ficsofský ústav AV ČR, v. i.			
2015 WOS.000	Crayfish	KNCWLE	Fatoka, J	Patkova, J	NEPRAVDA	Q3	4.1	Agricr Ficsofský ústav AV ČR, v. i.			
2017 WOS.000	Edokine	PHILICS	M AM, Rie	AM, Rie	NEPRAVDA	Q3	6.3	Philor Ficsofský ústav AV ČR, v. i.			
2015 WOS.000	Epistemj	JOURNALS	Bilkova, P	Pels, M	NEPRAVDA	Q3	1.2	Comp Ficsofský ústav AV ČR, v. i.			
2015 WOS.000	Epistemj	JOURNALS	Bilkova, P	Pels, M	NEPRAVDA	Q3	1.1	Mathy Ficsofský ústav AV ČR, v. i.			
2015 WOS.000	The Deyar	PHILOSOF	Paruskrir	Perusnik	NEPRAVDA	Q3	6.3	Philor Ficsofský ústav AV ČR, v. i.			
2017 WOS.000	The Sublr	EARLY SC	Kocanovz	Kocanovz	NEPRAVDA	Q3	6.3	Philor Ficsofský ústav AV ČR, v. i.			
2015 WOS.000	"One Big"	SOLO	Ferberg,	Feiberg,	PRAVDA	Q4	5.4	Socio Ficsofský ústav AV ČR, v. i.			
2015 WOS.000	Grafice	KNCWLE	Fatuka, J	Patkova, J	NEPRAVDA	Q4	1.6	3 cloy Ficsofský ústav AV ČR, v. i.			
2017 WOS.000	Czechoslo	STUD F	Mervart,	Mervart,	NFPAVDA	Q4	6.3	Philor Ficsofský ústav AV ČR, v. i.			
2015 WOS.000	Exploring	JOURNAL	Kasanda,	Kasanda,	NEPRAVDA	Q4	5.9	Other Ficsofský ústav AV ČR, v. i.			
2017 WOS.000	Function	HIST PHIL	Badesa,	C Badesa,	PRAVDA	Q4	1.2	Comp Ficsofský ústav AV ČR, v. i.			
2017 WOS.000	Function	HIST PHIL	Badesa,	C Badesa,	PRAVDA	Q4	6.3	Philor Ficsofský ústav AV ČR, v. i.			
2017 WOS.000	Function	HIST PHIL	Badesa,	C Badesa,	PRAVDA	Q4	1.1	Mathy Ficsofský ústav AV ČR, v. i.			
2015 WOS.000	How co	ČESKOSL	Nitsche,	Nitsche,	NEPRAVDA	Q4	5.1	Psych Ficsofský ústav AV ČR, v. i.			
2015 WOS.000	Logically	ARGUME	Peregrin,	Peregrin,	NEPRAVDA	Q4	6.3	Philor Ficsofský ústav AV ČR, v. i.			
2015 WOS.000	Logically	ARGUME	Peregrin,	Peregrin,	NEPRAVDA	Q4	5.8	Mecir Ficsofský ústav AV ČR, v. i.			
2015 WOS.000	Ludic pr	ČESKOSL	Koubkova,	Koubkova,	NFPAVDA	Q4	5.1	Psych Ficsofský ústav AV ČR, v. i.			
2017 WOS.000	Paul of V	HIST PHIL	Hanke, M	Hanke, M	NEPRAVDA	Q4	1.2	Comp Ficsofský ústav AV ČR, v. i.			
2017 WOS.000	Paul of V	HIST PHIL	Hanke, M	Hanke, M	NEPRAVDA	Q4	6.3	Philor Ficsofský ústav AV ČR, v. i.			
2017 WOS.000	Paul of V	HIST PHIL	Hanke, M	Hanke, M	NEPRAVDA	Q4	1.1	Mathy Ficsofský ústav AV ČR, v. i.			
2015 WOS.000	The scier	ČESKOSL	Urban, P	Urban, P	NEPRAVDA	Q4	5.1	Psych Ficsofský ústav AV ČR, v. i.			
2015 WOS.000	Well-bel	ČESKOSL	Pokorny,	Pokorny,	NEPRAVDA	Q4	5.1	Psych Ficsofský ústav AV ČR, v. i.			



# Nejistota v poskytování dat pro vědecké účely

Legitimní zájem na ochraně soukromí a citlivých dat ✗ zájem na pohodlí

- 106/1999 Sb. Zákon o svobodném přístupu k informacím - pro výzkumné účely není optimální řešení (pro výzkumníky ani pro poskytovatele dat)
- občanskoprávní cesty - např. smlouvy o mlčenlivosti, licenční smluvní vzory, rámcové smlouvy mezi veřejnými a výzkumnými institucemi
- Směrnice Evropského Parlamentu a Rady (EU) 2019/790
  - „Tuto právní nejistotu v oblasti vytěžování textů a dat je třeba řešit stanovením povinné výjimky pro vysoké školy a jiné výzkumné organizace i pro instituce kulturního dědictví z výhradního práva na rozmnožování a rovněž z práva zakázat vytěžování databáze.“
- Nové legislativní a prováděcí úpravy? (vliv rovněž na kvalitu)
- **Otevřená data a API pod veřejnou licencí**

# Otevřená věda - inspirace pro veřejnou správu



správa dat dle principů FAIR

- Findable, Accessible, Interoperable, Reusable, „as open as possible, as closed as necessary“



kurátoři dat / data stewards

- na koho se mohou referenti obracet při vyřizování žádostí o data?



sdílení dat uvažováno ex-ante, nikoliv ex-post



pečlivé plánování datových modelů, vyvažování konzistence a přesnosti



automatizované systémy pro detekci (potenciálně) chybných záznamů



citační vzory pro datasety

- umožňují dohledat práci s daty v odborných publikacích



dialog s vědeckou komunitou

- potřeby vědců a vědkyň jsou specifické
- motivace pro práci s administrativními daty vychází z výzkumných otázek
- způsoby využití nelze předjímat (viz např. frekvence křestních jmen)