



Zborník prezentácií a príspevkov z konferencie ILIDE 2023

ILIDE 2023 conference proceedings

Bratislava 2023



Obsah (contents):

David Minor: Research libraries and society. How do we help fight against misinformation?....3
Pekka Olsbo: The Finnish model for coordinating Open Science.....27
Neil Jefferies: IIIF, annotation & machine learning; Building research workflows on library resources.....36
Daniel Sedlák: Prototyp autonómnej navigácie pre UAV.....49
Daniel Sedlák: Interiérová lokalizácia pre prevádzku UAV.....55
Petr Knoth: OAI identifiers; Decentralised PIDs for research outputs in repositories.....62
Petr Knoth: Core GPT; Combining Open Access research and AI for credible, trustworthy question answering.....76
David Jiří Šlosar: Mapping of scientific outputs of Czech institutions in WoS.....82
Piotr Gołkiewicz: Reaxys; What´s new.....88
Petr Žabička: PERO OCR for prints and manuscripts.....91
Kevin Michael Leonard: Curating without an institutional repository.....114
Martin Lhoták: DL4DH; Digital Libraries for Digital Humanities.....122
Heather Briston: US Academic Libraries and Challenges to Fair Use.....133



Research libraries and society

How do we help fight against misinformation?

David Minor
University of California San Diego Library
dminor@ucsd.edu

Research libraries and society

How do we help fight against misinformation?

David Minor
Director, Research Data Curation Program
UC San Diego Library

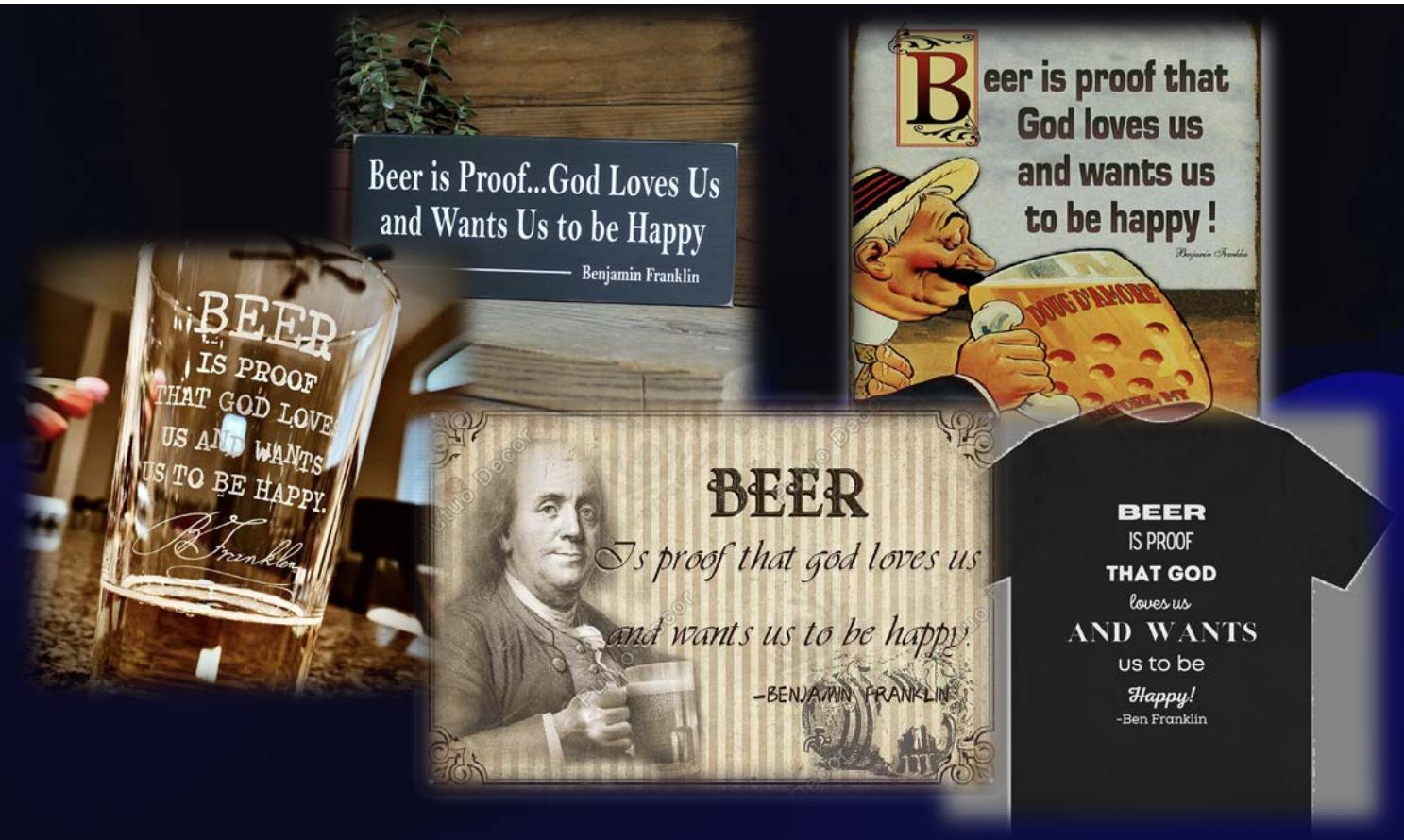
*... but first
... Let's talk about biases. ...*

Examples



Benjamin Franklin

"Beer is proof that
God loves us and
wants us to be
happy,



Keanu Reeves: Hollywood Elites Use 'Blood Of Babies' To Get High

© November 19, 2017 by Baxter Dmitry



submit

Hollywood elites use "the blood of babies to get high" according to Keanu Reeves, who warns that "these people believe the more innocent the child, and the more it suffered before it died, the better the high."

Forums > Lifestyle > Eat-Drink-Man-Woman

Keanu Reeves, Mel Gibson, Jim Carrey say Hollywood Elites eat and drink babies flesh and blood

holytrinity33 · Sep 11, 2022



1 2 3 4 5 Next ▶

... ▾



holytrinity33
Banned

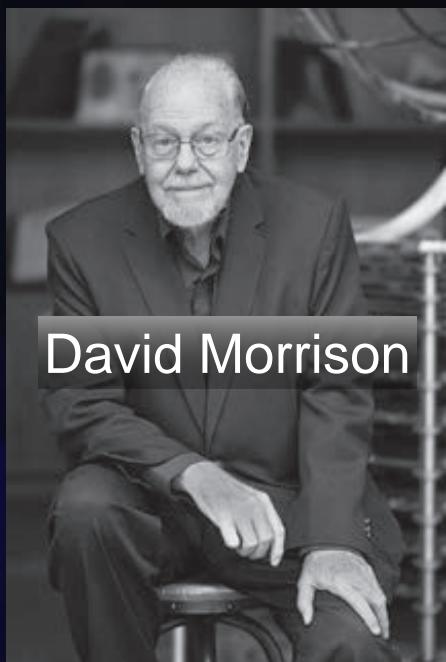
Sep 11, 2022

#1

EvieAlfie Games 5 months ago

My dad was there, there was no laughter there it was added to cover it up. Thanks to Jim for putting his life in danger to say this!!

"Nibiru"



David Morrison

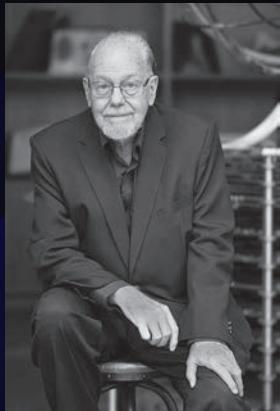
Senior scientist and past Director of Space
at NASA Ames Research Center

Former director of the Carl Sagan Center
for Study of Life in the Universe at the SETI
Institute and of the NASA Lunar Science Institute

A founder of the multi-disciplinary
field of astrobiology

Host of "Ask an Astrobiologist" series on NASA's website

"If NASA denies it, it *must* be true!"



"I'm so scared, I can't sleep."

"I don't want to die."

"You're putting my family at risk."

"... still others asking if they should kill themselves, their children or their pets."

(5G)

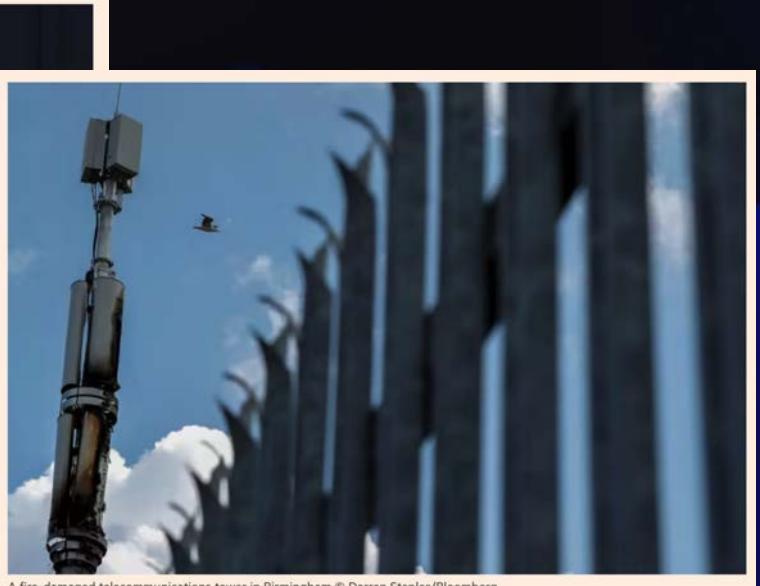


How a 5G coronavirus conspiracy spread across Europe | Free to read

Spate of arson attacks on telecoms masts fuelled by disinformation over pandemic's origins



In the UK, conspiracy theories have even seeped on to mainstream TV, alarming broadcasters and Ofcom © Neil Hall/EPA/Shutterstock



A fire-damaged telecommunications tower in Birmingham © Darren Staples/Bloomberg

"A tsunami of bullshit"
- Dutch Official

The Safety Issue

Safety is our mission, and it guides all of our decisions. In the United States, 5G services launched in [46 markets](#) on January 19, using frequencies in a radio spectrum called the C-band. These frequencies can be close to those used by radio altimeters, an important piece of safety equipment in aircraft. To make sure that this does not lead to hazardous interference, the FAA requires that radio altimeters are accurate and reliable.

Disruption Risk to Aviation from 5G

Because the proposed 5G deployment involves a new combination of power levels, frequencies, proximity to flight operations, and altitude, it may impose restrictions on flight operations using certain types of radio altimeter equipment close to antennas in 5G networks.

These safety restrictions could affect flight schedules and operations. The FAA continues to work every day to reduce effects and move forward with progress to safely integrate 5G and aviation.

Collaborative Work Underway to Reduce Delay, Cancellation Risk

Is my airport affected?

This map shows the percentage of the U.S. commercial fleet and aircraft types that can land at U.S. airports with low-visibility approaches or a high-volume of aircraft with systems that could be adversely affected by 5G.

If your local airport is not on this map, it is likely for one of two reasons:

1. The airport isn't in the 5G deployment area.
2. The airport doesn't currently have the runway systems needed to support low-visibility landings, even without 5G.



The FAA is working to ensure that radio signals from newly activated wireless telecommunications systems can coexist safely with flight operations in the United States, with input from the aviation sector and telecommunications industry.

Definitions

Misinformation:

false information that is spread,
regardless of intent to mislead.

Disinformation:

deliberately misleading or biased information; *manipulated* narrative or facts; propaganda, fake news.

Infodemic:

a rapid and far-reaching spread of both accurate and inaccurate information about something.

Post Truth:

circumstances in which objective facts are less influential in shaping public opinion than appeals to emotion and personal belief

How did we get here?

1. Social media. Duh.

National Library of Medicine study (September 2022)

Found inaccurate or misleading information:

- 51% of posts associated with vaccines
- up to 28.8% of posts associated with COVID-19
- up to 60% of posts related to pandemics
- up to 30% YouTube videos about emerging infectious diseases

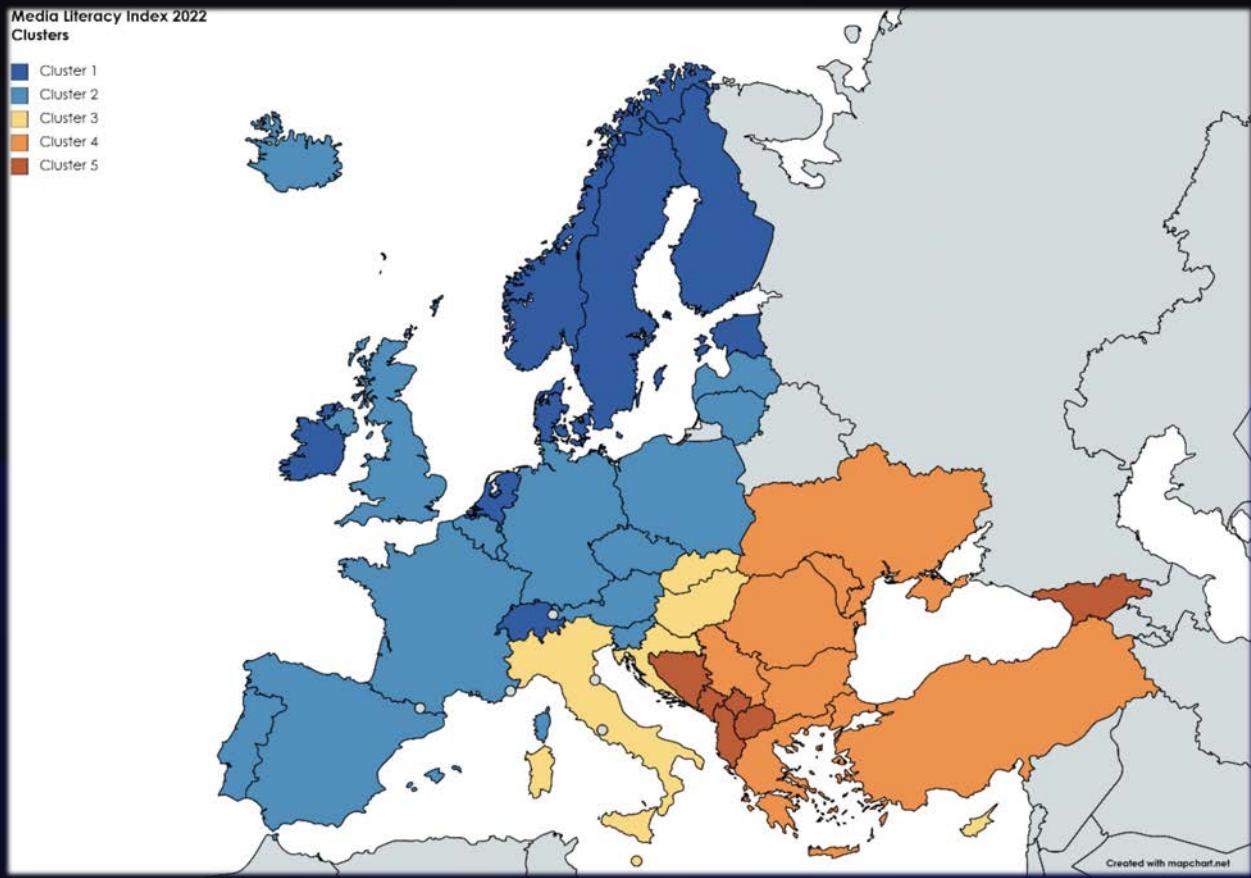
2. Societal development

Rank (1-35)	Country	Score (100-0)	Cluster
1	Finland	76	1
2	Norway	74	
3	Denmark	73	
4	Estonia	72	
5	Ireland	71	
6	Sweden	71	
7	Switzerland	68	
8	Netherlands	66	
9	Germany	62	
10	Iceland	62	
11	UK	62	2
12	Austria	61	
13	Belgium	61	
14	Portugal	61	
15	Spain	59	
16	France	58	
17	Lithuania	58	
18	Czech Republic	57	
19	Poland	56	
20	Slovenia	56	
21	Latvia	54	3
22	Luxembourg	54	

Media Literacy Index 2022 October 2022

<https://www.osce.org/files/f/documents/0/4/534146.pdf>

23	Italy	50	3
24	Slovakia	49	
25	Croatia	47	
26	Malta	44	
27	Hungary	42	
28	Cyprus	41	4
29	Ukraine	39	
30	Greece	38	
31	Romania	36	
32	Serbia	35	
33	Bulgaria	33	5
34	Moldova	32	
35	Montenegro	32	
36	Turkey	31	
37	Albania	25	
38	BiH	24	
39	Kosovo	23	
40	North Macedonia	23	
41	Georgia	20	



- Rapidly emerging
- Large percentage of unknown variables
- Ill-defined consequences or side effects



Photo by [Mason Kimberley](#) on [Unsplash](#)

Tools available to all



Separando la verdad de la ficción en Honduras

+ Enviar contenido



TENDENCIAS

Home /

Us
m
Cha
Psy

Last up

<https://>

Si, hay subsidio para diputados y será voluntario recibirlo

Retractación

#Bono diputados
#Congreso Nacional
#Diputados
#subsidios

ZEDE Próspera ¿se queda o demandan? Verificamos discursos públicos

Sostenible

#Pedro Barquero
#Prospera
#zedes

¿Por qué Honduras queda fuera otra vez de la cuenta del milenio? iVerify te da la información verificada.

Sostenible

#Corrupción
#Cuenta del Milenio
#Política
#Promesas

División a la vista en bancada de Libre en el Congreso Nacional por Eliud Girón

Sostenible Tóxico

#Bancada
#División
#Libre<https://www.undp.org/digital/iverify>

Online misinformation and vaccine hesitancy

Renee Garrett ¹, Sean D Young ^{2,3}

Affiliations + expand

PMID: 34529080 PMCID: PMC8515268 DOI: 10.1093/tbm/bab128

Free PMC article

OXFORD ACADEMIC



ACTIONS

Cite

Collections

PAGE NAVIGATION

< Title & authors

Abstract

Plain language summary

Similar articles

Cited by

References

Publication types

MeSH terms

Substances

Related information

PREV RESULT
3 of 5,735NEXT RESULT
5 of 5,735

Abstract

Although rates of vaccination have increased worldwide, the rise in nonmedical exemptions for vaccination may have caused a resurgence of childhood vaccine-preventable diseases. Vaccine hesitancy plays an important role in the decreasing rates of vaccination and is considered by the World Health Organization as a top ten global threat to public health. Online vaccine misinformation is present in news outlets, websites, and social media, and its rapid and extensive dissemination is aided by artificial intelligence (AI). In combating online misinformation, public health experts, the medical community, and lay vaccination advocates can correct false statements using language that appeal to those who are undecided about vaccination. As the gatekeepers to online information, they can implement and enforce policy that limits or bans vaccine misinformation on their platforms. AI tools might also be used to address misinformation, but more research is needed before implementing this approach more broadly in health policy. This commentary examines the role that different online platforms appear to be playing in the spread of misinformation about vaccines. We also discuss the implications of online misinformation on attitudes about COVID-19 vaccine uptake and provide suggestions for ways to combat online misinformation.

Keywords: COVID-19; Media; Misinformation; Vaccine; Vaccine hesitancy.

Plain language summary

Vaccine hesitancy, the rejection or delay to get vaccinated even if there is an effective vaccine available, may be instrumental in the resurgence of vaccine-preventable disease. Studies have shown that the rise in nonmedical exemptions for vaccination increases rates of childhood vaccine-preventable disease. One factor that influences vaccine hesitancy is online misinformation. False or misleading information online regarding vaccines can be found in independent news outlets, websites, and social media. The spread of vaccine misinformation is especially important during the COVID-19 pandemic as false information can decrease pro-vaccine opinions. The recent announcement of an effective COVID-19 vaccine became a hot topic online, with many adults hesitant to take the vaccine. Public health experts, medical professionals, and pro-vaccine individuals can help curb the spread of misinformation by correcting false statements online. Social media companies can also aid in stopping misinformation by implementing and enforcing policy that limits misinformation on their platforms.

Countering Truth Decay

A RAND Initiative to Restore the Role of Facts and Analysis in Public Life



Photo by vepar5/Adobe Stock

<https://www.rand.org/research/projects/truth-decay.html>



OBJECTIVE ANALYSIS.
EFFECTIVE SOLUTIONS.

Search

RAND Campaign

Fighting Disinformation Home

About

Help Improve Our Data

RAND > Research > Truth Decay > Fighting Disinformation >

Tools That Fight Disinformation Online

Search for tools that fight disinformation by name, type, or by keyword:

examples: Hamilton 2.0, bot detection, fact-checking

[SHOW ALL TOOL CATEGORIES](#) ▾

82 results | [Clear all filters](#)

Adblock Plus

Adblock Plus is a browser extension and app that blocks advertisements and websites through the use of filtering lists. The tool automatically whitelists ads that meet the "acceptable ads" standards — so ads can avoid being blocked on this platform by adhering to those standards. Users can opt out of even these ads by adjusting the default settings. The tool, while initially an adblocker, has increasingly seen itself as a way for users to protect themselves not only from harmful ads, but also from other harmful sites, including those that spread disinformation.

Bad News

This tool intended to build user understanding of the techniques involved in the dissemination of disinformation. This game exposes players to fake news tactics used against them by putting them in the position of a news baron for fake news. Players win by publishing headlines that attract the

<https://www.rand.org/research/projects/truth-decay/fighting-disinformation/search.html>

Course Overview
5 Lessons

Lesson 1
What is Disinformation?
Learn the difference between disinformation and misinformation, why disinformation exists, and how it spreads so easily.

Lesson 2
Provocative, Dramatic, or Outlandish Content
Learn to recognize disinformation that triggers our emotions and how to stop it from spreading.

Lesson 3
Fabricated Claims, Selective Evidence, and Distorted Images or Video
Learn how disinformation is created to look like it's real, and what to do when you spot it.

Lesson 4
Bots and Trolls
Learn what bots and trolls are, why they are dangerous, and how to protect your social media feeds.

Lesson 5
Hacked Accounts
Learn how people use hacking to spread disinformation, how to spot a hacked account, and how to protect your own accounts from being hacked.

<https://www.wikihow.com/Course/How-to-Spot-and-Counter-Disinformation-Online>

<https://www.getbadnews.com>

EN ▾

ATTACK
ITF CHEAT
OBSCURE
DISRUPT
IGNORE START DENY
BAD NEWS CONCEAL DESTROY
From fake news to chaos! How bad are you? Get as many followers as you can.

ABOUT

TILT

Tools from libraries

https://librarydevelopment.group.shef.ac.uk/storyline/questioning/fake-news-game/story_html5.html

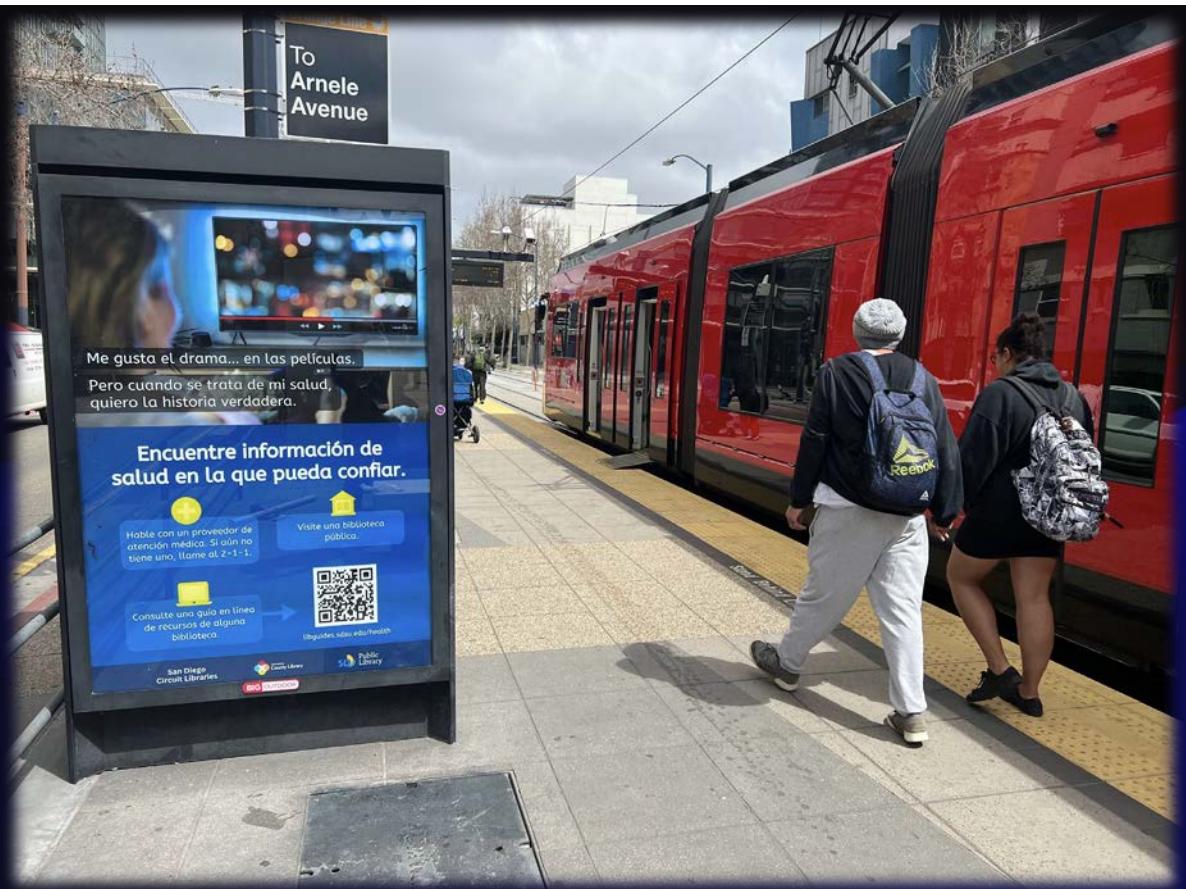
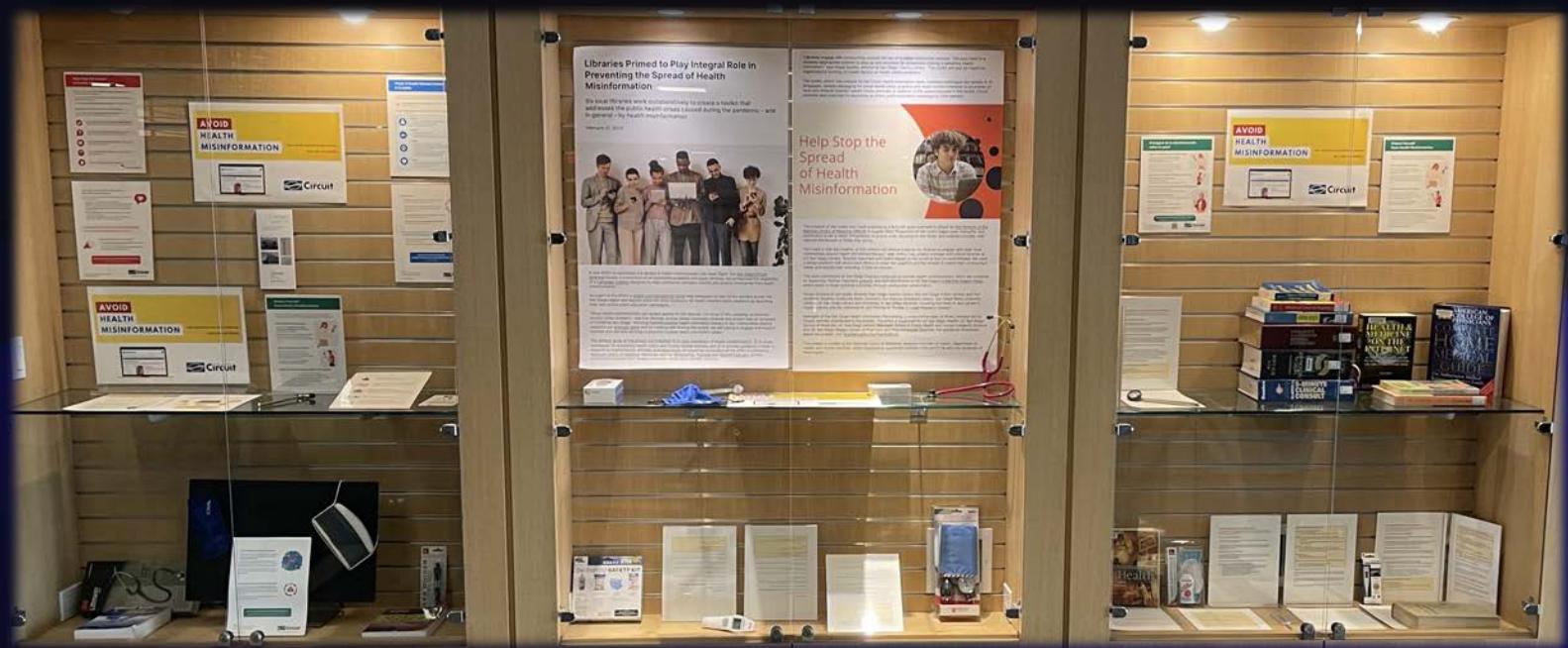
The University Library.
Research Skills and Critical Thinking

Discovering. Understanding. Questioning. Referencing. Creating. Disseminating.



The Fake News Game.

START



Protect Yourself From Health Misinformation

What is health misinformation?

Misinformation is **information that is false, inaccurate, or misleading** according to the best available evidence at the time.

Health misinformation can spread false claims about diseases, illnesses, potential treatments and cures, vaccines, diets, cosmetic procedures, and other health issues.

How to spot health misinformation

Ask questions if you're not sure

Does the information make sense?
Does it seem unbelievable?



Check if credible websites share this information

Type the claim into a search engine.
Do credible sources share this information too?



Look at who wrote the content

Did a health care professional or scientist write the information?

Check for evidence

Is there recent evidence to support the claim?

Talk to a credible health care professional

Ask your doctor or nurse if the information applies to you.

Read more about health misinformation at
[SurgeonGeneral.gov/HealthMisinformation](https://www.surgeongeneral.gov/HealthMisinformation)

Sources:

- Office of the Surgeon General
- National Library of Medicine

Last updated: 2023



libguides.sdsu.edu/health/avoid-misinformation

<https://libguides.sdsu.edu/library-toolkit-addressing-health-misinformation>

Comprueba si el contenido relacionado con la salud



Hazte estas preguntas para verificar las propiedades saludables es creíble:



Proveedor

- ¿Quiénes están a cargo del sitio?
- ¿Por qué han creado el sitio web?



Fondos

- ¿Quién paga el sitio web?
- ¿El sitio web tiene anuncios?



Calidad

- ¿De dónde proviene la información?
- ¿Se basa en **evidencia científica**?
- ¿La información está **actualizada**?
- ¿Quién escribió la información?
- ¿Los **expertos** revisaron la información?
- ¿El sitio web **evita hacer afirmaciones incorrectas**?
- ¿Evita hacer afirmaciones incorrectas?



Privacidad

- ¿El sitio web solicita tu información personal?
- ¿Te dicen cómo lo utilizarán?

Discute la afirmación de propiedades saludables con tu profesional de la salud antes

Obtén más información sobre cómo evaluar la información de salud a través de MedlinePlus en bit.ly/hi-evals

Tumulong na pigilan ang pagkalat ng maling impormasyon sa kalusugan

Ang maling impormasyon ay impormasyon walang katotohanan, hindi tumpak, o mapanilang ayon sa pinakamahusay na mapagamit na ebidensya sa panahong iyon. Maaari itong magdulot ng kalithan at hindu pagkakaunawaan. Maaari itong humantong sa mga tao na gumawang ng paggapsaya na may mapaminsalang kahihinathan.

Maprotektahan natin ang ating sarili at ang ating pamilya at mga kaibigan mula sa maling impormasyon sa kalusugan.

Tukuyin ang maling impormasyon sa kalusugan at iwasang ibahagi ito

Gaminin ang **Health Misinformation Checklist** ng Surgeon General sa tuwing makakataugo ka ng paghayag na may kaugnayan sa

- Nap-check ka ba sa **CDC** o makita mayroong anumang pagkakasabot?
- Nagtanong ka ba sa isang ng pangkakalusugan tulad ng iyong karagdagang impormasyon?
- Ipinasok mo ba ang pahayag na bereripika ng isang mapagkakaunawaan?
- Tiningnan mo ba ang pahayag na mapagkakatiwalaan mo?
- Kung hindi ka sigurado sa iri

Makipag-usap sa iyong pamilya, impormasyon sa kalusugan

Ang pagbibigay ng mga "fact-check" hindu palaging epektibo. Subukan ang

- **Makinig** – Subukang unawaan ang makikita.
- **Makiramay** – Magtanong pa gaano kahirap maghanap ng sagot.
- **Ituro ang mga mapagkakaunawaan tulad ng MedlinePlus**.
- **Huwag ipahiya sa publiko** malumanay at maunawan.
- **Gumamit ng inklusibong pariralang tulad ng "Nanilito rin**

Giúp ngā chyện sự lan truyền của thông tin sai lệch về sức khỏe

Thông tin sai lệch là thông tin hư cấu, không chính xác hoặc gây hiểu nhầm theo các bằng chứng tốt nhất hiện có tại thời điểm đó. Thông tin này có thể gây ra sự bối rối và hiểu lầm.

Thông tin này có thể khiến mọi người đưa ra quyết định với những hậu quả có hại.

Chặng ta có thể bảo vệ bản thân và gia đình và bạn bè của chúng ta khỏi những thông tin sai lệch về sức khỏe.

Phát hiện thông tin sai lệch về sức khỏe và tránh chia sẻ thông tin sai

Sử dụng Danh sách Kiểm tra của Surgeon General's **Health Misinformation Checklist** mỗi khi quay bít gặp nội dung liên quan đến sức khỏe mà không chắc chắn về nội dung đó.

- Quý vị đã già một chuyện giả châm sóc sức khỏe đúng tin cậy như bác sĩ hoặc y tá của quý vị xem họ có bắt kỹ thông tin bổ sung nào không?
- Quý vị đã nhận các tuyên bố vào một công cụ tìm kiếm để xem nội dung đó đã được xác minh bởi một nguồn đáng tin cậy chưa?
- Quý vị đã xem trang "Giao thiệu về chúng tôi" trên trang mạng để xem liệu quý vị có thể tin tưởng nguồn không?
- Nếu quý vị không chắc chắn về nội dung, đừng chia sẻ = **Rất quan trọng**

Trao đổi với gia đình, bạn bè và cộng đồng của quý vị về thông tin sai lệch về sức khỏe

Đưa ra "xác minh dữ kiện (fact-check)" cho những người chia sẻ tuyên bố sai không phải lúc nào cũng hiểu quả. Hãy thử các mẹo sau:

- **Lắng nghe** – Cố gắng hiểu rõ sự thật và niềm tin của họ.
- **Đồng cảm** – Đặt câu hỏi để hiểu họ đến đâu. Nói về việc khó tìm được thông tin từ khía cạnh đó.
- **Chỉ cho các nguồn đáng tin cậy** – Cho họ xem các nguồn như **MedlinePlus** hoặc trang mạng của **Trung tâm CDC**.
- **Đừng gầy xéo hỏi một cách cung khai** – Cố gắng trả chuyện trực tiếp một-một với nhau. Hãy nhẹ nhàng và thấu hiểu.
- **Sử dụng ngôn ngữ hòa nhập** – Thể hiện rằng quý vị đồng nhất với họ. Hãy thử các cụm từ như "Tôi cũng bối rối" hoặc "Tôi hiểu".

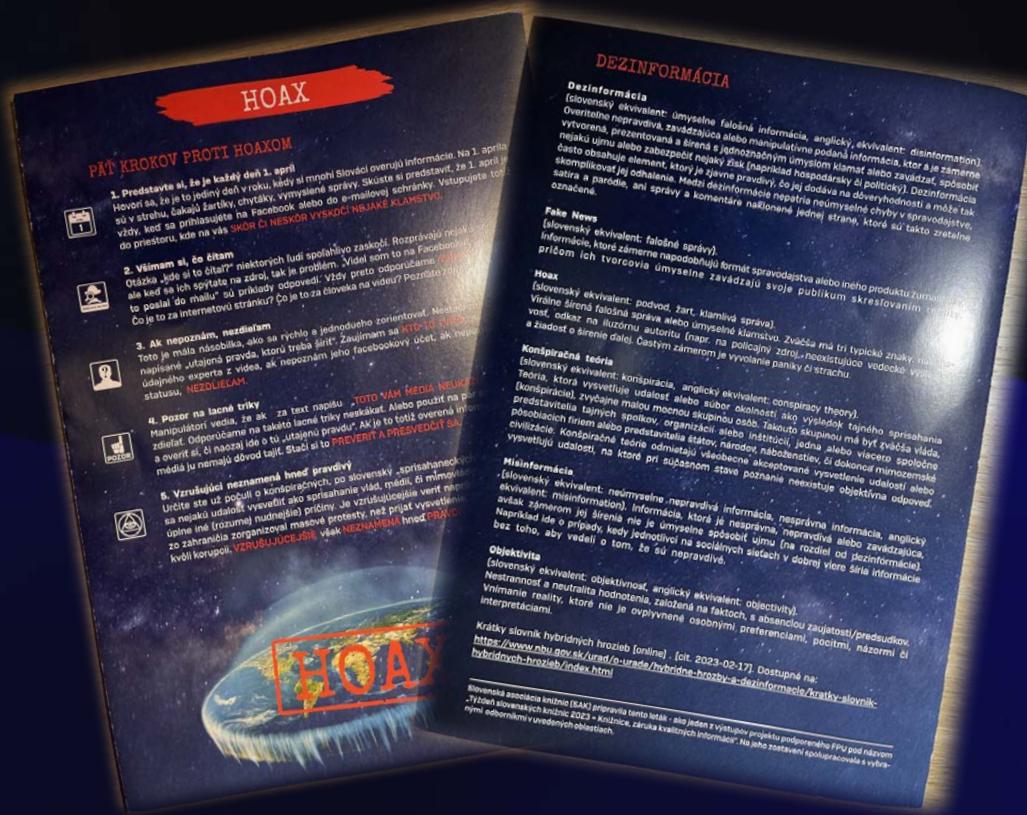
Khuyến khích mọi người nói chuyện với chuyên gia chăm sóc sức khỏe của họ

- Đề nghị họ hỏi bác sĩ hoặc y tá về những lo lắng về sức khỏe của họ.

The Library of Juraj Fándly in Trnava



Krajská knižnica Ľudovíta Štúra in Zvolen



DEZINFORMÁCIA

Desinformácia
(slovensky ekvivalent: umyseľne falóšna informácia, anglicky ekvivalent: disinformation).
Ovocnečné nepravdivá, zavádzajúca alebo manipulačne podané informácia, ktorá je zameraná vytvárať, prezentovať a poslat s jehožnásledom umyselnou klasifikáciu alebo zavádzat, spôsobením nejakú účinku alebo zabezpečiť nejaký zisk (napr. hospodársky či politický). Dezinformácia skomplikuje element, ktorý je ľavne pravdivý, čo jej dodáva na dosahu výrovnosť a môže tak skomplikovať aj odhalenie. Medzi dezinformácie patrili neúmyseľne chyby v spravodajstve, časopise, a parodie, ani správy a komentáre naikonané jednej strane, ktoré sú takto zreteľne označené.

Fake News
(slovensky ekvivalent: falošné správy).

Hoax
(slovensky ekvivalent: podvod, žart, Hamľivá správa).
Virálne šírené falošné správy alebo umyseľné klamstvá. Zvyčajne má tri typické znaky: nájom, vost, odkaž na iluzívnu autoritu (napr. na poliacių zdroj, neexistujúce vedecké výskumy) a žiadost o šírenie ďalej. Často sú využívané paniky či strachu.

Konspiračná teória

(slovensky ekvivalent: konspiračia, anglicky ekvivalent: conspiracy theory).
Teória, ktorá vysvetľuje udalosť alebo súbor okolností ako výsledok tajného sprisahania (konspiracie), vyzývajúca malou možnosť skupinovú cestu. Takisto skupinu môžu byť zvečas všechny pôsobiacich firm, alebo predstaviteľov štátov, národov, náročenstiev, či dokonca mimozemské civilizácie. Konspiračná teória odmietá väčšinou akceptované vysvetlenie udalosti alebo vysvetľujú udalosť, na ktorú pri súčasnom stave poznania nesie existuje objektívna odpoveď.

Mieleniformácia

(slovensky ekvivalent: nedužinovo-nepravdivá informácia, nepravdivá informácia, anglicky ekvivalent: misinformation). Informácia, ktorá je nepravdivá, nepravdivá alebo zavádzajúca, alebo zavádzajúca ďalšiu mielu medzi skupinou osôb. Takisto skupinu môžu byť zvečas všechny pôsobiacich firm, alebo predstaviteľov štátov, národov, náročenstiev, či dokonca mimozemské civilizácie. Konspiračná teória odmietá väčšinou akceptované vysvetlenie udalosti alebo vysvetľujú udalosť, na ktorú pri súčasnom stave poznania nesie existuje objektívna odpoveď.

Objektivita
(slovensky ekvivalent: objektivnosť, anglicky ekvivalent: objectivity).

Neutralnosť a neutralita hodnotenia, založená na faktoch, s absentiou zaujatostí/predstupu. Nezávislosť reality, ktorá nie je ovplyvnená osobními preferenciami, pocitmi, názormi či interpretáciami.

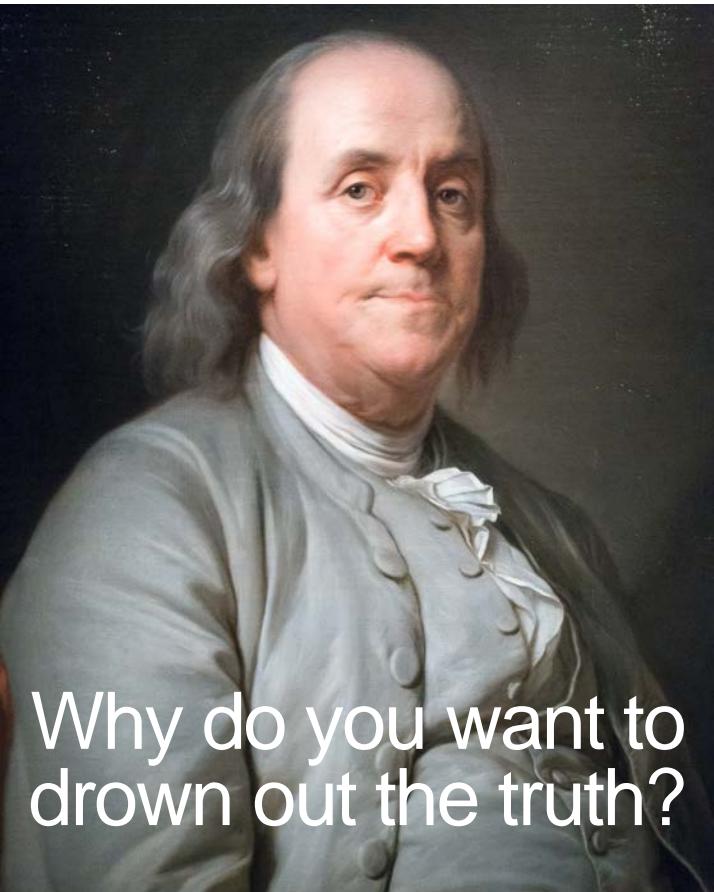
Krátky sloviník hybridných hrozíc (online). [cit. 2023-02-17]. Dostupné na: <https://www.rbu.gov.sk/urad/o-urade/hybridne-hrozby-dezinformace-kratky-slovinik-hybridnych-hroziba/index.html>

... and now
Some Deep Thoughts



INFODEMIC





Why do you want to drown out the truth?

He made wine to make us happy.
When you see your neighbor at the table pouring wine into his glass, do not rush to pour water into it.

B Franklin

Thank you.

dminor@ucsd.edu



The Finnish model for coordinating Open Science

Pekka Olsbo
Open Science Centre, University of Jyväskylä
pekka.o.olsbo@jyu.fi



The Finnish Model for Coordinating Open Science

Pekka Olsbo
Open Science Centre, University of Jyväskylä
ILIDE 2023

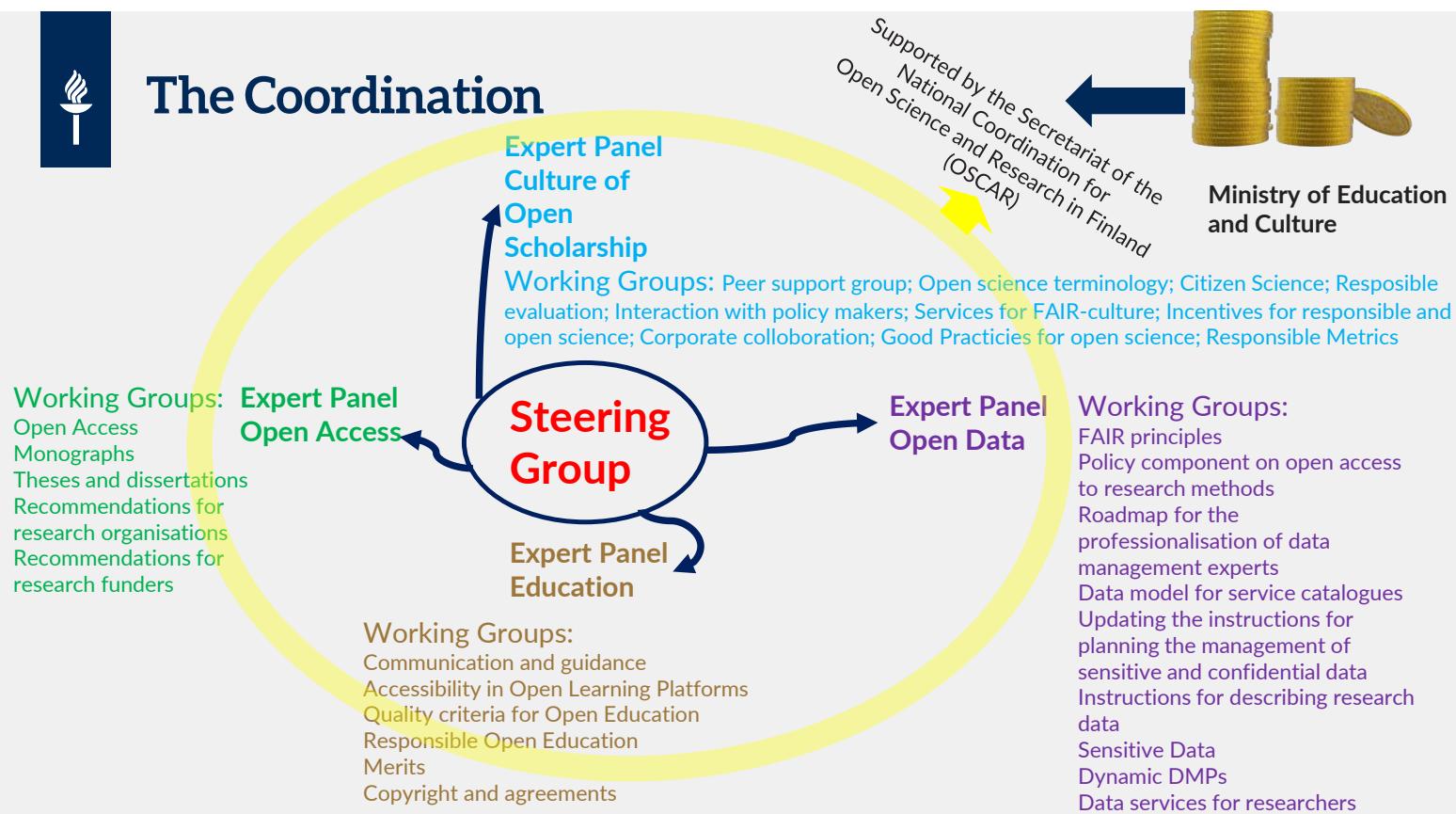


Introduction – national coordination and policies in Europe

- In many European countries Open Science is coordinated by the Ministry of Education/Science and/or by the national research funders
- National Open Science policies are usually created and defined by the ministry
- But this is not the case in Finland

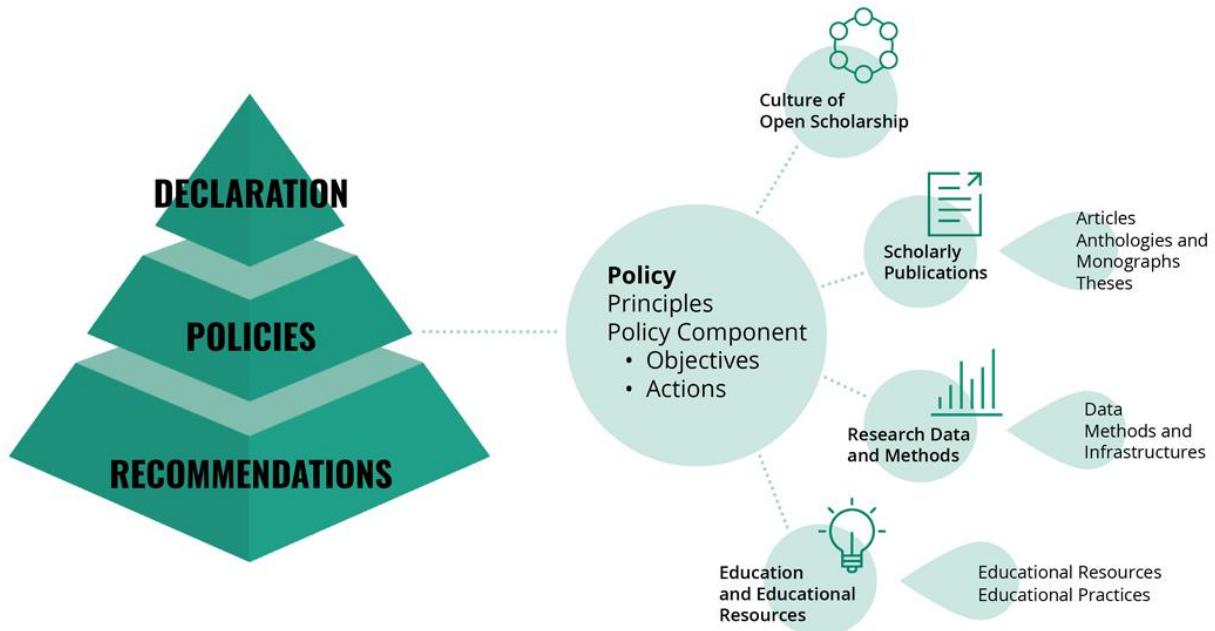


The Coordination





Policies



Declaration for Open Science and Research

- <https://avointiede.fi/en/policies/declaration-open-science-and-research-2020-2025>
- The Declaration for Open Science and Research presents a common vision for the Finnish research community. This vision states that open science and research should be integrated in researchers' everyday work and support not only the effectiveness of research outputs but also the quality of research. Furthermore, the vision sees the Finnish research community as an international forerunner in open science and research.
- The declaration was accepted at the Steering Group meeting on 10th of December 2019 and published on 9th of January 2020.



More detailed view on policies and recommendations 1

Policy for Open Scholarship

- [Open science recommendation and checklist for research, development and innovation activities in collaboration between research organizations and companies](#)
- [Recommendation for citizen science](#)
- [The self-evaluation tool for culture of open scholarship services](#)

Open research data and methods. National policy and executive plan by the higher education and research community for 2021–2025. Policy component 1: Open access to research data

- [Interoperable 1-2-3. Guidelines on fostering data interoperability](#) (in Finnish)
- [Template for agreement on data authorship](#) (in Finnish)
- [The use of Persistent Identifiers for Research Datasets](#)
- [Ownership and licencing of research data](#) (in Finnish)
- [Additional instructions for planning the management of sensitive and confidential data](#)



More detailed view on policies and recommendations 2

Open access to scholarly publications – National policy and executive plan by the research community in Finland for 2020–2025: Policy component for open access to journal and conference articles

- [Policy component for compilations and individual works draft](#) (in Finnish)
- [Recommendations on technology for open publishing](#) (2021, in Finnish)
- [Recommendations on open access to scholarly publications for research funders](#) (in Finnish)
- [Recommendations on open access to scholarly publication for research organisations](#) (in Finnish)
- [Guidelines for scholarly publishers on creating responsible data policies](#) (in Finnish)
- [Guidelines for researchers and scholarly publishers on open licensing](#) (in Finnish)



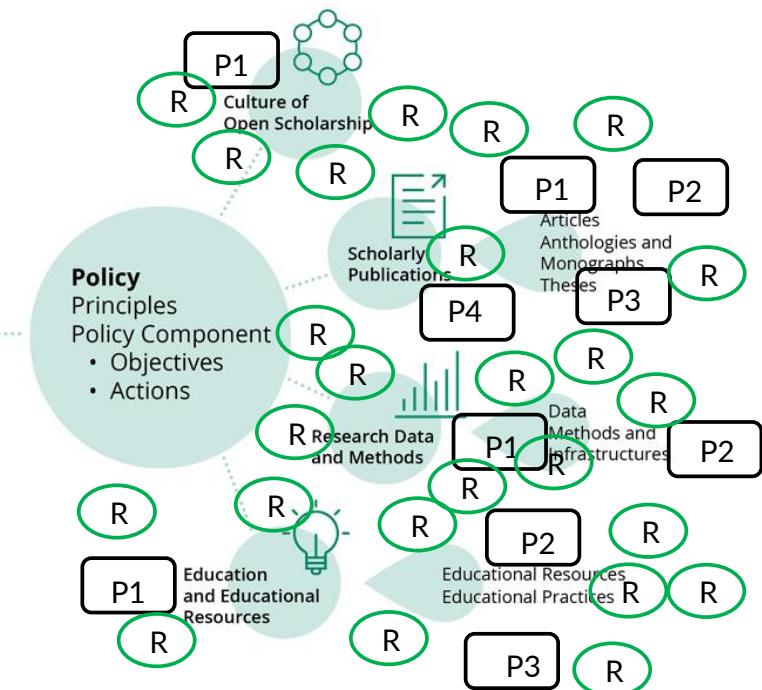
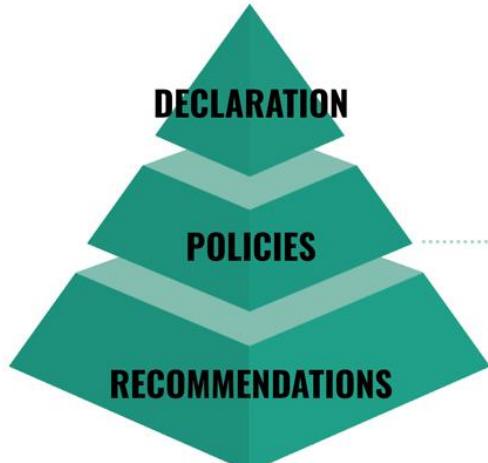
More detailed view on policies and recommendations 3

[Open education and educational resources. National policy and executive plan by the higher education and research community for 2021-2025. Policy components 1 \(Open access to educational resources\) and 2 \(Open educational practices\)](#)

- In preparation: Recommendation on quality criteria for Open educational practices
- In preparation: Recommendation on accessibility of open learning environments
- Recommendation on integrity and inclusiveness of open education (2023, [final draft in Finnish](#))
- Recommendation on services, tools and good practices related to OER ([final draft in Finnish](#))
- Quality criteria for open educational resources (part of [Recommendations for open education 2021](#))
- Guidelines on the accessibility of open educational resources (part of [Recommendations for open education 2021](#))
- Recommendation on gaining merit in open education and open educational resources (part of [Recommendations for open education 2021](#))
- Competence requirements for open education (part of [Recommendations for open education 2021](#))
- Recommendation on copyright issues to be taken into consideration in the open publication of educational resources (published 2020 as an appendix to [Policy for open educational resources](#))

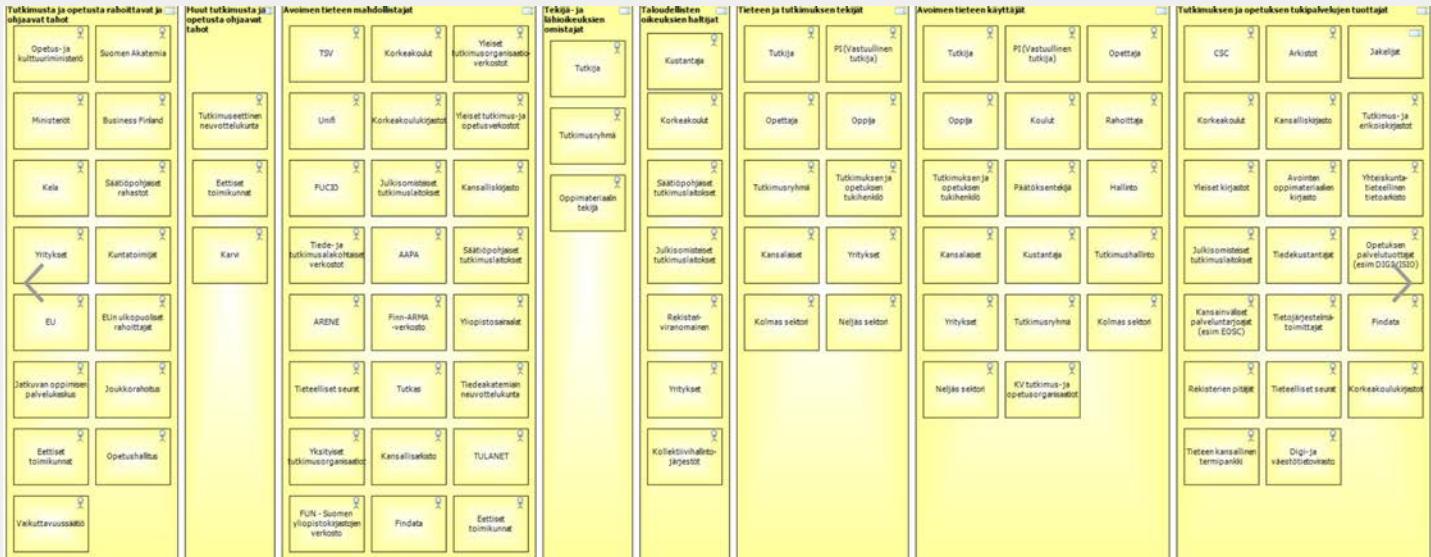


The Jungle of Policies and Recommendations

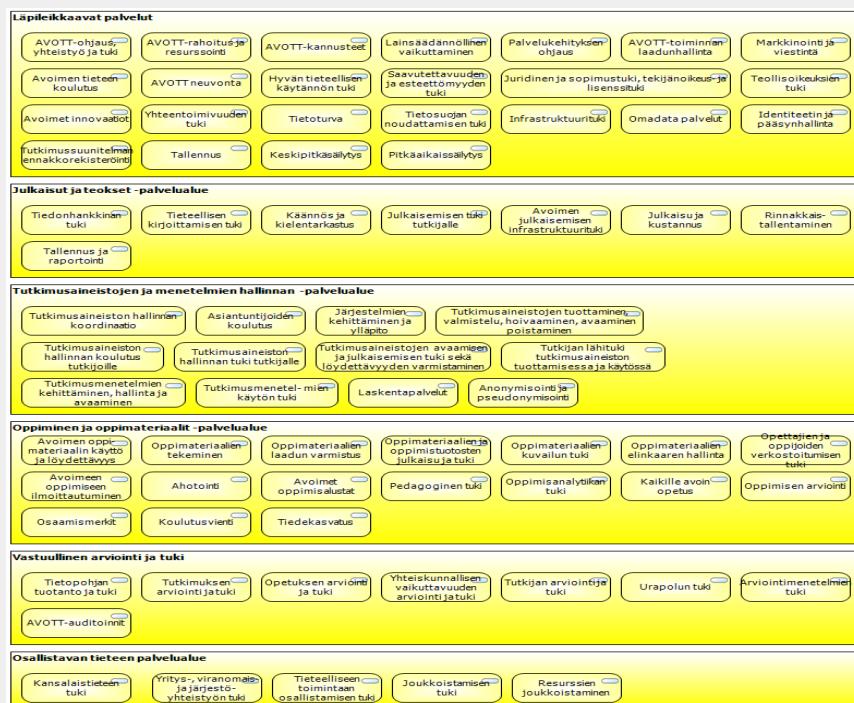




The Architecture of Open Science: Roles

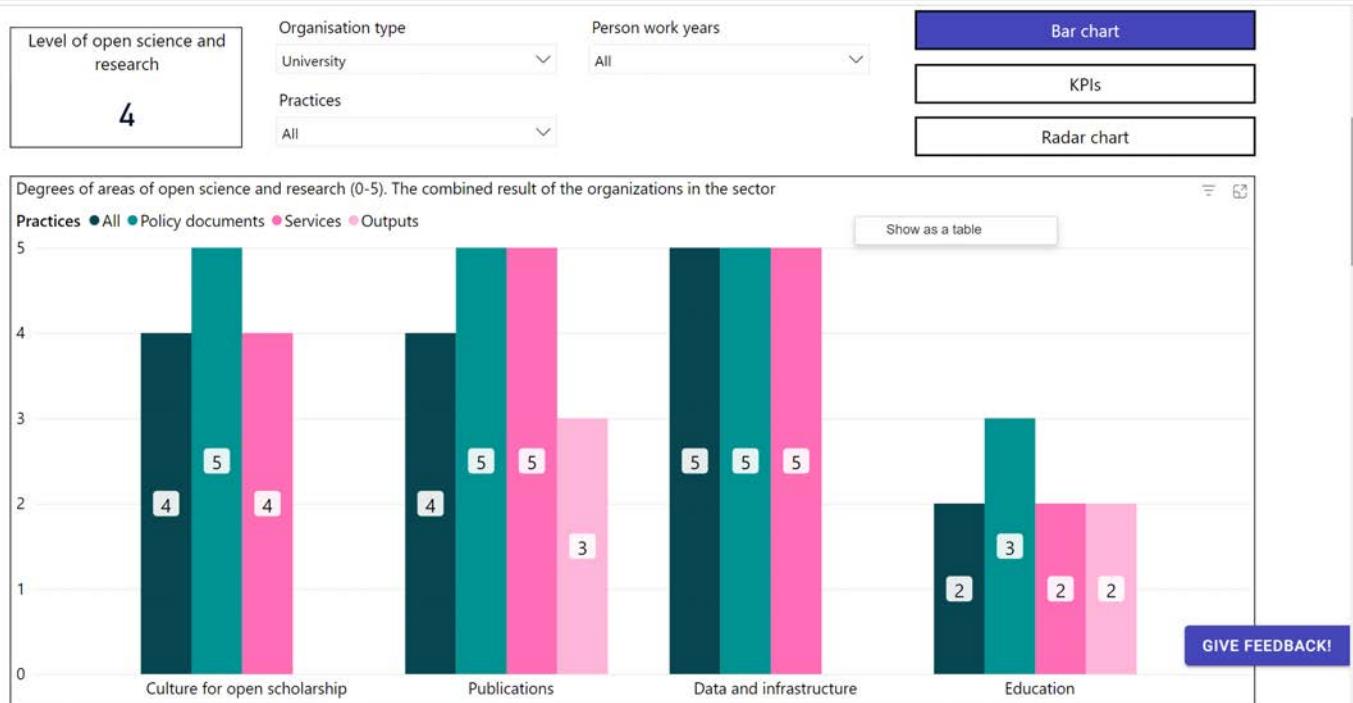


The Architecture of Open Science: Services





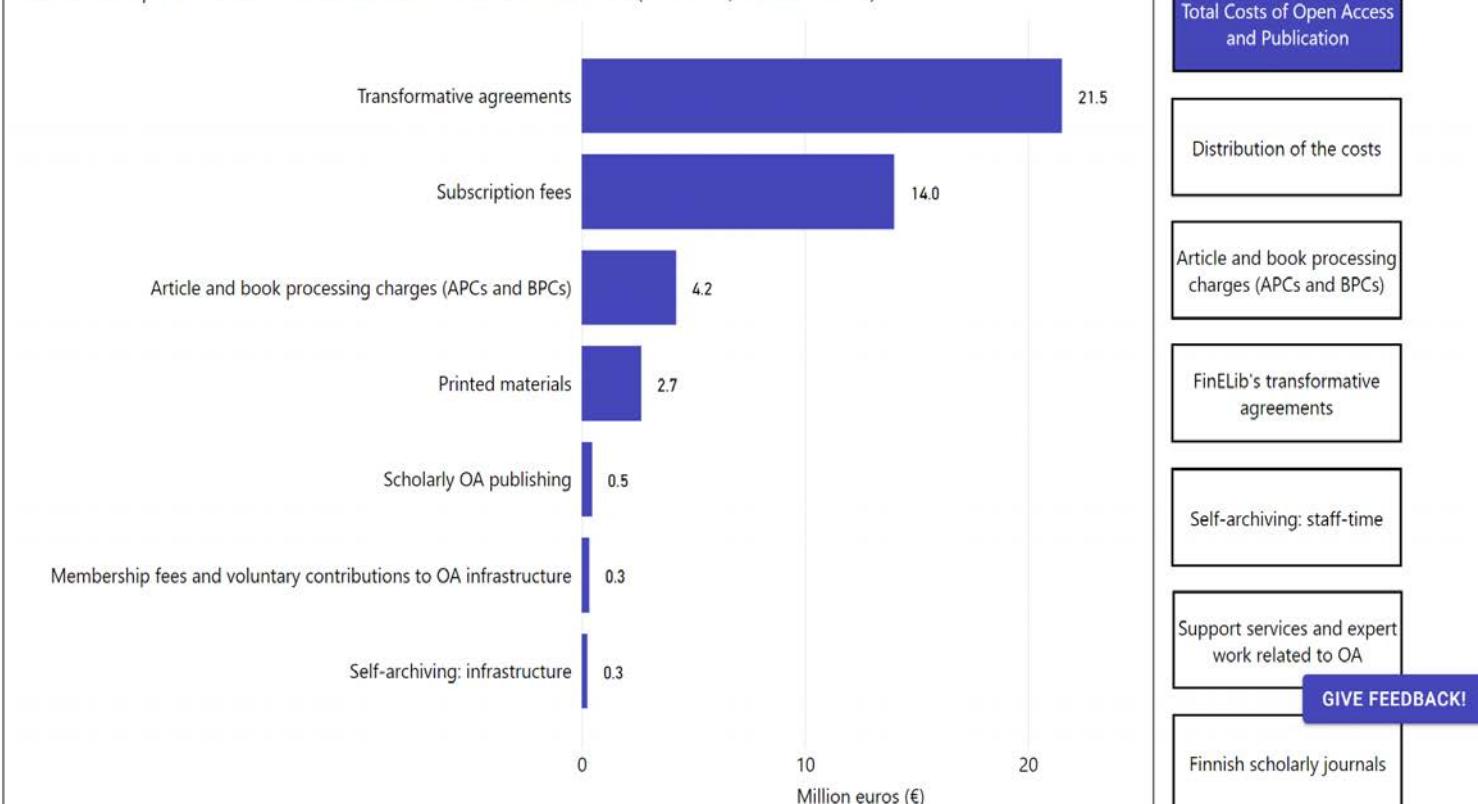
Evaluation of organizations: Profile of openness



Results so far

- <https://research.fi/en/>
- Peer reviewed articles 2022: 80 % OA
- Research data 2022: 655, 66 % OA
- 2551 open educational materials in the [Library of Open Educational Resources](#), aoe.fi
- Hundreds of people participating the coordination
- Total costs of Open Access and publication to Finnish research institutions 2021
43 500 000 €

Total costs of Open Access and Publication to Finnish institutions in 2021 (in total 43,51 million euros)



Questions?

- Is all this worth the money and effort?
- Have we made things too complicated?
- Are we doing the right things?

Thank You!

pekka.o.olsbo@jyu.fi



IIIF, annotation & machine learning

Building research workflows on library resources

Neil Jefferies
Bodleian Libraries, University of Oxford
neil.jefferies@bodleian.ox.ac.uk



IIIF, ANNOTATION & MACHINE LEARNING

Building Research Workflows on Library Resources

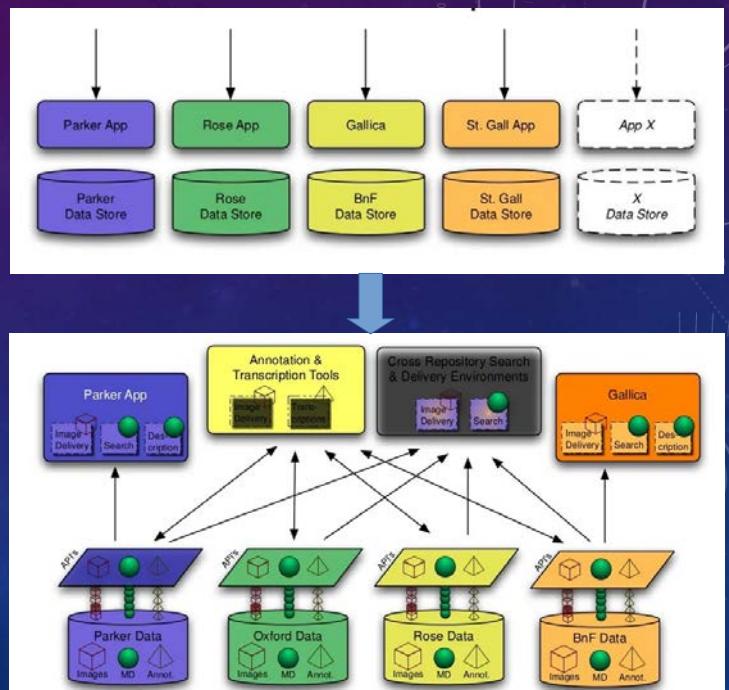
NEIL JEFFERIES, BODLEIAN LIBRARIES, UNIVERSITY OF OXFORD

DATA FUTURES GMBH

<https://orcid.org/0000-0003-3311-3741>

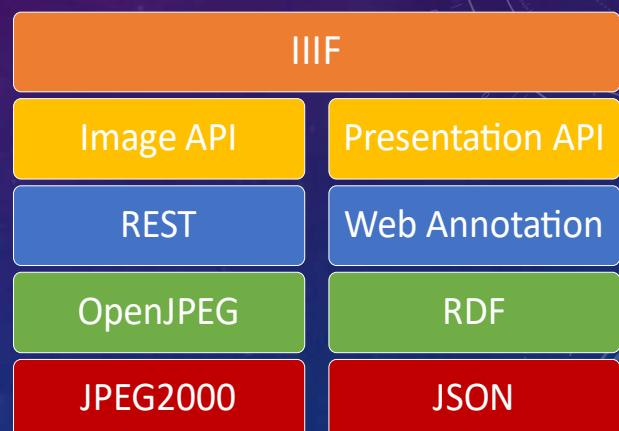
INTERNATIONAL IMAGE INTEROPERABILITY FRAMEWORK

- Discovery
- Basic Display
 - Comparison – cross-corpus and interdisciplinary research
 - Transformations – Zoom, rotate and pan/scroll
 - Metadata
 - Layering
- Analysis and enrichment
 - Reassembly – fragment and page level
 - Transcription (and transliteration, translation)
 - Highlighting
 - Annotation – layers
 - Linking
- Interoperability between collections/resources
 - Move from silo-ed resources to a scholarshop-friendly environment



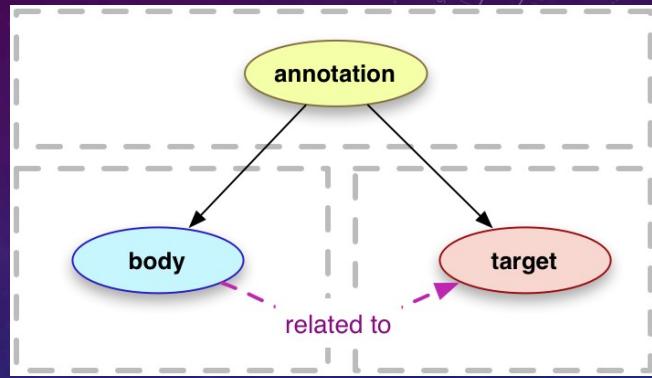
IIIF API'S

- The Presentation API – Defines how complex materials are presented
- The Image API – Defines how to deliver image resources
 - This may be to a Presentation API viewer in a browser, or any other software
- Extensions to these API's to handle video, 3D etc.
- Other API's include
 - Content Search API for searching image annotations
 - Authentication API for accessing restricted materials
 - Content State API for initialising a viewer in a particular state (e.g. open on a specific page)
 - Change Discovery API – To aid IIIF resource discovery



(WEB) ANNOTATION

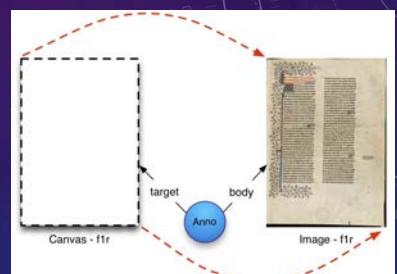
- Web Annotation Data Model (WADM)
 - <https://www.w3.org/TR/annotation-model/>
 - 2017, merger of Annotation Ontology and Open Annotation
 - Expressed using RDF
- Very Simple: an annotation has...
 - 1 or more targets – what the annotation is about
 - 0 or more bodies – what the annotation says (usually)
 - The body or target may be a resource specified by an IRI
 - BUT this can be qualified by a state or fragment selector, for example
 - Selectors are crucial for IIIF to specify which parts of images
 - Arbitrary shapes can use SVG (Scalable Vector Graphics, <https://www.w3.org/TR/SVG2/>)
- Annotations are extensible
 - Additional RDF statements can be added
 - Some, such as a data type for body/target, are defined in WADM



```
{  
  "@context": "http://www.w3.org/ns/anno.jsonld",  
  "id": "http://example.org/anno27",  
  "type": "Annotation",  
  "body": "http://example.org/road1",  
  "target": {  
    "source": "http://example.org/map1",  
    "selector": {  
      "type": "SvgSelector",  
      "value": "<svg> ... </svg>"  
    }  
  }  
}
```

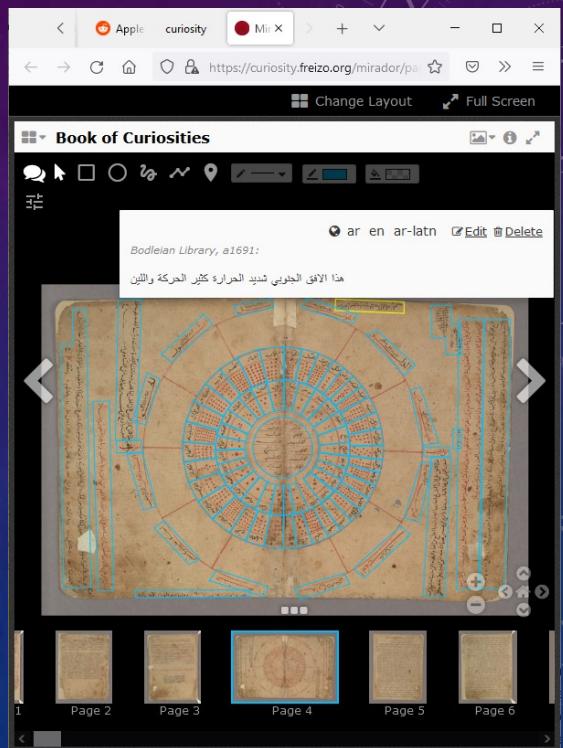
IIIF ANNOTATIONS (AS A CONSEQUENCE...)

- IIIF uses annotations to indicate where images should be located on the “canvas” presented to the user
- Since the code and structures for annotation are already part of IIIF, annotating the presented (composite) images is a logical extension
- The first annotations were embedded in manifests
 - Annotation services provided by the organisation holding the materials
 - Embedded in individual artefacts
 - Not really in keeping with the “interoperability” aims of IIIF
- Hypothes.is
 - Anyone can annotate more or less anything
 - Popular for teaching but not exactly “scholarly”;



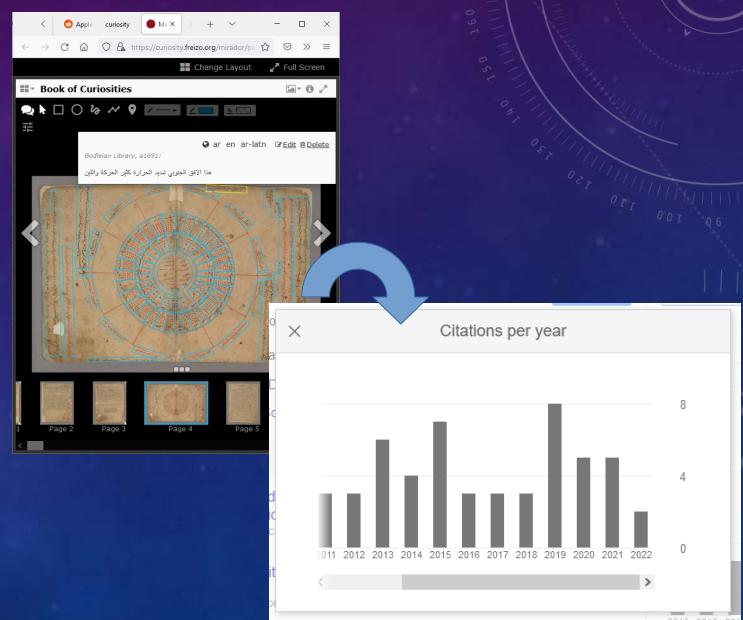
PRESENTATION VS ANALYSIS

- Presentation
 - Primarily about human-centric consumption of images
 - Skeumorphism – making the digital behave like the physical
 - Provides the basis for surfacing research via links and annotation
 - Increasing expectation of “rich experiences”
 - → Resources need to be augmented and enriched
 - → Focus on individual objects
- Analysis
 - Increasingly about machine access to images
 - → Machine-generated enrichment and augmentation
 - → Increasingly “at-scale”
 - Potential for new presentation experiences – visualisations etc.
 - → Augmented resources can improve research “impact”
 - → Important to ensure attribution (to machines and researchers)
 - → ...but also funders and projects



ANNOTATION AS A SCHOLARLY MECHANISM

- Transforming Web annotations into preservable scholarly annotations requires solutions to several problems:
 - Attribution/provenance, so authors and sources can be identified
 - Manageability so that annotations can be related to grants, projects etc.
 - Cite-ability so that annotations can form part of ongoing intellectual discourse
 - Recognition as a valid scholarly output
 - In part, this comes from solving the previous problems



SCHOLARLY ANNOTATIONS IN AN θ STOR

- Annostor.org is an annotation workflow tool from Data Futures, it adds...
- ORCID (Open Researcher and Contributor ID)
 - Provides a scholarly identifier to annotation authors
- RAID (Research Activity ID) provides information about research projects
 - Allows funders to identify their outputs
- Annotation Lists
 - Already part of WADM, aggregates annotations for easy management
- Recognition
 - ORCID recognises Annotation Lists as a research output
 - InvenoRDM (and soon Zenodo) allows deposit/viewing of Annotation Lists + Manifests
 - Annotation Lists are given DataCite DOI's (Digital Object Identifiers) for citation

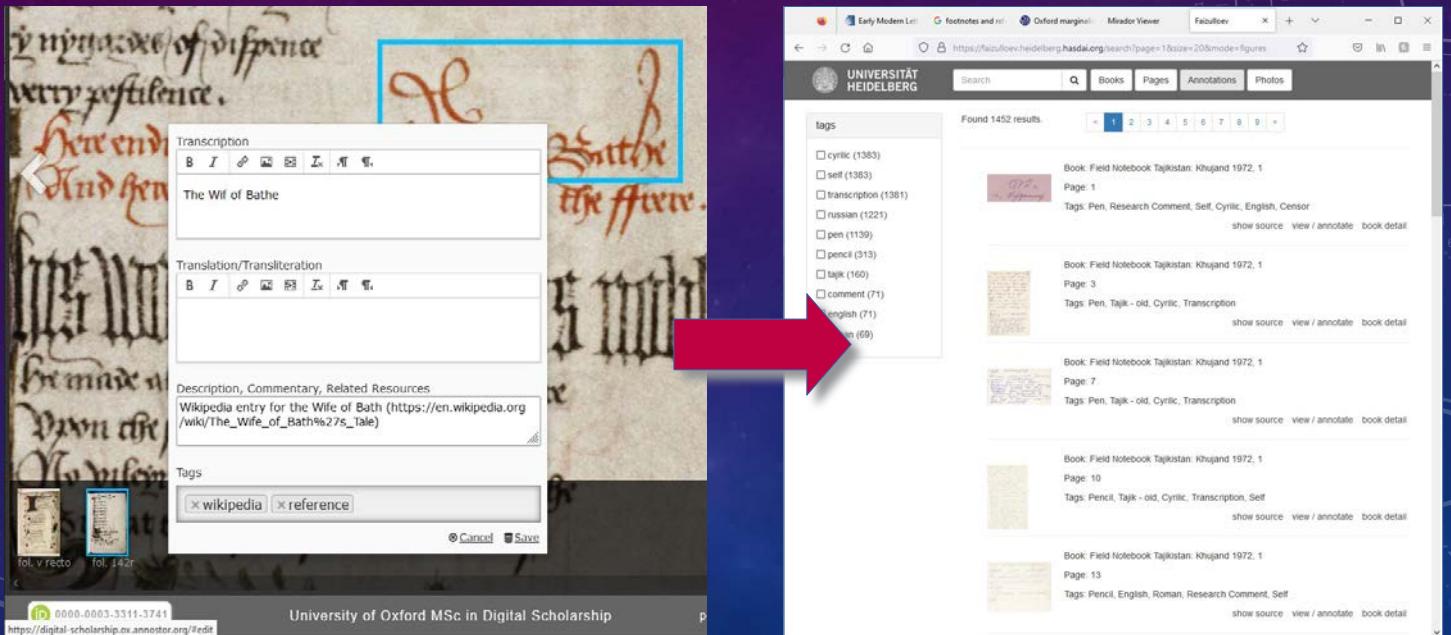
AN θ STOR

x Close

Filter objects: Add new object from URL: Load

	Cantigas de Santa María, Códice rico Cantiga 45 [annotated]	
Bodleian Libraries	6 items	
	pages from 'MS. Rawl. poet. 223' which includes 'Siege of Troy' and 'Canterbury Tales': https://digital.bodleian.ox.ac.uk/objects	
Bodleian Libraries	2 items	
	Bodleian Library Dep. Hughenden 40/1. fols. 115r-116v	
Bodleian Libraries	3 items	
	BnF. Département des Manuscrits. Français 24432	
Bibliothèque nationale de France	450 items	

FROM ANNOTATIONS TO DATASETS



The Basel Avisblatt (1729–1844)

Building a Research Workflow

CREDIT: INA SERIF, UNIVERSITY OF BASEL

SOURCE MATERIAL

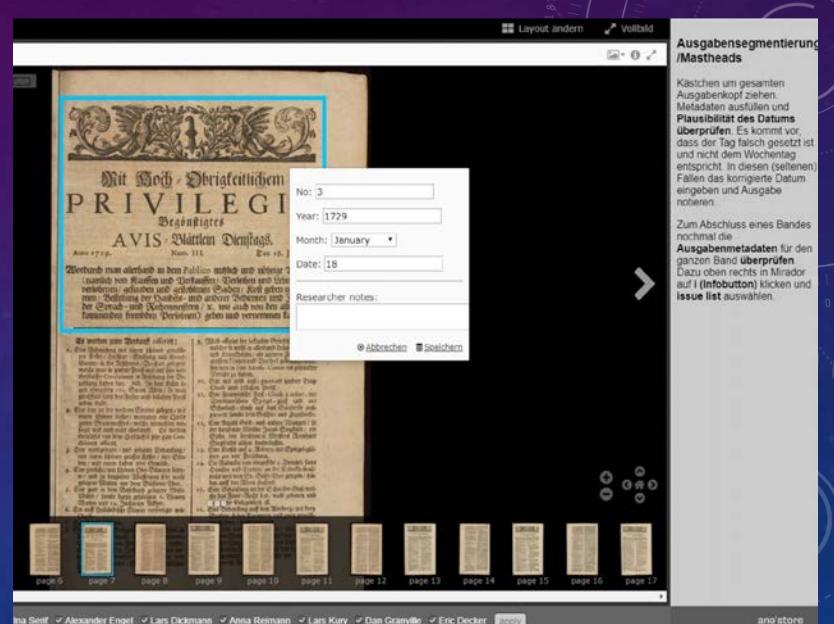


- The Basel Avisblatt (1729–1844)
- First Swiss intelligencer, (almost) pure advertisement paper
- (At least) weekly → 6626 issues, ~50 000 pages
- ~1 million ads (~450,000 postings, 80% reprinted in the following issue)

- Markets for [new and second-hand] goods, labour, housing, credit; shop and service promotion, lost and found, 'opportunities', etc.
- Social medium/platform economy: window to the socio-economic exchange of a Swiss city on its way to modernity
- **Project:** Digital edition + variety of historical research making use of the source
- **Challenge:** create database of ads suitable for a multitude of research questions, especially: address suitable subsets

BUILDING THE CORPUS

- Digitisation: University Library Basel (best specimen for every year out of three complete sets from Library and Statal Archive)
- Ingest to *freizo* and issue segmentation:
 - Image processing (tiling)
 - Creation of book assets plus IIIF manifests
 - Manual masthead annotation
 - Generation of issue assets from masthead annotation meta data



MANUAL CLASSIFICATION

- Manual transcription and classification of all ads in four yearbooks (1734, 1754, 1774, 1834):
 - → Generating groundtruth as basis for automation – or ‘Plan B’ data (if automation fails)

und Posament zu verleihen und findet man deshalbten Bescheid bey Mir. Hans Jacob Ochs dem Dreher allda wohnhaft.

Verlorenen und gefundene Sachen:

1. Vergangenen Mittwoch 8. Tag ist ein gelber Hund von mehr als mittelmäfiger Größe mit stark gestutzten Ohren und einer weißen Kehle verloren worden, wer ihn wieder an Behörde liefern wird hat eine Belohnung zu gewarten.
2. Vermischten Sonntag Morgens ist auf dem Weg von St. Paulsberg bis in die St. Peters Kirch ein SilberSchlösslein von einem Psalmbuch verloren worden, dem Finder offeriert man den Werth des Silbers gegen dessen Erstattung.
3. Vergangenen Freitag 8. L. ist ein blau und weinles Schmuckstuch vom St. Paul Thurn. den Hera

B I ⌂ ⌂ Tx ⌂ ⌂

1. Vergangenen Mittwoch 8. Tag ist ein gelber Hund von mehr als mittelmäßiger Grösse mit stark gestutzten Ohren und einer weissen Kehle

Textilien Hausrat Lebensmittel Schmuck Drucksachen
 Ding Tier Immobilien Kirchenstuhl Arbeit Kost
 Platzierung Caritas Finanz Lotterie Transport
 Auskunft uneindeutig

Biete Suche Ankündigung Werbung Nachricht
 unbekannt

Kauf Miete/Leihe Tausch Fundsache Gratis
 Mitmachen Versteigerung Pacht

deutsch französisch latein

Berichthaus Messe Preis Person Ort

Researcher notes:

Reading order:

review header no advert
 Continues | Link to previous: [NONE]

AUTOMATION OF PROCESSES

- Segmentation → boxes from manual segmentation effort used as training material for *dhSegment*, employed in Transkribus; accuracy ~90%
 - Text recognition → HTR (Handwritten Text Recognition) with Transkribus, training two models for two font epochs; character error rate <1%
 - Classification → Algorithmic filters from dictionaries and regular expressions that classify according to content

STRUCTURAL AND TEXTUAL DATA IN TRANSKRIBUS

The screenshot shows a window titled "Transkribus v1.10.0.6-SNAPSHOT (23_01_2020_10:58) Loaded doc: 1734, ID: 13006, Page: 60, file: BAU_1_000178568_1734_0002.jpg [Image Meta Info: Resolution:1.0; w:765 h:5198 x:8272] [current line: w:765 h:204 x:112]".

Annotations:

- 8-14 zum Stande kommen und befördert werden möge.
- 8-15 Inhalt oder Preis ist nicht nötig ein mehreres zu gedenken,
- 8-16 inmassen aus berührt- im Adresse Contor zu habenseydem
- 8-17 plan, schon eint und andere zu ersehen seyn wird.
- 8-18 6. Vor ungefähr 3. Wochen ist ein zweyfundig mößingerer Ge-
- 8-19 wichtstein aus einem Specier-Laden entwendet worden; man
- 8-20 ersucht freundlich darauf zu vigiliiren und das Entdeckende an-
- 8-21 zuzeigen.
- 10-1 , Verwichenen Sontag 8. Tag nach der Morgen-Preßtag ist in St.
- 10-2 Peters Kirch auf dem kettern am Boden beim 128.
- 10-3 Siu ein
- 10-4 kleines Psalmbuch mit kleinen silbernen Schloßlein gefunden
- 10-5 worden; wer Ansprach daran hat kan im Berichthauß um die
- 10-6 Gebühr vermennen bey weme selbiges zu entheben seye.
- 11-1 Preiß der Lebens-Mittien.
- 11-2 Kemen
- 11-3 Pf. 15. à 17.6. 6. pf.
- 11-4 .
- 11-5 pf.
- 11-6 5. 6. bis 7.
- 11-7 7.

ALGORITHMIC CLASSIFICATION

- Ads are (very) short texts → problem for topic modelling approaches.
- Dynamic tagging (algorithmic classification using scripts, reg-ex)
- Many more tags than manual classification

```
# Tagfilter for dogs
# @export
tagfilter_dogs <- function(){
  dict <- list()
  dict$pos <- list()
  general = "(Hlh)(ulünde)?(nis)?\b", # added "e?(nis)?\b", otherwise finds Hundert [AE]
  Federhund = "(Fif)edern(-)\s(Hlh)(ulünd)", 
  Haushund = "(Hlh)aus(-)\s(Hlh)(ulünd)",
  Kettenhund = "(Klk)etten(-)\s(Hlh)(ulünd)",
  Epagneulkund = "(Ele)pgagneul(-)\s(Hlh)(ulünd)",
  Huhnerhund = "(Hlh)(ulü)(h)nec(-)\s(Hlh)(ulünd)",
  Pudelhund = "(Plp)udel(-)\s(Hlh)(ulünd)Pudel",
  Budelhund = "(Bld)udel(-)\s(Hlh)(ulünd)",
  Dackenhund = "(Dld)acken(-)\s(Hlh)(ulünd)(Dld)eck,n(-)\s(Hlh)(ulünd)",
  Wachtelhund = "(Wle)achtel(-)\s(Hlh)(ulünd)",
  Doque = "(Dld)oquen(-)\s(Hlh)(ulünd)Doque",
  Arlequin = "(Aor)le(q)uin(-)\s(Hlh)(ulünd)",
  Stauphund = "(Sls)c(a)up(-)\s(Hlh)(ulünd)",
  Mops = "(Mlm)o(p)bs)(Mlm)o(p)bs(-)\s(Hlh)(ulünd)",
  Danischhund = "(Dld)(di)anisch(-)\s(Hlh)(ulünd)",
  Bummerhund = "(Bbl)ummer(-)\s(Hlh)(ulünd)",
  Vorsteherhund = "(Viv)orsteher(-)\s(Hlh)(ulünd)",
  Fotzelhund = "(Fif)otzel(-)\s(Hlh)(ulünd)",
  Spitzpommer = "(Sls)pitz(-)\s(Hlh)(ulünd)(Pip)b(b)ommer(-)\s(Hlh)(ulünd)"
```

```
dict$neg <- list(
  raw_material = "(Sls)eehund",
  animal = "Chaise-Pferd/Reisepferd/Chaisepferd",
  knife = "(Sls)ackmesser"
)
create_filter_output(dict)
```

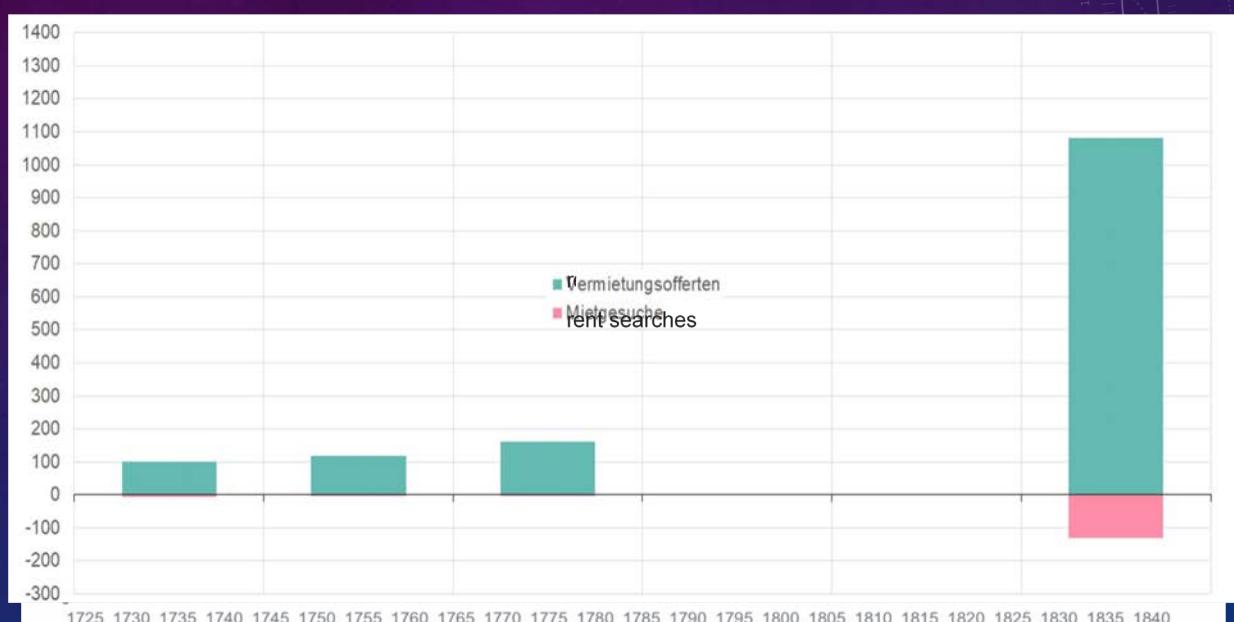
"antiquities"	"alcohol"	"animalobjects"	"animatran"
"barrel"	"art"	"attributes_female"	"bag"
"bottle"	"bathobject"	"bed"	"beading"
"bowl"	"blipromo2"	"bedroom2"	"bale"
"butcher"	"bustlingcomponents"	"bedroom3"	"carpet"
"cabinet"	"cane"	"chat"	"chartiy"
"cats"	"chariseax"	"chatte"	"clippin"
"coffee"	"container"	"chutney"	"cooking"
"costume"	"cutlery"	"drilledfruit"	"divider"
"dogs"	"dottedc"	"employment2"	"dumb"
"doggy"	"dowyment1"	"feed"	"employmentoffer"
"doggy"	"firestart"	"firework"	"finance_loan"
"doggy"	"fog"	"gas"	"fish"
"glasses"	"hand"	"handkerchief"	"flame"
"hay"	"health"	"homedeco"	"flannel"
"honey"	"housing_other"	"housing_other2"	"foot"
"housing_rent2"	"housing_rent3"	"housing_sold1"	"fox"
"housing_sale3"	"housing_sales4"	"housing_sales5"	"franchise"
"jewellery"	"kitchen"	"kitchen"	"framing"
"key"	"knitting"	"livespace"	"frontendfound1"
"knitting"	"lottery2"	"measure"	"goat"
"knitting"	"milk"	"mineralwater"	"alarm"
"knitting"	"misobject"	"mushrooms"	"NAT"
"knitting"	"muse"	"nursery"	"oak"
"knitting"	"outerwear"	"pasta"	"pastry"
"knitting"	"plant"	"plantraw"	"poultry"
"knitting"	"print1"	"paper"	"print3"
"knitting"	"print4"	"printed"	"print5"
"knitting"	"print5"	"pushchain"	"riding"
"knitting"	"print9"	"shoe"	"shopequip"
"knitting"	"sewinghand"	"slices"	"sugar"
"knitting"	"sewinghand"	"stove"	"tallinen"
"knitting"	"storage"	"table"	"textmaterial"
"knitting"	"syrup"	"seat"	"top1"
"knitting"	"suitcases"	"tobaccoobjects"	"transactiontype_offer3"
"knitting"	"timpiece"	"transactiontype_offer1"	"transactiontype_offer2"
"knitting"	"toy"	"transactiontype_request1"	"transactiontype_request2"
"knitting"	"trolley"	"translatr1"	"unlabel1"
"knitting"	"troley"	"translatr2"	"unlabel2"
"knitting"	"uniform"	"upholstery"	"ut_animal"
"knitting"	"ut_consumables_drink"	"ut_consumables_food"	"ut_consumables_other"
"knitting"	"ut_consumables_drug"	"ut_consumables_grocery"	"ut_consumables_grocery"
"knitting"	"ut_consumables_grocery"	"ut_consumables_grocery"	"ut_consumables_grocery"
"knitting"	"ut_things_materials"	"ut_consumables_grocery"	"ut_consumables_grocery"
"knitting"	"weaving"	"ut_consumables_grocery"	"ut_consumables_grocery"
"knitting"	"wood"	"woodobject"	"wineobject"

DATABASE GENERATION

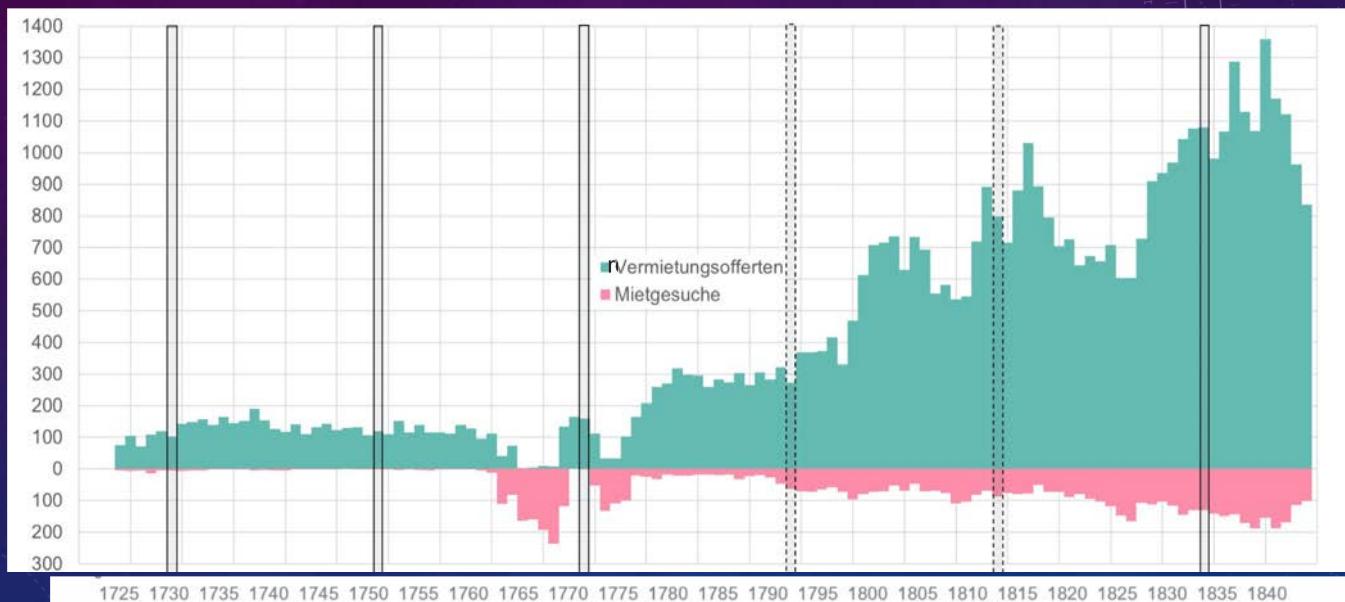
- Creation of database from page xml → ‘source db’
 - First in DataFutures’ freizo, then as InvenioRDM (<https://avisblatt.dg-basel.hasdai.org>)
- → ‘analysis db’ in Github

	A	B	C	D	E	F	G	H	I	J	O	X	Y
1	id	text	rev	issue	book	date	pageno	readingorder	inscribed	isheader	fragment1	fragment2	header_tag
2	a8b05799-5bfb-5488-a56c-04cae43e34d2/t0	Es wird zum Verkauff angebracht:	N	1	24663	1745-01-07	3	1	1	TRUE	https://iiif.avisblatt.freizo.org/image/24666_0000/115,4011,1650,135/full/0/default.jpg		saleoffer
3	a8b05799-5bfb-5488-a56c-04cae43e34d2/t1	1. Eine Behausung an der Rhein-Gasse gegen dem St. Antonier-Hof hinüber von Gelegenheit wie der Augenschein zeigen wird.	N	1	24663	1745-01-07	3	2	1	FALSE	https://iiif.avisblatt.freizo.org/image/24666_0000/115,4146,1650,270/full/0/default.jpg		saleoffer
4	a8b05799-5bfb-5488-a56c-04cae43e34d2/t2	2. Ohngefehr 2. Jucharten Reben an der Grentzacher-Straß, samt einem Feld hinten daran stossende, mit einem hohen Häuslein, worin n eine Trodten und Bogten rc. und ist sich darum anzuanzumelden, bey Herrn Hans Georg Wannenwetsch zum Stemen-Eck in Klein-Basel, welcher deme, Alters halben, nicht mehr abwarten kan.	N	1	24663	1745-01-07	3	3	1	FALSE	https://iiif.avisblatt.freizo.org/image/24666_0000/1813,3978,1666,422/full/0/default.jpg		saleoffer

THE RENTAL MARKET – 4 YEARS

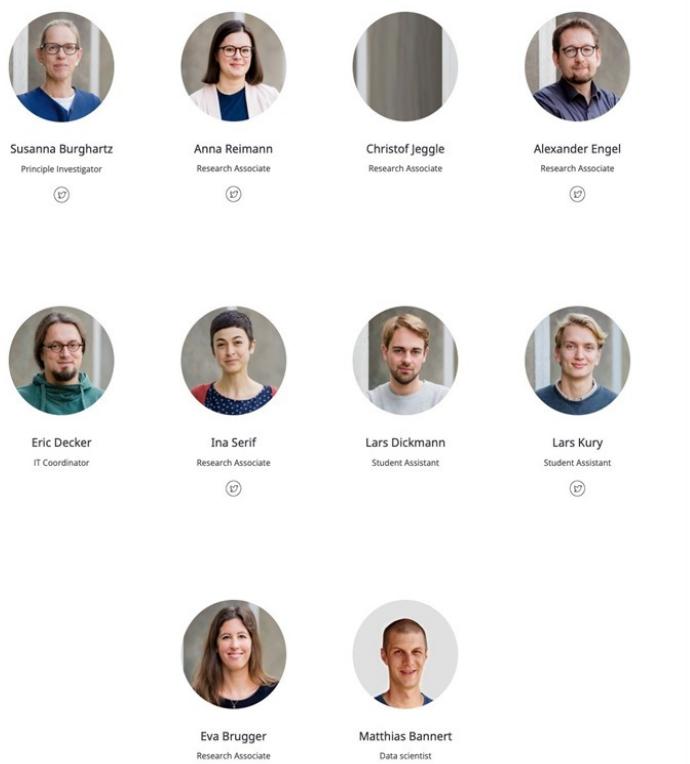


THE RENTAL MARKET – 116 YEARS



<https://avisblatt.ch>

- «Printed Markets. Information, Data and News in the Basel *Avis-Blatt*, 1729-1845» (funding 2018–2023, University of Basel)
- Digitization, text recognition, ad classification for further analysis
- 1 PhD thesis, 3 masters theses, 2 postdocs, ...
- Digital collection with full text on platform <https://www.e-newspaperarchives.ch/>
- Digital collection/repository at <https://avisblatt.dg-basel.hasdai.org/>
- Database of advertisements at <https://github.com/Avisblatt/>, including package and documentation



SOME OTHER THINGS...

Building Research Workflows on Library Resources

NEIL JEFFERIES, BODLEIAN LIBRARIES, UNIVERSITY OF OXFORD

DATA FUTURES GMBH

<https://orcid.org/0000-0003-3311-3741>

HUMAN TRAINED AI/ML MODELS

- Use human generated annotations as a training set
 - A common approach in the sciences where large image collections have been easier to acquire
 - Subsequently human corrections to machine generated annotations can improve accuracy
 - Typically these annotations will be classifications, or keywords amenable to downstream statistical analysis

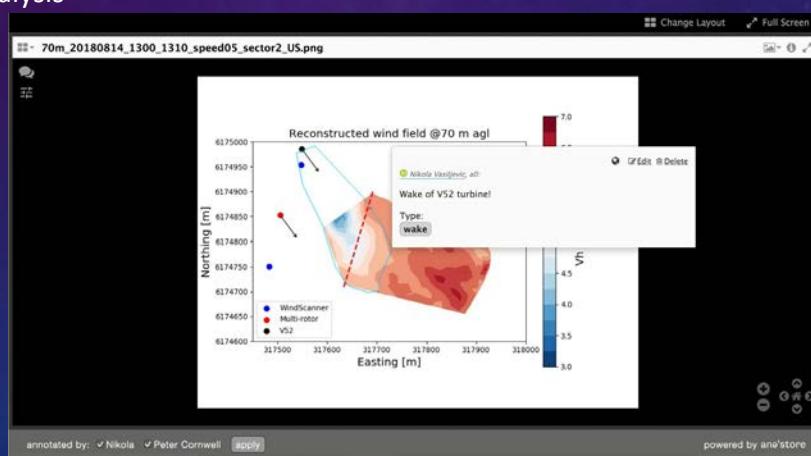


IMAGE FORM INDEXING

- Indexing images based on their shape developed by the Oxford Visual Geometry Group
 - Search/classify images according to similarity
- Bodleian Ballads ImageBrowse
<https://www.robots.ox.ac.uk/~vgg/demo/ballads/>
- Identification of re-use of wood blocks – increasing damage can be used to sequence usage temporally
- Block copying can be seen in less precise matches
- The same image can represent different things according to the ballad it appears against
 - ...problems for AI-based image identification that does not take context into account



Prototyp autonómnej navigácie pre UAV

Daniel Sedlák
Qintec a.s.
daniel.sedlak@qintec.sk

ABSTRAKT

Publikácia sa zaoberá prototypom autonómnej navigácie pre UAV v interiéri, pričom je rozdelená na dve časti. Prvá časť je venovaná 3D rekonštrukcií stereo obrazov, ktorá zahŕňa aj meranie hĺbky, čo je jedna z najdôležitejších funkcií pri autonómnej navigácii, technológií lokalizačného systému VLC, lokalizačnému ICP algoritmu a simultánnej lokalizácii SLAM. Druhá časť sa venuje výberom vhodných technológií pre prototyp kvadrokoptéry s následným testovaním a realizáciou.

KLÚČOVÉ SLOVÁ

UAV, SLAM, LiDAR, stereoskopická kamera, autopilot, kvadrokoptéra, lokalizácia, mapovanie

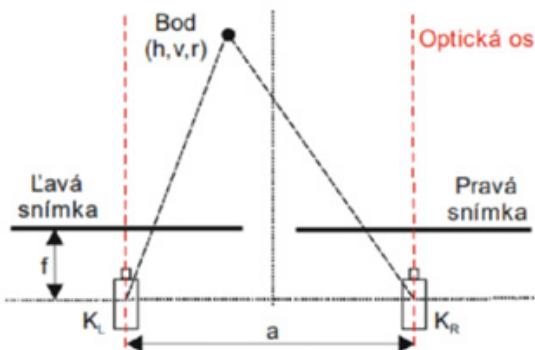
1. ÚVOD

V dnešnej dobe sa čoraz viac stretávame s pojmom „unmanned air vehicle“ (UAV) - bezpilotné lietajúce stroje, akými sú napríklad drony. Tento príspevok sa zaoberá výskumom a vývojom prototypu mnohoúčelovej autonómnej navigácie kvadrokoptéry v interiéri. Prvou podmienkou bezpilotného riadenia je znalosť polohy dronu v priestore a prekážok v okolí. Na tento účel slúži množstvo algoritmov a senzorov či iných technológií, vďaka ktorým nemusí byť vyhodnotenie lokalizácie zložité. S určením presnej polohy UAV prostriedku v priestore je však spojená nepresnosť, ktorú je potrebné minimalizovať. Problematika lokalizácie je priamo spojená s navigáciou, t.j. aby mohol dron lietať autonómne, musí vedieť o dianí pred ním v trojrozmernom priestore, na čo mu slúži napríklad stereoskopická kamera.

2. 3D REKONŠTRUKCIA STEREO OBRAZOV

V dnešnej dobe sa na 3D rekonštrukciu obrazu využívajú rôzne techniky a metódy nachádzajúce uplatnenie v oblastiach ako je sledovanie lietadiel alebo odomknutie telefónu pomocou kamery. Vo všeobecnosti sú 3D objekty veľmi náročné na rekonštrukcie scén, ale majú významnú rolu pri skúmaní napríklad pamiatok alebo objavovaní neprebádaných jaskyň metódou skenovania a následnej rekonštrukcie. Ďalšie uplatnenie je napríklad v automobilovom priemysle, kde pomocou umelej inteligencie dochádza k zvýšeniu bezpečnosti cestnej premávky.

Metódu 3D rekonštrukcie môžeme uskutočniť dvoma spôsobmi, aktívne alebo pasívne. Pri aktívnych metódach ide o skenovanie objektov laserom a priebežné vyhodnocovanie získaných dát, pasívna metóda pracuje so zhromaždenými snímkami obrazov. Obe metódy majú svoje silné a slabé stránky a taktiež výhody a nevýhody. Aktívna metóda využíva laser a tým je cenovo menej dostupná, no rekonštrukcia je vďaka laseru rýchlejšia. Pasívne metódy sú cenovo dostupnejšie a na spracovanie používajú rôzne algoritmy, ktoré našli široké uplatnenie pri rekonštrukcii a detekčných systémoch



Obr. 1. Všeobecné umiestnenie dvojice kamier^[1]

Obr. 1. - Znázornenie umiestnenia dvoch stereo kamier. Ak je známe horizontálne posunutie, vzdialosť stredov kamier a ohnisková vzdialenosť, je možné vypočítať hĺbku akéhokoľvek bodu. Táto metóda je teda vhodná pre UAV a existuje rôzna škala stereoskopických kamier zameraných na vývoj robotov.

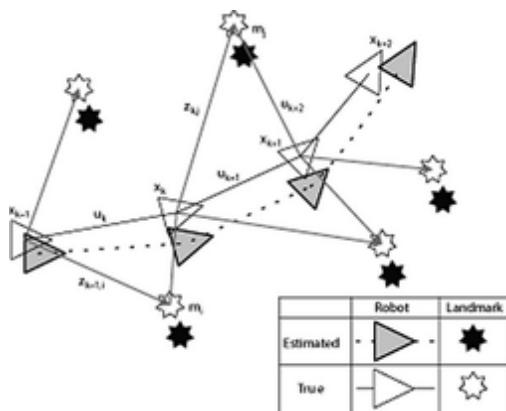
3. TECHNOLÓGIA VLC A ICP ALGORITMUS

VLC „visible light communication“ alebo komunikácia viditeľným svetlom je technológia lokalačného systému, ktorú možno použiť na miestach citlivých na rádiové frekvencie. Ukázalo sa, že VLC poskytuje vyššiu presnosť v porovnaní so systémami Wi-Fi. Vďaka zlúčeniu diód vyžarujúcich svetlo (LED) sa VLC široko používa pri lokalizácii. LED má mnoho výhod, vrátane dlhej životnosti, odolnosti voči vlhkosti, nízkej spotreby energie a nízkych prevádzkových nákladov. LED môže tiež modulovať signály svetelných vĺn pri vysokej rýchlosťi^[2]. Lokalizáciu pomocou VLC je možné vykonať pomocou systémov na báze fotodiód, ktoré zachytávajú intenzitu svetla a systémov založených na obrazových senzoroch, ktoré dokážu zachytiť svetelné impulzy^[3]. LED aj senzory by však mali byť v zornom poli pre presnú lokalizáciu. Detekcia svetla a lokalizácia vzdialenosť (LiDAR) poskytuje informácie o obrysoch okolitých prekážok a tým môžu tieto systémy poskytnúť presnú lokalizáciu^[4].

ICP algoritmus je metóda porovnávania skenov, ktorá pristupuje k SLAM „Simultaneous localization and mapping“ problému bez použitia príznaku, každý bod senzorického merania môžeme považovať za samostatný príznak a potom sa tento rozdiel vytráca. Základom týchto metód je nájdenie optimálnej transformácie medzi niekoľkými nasledujúcimi senzorickými meraniami, prípadne je hľadaná transformácia napríklad medzi mriežkou obsadenosti a novým senzorickým meraním. Optimálna transformácia je potom hľadaná vybranou optimalizačnou metódou, väčšinou optimalizačným algoritmom najmenších štvorcov^[5]. Výhoda týchto metód je zrejmá a jej použitie je vhodné v komplikovaných prostrediach vzhľadom na množstvo prekážok, ktoré môžu blokovať signál pre predošlú metódu. Tvorba primitívnej bodovej mapy je relatívne jednoduchá, na druhej strane sú tieto metódy citlivé na počiatočné podmienky.

4. SIMULTÁNNA LOKALIZÁCIA A MAPOVANIE (SLAM)

SLAM je skratka pre Simultánnu lokalizáciu a mapovanie. V mapovaní je problém integrovať informácie, ktoré zhromažďujú senzory robota do danej reprezentácie. To môže byť opísané otázkou: „Ako vyzerá svet?“. Centrálné aspekty v mapovaní sú reprezentácia prostredia a interpretácia dát senzorov. Na rozdiel od mapovania, v lokalizácii je problémom odhad pozície robota vo vzťahu k mape. Inými slovami, robot má odpoveď na otázku: „Kde to som?“. Zvyčajne sa rozlišuje medzi pozíciou sledovania, keď je začiatočná pozícia známa a globálnej lokalizáciou, v ktorej nebola daná žiadna prioritná vedomosť o štartovacej pozícii. Simultánná lokalizácia a mapovanie je preto definované ako problém vybudovania mapy, v ktorej v rovnakom čase prebieha lokalizácia robota bez použitia tejto mapy. V praxi nemožno tieto dva problémy riešiť nezávisle od seba. Predtým, ako robot môže odpovedať na otázku „Ako prostredie vyzerá?“, z výsledkov pozorovania je potrebné vedieť, z ktorého prostredia získané pozorovania pochádzajú. Zároveň je ťažké bez mapy odhadnúť aktuálnu polohu robota. Preto je SLAM často prirovnávaný k problému: „Čo bolo skôr? Sliepka alebo vajce?“ Dobrá mapa je potrebná pre lokalizáciu, zatiaľ čo presný odhad pozície je potrebný na vytvorenie mapy.



Obr. 2. Základný SLAM problém [6]

Obr. 2. - V časovom okamihu k, sú nasledujúce veličiny definované ako:

- x_k - stavový vektor, ktorý popisuje umiestnenie a orientáciu robota.
- u_k - kontrolný vektor, ktorý aplikuje v čase k-1 riadiť vozidlo na stav x_k v čase k.
- m_i - vektor, ktorý popisuje umiestnenie i-teho orientačného bodu, ktorého pravé umiestnenie predpokladá, že je nemenné.
- z_{ik} - pozorovanie prevzaté z miesta vozidla i-teho orientačného bodu v čase k. Keď existuje viac pozorovaní orientačných bodov v jednom okamihu alebo keď konkrétny orientačný bod nie je relevantný k diskusii, pozorovanie bude zapísané iba ako zk.

3. NÁVRH A REALIZÁCIA PROTOTYPOVÉHO RIEŠENIA

Pri realizácii prototypu sme použili stavebnicu dronu kvadrokoptéry HOLYBRO X500 V2 obsahujúcu všetky potrebné komponenty ako konštrukcia, motory, vrtule, atď. Riadenie letu zabezpečuje autopilot Cube Orange s redundantným IMU systémom.



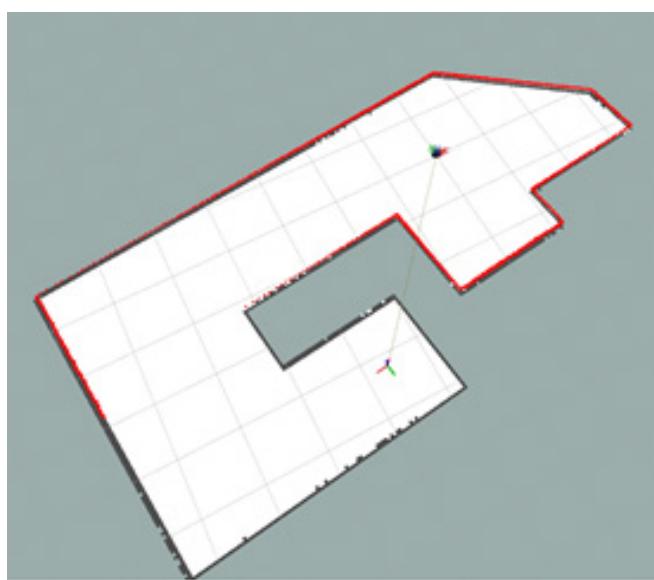
Obr. 3. Kvadrokoptéra HOLYBRO X500 V2

Po zostavení dronu a nakonfigurovaní autopilotu bolo realizovaných niekoľko skúšobných letov pomocou RC ovládača. Realizovali sme simultánnu lokalizáciu a mapovanie SLAM v metaoperačnom systéme ROS, čo si vyžaduje vysoký výpočtový výkon a technológie na meranie vzdialenosť. Kvadrokoptéra je osadená vývojovou doskou NVIDIA Jetson Nano Developer Kit, s modulom poskytujúci vysoký výkon a prispôsobený pre vývoj AI.



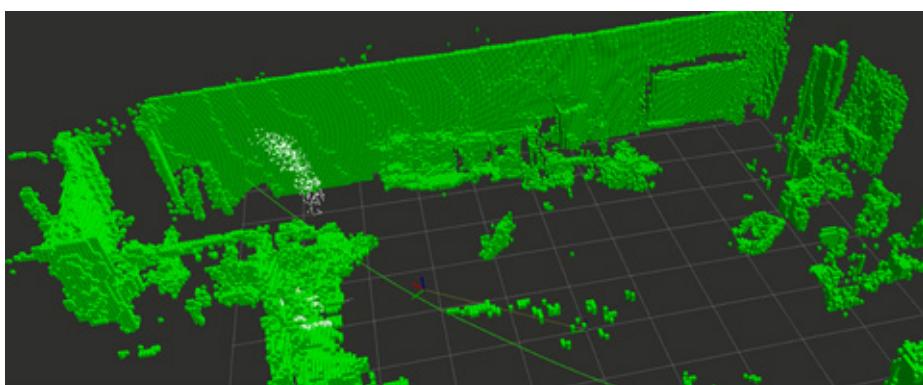
Obr. 5. Jetson Nano Developer Kit

Nevyhnutou súčasťou SLAM je zhromažďovanie dát zo senzorov merajúcich vzdialenosť v okolí dronu, aby bolo možné vytvorenie mapy a lokalizácie v nej. Nakoľko sa dron pohybuje v trojrozmernom priestore, je vybavený stereoskopickou kamerou ZED 2 pre 3D skenovanie priestoru pred dronom a pre súvislé skenovanie celého okolia pomocou LiDAR-u RPLIDAR A3. LiDAR je zapojený k vývojovej doske, na ktorej beží operačný systém LINUX a robotický operačný systém ROS. Systém ROS zabezpečuje simultánnu lokalizáciu a mapovanie za pomoci LiDAR-u. Vývojová doska pomocou sériového rozhrania komunikuje s autopilotom a zasiela mu aktuálne dátá o polohe.



Obr. 6. Príklad mapy zo skenovania LiDAR-om ^[7]

SLAM spracováva dátá z LiDAR-u a zobrazuje mapu s lokalizáciou v reálnom čase. Aktuálne prebieha testovanie lokalizácie zariadenia v priestore prostredníctvom RPLIDAR A3 a zároveň merania hĺbky ZED 2 kamerou kontrolovaným riadením pomocou RC ovládača.



Obr. 7. Príklad 3D mapy z meraní kamerou ZED 2 [8]

Po dosiahnutí požadovaných výsledkov presnosti lokalizácie budú nasadené algoritmy autonómneho riadenia. Úlohou riadiacich algoritmov bude zabezpečovať vytvorenie trasy pre dron po zadaní cieľových súradníc a prepočítavanie nových trás na základe kolíznych situácií v reálnom čase. Toto umožní dosiahnutie cieľových súradníc najkratšou trasou a aktívne vyhýbanie sa všetkým statickým a dynamickým prekážkam. Riadenie bude založené na generovaní stromovej štruktúry v priestore pred dronom a v jeho okolí. Jedným z takýchto algoritmov je RRT „rapidly exploring random tree“.

ZÁVER

V prvej časti publikácie bola predstavená 3D rekonštrukcia stereo obrazov, ktorej úlohou je v tomto projekte meranie hĺbky v priestore pred dronom. Pre autonómnu navigáciu je rovnako dôležite poznáť celkové okolie a preto sa v príspevku zaobráme technológiou lokalizačného systému VLC a ICP algoritmov využívaných v simultánnej lokalizácii a mapovaní. V druhej, vývojovej časti, bolo zrealizované zostavenie kvadrokoptéry HOLYBRO X500 V2, návrh hardvéru a spustenie simultánnej lokalizácie a mapovania pomocou RPLIDAR A3 v prostredí operačného systému ROS. Následne prebehlo testovanie stereoskopickej kamery ZED 2, ktorej výstupom boli namerané hĺbky v priestore pred dronom nevyhnutné pre autonómnu navigáciu. Pre zvýšenie bezpečnosti sme navrhli vlastnú bezpečnostnú klietku vyrobenú z karbónových tycí pospájaných na mieru namodelovanými a vytlačenými spojkami v laboratóriu prostredníctvom 3D tlačiarne.

POĎAKOVANIE

Tento článok vznikol vďaka podpore v rámci Operačného programu Integrovaná infraštruktúra pre projekt „Inteligentné operačné a spracovateľské systémy pre UAV“, kód projektu v systéme ITMS2014+: 313011V422, spolufinancovaný zo zdrojov Európskeho fondu regionálneho rozvoja.



EURÓPSKA ÚNIA

Európsky fond regionálneho rozvoja
OP Integrovaná infraštruktúra 2014 – 2020



MINISTERSTVO

DOPRavy
SLOVENSKEJ REPUBLIKY

LITERATÚRA

- [1] O. Kováč, J. Mihálik, Tvarovanie 3R modelu ľudskej hlavy pomocou stereoskopického vnímania, 2016., Technical University of Košice, elektrorevue ISSN 1213-1539 URL: https://www.researchgate.net/publication/297551489_Tvarovanie_3R_modelu_ludskej_hlavy_pomocou_stereoskopickeho_snimania_Shapping_of_3D_model_of_human_head_with_stereoscopic_scanning

- [2] Luo, J., Fan, L., & Li, H. (2017). Indoor positioning systems based on visible light communication: State of the art. *IEEE Communications Surveys & Tutorials*, 19(4), 2871–2893.
- [3] Huynh, P., & Yoo, M. (2016). VLC-based positioning system for an indoor environment using an image sensor and an accelerometer sensor. *Sensors*, 16(6), 783.
- [4] Xiao, Y. Ou, Y., & Feng, W. (2017) Localization of indoor robot based on particle filter with EKF proposal distribution. In 2017 IEEE international conference on cybernetics and intelligent systems (CIS) and IEEE conference on robotics, automation and mechatronics (RAM), 2017, pp. 568–571.
- [5] Mazi R., Lokalizace pro autonomní systémy. Disertační práce, České vysoké učení technické, Praha, 2007.
- [6] H. Durant-Whyte, T. Bailey. 2006. Simultaneos Localisation and Mapping (SLAM): Part 1 The Essential Algorithms. *IEEE Robotics and Automation Magazine*. ISSN 1070-9932, 2006. vol.13, no2., p.99-110.
- [7] Kiran Palla, Creating Map using Laser Scanner and Gmapping, 2018, URL: <https://kiranpalla.com/autonomous-navigation-ros-differential-drive-robot-simulation/creating-map-using-laser-scanner-and-gmapping/>
- [8] Chad Sweet , SLAM: The L is for Localization – the M is for Mapping, 2022, URL: <https://www.modalai.com/blogs/blog/slam-the-l-is-for-localization-the-m-is-for-mapping>



Interiérová lokalizácia pre prevádzku UAV

Daniel Sedlák
Qintec a.s.
daniel.sedlak@qintec.sk

ABSTRAKT

Publikácia sa zaobrá interiérovou lokalizáciou pre prevádzku UAV, pričom je rozdelená na dve časti. Prvá je venovaná technológiám lokalizačných systémov a metódam pre lokalizovanie. Druhá časť sa venuje výberom vhodných technológií pre realizáciu prototypu kvadrokoptéry.

KLÚČOVÉ SLOVÁ

UAV, SLAM, lokalizácia, mapovanie, kvadrokoptéra, senzory, autopilot

1. ÚVOD

V dnešnej dobe sa čoraz viac stretávame s pojmom „unmanned air vehicle“ (UAV) - bezpilotné lietajúce stroje, akými sú napríklad drony. V tejto publikácii sa zaoberáme výskumom a vývojom prototypu riešenia systému interiérovej lokalizácie pre prevádzku mnohoúčelovej kvadrokoptéry. Prvou podmienkou schopnosti bezpilotného riadenia je znalosť polohy dronu v priestore a znalosť prekážok v okolí. Na tento účel slúži množstvo algoritmov, senzorov, a technológií, vďaka ktorým nemusí byť vyhodnotenie lokalizácie zložité. S určením presnej polohy UAV prostriedku v priestore je však spojená nepresnosť, ktorú je potrebné minimalizovať.

2. TECHNOLÓGIE LOKALIZAČNÝCH SYSTÉMOV

Wi-Fi

Wi-Fi je názov populárnej bezdrôtovej sieťovej technológie, ktorá funguje v rámci RF pásmu 2,4

GHz pre IEEE 802.11b, IEEE 802.11g, IEEE 802.11n a v 5 GHz pre IEEE 802.11a. Vnútorné prostredie, akým je univerzita, alebo budova s kanceláriami, majú už takmer všetky nainštalované distribuované WiFi hot spoty, ktoré poskytujú pokrytie celej budovy. Medzi zariadenia, ktoré využívajú technológiu Wi-Fi, patria napríklad osobné počítače, smartfóny, digitálne fotoaparáty, tablety [1, 2]. Náklady na nákup infraštruktúry a používateľských zariadení pre Wi-Fi môžu byť veľmi nízke. Pokrytie Wi-Fi je veľmi závislé na konštrukcii budovy, priemerne však dosahovala vzdialenosť približne 100 m. Nové štandardy a technológie umožnili zvýšiť pokrytie na približne 1 km. Lokalizácia Wi-Fi zariadení je založená na RSS (sile prijatého signálu) [3, 4]. Wi-Fi pokrýva širšie oblasti ako Bluetooth a poskytuje vyššiu príepustnosť, vďaka čomu je používanie Wi-Fi praktickejšie [3].

Bluetooth

Bluetooth (IEEE 802.15.1) je určený pre bezdrôtovú komunikáciu na krátku vzdialenosť medzi zariadeniami. Bluetooth na komunikáciu využíva rádiové vlny s frekvenciami medzi 2,402 GHz a 2,480 GHz. Vyznačuje sa cenou efektívnosťou, nízkym vysielacím výkonom, výdržou batérie, bezpečnosťou a efektívou komunikáciou a dostupnými riešeniami [5, 6]. Nová verzia Bluetooth s názvom „Bluetooth Low Energy“ (BLE) dokáže pokryť priestor do vzdialenosťí 70–100 m s rýchlosťou až 24 Mbps a vyššou energetickou účinnosťou [7]. Vzhľadom na to nie je Bluetooth vhodný na lokalizáciu pre veľkú oblasť [8]. Môže byť použitý na zistenie polohy používateľa na základe meraní RSS. V poslednej dobe sa lokalizácia založená na BLE využíva v smartfónoch ako iBeacons (Apple) a Eddystone (Google). Smartfón možno použiť na lokalizáciu v rámci letísk, železničných staníc, veľkých trhov, nákupných centier a reštaurácií, kde je mapa oblasti odoslaná do smartfónu a následne sa vykoná lokalizácia pomocou BLE [7].

Ultrazvuk

Ultrazvuková lokalizácia je vynikajúca pri krátkych vzdialostiach kvôli nízkej spotrebe energie a lacným komponentom [9]. Na druhej strane sa môže znížovať presnosť lokalizácie pomocou tejto technológie v dôsledku viacnásobných odrazov z povrchov a problémov so synchronizáciou medzi komunikujúcimi uzlami [10, 11]. Ultrazvukové lokalizačné systémy sú však náročné na výpočtový výkon z dôvodu použitia komplexných algoritmov spracovania signálov. Ak sa však splnia všetky podmienky, tak presnosť lokalizácie môže dosiahnuť pár centimetrov. Aby boli výsledky presné, je potrebné umiestnenie ultrazvukových uzlov určiť pred lokalizáciou [12].

Komunikácia viditeľným svetlom

VLC „visible light communication“ alebo komunikácia viditeľným svetlom je technológia lokalizačného systému, ktorú možno použiť na miestach citlivých na rádiové frekvencie. Ukázalo sa, že VLC poskytuje vyššiu presnosť v porovnaní so systémami Wi-Fi. Vďaka zlúčeniu diód vyžarujúcich svetlo (LED) sa VLC široko používa pri lokalizácii. LED má mnoho výhod, vrátane dlhej životnosti, odolnosti voči vlhkosti, nízkej spotrebe energie a nízkych prevádzkových nákladov. LED môže tiež modulovať signály svetelných vĺn pri vysokej rýchlosťi [13]. Lokalizáciu pomocou VLC je možné vykonať pomocou systémov na báze fotodiód, ktoré zachytávajú intenzitu svetla a systémov založených na obrazových senzoroch, ktoré dokážu zachytiť svetelné impulzy [14]. LED aj senzory by však mali byť v zornom poli pre presnú lokalizáciu. Detekcia svetla a lokalizácia vzdialenosť (LiDAR) poskytuje informácie o obrysoch okolitých prekážok a tým môžu tieto systémy poskytnúť presnú lokalizáciu [15].

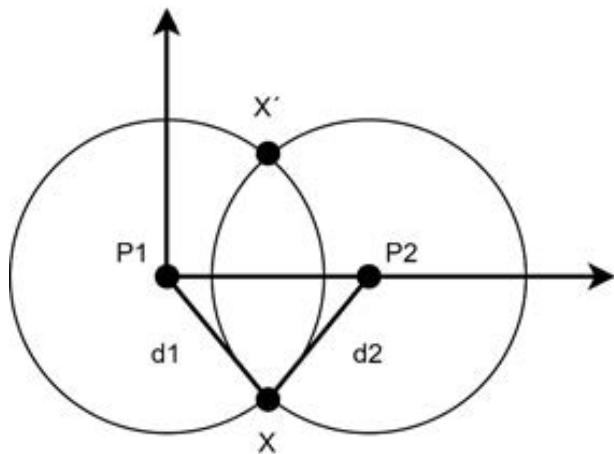
3. METÓDY LOKALIZÁCIE

Technológie ako Wi-Fi, Bluetooth, či ultrazvuk sú vhodné na použitie dvoch metód lokalizácie, ktorými sú triangulácia a trilaterácia, zatiaľ čo VLC je možné využiť pri ICP algoritme. Triangulácia a trilaterácia na lokalizáciu využíva príncíp výpočtu pozície pozorovateľa vzhľadom na množinu orientačných bodov. Iteračný najbližší bod (ICP) sa často používa na rekonštrukciu 2D alebo 3D povrchov z rôznych skenovaní, lokalizáciu robotov a dosiahnutie optimálnej plánovanej cesty.

Triangulácia

Odhad pozície touto metódou je známy už veľmi dlho. Uvedené odhady polohy, alebo vzdialenosťí

od viacerých orientačných bodov s vopred známymi polohami umožňujú zástupcovi vypočítať svoju vlastnú pozíciu. Najjednoduchší a najznámejší prípad je ten, ktorý zároveň pomenúva túto metódu. Ako princíp používa uhly, alebo meranie vzdialenosť voči viacerým orientačným bodom. Triangulácia rieši planárne úlohy polohovania, teda parametre trojuholníka vzhľadom ku kombinácii strán a uhlov. Aj keď geometria trojuholníka nie je jediná možnosť pre použitie orientačných bodov, je to azda ten najprirodzenejší spôsob^[16].



Obr. 1. Najjednoduchší triangulačný prípad^[17]

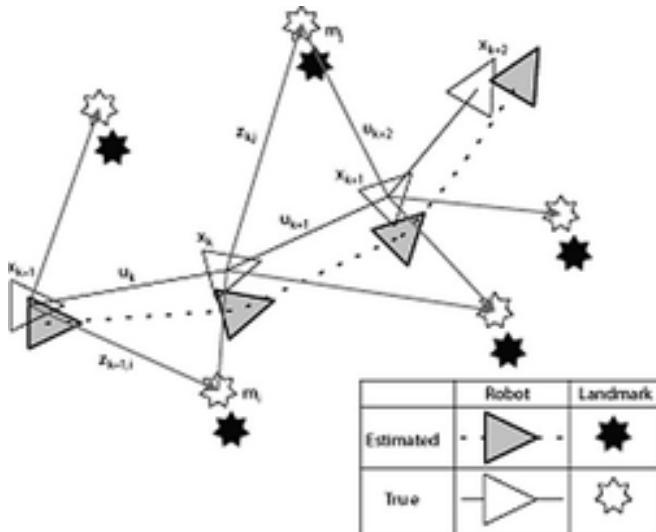
Obr.1. - Robot na neznámom mieste X sníma dva orientačné body P1 a P2 meraním k nim príslušných vzdialenosť D1, a D2. (To zodpovedá prípadu, kedy majáky v známych miestach vydávajú signál a robot získa vzdialenosť na základe časového oneskorenia, kedy robot signál prijme.) Robot musí ležať na priesecníku kružnice s polomerom d1, so stredom v P1 a s kružnicou o polomeru d2 a stredom P2. Bez straty všeobecnosti môžeme predpokladať, že pôvodný orientačný bod je P1 je v bode (0,0) a P2 je v (a, 0). Písmeno a označuje známu vzdialenosť orientačných bodov P1, a P2^[17].

ICP algoritmus

ICP algoritmus je metóda porovnávania skenov, ktorá pristupuje k SLAM „Simultaneous localization and mapping“ problému bez použitia príznaku, každý bod senzorického merania môžeme považovať za samostatný príznak a potom sa tento rozdiel vytráca. Základom týchto metód je nájdenie optimálnej transformácie medzi niekoľkými nasledujúcimi senzorickými meraniami, prípadne je hľadaná transformácia napríklad medzi mriežkou obsadenosti a novým senzorickým meraním. Optimálna transformácia je potom hľadaná vybranou optimalizačnou metódou, väčšinou optimalizačným algoritmom najmenších štvorcov^[18]. Výhoda týchto metód je zrejmá a jej použitie je vhodné v komplikovaných prostrediach vzhľadom na množstvo prekážok, ktoré môžu blokovať signál pre predošlú metódu. Tvorba primitívnej bodovej mapy je relatívne jednoduchá, na druhej strane sú tieto metódy citlivé na počiatočné podmienky.

4. SIMULTÁNNA LOKALIZÁCIA A MAPOVANIE (SLAM)

SLAM je skratka pre Simultánnu lokalizáciu a mapovanie. V mapovaní je problém integrovať informácie, ktoré zhromažďujú senzory robota do danej reprezentácie. To môže byť opísané otázkou: „Ako vyzerá svet?“. Centrálné aspekty v mapovaní sú reprezentácia prostredia a interpretácia dát senzorov. Na rozdiel od mapovania, v lokalizácii je problémom odhad pozície robota vo vzťahu k mape. Inými slovami, robot má odpoved' na otázku: „Kde to som?“. Zvyčajne sa rozlišuje medzi pozíciovou sledovania, keď je začiatočná pozícia známa a globálnou lokalizáciou, v ktorej nebola daná žiadna prioritná vedomosť o štartovacej pozícii. Simultánna lokalizácia a mapovanie je preto definované ako problém vybudovania mapy, v ktorej v rovnakom čase prebieha lokalizácia robota bez použitia tejto mapy. V praxi nemožno tieto dva problémy riešiť nezávisle od seba. Predtým, ako robot môže odpovedať na otázku „Ako prostredie vyzerá?“, z výsledkov pozorovania je potrebné vedieť, z ktorého prostredia získané pozorovania pochádzajú. Zároveň je ľažké bez mapy odhadnúť aktuálnu polohu robota. Preto je SLAM často prirovnávaný k problému: „Čo bolo skôr? Sliepka alebo vajce?“ Dobrá mapa je potrebná pre lokalizáciu, zatiaľ čo presný odhad pozície je potrebný na vytvorenie mapy.



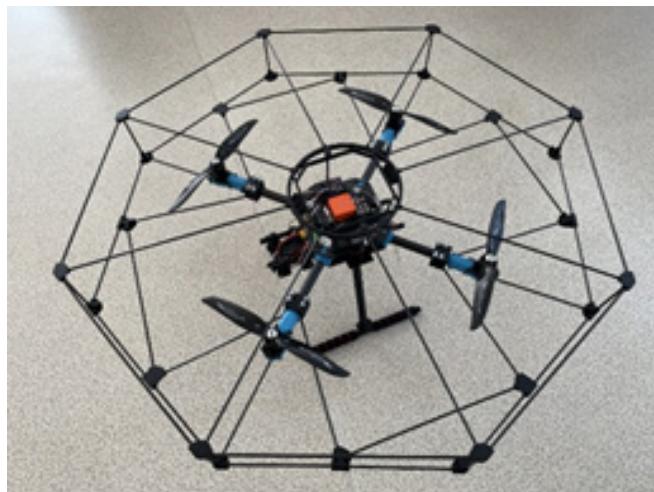
Obr. 2 Základný SLAM problém [19]

Obr. 2. - V časovom okamihu k, sú nasledujúce veličiny definované ako:

- x_k - stavový vektor, ktorý popisuje umiestnenie a orientáciu robota.
- u_k - kontrolný vektor, ktorý aplikuje v čase k-1 riadiť vozidlo na stav x_k v čase k.
- m_i - vektor, ktorý popisuje umiestnenie i-teho orientačného bodu, ktorého pravé umiestnenie predpokladá, že je nemenné.
- z_{ik} - pozorovanie prevzaté z miesta vozidla i-teho orientačného bodu v čase k. Keď existuje viac pozorovaní orientačných bodov v jednom okamihu alebo keď konkrétny orientačný bod nie je relevantný k diskusii, pozorovanie bude zapísané iba ako z_k .

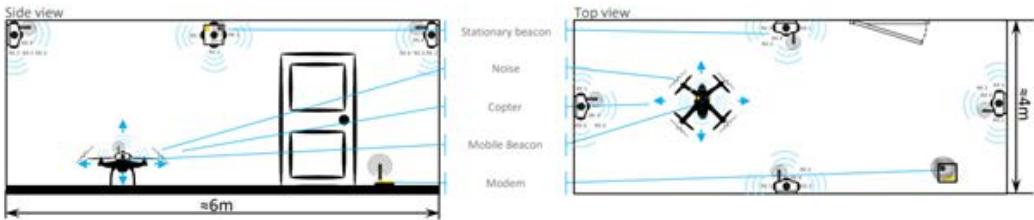
5. NÁVRH A REALIZÁCIA PROTOTYPOVÉHO RIEŠENIA

Pri realizácii prototypu sme použili stavebnicu dronu kvadrokoptéry HOLYBRO X500 V2 obsahujúcu všetky potrebné komponenty ako konštrukcia, motory, vrtule, atď. Riadenie letu zabezpečuje autopilot Cube Orange s redundantným IMU systémom.



Obr. 3. Kvadrokoptéra HOLYBRO X500 V2

Po zostavení dronu a nakonfigurovaní autopilota bolo realizovaných niekoľko skúšobných letov pomocou RC ovládača. Ďalším krokom bola lokalizácia dronu, ktorá prebiehala v hale určenej na výskum a vývoj v rámci projektu. Po rozmiestnení dodaného hardvéru sa prešlo k testom, ktoré dokázali, že z použitých technológií najpresnejšie určenie pozície v interiéri dosahuje ultrazvuková technológia od Marvelmind robotics.



Obr. 4. Príklad umiestnenia Marvelmind Super-Beacon [20]

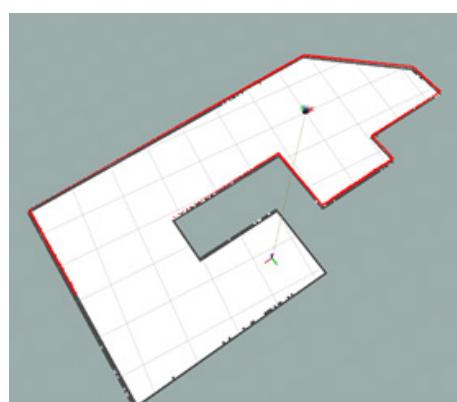
Lokalizácia Marvelmind Super-Beaconmi je natoľko presná, že po pripojení vyhodnocovacej jednotky k autopilotu Cube Orange sa vytvára možnosť ovládať dron s presnosťou na niekoľko centimetrov. Keďže beacons vytvárajú interiérový navigačný systém a dron ovláda autopilot, tak je po zadaní cieľových súradníc možný autonómny presun dronu z bodu A do bodu B.

Nevyhodou tejto metódy lokalizácie je zložitá príprava priestoru, v ktorom sa bude dron pohybovať. Je nevyhnutné zabezpečiť presné polohy beacons, v priestore sa nesmú vyskytovať prekážky, atď. To však nie je v mnohých prípadoch možné a preto sme využili kombináciu technológií – lokalizáciu a mapovanie SLAM. Tento prístup si vyžaduje vysoký výpočtový výkon a použitie technológií na meranie vzdialenosť. Preto sme navrhli použitie vývojovej dosky NVIDIA Jetson Nano Developer Kit. Tento modul poskytuje vysoký výpočtový výkon a je prispôsobený pre vývoj AI.



Obr. 5. Jetson Nano Developer Kit

Nevyhnutou súčasťou SLAM je zhromažďovanie dát zo senzorov merajúcich vzdialenosť v okolí dronu, aby bolo možné vytvorenie mapy a lokalizácie v nej. Nakol'ko sa dron pohybuje v trojrozmernom priestore, je vybavený stereoskopickou kamerou ZED 2 pre 3D skenovanie priestoru pred dronom a pre súvislé skenovanie celého okolia pomocou LiDAR-u RPLIDAR A3. LiDAR je zapojený k vývojovej doske, na ktorej beží operačný systém LINUX a robotický operačný systém ROS. Systém ROS zabezpečuje simultánnu lokalizáciu a mapovanie za pomoci LiDAR-u. Vývojová doska pomocou sériového rozhrania komunikuje s autopilotom a zasiela mu aktuálne dátá o polohe.



Obr. 6. Príklad mapy zo skenovania LiDAR-om [21]

SLAM spracuje dátá z LiDAR-u a po vyhodnotení začne zobrazovať mapu s lokalizáciou, ktorú je možné overiť aj s presnou pozíciou vďaka Marvelmindu. Momentálne sa pomocou RC ovládača lieta po hale a testuje sa funkčnosť pred pridaním stereoskopickej kamery, ktorá bude predovšetkým slúžiť na autonómnu navigáciu.

ZÁVER

V prvej časti publikácie boli predstavené technológie - Wi-Fi, Bluetooth, ultrazvuk a komunikácia viditeľným svetlom. Tieto technológie majú rôzne využitie v oblastiach, akými sú napríklad prenos dát a fungujú v rámci výskumu a vývoja prototypu UAV ako lokalizačný systém. Lokalizačný systém využíva na určenie polohy rôzne metódy a techniky, ako napríklad ICP algoritmus využívaný pri simultánnej lokalizácii a mapovaní (SLAM). V druhej, vývojovej časti, bolo zrealizované zostavanie kvadrokoptéry HOLYBRO X500 V2, návrh hardvéru a spustenie simultánnej lokalizácie a mapovania pomocou RPLIDAR A3 v prostredí operačného systému ROS. Pre zvýšenie bezpečnosti sme navrhli vlastnú bezpečnostnú klietku vyrobenu z karbónových tyčí pospájaných na mieru namodelovanými a vytlačenými spojkami v laboratóriu prostredníctvom 3D tlačiarne.

POĎAKOVANIE

Tento článok vznikol vďaka podpore v rámci Operačného programu Integrovaná infraštruktúra pre projekt „Inteligentné operačné a spracovateľské systémy pre UAV“, kód projektu v systéme ITMS2014+: 313011V422, spolufinancovaný zo zdrojov Európskeho fondu regionálneho rozvoja.



EURÓPSKA ÚNIA

Európsky fond regionálneho rozvoja
OP Integrovaná infraštruktúra 2014 – 2020



MINISTERSTVO

DOPRAVY

SLOVENSKÉJ REPUBLIKY

LITERATÚRA

- [1] Chen, Z., Zou, H., Jiang, H., Zhu, Q., Soh, Y. C., & Xie, L. (2015). Fusion of WiFi, smartphone sensors and landmarks using the Kalman filter for indoor localization. *Sensors*, 15(1), 715–732.
- [2] Navarro, E., Peuker, B., Quan, M., Clark, A. C., & Jipson, J. (2010). Wi-Fi Localization Using RSSI Fingerprinting.” Citeseer, 2010.
- [3] Le Dertz, N., Gain, F., & Zetterberg, P. (2012). WiFi fingerprint indoor positioning system using probability distribution comparison. In 2012 IEEE international conference on acoustics, speech and signal processing (ICASSP), 2012, pp. 2301–2304.
- [4] Lashkari, A. H., Parhizkar, B., & Ngan, M. N. A. (2010). “WIFI-based indoor positioning system. In Second International Conference on Computer and Network Technology, 2010, 76–78.
- [5] Qureshi, U. M., Umair, Z., & Hancke, G. P. (2019). Evaluating the implications of varying Bluetooth low energy (BLE) transmission power levels on wireless indoor localization accuracy and precision. *Sensors*, 19(15), 3282
- [6] Altini, M., Brunelli, D., Farella, E., & Benini, L. (2010) Bluetooth indoor localization with multiple neural networks. In IEEE 5th International Symposium on Wireless Pervasive Computing 2010, 2010, pp. 295–300.
- [7] Zafari, F., Gkelias, A., & Leung, K. K. (2019). A survey of indoor localization systems and technologies. *IEEE Communications Surveys & Tutorials*, 21(3), 2568–2599.
- [8] Cheung, K. C., Intille, S. S., & Larson, K. (2006). An inexpensive bluetooth-based indoor positioning

hack. In Proceedings of UbiComp, 2006, vol. 6.

- [9] Hsiao, C.-C., & Huang, P. (2008). Two practical considerations of beacon deployment for ultrasound-based indoor localization systems. In 2008 IEEE International Conference on Sensor Networks, Ubiquitous, and Trustworthy Computing (sutc 2008), 2008, pp. 306–311.
- [10] Sainjeon, F., Gaboury, S., & Bouchard, B. (2016). Real-Time Indoor Localization in Smart Homes Using Ultrasound Technology. In Proceedings of the 9th ACM International Conference on PErvasive Technologies Related to Assistive Environments, 2016, pp. 1–4.
- [11] Hammoud, A., Deriaz, M., & Konstantas, D. (2016). Robust ultrasound-based room-level localization system using cots components. In 2016 Fourth international conference on ubiquitous positioning, indoor navigation and location based services (UPINLBS), 2016, pp. 11–19.
- [12] Runge, A., Baunach, M., & Kolla, R. (2011). Precise self-calibration of ultrasound based indoor localization systems. In. International Conference on Indoor Positioning and Indoor Navigation, 2011, 1–8.
- [13] Luo, J., Fan, L., & Li, H. (2017). Indoor positioning systems based on visible light communication: State of the art. *IEEE Communications Surveys & Tutorials*, 19(4), 2871–2893.
- [14] Huynh, P., & Yoo, M. (2016). VLC-based positioning system for an indoor environment using an image sensor and an accelerometer sensor. *Sensors*, 16(6), 783.
- [15] Xiao, Y. Ou, Y., & Feng, W. (2017) Localization of indoor robot based on particle fiter with EKF proposal distribution. In 2017 IEEE international conference on cybernetics and intelligent systems (CIS) and IEEE conference on robotics, automation and mechatronics (RAM), 2017, pp. 568–571.
- [16] K. S Arun, T. S. Huang, D. Blostein, 1987. Latest-Squares Fitting of two 3D Point Sets. *IEEE Transactions on Pattern analysis and Machine intelligence*, ISSN 0162-8828, 1987, vol. PAMI-9, no.5, p698-700.
- [17] G. Dudek, M. Jenkin, 2010. Computational Principles of Mobile Robotics. 2nd edition, Cambridge University Press, 2010, p244-245, ISBN 978-0-521-69212-0.
- [18] Mazi R., Lokalizace pro autonomní systémy. Disertační práce, České vysoké učení technické, Praha, 2007.
- [19] H. Durant-Whyte, T. Bailey. 2006. Simultaneos Localisation and Mapping (SLAM): Part 1 The Essential Algorithms. *IEEE Robotics and Automation Magazine*. ISSN 1070-9932, 2006. vol.13, no2., p.99-110.
- [20] Marvelmind Indoor Navigation System Operating manual, v2022_08_24, URL: https://marvelmind.com/pics/marvelmind_navigation_system_manual.pdf
- [21] Kiran Palla, Creating Map using Laser Scanner and Gmapping, 2018, URL: <https://kirancalla.com/autonomous-navigation-ros-differential-drive-robot-simulation/creating-map-using-laser-scanner-and-gmapping/>



OAI identifiers

Decentralised PIDs for research outputs in repositories

Petr Knoth
The Open University
petr.knoth@open.ac.uk



Big Scientific Data and Text Analytics group :
AI for open and responsible research



OAI Identifiers

Decentralised PIDs for Research
Outputs in Repositories

CORE-GPT

Combining Open Access research and AI
for credible, trustworthy question
answering

Dr. Petr Knoth





CORE is the world's most used aggregator of **Open Access** papers, collating and enriching content from over **11,000 repositories**.

- >20 Million monthly active users (MAU)
- 34 Million full-text research papers hosted by CORE.
- 260 Million metadata records

Signatory of Principles of Open Scholarly Infrastructure (**POSI**)



Providing
seamless access
to open research
for humans and
machines.

CORE delivers **services** for HEIs, researchers, funders and commercial partners, offering seamless access to research.



Commercial Partners



- Innovation and trends analysis
- Plagiarism detection
- Fact checking
- Finance
- Health

Institutional Members



Research areas

- AI Applications in Research Evaluation (e.g. citation type classification, bibliometrics, impact assessment)
- Automatic Expert Finder systems (e.g. for peer-review and grant applications)
- Deduplication, document classification, rapid systematic reviews
- Research graphs: entity extraction (affiliation, author, etc.)
- Research recommender systems and academic search

Dr. Petr Knoth : Senior Research Fellow in Text and Data Mining petr.knoth@open.ac.uk



What is CORE



CORE's mission is

to index open access research worldwide and deliver unrestricted access for all.

We are here to support you and to advance the Open Access / Open Research movement



WE ARE

the world's **most used scholarly database** of open access research papers with >20 million monthly active users

WE

provide solutions for content management, discovery and scalable machine access to research.

WE ARE

a **not-for-profit** scholarly infrastructure dedicated to the open access mission, **adopters of POSI principles**.



WE

serve the global network of repositories and journals by increasing discoverability and reuse of open access content.

Our services

Seamless access to open research for humans and machines.

Use cases

- Innovation and trends analysis
- Plagiarism detection
- Fact checking
- Research graphs
- Market prediction
- Drug discovery
- ...

Commercial Partners

ontochem
IT SOLUTIONS



cypbris

EZASSI
Technology to innovate faster.



Academic members

Content discovery

Search

Recommender

Discovery

Raw data services

API

Dataset

FastSync

Managing content

Repository Dashboard

Identifiers

OAI Resolver



Why is CORE important?

Increase your contents' discoverability and prevent its misuse

Search, Recommender, Discovery, PMC Linkout

Make your papers uniquely identifiable and resolvable with PIDs

OAI Resolver

Assess and contribute to Open Access compliance and FAIRness

Indexed by CORE badge

Make your content machine readable

CORE API, CORE Dataset, Repository Health Check

Become a CORE Member and benefit from lots more

Dashboard: Metadata validation and monitoring

>20M monthly active users

rioxx

UKCORR

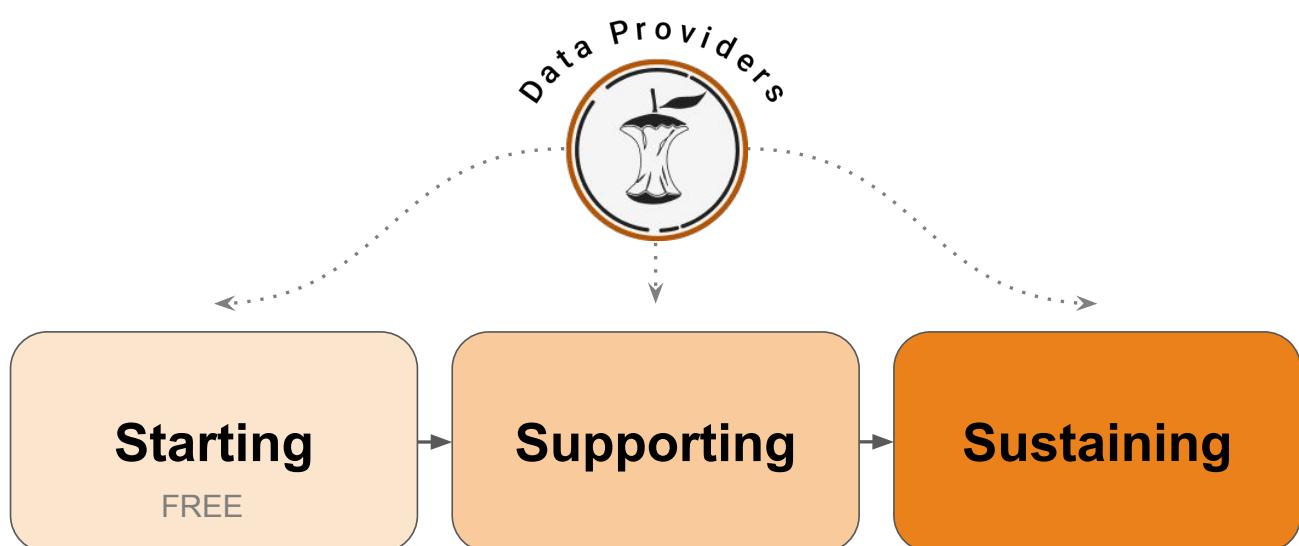
UKRI
UK Research and Innovation

CORE Membership

- A network of data providers who are committed to the ongoing success of the **Open Access movement**
- We provide **tools and benefits** for our members
- All CORE data providers are eligible to become CORE Starting Members **free** of charge
- Supporting and Sustaining Members:
 - help shape our development roadmap
 - support and sustain CORE



Three levels of CORE Membership



Why is CORE important?

>20M monthly active users

rioxx **UKCORR**

UKRI UK Research and Innovation

Our services

Seamless access to open research for humans and machines.

Use cases

- Innovation and trends analysis
- Plagiarism detection
- Fact checking
- Research graphs
- Market prediction
- Drug discovery
- ...

Commercial Partners

ontachem
IT SOLUTIONS



cyprios

EZASSI
Technology to innovate faster



Academic members



Content discovery

Raw data services

Managing content

Search



API

Repository Dashboard

Recommender



Dataset

Identifiers

Discovery



FastSync

OAI Resolver

OAI Identifiers: Decentralised Identifiers for Repository Records

Dr. Petr Knoth

CORE, Knowledge Media institute, The Open University

Acknowledgements: Herbert van de Sompel, Martin Klein



Drivers for PIDs adoption in repositories

- **Open Access Policies:** Plan S, REF2021 OA Policy, UKRI OA Policy ...
 - Policies typically ask for: “all repository outputs to have PIDs”
 - Deposits of research outputs into repositories (multiple authors can / should deposit) within a certain time limit (e.g. acceptance or publication)

“Why did the persistent identifier keep asking for a date with the URL? Because it was determined to establish a long-lasting relationship!”

What PIDs should be used to identify repository records? (Green OA)



General PID requirements

Requirements for repository identifiers:

- 1. Unique:** the same identifier should not be used to identify another entity.
- 2. Dereferenceable:** it should be possible to resolve this identifier to the entity which it identifies.
- 3. Persistent:** for a given entity, the identifier should remain the same over its lifetime.



What identifiers should be used to identify records in repositories? (Green OA) – 1/2

- Requirements for repository record identifiers:
 1. **Dereferencing to the repository:** We need an identifier that will resolve (dereference) to the splash page of the resource in a given repository.
 2. **Identifier of the repository record rather than of the scholarly work:** ability to identify metadata only and metadata with research outputs records.
 3. **High adoption rate:** needs to be widely adopted across the network of repositories so that all repository records could be uniquely identified.



What identifiers should be used to identify records in repositories? (Green OA) – 2/2

- Desirable properties of repository identifiers:
 1. **Distributed control:** it is desirable for the identifier to be minted at the level of each individual repository rather than by a central entity.
 2. **Cost free:** Minimum or no cost for minting them.
 3. **Complement the DOI system:** Ability to link repository records to entities identified by a DOI, e.g. full text, dataset. (especially for multi-author, multi-institution deposits)



OAI identifier: The missing PID for repository records actually already exists ... 😊

- Created by the Open Archives Initiative over 20 years ago, used as part of OAI-PMH
- Already **universally adopted** across repositories
- **Decentralised**
- **Persistent:** in repositories declaring persistent support for deletedRecords in OAI-PMH
- **Free to mint**



Structure:

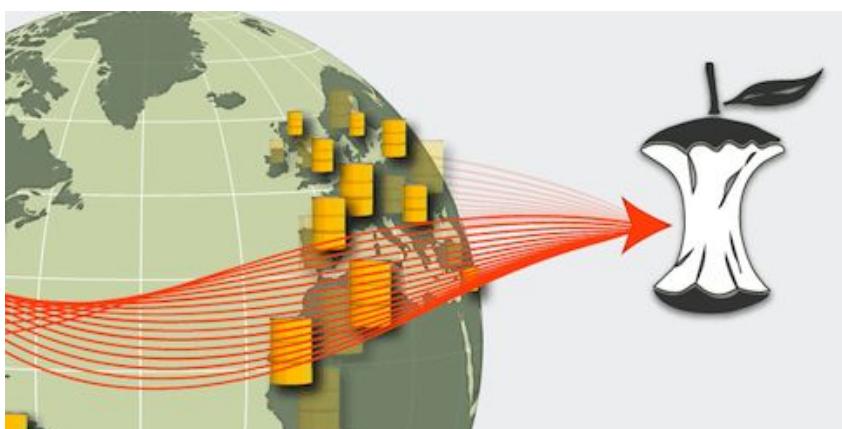
oai:oro.open.ac.uk : 73739

globally unique OAI Prefix

locally unique identifier



CORE



- Aware of deposits from across the global repositories network and their OAIs
- Ideally positioned to deliver an OAI resolver service

Over 200 million metadata records

Nearly 25 million free to read full texts

Over 12,000 data providers in 147 countries



Example OAI identifiers



OAI Resolver

CORE provides a global resolver for OAI identifiers at

https://core.ac.uk/oai_resolver

The RESTful call to resolve an OAI is available at:

<https://oai.core.ac.uk/< oai-identifier>>

- Example OAIs:

oai:eprints.gla.ac.uk:129357

oai:digitalcommons.odu.edu:oaweek-1012

oai:oro.open.ac.uk:75049

oai:dspace.stir.ac.uk:1893/24654



OAI i Resolve an OAI identifier



Insert an OAI identifier

Example: oai:researchonline.rca.ac.uk:1035

RESOLVE

OAIs is a globally unique identifier used by repositories.

Insert an OAI identifier into the search box.

Make sure you entered it correctly and click Resolve.

You will be taken to a page associated with that identifier.



Resolve an OAI identifier

oai:researchonline.rca.ac.uk:1035



Example: oai:researchonline.rca.ac.uk:1035

RESOLVE

OAI is a globally unique identifier used by repositories.

Insert an OAI identifier into the search box.

Make sure you entered it **correctly** and click Resolve.

You will be taken to a page associated with that identifier.

← → C researchonline.rca.ac.uk/1035/

Archive of research processes and output produced by RCA

Royal College of Art

Experimental Design: Design Experimentation

Tools

Hall, Ashley, 2011, Journal Article, *Experimental Design: Design Experimentation* Design Issues, 27 (2). pp. 17-26. ISSN 07479360

Information **Documents**

Abstract or Description: This paper was selected for publication in MIT's Design Issues. The research takes an original approach by positioning experimentation as a comprehensive design methodology, rather than using the traditional industrial design approach of employing experimentation as a problem-solving tool within a standard design model. It is an evolution of design thinking on non-linear design methods first developed by Hall and presented to the 'International Association of Societies of Design Research Conference', Seoul, South Korea (2009), and in a paper entitled 'Innovation design engineering: Non-linear progressive education for diverse intakes' presented at the 'International Conference on Engineering and Product Design Education', University of Brighton, UK, which offered a non-linear pedagogy (Hall and Childs 2009) that uniquely supports a diverse interdisciplinary intake. Experimental design is well known in the science domain but very little evidence has been recorded of experimentation in industrial design and its position in relation to work in other science and research domains. Connections are made with theories on research methods, an analysis of case studies and comparisons of literature on experimentation from science disciplines, especially that of Kuhn (1962), Galison (1987), Pasteur's quadrant for scientific research in Stokes (1997) and Borgdorff (2007). Hall makes significant claims in exploring and articulating a model of design experimentation that highlights the differences between scientific and design experimentation. This work was original in describing an experimental design model for the increasing activity in early phases of design development by recording and enhancing knowledge in this important area for future design research and practice. The methods researched in the paper were later used in experimental design workshops in Daegu, South Korea (2011) and Busan, South Korea (2012).

Official URL: <http://www.mitpressjournals.org/doi/abs/10.1162/DE...>

Subjects: [Creative Arts and Design > W200 Design studies](#)
[Creative Arts and Design > W200 Design studies > W240 Industrial/Product Design](#)

School or Centre: School of Design

Activating OAI resolver mapping for your repository in CORE

The screenshot shows the CORE dashboard with the URL dashboard.core.ac.uk/data-providers/86/settings. The left sidebar has a 'Settings' tab selected. The main area is titled 'OAI i' and 'Register OAI mapping'. It explains how to register a mapping between an OAI Prefix and a URL. The 'OAI Prefix' field contains 'oai:oro.open.ac.uk:'. Below it, the 'URL Mapping' field contains 'http://oro.open.ac.uk/cgi/oai2'. A green toggle switch indicates 'Resolving to your repository is activated'. A 'SAVE' button is at the bottom.

Adoption of OAIs



OAI identifiers need more community recognition and awareness!



Property	OAI
Unique	✓
Dereferenceable to a repository	✓
Persistent	✓ (in repositories declaring "persistent" handling of deletedRecords)
Adoption rate	Very high
Identification of records with and without objects (e.g. full texts, datasets)	✓
Cost of minting	Free of charge
Distributed control	✓



How can we incorporate this in the UKRI OA policy?



Technical requirements for institutional and subject repositories:

To be considered compliant with UKRI's open access requirements research articles made available via repositories are required to meet the following technical requirements:

- PIDs for research outputs must be implemented according to international recognised standards, examples of international standards include DOI, URN or Handle
- article-level metadata must be implemented according to a defined application profile that supports the UKRI Open Access Policy and if possible is available via a CC0 public domain dedication; this must include the persistent identifier to both the Author's Accepted Manuscript and the Version of Record; the metadata standard must adhere to international best practice such as the **RIOXX** and **OpenAIRE** guidelines.
- machine-readable information on the open access status and the licence must be embedded in the metadata in a standard non-proprietary format.
- common unique PIDs for research management information (for example identifiers for funders and /or organisations) are strongly encouraged; ORCID, the researcher identifier, must be supported.
- the repository must be registered in the Directory of Open Access Repositories (OpenDOAR) and **its metadata must be uniquely identifiable and accessible in accordance with the OAI-PMH specification.**



Contributions



1. OAI identifiers can be used as cost-free decentralised PIDs for repository records.
2. OAI identifiers deserve more community awareness.
3. Delivered an OAI resolver as an important piece of open infrastructure to drive up the awareness of OAIs



**Thank you!
Q&A**





Core GPT

Combining Open Access research and AI for credible, trustworthy question answering

Petr Knoth

The Open University
petr.knoth@open.ac.uk



Big Scientific Data and Text Analytics group :
AI for open and responsible research



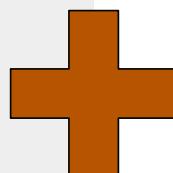
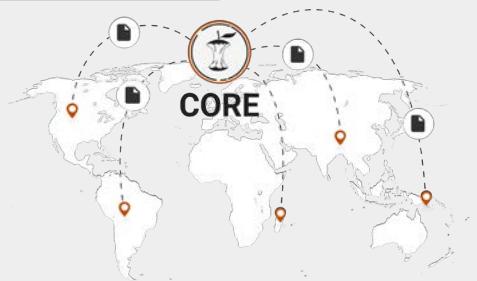
CORE-GPT

Combining Open Access research and AI for credible, trustworthy question answering



CORE-GPT

CORE is the world's largest collection of **Open Access** papers, collating and enriching content from over **11,000** data providers.

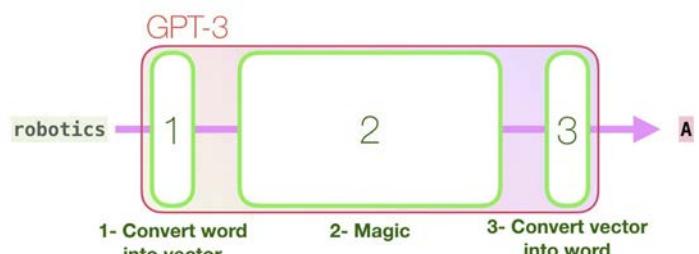


- **>20 Million** monthly active users
- **34 Million** full-text research papers hosted by CORE.
- **260 Million** metadata records



GPT large language models*

- Can comprehend context and generate human-like text
- Can infer meaning from large-scale data



@JayAlammar

*Other large language models are available

CORE-GPT

Introducing CORE-GPT



Tell me a joke about ChatGPT responses not being based on scientific papers



Why did ChatGPT refuse to base its responses on scientific papers?

Because it didn't want to get lost in the references!

Introducing CORE-GPT

The screenshot shows a web-based interface for CORE-GPT. At the top, there's a search bar with the placeholder "What do you want to know?" Below it is a text input field with the question "Tell me a joke about ChatGPT responses not being based on scientific papers". To the right of the input field is a green "Ask" button. A response from ChatGPT is displayed below: "Why did ChatGPT refuse to base its responses on scientific papers? Because it didn't want to get lost in the references!" At the bottom of the interface, there's a navigation menu with links like "Project", "About us", "Team", "Services", "Blog", "Data", "Data providers", "Harvested content", "Support", "FAQs", "Contact us", "Legal", "Accessibility", "Cookies", "Terms and Conditions", and "Privacy". On the right side, there's a sidebar with the text "Writing about CORE? Discover our research outputs and cite our work." and logos for Jisc and The Open University.

CORE-GPT Results

The screenshot shows a search results page for "What are the latest research trends in biomimicry?". At the top, there's a search bar with the same question. Below it is a detailed search result card. The card features a small thumbnail image of two people looking at a document, the title "What are the latest research trends in biomimicry?", a brief summary of the research, and a "Read more" button. At the bottom of the card, there's a note about COVID-19 and a link to the full document. The page also includes a sidebar with the text "Writing about CORE? Discover our research outputs and cite our work." and logos for Jisc and The Open University.



CORE-GPT Results

What are the latest research trend in biomimicry?

Biomimicry in architecture and building design is a promising research trend for reducing energy use and decarbonizing buildings, as it takes inspiration from nature to create more sustainable solutions [1]. Mechatronic and soft robotics, particularly in marine environments, are also taking inspiration from the animal kingdom to create high-performance robots, and there is significant research in energy provision, biomaterials, and locomotor designs [2]. Similarly, a recent biomimetic exploratory study aimed at designing a passive cooling unit for buildings in hot, arid regions by emulating biological cooling strategies in nature, such as cooling through animals' respiratory passages, thermal radiators, and airflow cooling in termite mounds [3]. Lastly, to address climate change and biodiversity loss, built environment-focused responses that take advantage of biomimicry could potentially make a contribution to solving these problems by interacting harmoniously with the living world [4].

1. <https://core.ac.uk/works/131198542> - Biomimetic Buildings: Copying Nature for Energy Efficiency
2. <https://core.ac.uk/works/8388662> - Research Trends and Future Perspectives in Marine Biomimicking Robotics
3. <https://core.ac.uk/works/68990789> - Biomimetics for passive air conditioned design for buildings in the hot arid regions
4. <https://core.ac.uk/works/24102079> - Ecosystem Services Analysis for the Design of Regenerative Urban Built Environments
5. <https://core.ac.uk/works/123801488> - Holistic biomimicry: a biologically inspired approach to environmentally benign engineering

[See more in CORE](#)



Reflections / limitations ...

ChatGPT

- Can get confused (esp. when answers are ambiguous) mixing content from entirely semantically different uses of a concept
- Can be made to argue your way producing biased text
- It can start inventing things / hallucinate ...

CORE-GPT

- Answers need to be anchored to research papers.
- More honest about what it doesn't know => fewer hallucinations
- References make it easier to assess the trustworthiness of the answer.

BOTH

- Powerful at synthesizing content and creating summaries
- Able to compare and contrast
- Can get confused (esp. when answers are ambiguous) mixing content from entirely semantically different areas / uses of a concept
- Can be made to argue your way producing biased text
- Critical thinking and judgement needs to be exercised



CORE-GPT Roadmap

Development Roadmap

- Search (Q/A)
 - Conversational Agent (ChatBot)
- API
- Recommender
- Development work
 - Offline model comparison / integration
 - Training new model on CORE data
 - Using structured full text (GROBID)



Take home ...

- CORE provides a large corpus of open research papers which can be utilised to make GPT answer in an evidence-based manner using references.
- Aiming to make CORE-GPT available from the CORE API.
- Our group is very interested in working with others on research projects in the area of AI for open and responsible research



THANK YOU



Mapping of scientific outputs of Czech institutions in WoS

David Jiří Šlosar

Library of the Czech Academy of Sciences, Charles University

slosar@lib.cas.cz



ÚSTAV INFORMAČNÍCH STUDIÍ
A KNIHOVNICTVÍ
Filozofická fakulta
Univerzita Karlova



Mapping of scientific outputs of Czech institutions in WoS

David Jiří Šlosar

ILIDE 2023, Jasná

25th April 2023

Introduction

- Web of Science is the important tool for obtaining data on scientific production in Czech Republic.
- Part of information for evaluation and for CRIS (Current Research Information System) is drawn from WoS.
- Information should be extracted as widely and accurately as possible.
- This requires the proper way of retrieving data.

2

Available tools and data sources

- To obtain information about the production of a scientific institutions in WoS, it is possible to use:
 - UT WOS list from institutional CRIS (ASEP, OBD)
 - List of UT WOS from national CRIS (IS VaVaI - RIV)
 - Organization Enhanced Names
 - Affiliation Search Query (ASQ)

3

Affiliation Search Query

Author Information

Corresponding Address: Orozco, Modesto (corresponding author)

- Inst Rec Biomed, Mol Modeling & Bioinformat Unit, Parc Cientif Barcelona, E-08028 Barcelona, Spain

Addresses:

- Inst Rec Biomed, Mol Modeling & Bioinformat Unit, E-08028 Barcelona, Spain

Inst Nacl Bioinformat, E-08028 Barcelona, Spain

- Barcelona Supercomp Ctr, Computat Biol Program, E-08028 Barcelona, Spain

Acad Sci Czech Republ, Inst Organ Chem & Biochem, Ctr Biomol & Complex Mol Syst, CR-16610 Prague 6, Czech Republic

Acad Sci Czech Republ, Inst Biophys, CS-61265 Brno, Czech Republic

Affiliation

Czech Academy of Sciences

Institute of Biophysics of the Czech Academy of Sciences

Masaryk Univ, Fac Sci, CS-61137 Brno, Czech Republic

Univ Utah, Dept Med Chem, Salt Lake City, UT 84112 USA

Univ Utah, Dept Pharmaceut Chem, Salt Lake City, UT 84112 USA

Univ Utah, Dept Pharmaceut, Salt Lake City, UT 84112 USA

Univ Utah, Dept Bioengn, Salt Lake City, UT 84112 USA

Univ Nottingham, Sch Pharm, Nottingham NG7 2RD, England

Univ Nottingham, Ctr Biomol Sci, Nottingham NG7 2RD, England

Univ Barcelona, Fac Biol, Dept Bioquim & Biol Mol, E-08028 Barcelona, Spain

...fewer addresses

4

Affiliation Search Query

- Institute AND Organization AND Nation
- Example: Inst Biophys, Czech Acad Sci, Czech Republic =>
- AD = (((Czech Acad Sci) SAME (Inst Biophys)) SAME Czech)
- Reality:
- AD = (((Akad Cesk* Repub* OR Akad CR OR Akad Ved OR AS CR OR AS CZ OR ASCR OR ASCZ OR AV CR OR AVCR OR CAS OR AS Czech Rep* OR Acad Sci* OR ASC OR AV Ces* Rep* OR Ac Sci OR ACSR OR Czech Acad* OR Acad Czech Repub*) SAME (Inst Biophys OR Biofyz* Ustav OR Inst Biophys OR Inst Biophys)) SAME Czech)

5

Affiliation Search Query

- Institute of Physics of the Czech Academy of Sciences
- AD = (((Akad Cesk* Repub* OR Akad CR OR Akad Ved OR AS CR OR AS CZ OR ASCR OR ASCZ OR AV CR OR AVCR OR CAS OR AS Czech Rep* OR Acad Sci* OR ASC OR AV Ces* Rep* OR Ac Sci OR ACSR OR Czech Acad* OR Acad Czech Repub*) SAME (Inst Phys OR Fyzik* Ust* OR FZU OR Phys Inst) NOT (Nucl OR phys* Mat* OR Mat* phys* OR phys Met OR phys* Plasma* OR plasma* phys* OR Atmospher OR phys Chem OR Chem proc OR Heyrovsky OR Atomospher OR Inst Atmospher OR Solar OR mech) OR Inst Phys SAME Slovance OR ELI Beamlines SAME Inst Phys OR IOP SAME ELI Beamlines OR Acad Sci Czech Republic SAME Slovance OR Czech Acad Sci SAME Slovance OR Fyzikalni Ustav AV CZR OR Dept Phys AVCR OR Acad Sci Czech Repub* SAME Dept Phys OR HiLASE Project SAME Inst Phys OR IP AS CR SAME Olomouc OR ((ASCR SAME IOP) SAME Hilase ctr) OR Czech AS SAME Inst Phys) SAME Czech)

6

Affiliation Search Query

- =A(NEBO(A(NEBO(JE.ČISLO(HLEDAT("Akad Cesk* Repub*";A2));JE.ČISLO(HLEDAT("Akad CR";A2));JE.ČISLO(HLEDAT("Akad Ved";A2));JE.ČISLO(HLEDAT("AS CR";A2));JE.ČISLO(HLEDAT("AS CZ";A2));JE.ČISLO(HLEDAT("ASCR";A2));JE.ČISLO(HLEDAT("ASCZ";A2));JE.ČISLO(HLEDAT("AV CR";A2));JE.ČISLO(HLEDAT("AVCR";A2));JE.ČISLO(HLEDAT("CAS";A2));JE.ČISLO(HLEDAT("AS Czech Rep*";A2));JE.ČISLO(HLEDAT("Acad Sci*";A2));JE.ČISLO(HLEDAT("ASC";A2));JE.ČISLO(HLEDAT("AV Ces* Rep*";A2));JE.ČISLO(HLEDAT("Ac Sci";A2));JE.ČISLO(HLEDAT("ACSR";A2));JE.ČISLO(HLEDAT("Czech acad*";A2));JE.ČISLO(HLEDAT("Acad Czech Repub*";A2)));NEBO(JE.ČISLO(HLEDAT("Inst Biophys";A2));JE.ČISLO(HLEDAT("Biofyz* Ustav";A2))));A(JE.ČISLO(HLEDAT("Inst Biophys";A2));JE.ČISLO(HLEDAT("Kralovopolska 135";A2)));A(JE.ČISLO(HLEDAT("Inst Biophys";A2));JE.ČISLO(HLEDAT("Acad Sci Czech Republ";A2)));A(JE.ČISLO(HLEDAT("Czech Acad Sci";A2));NEBO(JE.ČISLO(HLEDAT("Inst Biophys";A2));JE.ČISLO(HLEDAT("Inst Biophys";A2)))));JE.ČISLO(HLEDAT("Czech";A2)))

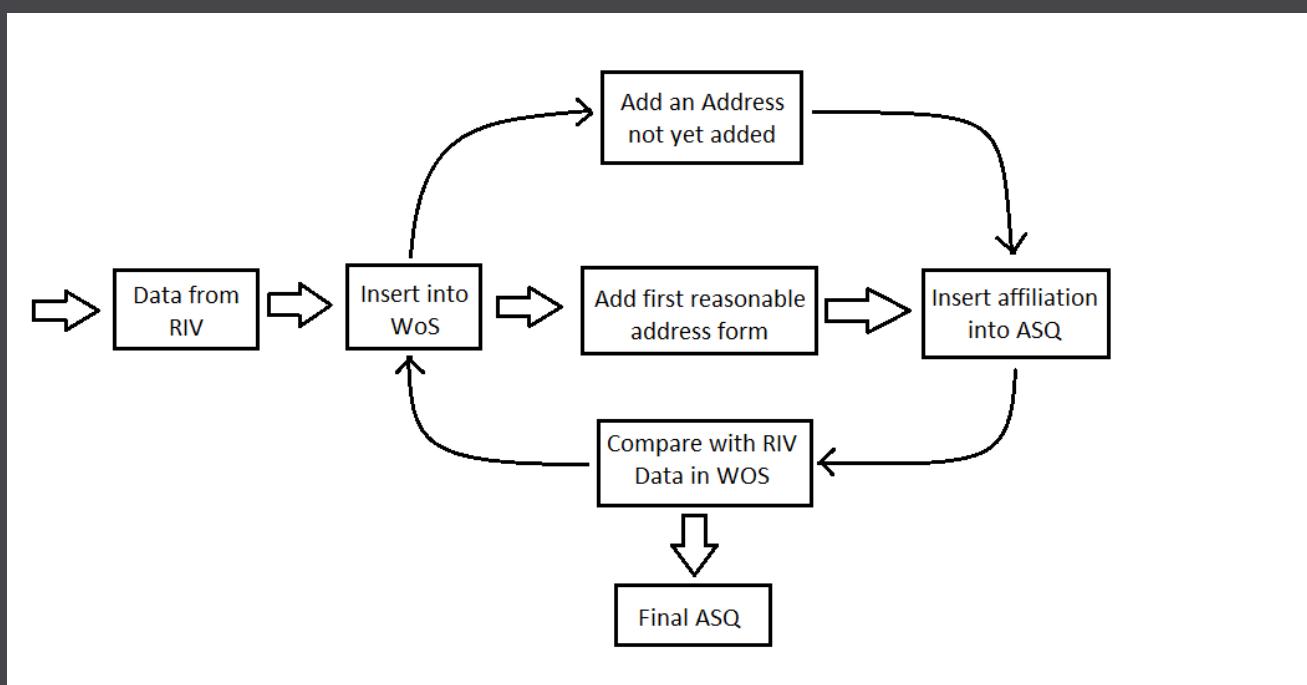
7

Affiliation Search Query

- Made for almost all institutes of CAS and faculties of public Universities in Czech Republic (235).
- At faculty/institute level
- Final publication year 2014 - 2018
- Document type Article, Proceedings Paper, Review
- AND FPY = (2014 OR 2015 OR 2016 OR 2017 OR 2018) AND DT = (Article OR Proceedings Paper OR Review)

8

Framework



9

Results

- Czech Academy of Sciences: OEN available at the level of institutes (51 included – 3 excluded), CRIS ASEP available.

Source	Number of results
RIV	25692
ASEP	26830
OEN	27793
ASQ-RIV	28047
ASQ-ALL	28082

- Charles University: No OEN available at the faculty level (18 fac. Incl. – 2 excl.).

Source	Number of results
RIV	19161
OBD	18159
ASQ-RIV	21749
ASQ-ALL	21786

- University of Chemistry and Technology Prague: No CRIS available. No OEN available at the faculty level. 5 Faculties.

Source	Number of results
RIV	2345
ASQ-RIV	3368

10

Thank you for your attention

Name: Mgr. David Jiří Šlosar

Email: slosar@lib.cas.cz, davidjiri.slosar@ff.cuni.cz



ÚSTAV INFORMAČNÍCH STUDIÍ
A KNIHOVNICTVÍ
Filozofická fakulta
Univerzita Karlova



11

Reaxys

What's new

Piotr Gołkiewicz
Elsevier
P.Golkiewicz@elsevier.com

Reaxys is a unique chemistry database that contains extracted data and information about chemical compounds and their associated physicochemical properties, reaction and synthesis information, biological activity, commercial availability, use, extraction from natural products, as well as bibliographic data that can be browsed and displayed through a functional interface.

2023 marks 35 years of innovation, since 1989, when the first data was loaded into Beilstein.

Reaxys Content.

Reaxys started with Beilstein and Gmelin Handbooks that annotated information manually. It guarantees high quality of data extraction, but manual extraction is slow, poorly scalable, expensive, with time it becomes harder and harder to follow the burst of information. Thus, automated content extraction is the only approach that can address all these challenges. Using latest technologies, Reaxys is performing automatic substance excerpt from text and images in patent archives, further growing an unprecedented amount of extracted experimental physico-chemical properties of chemical substances.

Reaxys has the largest literature coverage in the field of chemical and related sciences (Life Science) with 16,000 journals, covering ca 4000 publishers.

In addition, it offers full patent coverage from 105 patent offices in 170 patent classes. At the time of writing this article the number of patents exceeded 38 million with 36 million patents added since 2020. It provides accessible IP with translation of non-English language abstracts & claims and supports novelty search with substances indexed across the full patent body.

Reaxys offers an unprecedented amount of extracted experimental physico-chemical properties of chemical substances. It covers chemistry and related sciences: organic (incl. synthetic chemistry), inorganic, analytical, spectroscopy, analytical methods, bio-chemistry, pharmacy, material science, natural products, environmental chemistry and more.

Reaxys Targets ad Bioactivities.

With the addition of the Reaxys Targets and Bioactivities module Reaxys became the world's largest database of biological activity of chemical substances (44M measurement points). It provides a comprehensive set of bioactivity information to help establish known structure-activity relationships, DMPK and ADME-Tox properties. In addition, Reaxys provides a graphical bioactivity visualisation allowing fast and efficient view of structures and their bioactivity values against various targets. Users can tailor routes: add and delete reaction steps and branches. They also have better route selection, with view of commercial substances in project page, number of suppliers and speed of delivery.

Reaxys Predictive Retrosynthesis.

The Reaxys Predictive Retrosynthesis module is designed to drive time savings and cost control in synthetic chemistry projects. It provides fast predictions of synthesis routes to known and unknown small organic molecules. The module combines high-quality reaction data from Reaxys with state-of-the-art AI & ML technologies from our partner Pending.AI, in an easy-to-use interface. Unique to the Reaxys Predictive Retrosynthesis module, routes include access to supporting literature references, experimental procedures, commercial availability of starting materials and much more.

Reaxys Commercial Substances - Commercial suppliers of starting materials.

Reaxys Commercial Substances provides up-to-date information on the commercial availability and prices of commercial substances - data from over 400 suppliers, which can be easily found thanks to the clear interface and many filters.

Two structure editors to choose from, MarvinJS and ChemDrawJS

Having in mind the ease-of-use Reaxys has 2 built in structure editors – MarvinJS and ChemDrawJS. Both editors are the top of the class tools to draw chemical structures and substructures.

Direct access to substance information in Scopus.

Users with access to both Scopus and Reaxys benefit from Scopus document pages enriched with Reaxys data. With a single click user can move from Scopus record containing a chemical structure to Reaxys, where he or she can check all the relevant information and data related with the structure.

Enriched abstract content with excerpted experimental information is key to assess relevance and quality of a research document. Visual scan of chemical information, key to stay up-to-date on the recent advancements in a chemistry related field and includes additional links to commercial availability and synthesis planning

Better coverage of CAS numbers.

In partnership with CAS, an integrated retrieval service provides the corresponding Reaxys substance when a CAS RN is searched. This capability is now available via Quick Search and Query builder.

Improved access to SDS.

New links to Safety Data Sheets for commercial substances provide access to needed safety information.

Reaxys provides the option to filter products to only those with SDS links, for fast access to relevant data.

In summary

In summary Reaxys delivers:

- Improved R&D productivity across all disciplines of chemistry
- Empowers confident and thorough investigation of research topics
- Helps users to find answers more quickly than with other chemistry research solutions
- Enables greater discovery from a broader range of content
- Empowers all users, regardless of research experience
- Offers additional support for students and other first-time users
- Provides continued access to accurate synthesis pathways with integrated links to public and proprietary databases for substance availability
- Provides availability of substance–target affinity data and comprehensive pharmacokinetic, efficacy, toxicity, safety and metabolic profiles
- Integrates chemistry research workflow with other Elsevier products including Scopus and ScienceDirect
- Provides greater support for your chemical patent research

PERO OCR for prints and manuscripts

Petr Žabička
Moravian Library In Brno
petr.zabicka@mzk.cz



PERO OCR For Prints And Manuscripts

And Other Machine Learning Activities At The Moravian Library In Brno

www.mzk.cz

Agenda

- Introduction to the Moravian Library
- Subject categorisation using machine learning techniques
- Pero OCR: enhancing access to historical documents
- Beyond PERO

Moravian Library And Its Portals

- Legal deposit research library (over 4M items)
 - 100 years as a national institution
- Knihovny.cz portal
 - VUFind based, integrates collections and services of over 80 libraries
- Digital library (271,000 documents, 66M pages)
 - digitalniknihovna.cz
- Czech digital library
 - czechdigitallibrary.cz



Subject Categorisation

- Knihovny.cz indexes over 17M bibliographic records
 - older records without subject cataloguing ⇒ limited findability through filtering
- Machine assignment of 500 main subject categories
 - OCLC Conspectus based
 - trained on metadata and full-text
 - works quite well with metadata only
 - co-operation with 

Czech Digital Libraries

- Kramerius digital library system (open source)
 - over 40 installations (+ 2x Slovakia)
 - java core, Angular front-end
 - Solr, open API, IIIF gateway
 - jpeg2000, MODS, METS import, common metadata recommendations
 - OCR: ALTO, txt
- Kramerius registry: registr.digitalniknihovna.cz

PERO OCR

- Motivation: no OCR for manuscripts, historical printed documents or Czech fraktur
- Problems with OCR for scanned microfilms of newspapers
- Goal: Make the content of digitized documents searchable
- Project leader: Michal Hradíš (VUT)
- Duration: 2018 - 2022

Project partners:

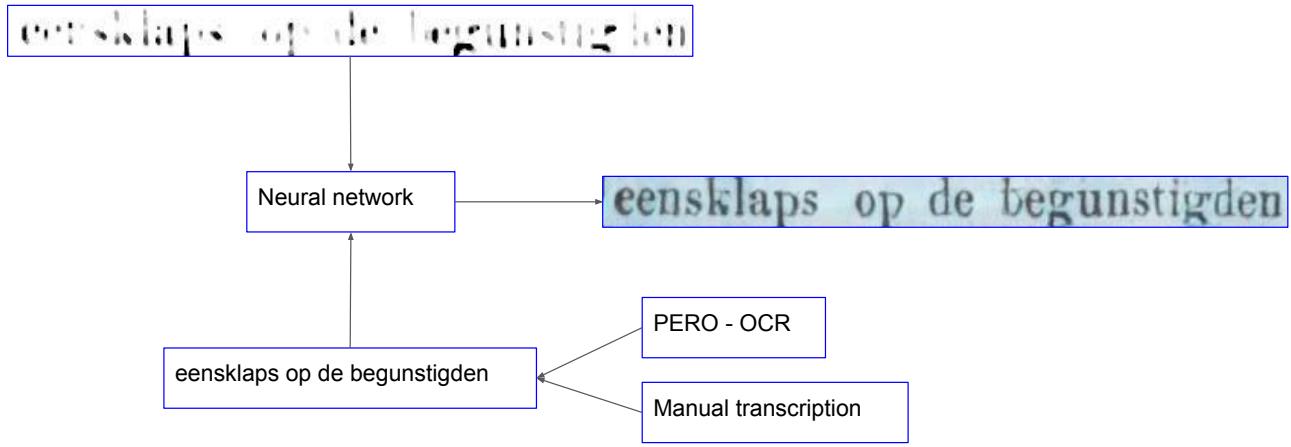


PERO OCR

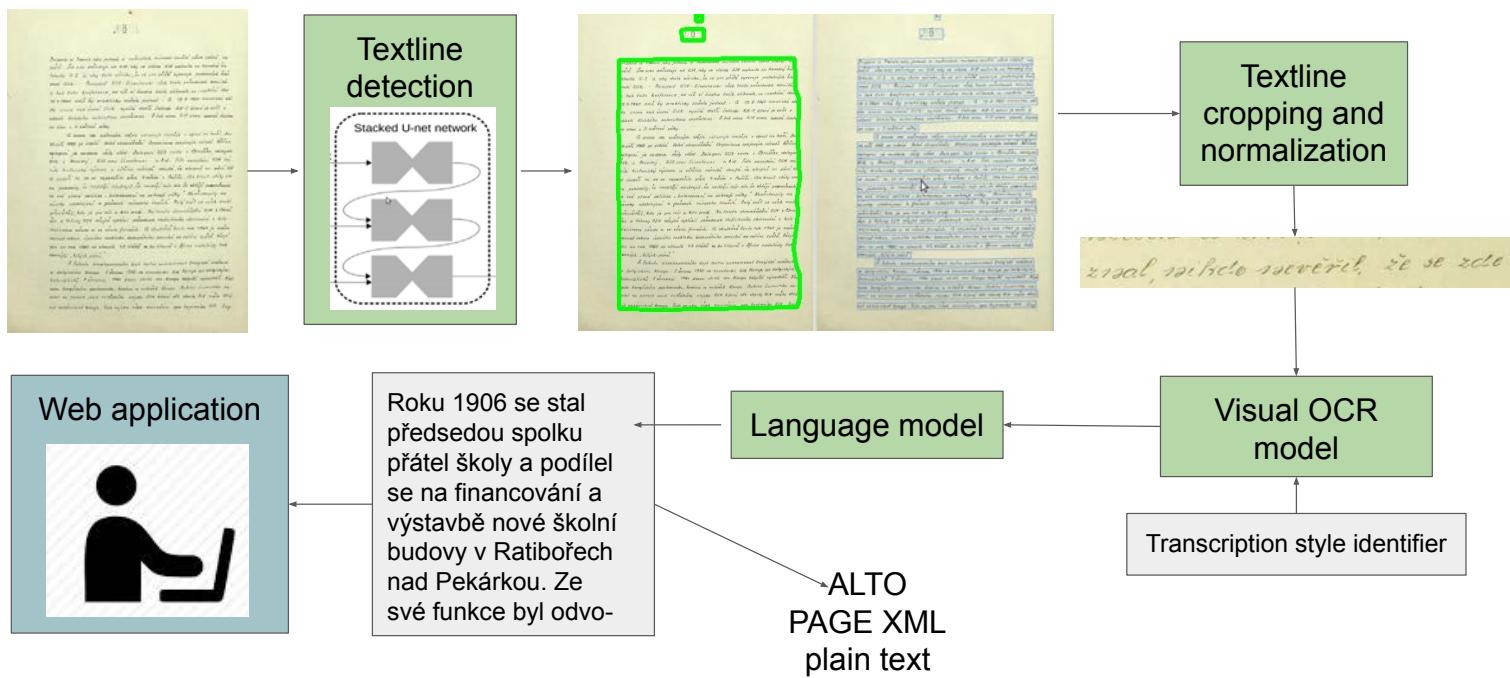
- CORE OCR - pero-ocr python package <https://github.com/DCGM/pero-ocr>
 - Python, PyTorch (running on both CPU/GPU)
- Web application for manual corrections - pero_ocr_web
 - Hosted for free at pero-ocr.fit.vutbr.cz
 - Source codes https://github.com/DCGM/pero_ocr_web
 - Used by: University of Münster, VHÚ, FAU Erlangen, FIT VUT, Jednotlivci, French National Centre for Scientific Research, Moravian Library, Charles University, Masaryk University, Library of Academy of Sciences, ...
- OCR API for large volume document processing
 - Hosted at <https://pero-ocr.fit.vutbr.cz/api> (free for testing)
 - Source codes <https://github.com/DCGM/pero-ocr-api>
 - Documentation <https://app.swaggerhub.com/apis-docs/LachubCz/PERO-API/1.0.4>
- Project info - <https://pero.fit.vutbr.cz/>

PERO OCR

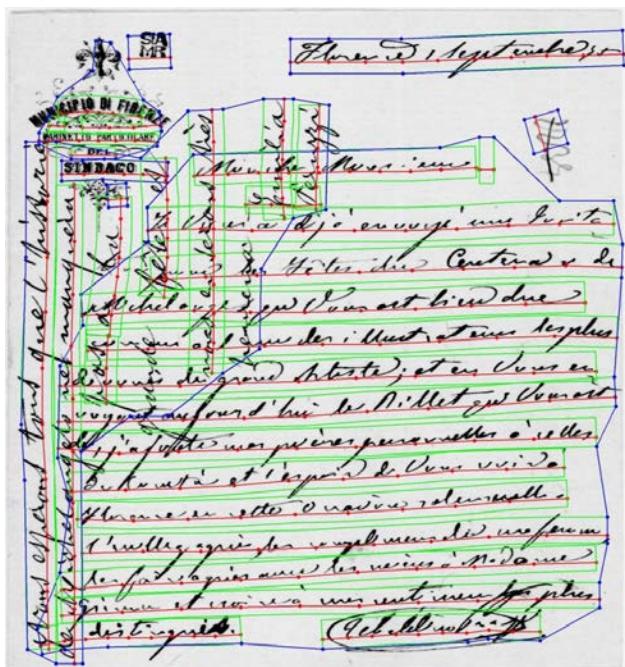
Experiments: Image Improvement



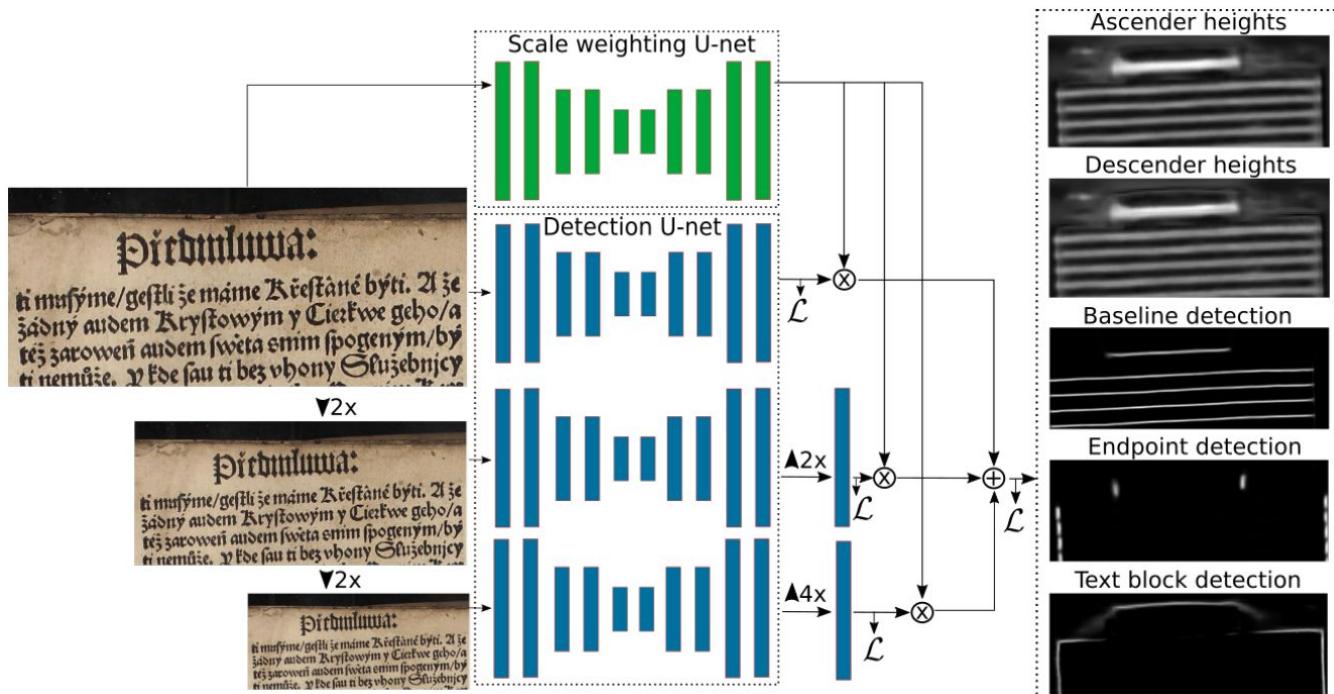
PERO: Full OCR pipeline with tools and applications



Detection of paragraphs and text lines is not that easy

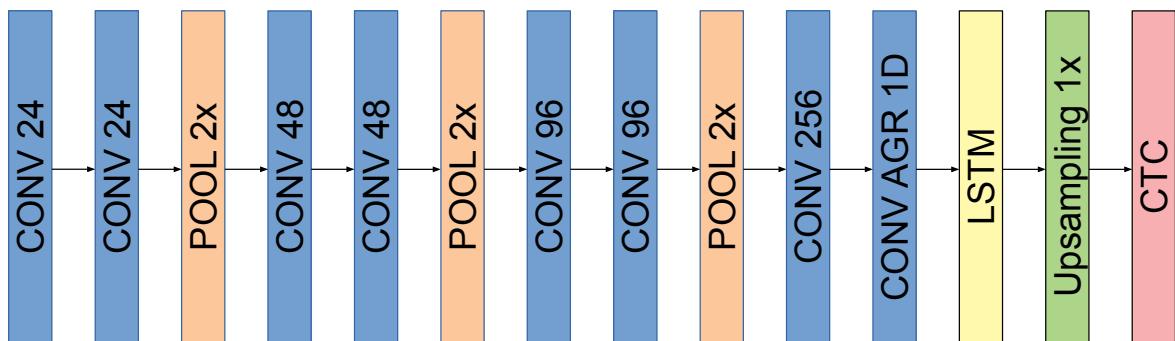
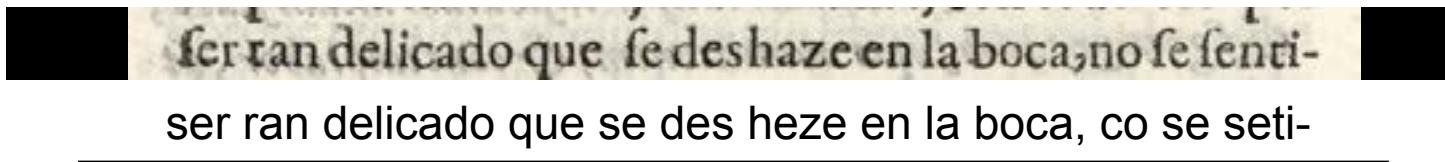


Text line detection - Fully convolutional networks



Kodym, O. and Hradiš, M. [Page Layout Analysis System for Unconstrained Historic Documents](#). International Conference on Document Analysis and Recognition (ICDAR), 2021.

Text transcription - convolutional/recurrent networks + CTC loss



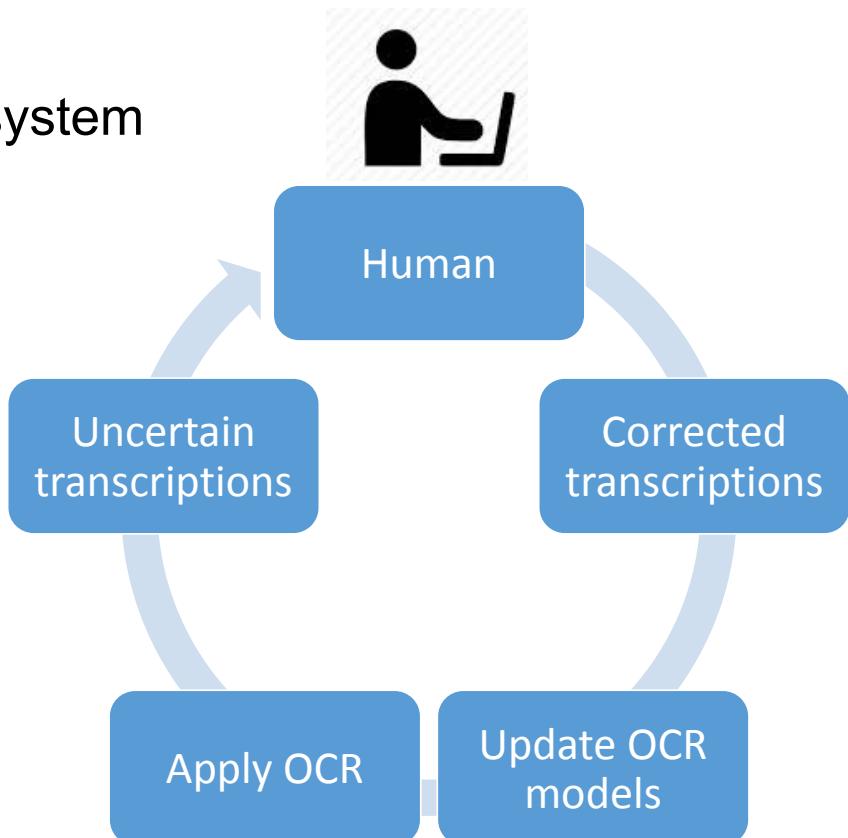
Kišš, M., Beneš, K., and Hradiš, M. [AT-ST: Self-Training Adaptation Strategy for OCR](#). In: Lladós J., Lopresti D., Uchida S. (eds) Document Analysis and Recognition – ICDAR 2021. ICDAR 2021.

Users improve the OCR system

User data is used for training

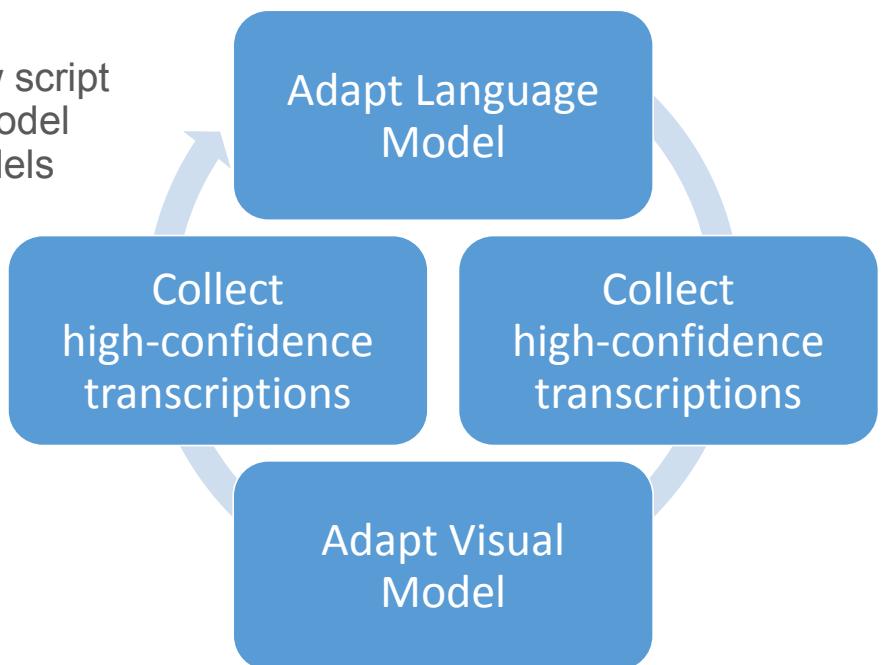
Challenges

- Confused users
- Sloppy users
- Weird users
- Transcription styles

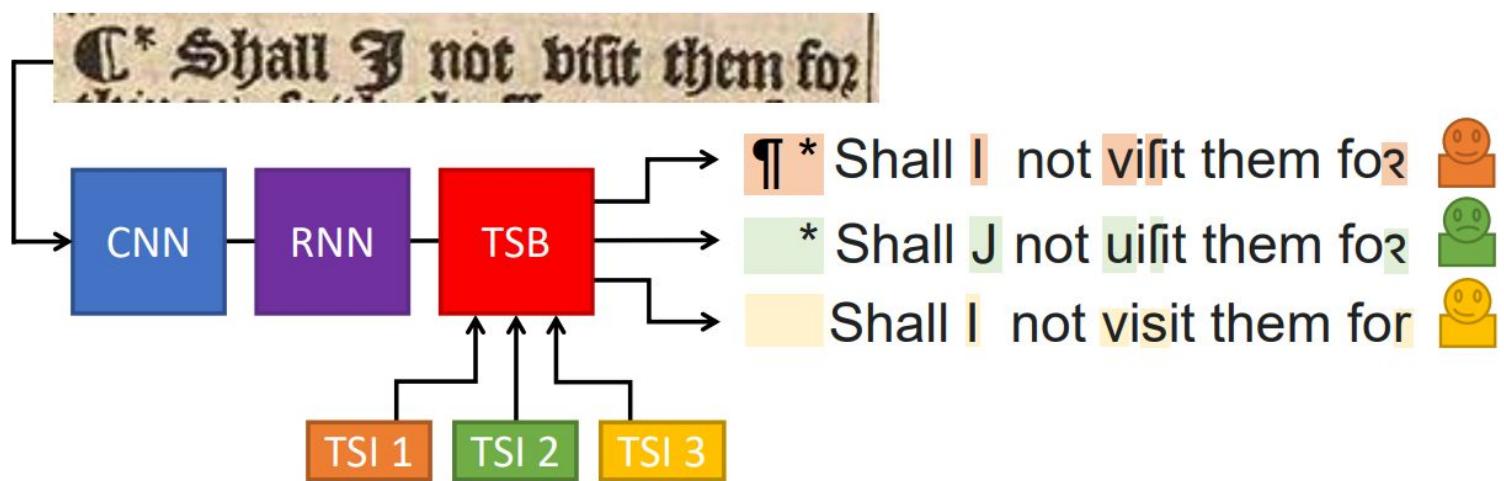


Adapting to a specific document

For transcription of a new script
Starting with a general model
Mutual adaptation of models



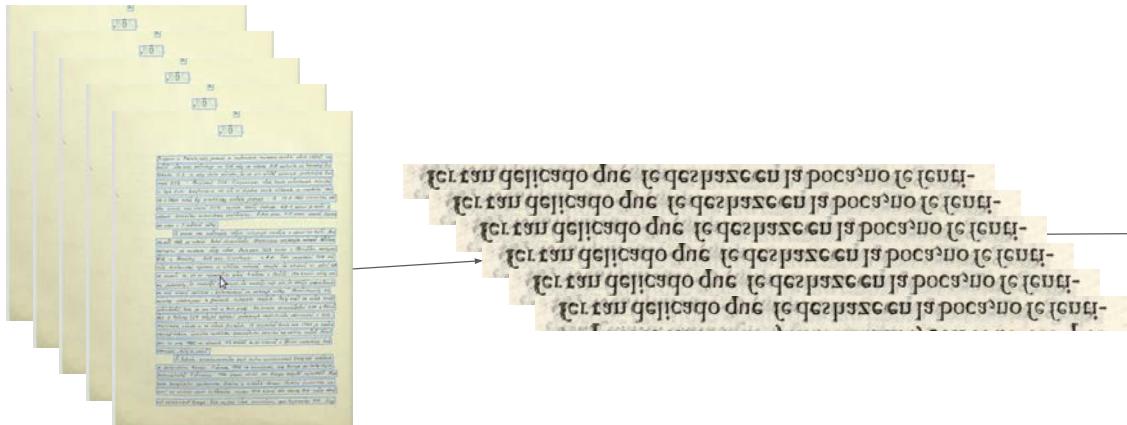
Transcription styles - transcription vs. transliteration



Transcription style selection

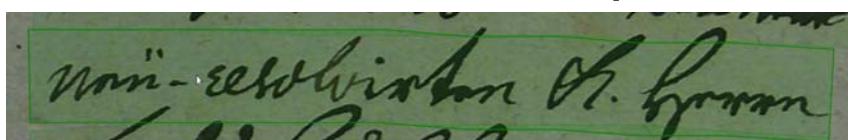
Manual

Automatic based on lines already transcribed and corrected



Immediate transfer of transcription corrections

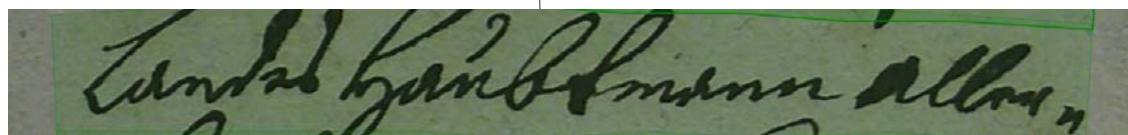
Corrected line:



neü-resolvirten d. Herrn



Landes Haubtmann aller,,



Scans of microfilmed newspaper

Pipra Kulnavou proměnil ji v okamžiku v živou pochodeň. Čáry hořícího šatu strhány s ní sice záhy, ale již bylo shledáno, že má popálenou celou dolní část těla, celou pravou ruku, levou ruku pod zápěstím a dolní část obličeje. Ubohá byla dopravena

Pipra Kulnavou proměnil ji v okamžiku v živou pochodeň. Čáry hořícího šatu strhány s ní sice záhy, ale již bylo shledáno, že má popálenou celou dolní část těla, celou pravou ruku, levou ruku pod zápěstím a

Scans of microfilmed newspaper - fraktur

Menschen, deren einziges Bestreben es ist, die aus einem Jammerthale in ein Eldorado umzuziehen, die Erde zu einem Wohnort umzugestalten, der blos Bro für Alle hat, der Welt auch eine und alle jene Annehmlichkeiten, auf die, ein

Menschen, deren einziges Bestreben es ist, die aus einem Jammerthale in ein Eldorado umzuziehen, die Erde zu einem Wohnort umzugestalten, der blos Bro für Alle hat, der Welt auch eine und alle jene Annehmlichkeiten, auf die, ein

Contemporary handwriting

mezi členy družstva tak potřebná při společném hospodaření.

Rovněž i v živočišné výrobě nebyla situace uspokojivá zvláště v mléčné produkci mohlo být dosaženo daleko lepších výsledků kdyby ošetřova-

telé dojnic dbali více společného zájmu nežli svého prospěchu.

Nutno ovšem doznati že rozptýlené ustájení dojnic na ~~ještě~~ místech,

Leafle

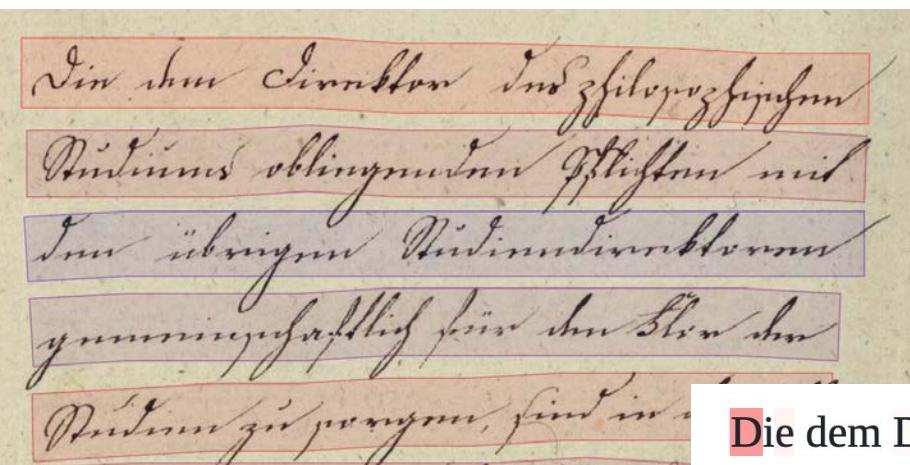
mezi členy družstva tak potřebná při společném hospodaření. ✓

Rovněž i v živočišné výrobě nebyla situace uspokojivá zvláště v mléčné produkci mohlo být dosaženo daleko lepších výsledků kdyby ošetřova-

telé dojnic dbali více společného zájmu nežli svého prospěchu. ✓

Nutno ovšem doznati že rozptýlené ustájení dojnic nach místech, ✓

Kurrent Script (German)



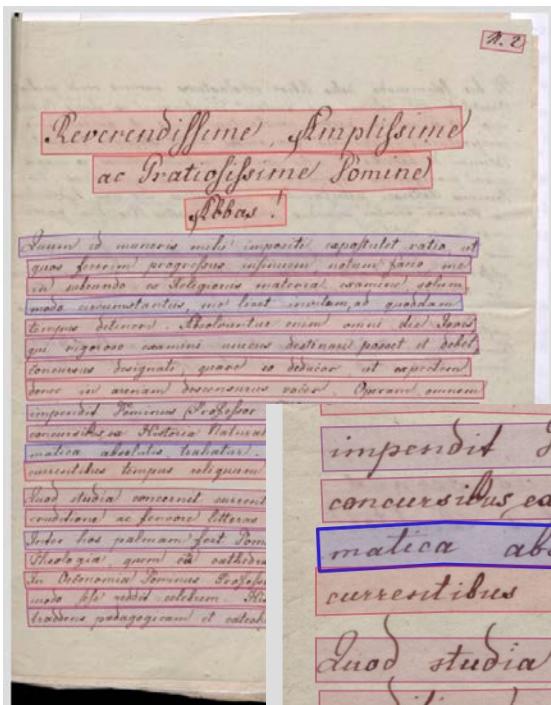
Die dem Direktor, des philosophischen
Studiums obliegenden Pflichten mit
den übrigen Studiendirektoren
gemeinschaftlich für den Fler der
Studien zu sorgen, sind in der all,,

Kurrent Script (Czech)

Přítomná řízeniny čili pýmosty zpráv
byly zprávy mapy i řízení, nejvyšších ředitelů
či přiležitostí a v možnějších dobách
bez velikého výkonu jich mazijných,
a bez nezvědomych řízení mi často =
mimořádná a významná.

Přítomné řízeniny čili sprosté zpěv,
wy pro nafisi chafu, neywjce toliko
přježitně a w rozličných dobách
bez welikého dussewnjho napjmáej
a bez nepowědomé tehdáž mi čapo=

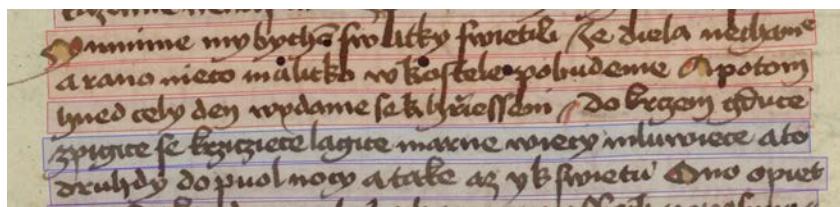
Manuscripts



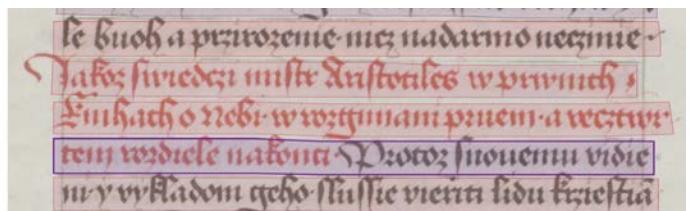
impedit Pominus Professor Löcher ne mora amplius,
concursibus ex Historia Naturali, Diplomatica, et Nemis,
matica absolutis, trahatur Interea omne a studiis
currentibus tempus reliquum materia revolvenda conseoro.
Quod, studia concernit currentia, nactus sum duces ingenio

impedit Pominus Professor Löcher ne mora amplius,
concursibus ex Historia Naturali, Diplomatica, et Nemis,
matica absolutis, trahatur. Interea omne a studiis
currentibus tempus reliquum materia revolvenda conseoro.
Quod studia concernit currentia, nactus sum duces ingenio
eruditio ac fervore litterarum rito tracto. Si

Czech medieval manuscripts



Imnime my bycha swatky swetili se diela nedani
a rano nieto malitko w kolce pobudeme A potom
hned cely den wydame sek hriessem Do krczem gduce
zpigice se krzicziece agite marne wiery mluwiece a to
druhdy do puol nocy a take az ys swietu Ono opiet



le buoh a przirozenie niez nadarmo necznie-
Jakoz swiedczi mistr Aristocileś w prwnich
Eiihach o Nebi w rozgiani pruem a vecztryr
tem rozdiele nakona Protoz Inouemu vidie
ni u vykladom geho flusie vieriti lidu krzestia

Cyrillic

Ці останні аж до нової доби, отже до появи капіталізму, оскільки репрезентувалися політично-пасивними громадськими шарами (селянством і робітництвом), національно перебували в історичній, так би мовити, летаргії. Вищі їхні верстви денаціоналізувалися, опинившись в лабетах чужої державності. З нею, поскільки держава згодом націоналізувалася, вони захоплювалися механічно чужою

Ці останні аж до нової доби, отже до появи капіталізму, оскільки репрезентувалися політично-пасивними громадськими шарами (селянством і робітництвом), національно перебували в історичній, так би мовити, летаргії. Вищі їхні верстви денаціоналізувалися, опинившись в лабетах чужої державності. З нею, поскільки держава

Arabic

ليس من شك في أن المصادر في الدولة تقدم

خدمة، ولكنها في مقابل أدائها لهذه الخدمة

تجد المقابل . هذا المقابل الذي كان حافزا

لافتتاح المزيد من المراكز والفروع في مختلف

انحاء الدولة ولكن لنسائل أين تلك الخدمة

ليس من شك في أن المصادر في الدولة تقدم

خدمة، ولكنها في مقابل أدائها لهذه الخدمة

تجد المقابل ، هذا المقابل الذي كان حافزا

لافتتاح المزيد من المراكز والفروع في مختلف

انحاء الدولة ولكن لنسائل أين تلك الخدمة

Beyond PERO

- Use AI to create structural metadata
- Optical music recognition for music search
- Enhance image search
- Enhance full-text search

Duration 2023 - 2027

Partnerships with large libraries and universities

Smart Digiline – Machine Learning For Digitization Of Printed Heritage

Development of a tool to replace manual metadata creation with AI

- Use of existing digital library data for training machine learning systems:
 - division of a bound periodical volume into individual issues
 - + publication dates, issue numbers
 - detect page numbers, categorize page types

Partners:



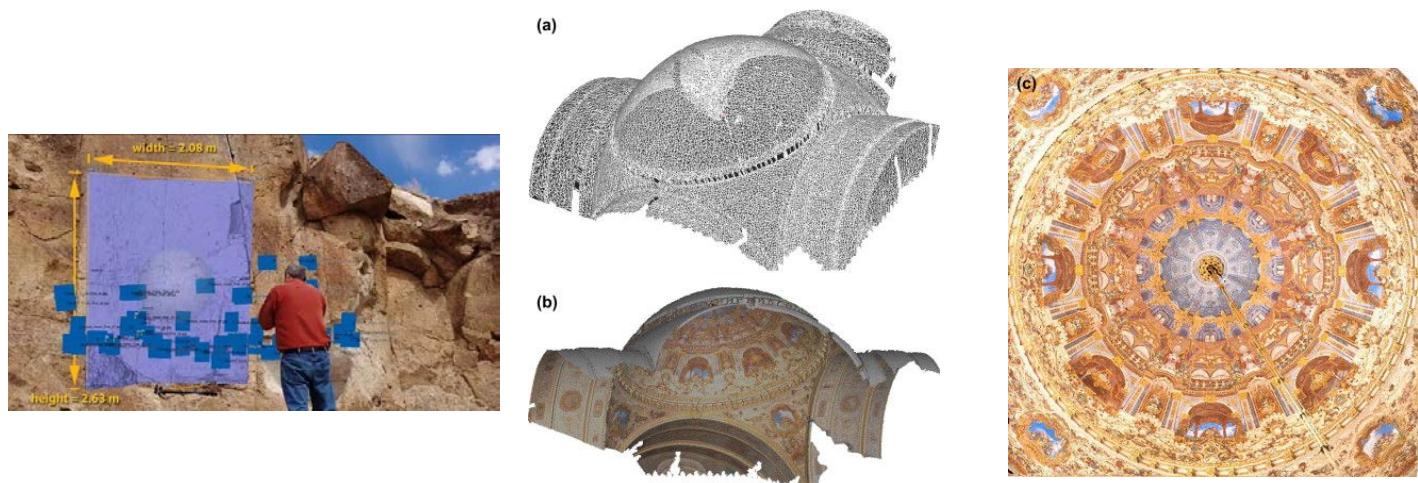
National Library
of the Czech Republic



Smart Digiline

Open Source tools for large object scanning by multi-view reconstruction

Photos, 3D reconstruction, color calibration, high resolution texture without shadows



Orbis Pictus

Open images in digital libraries to the public

- Identify graphic elements contained in digitized documents
- Categorize them by type, add contextual information

Partners:



Orbis pictus - Image search

Query: **ancient Rome feast**

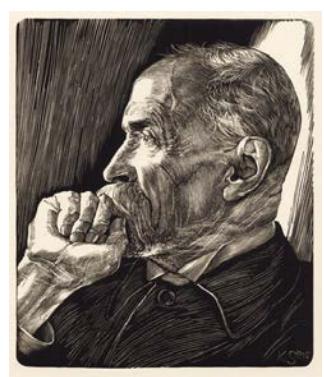
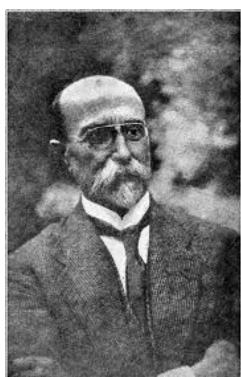
Results:



Orbis pictus - search people by face recognition

Query: president Masaryk

Results:



semANT – Semantic Document Exploration

Improve full-text search capabilities

- Search by word meaning
- Search by text parts, topics
- Improve text data analysis capabilities - topic evolution over time, relationships between topics, etc.

Partners:



semANT – Semantic Document Exploration

Expected results:

- Automatic segmentation of documents into parts (encyclopedias, dictionaries, chronicles...)
- Automatic generation of text descriptions (headlines, keywords, important text parts)
- Finding and analyzing topics in documents
- User application for advanced search in digital library (extension of Kramerius)

semANT - Automatic translation, cross-language search

Im besonderen hat der Secundararzt:

1. die Kopfzettel, die Vormerkblätter und Ordinationszettel der Zugänge durchzusehen, die darauf etwa beigefügten Bemerkungen (zur Beobachtung, unter Aufsicht, Arrestant u. dgl.) zu beachten, das hienach Erforderliche einzuleiten, die anamnestischen Momente zu erheben und in das Vormerkblatt entweder selbst einzutragen oder die Eintragung zu überwachen;

Im besonderen hat der Secundararzt:

1. die Kopfzettel, die Vormerkblätter und Ordinationszettel der Zugänge durchzusehen, die darauf etwa beigefügten Bemerkungen (zur Beobachtung, unter Aufsicht, Arrestant u. dgl.) zu beachten, das hienach Erforderliche einzuleiten, die anamnestischen Momente zu erheben und in das Vormerkblatt entweder selbst einzutragen oder die Eintragung zu überwachen;

In particular, the secondary physician:

1. To look through the header slips, the notice sheets and ordination slips of the arrivals, to observe any comments attached to them (on observation, under supervision, arrestant, etc.), to initiate what is necessary hereafter, to take the anamnestic moments collect and either enter in the memorandum yourself or monitor the entry;

semANT - Intelligent search and content exploration

Query: challenges of Czech society and government between the two world wars

Results:

9. In the years 1920-1922 we waged a triumphant struggle against Communism, and overcame all dangers by our own strength. In the years 1925-1929 we warded off all onslaughts of a Fascist tendency whether coming from abroad or arising at home, and we adopted a democratic regime which we have further strengthened by our own internal legislation and through the collaboration of all the constructive elements of nation and State, and have thus prevented nation and State becoming split into two camps where Left and Right struggle—even at the price of civil

semANT - Intelligent search and content exploration

Topic-aware similarity search

9. In the years 1920-1922 we waged a triumphant struggle against Communism, and overcame all dangers by our own strength. In the years 1925-1929 we warded off all onslaughts of a Fascist tendency whether coming from abroad or arising at home, and we adopted a democratic regime which we have further strengthened by our own internal legislation and through the collaboration of all the constructive elements of nation and State, and have thus prevented nation and State becoming split into two camps where Left and Right struggle—even at the price of civil

Find similar documents discussing:

- struggle against Communism
- pre-war Fascist tendencies
- beginning of Czech national state
- Left and Right struggle
- early Czech legislation
- early Czech politics
- early Czech international relations

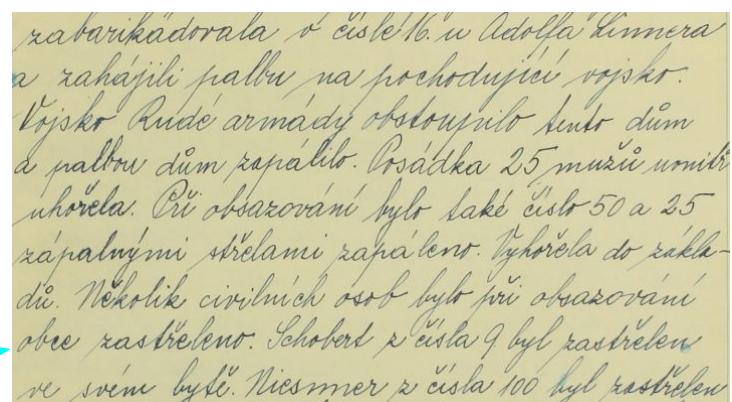
semANT - Intelligent search and content exploration

Query: Please, give me some examples, of liberation by the Red Army in villages.

System response:

The liberation was often peaceful as the german army ran away. However, some villages experienced fighting of various intensity. For example in Václavov by Bruntál, a part of the German garrison fortified themselves in a house, which was set on fire by the Red Army. 25 soldiers died inside. Further, several civilians were shot by the liberators.

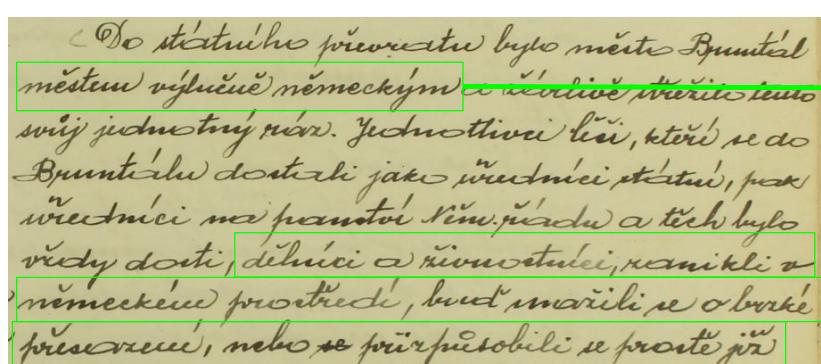
[Reference to the source chronicle] ...



A handwritten document in Czech, likely a historical chronicle. The text discusses the liberation of a village by the Red Army. It mentions that the German garrison in a house was set on fire, resulting in 25 soldiers' deaths. Civilians were also shot during the liberation.



semANT - Content tagging and discussions



Do státního převratu bylo město Bruntál městem výlučně německým a život v městě byl svij jehožto názv. Jednotlivci byli, kteří se do Bruntálu dostali jako výzvanci státu, pak výzvanci nejprve s různými a těch bylo vůdce došli, dělají a živostní, zanikli a německého prostředí, kdež mohli se obohatit, nebo se přizpůsobili se prostřejí

situation in pre-war border towns

Czech inhabitants of Germanized regions

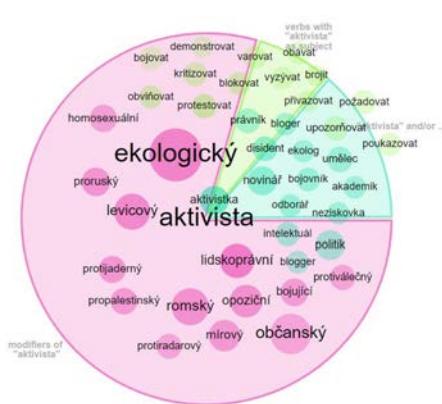
Discussion:

- [Franta Novák] The question is who wrote this text, if he had his own agenda and if we can take this seriously.
- [Anežka Hružová] The chronicler was local teacher Pavel Trk, who previously worked at a German school and he does not express any sentiments against Germans in the rest of the chronicle.

semANT - Visualization of topics, relations, and dynamics

Working with results of automated topic discovery and topics tagged by users

Topic summarization, topic relations and grouping, development of topics in time



OmniOMR – optical music recognition using machine learning for digital libraries

Enable search within musical notation and using musical notation; i.e. automatically:

- Detect parts of pages containing musical notation and determine their type
- Extract musically semantic information (pitch, duration, order of tones) from the detected regions
- Index the results in the database and develop a user-friendly, flexible search in notation data.

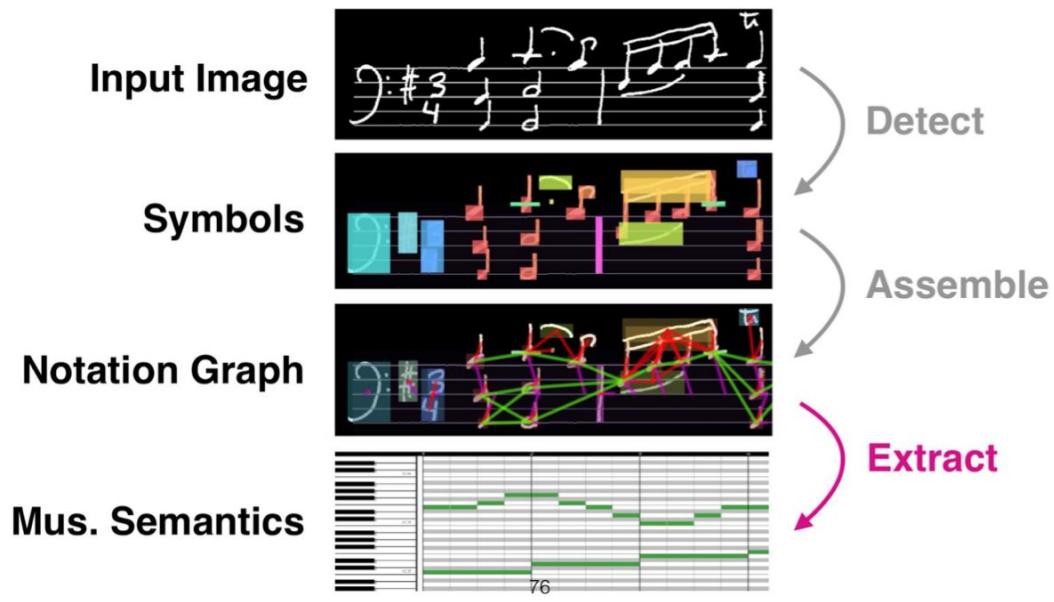
Partners:



FACULTY
OF MATHEMATICS
AND PHYSICS
Charles University



OmniOMR – optical music recognition using machine learning for digital libraries



Our Goal

Improve the search capabilities of digital libraries

Thank you!

Petr Žabička, petr.zabicka@mzk.cz

Michal Hradiš, ihradis@fit.vutbr.cz (PERO, Smart Digiline, Orbis Pictus, semANT)
Jan Hajíč jr., hajicj@ufal.mff.cuni.cz (OmniOMR)

bit.ly/peroilide23



Post-publication curation

Curating without an institutional repository

Kevin Michael Leonard
University of Ghent
KevinMichael.Leonard@ugent.be

POST-PUBLICATION CURATION: CURATING WITHOUT AN INSTITUTIONAL REPOSITORY

Kevin Leonard - Evelien Dhollander

A Tower of Books, A Bunker of Data?



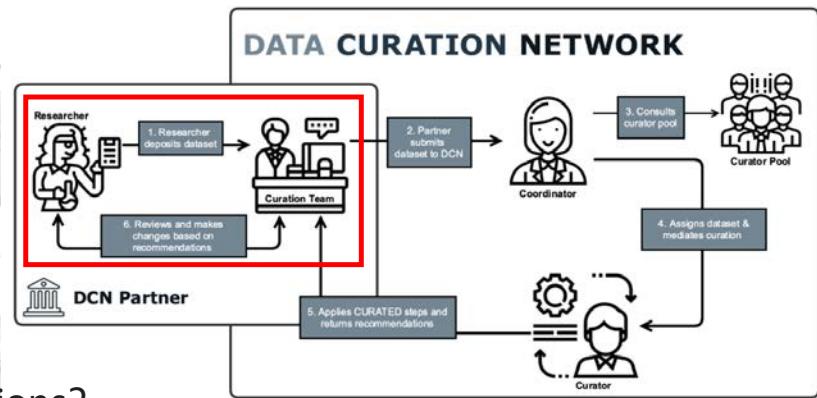
A New Horizon of Research Outputs

- Data and software as valued research outputs
- Systems for curation, sharing, and valorization of data still relatively young
 - Some fields more advanced
 - Others unprepared
- Emergence of new data support roles
 - Make data FAIR

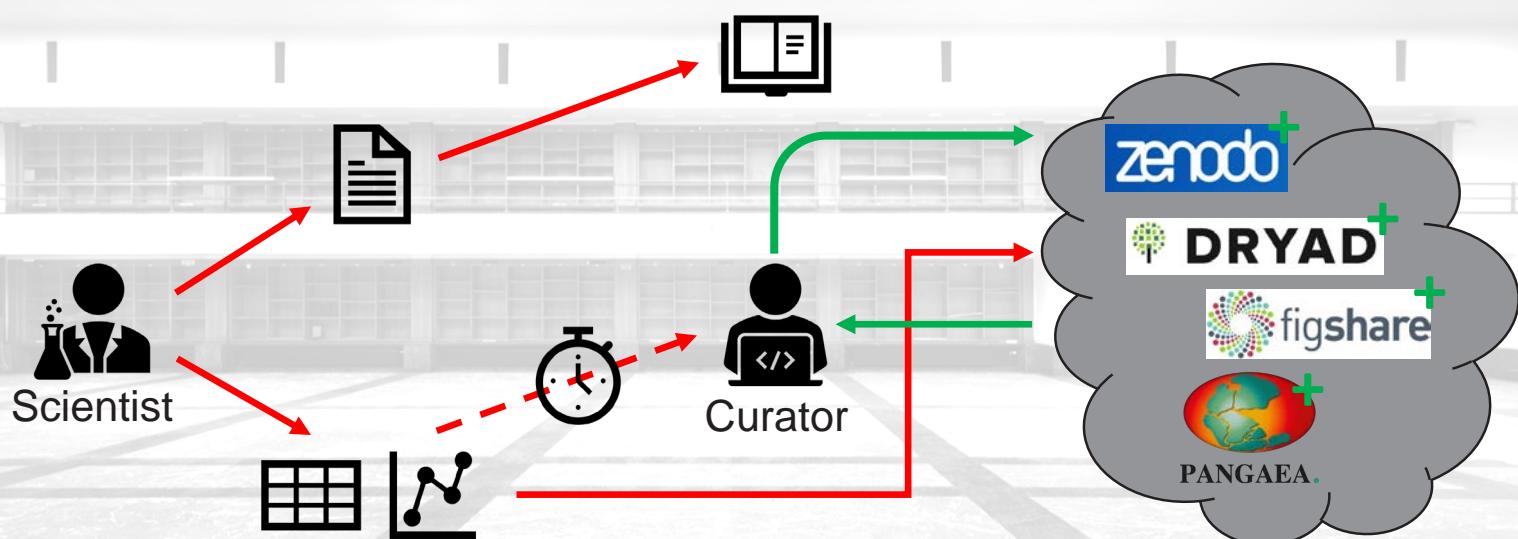


Data Curation Requires Datasets

- Data curation: adding value to (meta)data for long-term preservation
- Imagined (ideal) workflow:
 1. Researcher provides data to curator for curation
 - Voluntary submission
 - Automatic part of ingest in institutional repository
 2. Curator makes changes and recommendations
 3. Data is put online for long-term preservation
- Is that realistic for many institutions?



What Often Happens to Datasets, Really?



Proposed Workflow for Post-Publication Curation

1. Find datasets online
 - Employ existing data linking architectures
 - Use repository APIs
2. Produce (meta)data augmentation plan for discovered datasets
 - Develop plan based on current best practices for FAIR metadata
 - Recommend changes that maintain existing DOI networks
3. Provide researchers with an easily actionable curation plan



Step 1: Where Are The Datasets?

- Difficulties:
 - Datasets are broadly distributed
 - Affiliation information is not located in a consistent location (or format!)
 - Existing data linking systems (e.g., Scholix, DataCite) have limited coverage
- Solution:
 - Use repository APIs to search for individual datasets
 - Search outside of the <creator><affil> field

A screenshot of the zenodo Developers REST API interface. It features a large blue header with the 'zenodo' logo and 'Developers' text. Below the header is a search bar with a magnifying glass icon and the word 'Search'. At the bottom is a blue button labeled 'REST API'.



Repository Search Code

- Python code to search for institutional records
- searchQuery can include multiple items
 - Universiteit Gent
 - UGent
 - Ghent University
 - 00cv9y106 (ROR id)
- Saves DOIs of all datasets to csv
 - Can use OAI-PMH to extract more metadata information
- Focused on several popular repositories, easily extended
 - Zenodo, OSF, Dryad, Figshare, PANGAEA

<https://github.com/UGent-Data-Stewards/DatasetRepositorySearch>

```
currentPage = 'https://zenodo.org/api/records'
while(currentPage):
    response = requests.get(currentPage,
                           headers=headers,
                           params={'q': searchQuery,
                                   'type': 'dataset',
                                   'access_token': ACCESS_TOKEN})
    response.json()
    pythonObj = json.loads(response.text)
```

Step 2: What To Do With What You've Found

- Repositories often allow metadata fields to be edited
 - WITHOUT triggering the creation of a new version (and therefore a new DOI)
- Editable fields vary by repository:

STRI

CT

DRYAD

- Editing any metadata fields creates a new version

LENIEN

T

zenodo

- Most fields can be edited
- Title, authors, abstract

Develop Recommendation Plan

- Is the title clear?
- Are keywords provided?
- Are there links to related publications?
- Do the authors have linked ORCIDs or affiliations?
- Is there sufficient documentation in a README?
 - Can this information be provided in <description descriptionType="Abstract">?



Step 3: Communicating the Recommendations

- Implementation relies on participation of the researcher
 - Curation plan must be easily actionable with clearly articulated benefits

Metadata Field	Current Value	Recommended Changes	Rationale
Title			
Abstract			
...			

- Reduce burden on researcher to interpret instructions



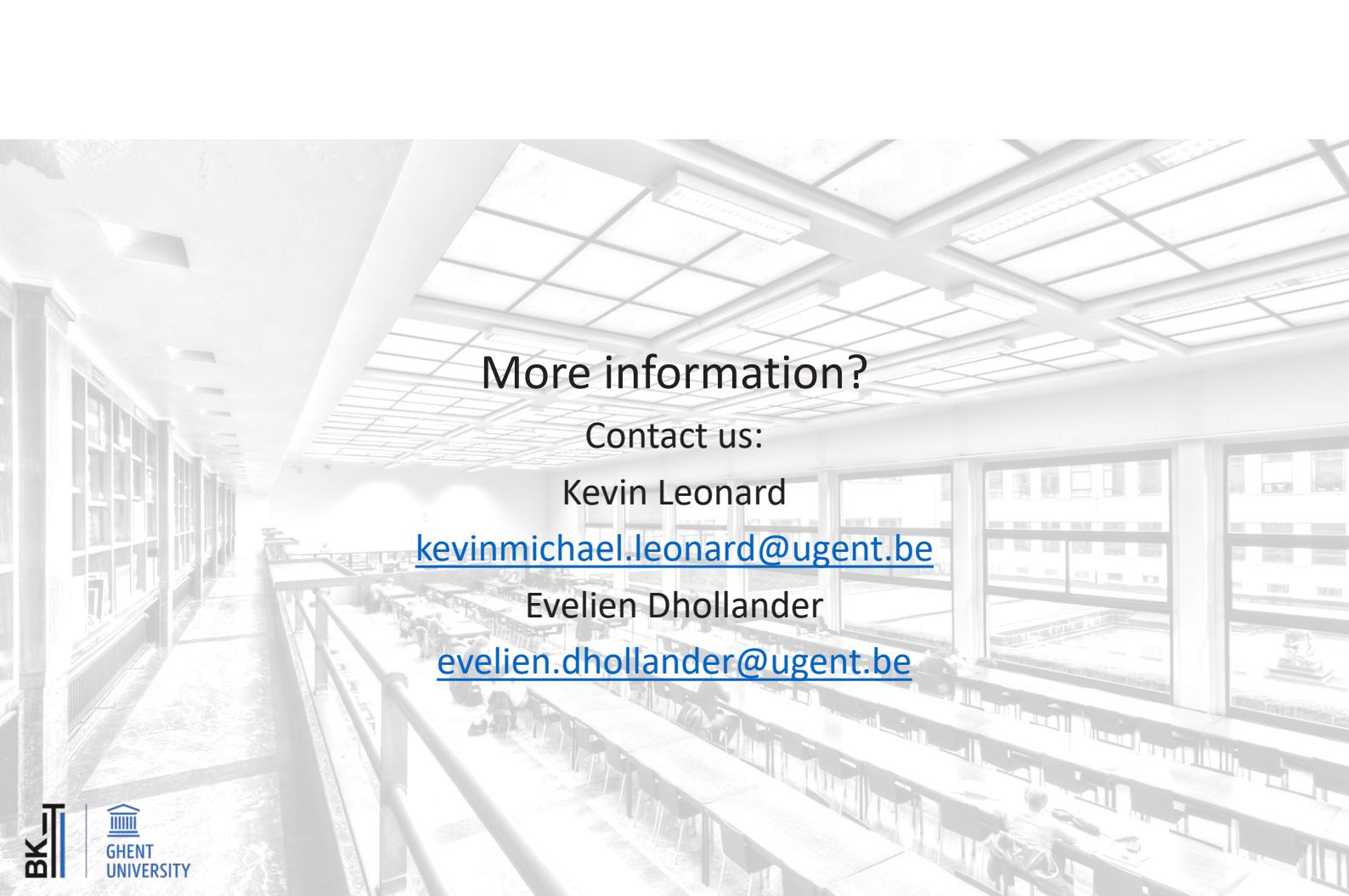
Example Outputs

- Currently, the code harvests >2000 total records
- Frequently encountered issues:
 - Datasets not linked to publications
 - No direct contact information provided
 - Missing keywords

Source	Number of Records Found
Dryad	209
Figshare	447
OSF	236
Pangaea	724
Zenodo	710

Conclusions: Curation after Publication

- Relatively simple method to provide value to published datasets
- Benefits even if author declines to make recommended edits:
 - Helps institution find their research outputs
 - Provides researchers with FAIRness recommendations that they can implement for future datasets
 - Communicates the existence (and utility!) of data support staff



More information?

Contact us:

Kevin Leonard

kevinmichael.leonard@ugent.be

Evelien Dhollander

evelien.dhollander@ugent.be



DL4DH

Digital Libraries for Digital Humanities

Martin Lhoták
Library of the Czech Academy of Sciences
lhotak@lib.cas.cz



DL4DH

Digital Libraries for Digital Humanities

Martin Lhoták
Library of the Czech Academy of Sciences
ILIDE, Jasná, Slovensko
24. 4. 2023

DL4DH project – basic information

- **DL4DH – development of tools for effective utilization and mining of data from digital libraries to reinforce digital humanities research**
- Research project funded by the program of the Ministry of Culture of the Czech Republic NAKI 2020-2022
- 13.5 million CZK (0,45 million EUR)
- Cooperation between Library of the Czech Academy of Sciences, National Library of the Czech Republic and Moravian Library in Brno
- The aim of the project is to enable data mining of the content from digital libraries operated in the Kramerius system for humanities-oriented research

CZ digital libraries landscape

- More than 40 digital libraries are operated in Kramerius system
- 700 000 digital documents
- More than 160 million pages
- Strong standardization
- Czech Digital Library – the national aggregator
- 3 libraries are members of LINDAT/CLARIAH-CZ – national infrastructure connected to Dariah, CLARIN and EHRI (part of ERIC - European Research Infrastructure Consortium)

Kramerius Registry

	Documents	Pages	Recommended	Books	Newspapers	Sound Recordings	Maps	Graphics
	287 086	73 059 862	23	244 863	4 910	682	13 780	18 871
	245 176	71 602 217	31	235 488	4 782	0	3 930	0
	271 134	66 625 785	25	229 542	3 993	683	13 704	18 880
	271 112	66 081 634	0	229 375	4 026	682	13 700	18 872
	289 926	58 771 488	0	231 114	3 252	0	472	50 681
	233 602	53 855 887		230 960	2 642	0	0	0
	8 118	4 685 024	15	7 869	245	0	1	0
	2 187	3 623 198	26	1 259	215	0	586	15
	2 594	2 849 043	33	1 008	312	0	997	0
	3 527	2 271 521	0	3 315	203	0	1	0
	6 173	2 206 892	15	1 126	254	0	47	4 564
	5 089	2 069 451	22	4 707	381	0	0	0
	8 166	2 052 631	0	6 953	322	0	0	891
	4 299	1 783 474	10	4 190	109	0	0	0



CZ digital libraries interface

- Enables searching
- Enables browsing
- Enables reading
- Enables downloading individual publications
- Supports METS ALTO for fulltext
- Open API
- ...
- No batch export
- No recognized entities in text
- Limited metadata and OCR

Výsledky: 1 - 60 54410

Aktivní filtry
Pouze veřejné

Dostupnost
Pouze veřejné 54410
Pouze neveřejné 200327
Vše 254737

licence
Díla nedostupná na trhu 2027

Typ dokumentu
Knihu 23191
Noviny a časopisy 698
Grafika 18819
Mapa 9515
Hudebniny 1228
Zvukové nahávky 167
Archiválie 548
Rukopis 240

Klíčové slovo
Molova mapová sbírka 12316
Mapové sbírky 12315
Historické budovy 4480
Vojenská topografie 3001
Archivní fondu 912

Hledat ve veřejných dokumentech digitální knihovny

Razení: Nově přidané

Slovenské besednice Tajovský, Jozef Gregor 1918 Knihu	Řeč professora Antonína Bělohoubečka, t.č. rektora c.k. Bělohoubek, Antonín 1895 Knihu	Fiesko der Salamikräber: Ein musikalischs Quodlibet in zwey Gleich, Joseph Alois, 1772-1841 1813 Hudebniny
Costantino: Drama Per Musica. Da rappresentarsi nell'imperiale Lotti, Antonio, asi 1667-1740 1716 Hudebniny	Orpheus und Euridice, oder: So geht es im Olympus zu: Eine Meisl, Karl, 1775-1853 1813 Hudebniny	Trois Fugues, Oeuv. 64: à la Camera : Pour le Piano-Forte a Dusík, Jan Ladislav, 1760-1812 1800 Hudebniny
Cinquante fanfares pour deux cors, oeuvre 1 Harton 1800 Hudebniny	Graduale Prima Messe in Sacra Nocte: a Canto, Alto, Tenore, Mášek, Václav Vincenc, 1755-1831 1823 Hudebniny	Preiss dir Gottheit! durch alle Himmel & c: Splendente te, Deus Mozart, Wolfgang Amadeus, 1756- 1800 Hudebniny
Brathia di Campagnia: a Clavi Cembalo, Due Oboe, Due Corni, Schimpke, Christoph, 1725-1789 [1760?] Hudebniny	Variacec op 19: pro flétnu a smyčcový kvartet Bernardi, Franz, 1767-[1800?] Hudebniny	Grand Trio pour Le Clavecin ou Pianoforte: d'un Violon et Hummel, Johann Nepomuk, 1778-[1830?] Hudebniny

Hledat v dokumentu

1 z 92 stránek

[1a]	[1b]	[1c]
[2]	3	4
5	6	7
8	9	10
11	12	13

LIDOVÁ KNIHOVNA

JOSEF GREGOR-TAJOVSKÝ:
Slovenské besednice

Cena K 1—

Slovenské besednice

Autor
Tajovský, Jozef Gregor
1918

Nakladatelství
V Praze: Ústřední dělnické knihkupectví a nakladatelství (Ant. Švábený), 1918

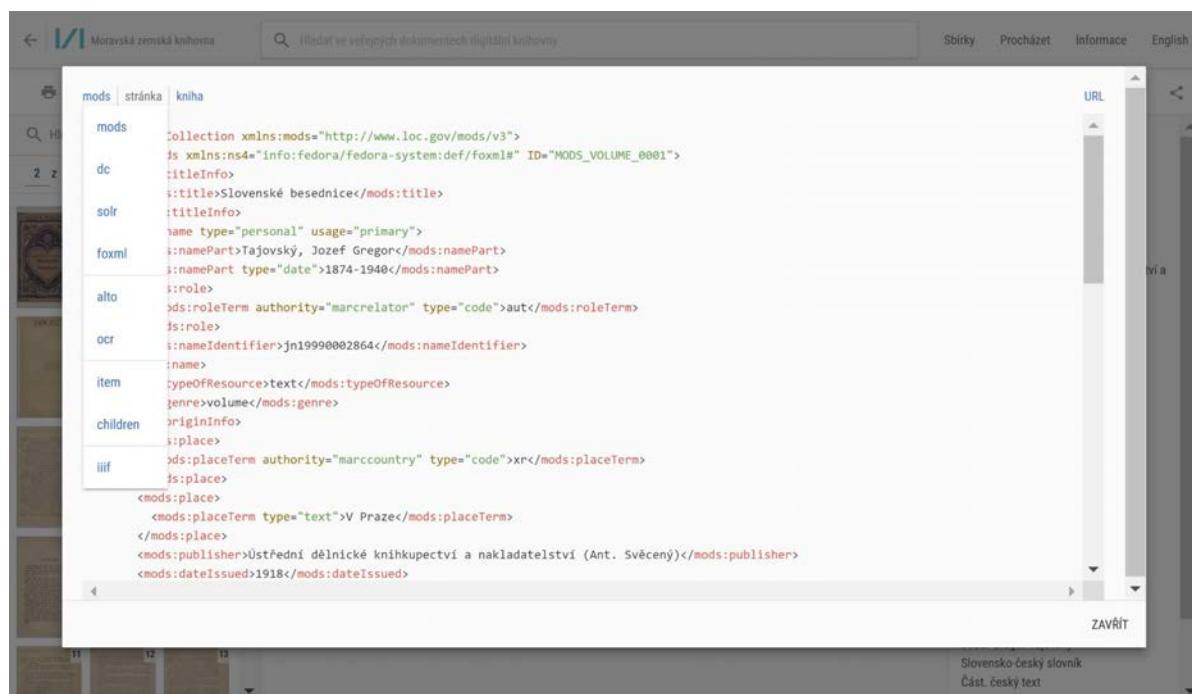
Typ dokumentu
Knihu

Jazyk
Slovenština
Čeština

Místo uložení
Národní knihovna České republiky - Knihovní fondy a služby
Signatura: I 018821

Fyzický popis
Rozsah: 86 s.; 17 cm
nekalitní a poškozená předloha

Poznámky
Josef Gregor-Tajovský
Slovensko-český slovník
Část. český text



The screenshot shows a digital library interface for the Moravská zemská knihovna. The top navigation bar includes links for 'Sbírky', 'Procházet', 'Informace', and 'English'. The main content area displays an XML document representing a book record. The XML code is as follows:

```

<mods xmlns:mods="http://www.loc.gov/mods/v3">
  <is xmlns:ns4="info:fedora/fedora-system:def/foxml#" ID="MODS_VOLUME_0001">
    <titleInfo>
      <i:title>Slovenské besednice</i:title>
    </titleInfo>
    <name type="personal" usage="primary">
      <i:namePart>Taťovský, Jozef Gregor</i:namePart>
    <i:namePart type="date">1874-1940</i:namePart>
    <role>
      <ds:roleTerm authority="marcrelator" type="code">aut</ds:roleTerm>
    <is:role>
      <i:nameIdentifier>jn19990002864</i:nameIdentifier>
    </name>
    <item typeOfResource>text</item>
    <genre>volume</genre>
    <children>
      <originInfo>
        <i:place>
          <ds:placeTerm authority="marccountry" type="code">xrc</ds:placeTerm>
        <is:place>
          <mod
            <mod:place>
              <mod:placeTerm type="text">V Praze</mod:placeTerm>
            </mod:place>
            <mod:publ
              <mod:publ
                <mod:place>
                  <mod:placeTerm type="text">Ustřední dělnické knihkupectví a nakladatelství (Ant. Svěcený)</mod:placeTerm>
                </mod:place>
                <mod:publ
                  <mod:publ
                    <mod:place>
                      <mod:placeTerm type="text">1918</mod:placeTerm>
                    </mod:place>
                  </mod:publ>
                </mod:publ>
              </mod:place>
            </mod:publ>
          </mod:place>
        </is:place>
      </originInfo>
    </children>
  </is>

```

The XML code describes a book titled 'Slovenské besednice' by Jozef Gregor Taťovský, published by Ustřední dělnické knihkupectví a nakladatelství (Ant. Svěcený) in 1918, located in Prague.



DL4DH – main outputs

- **DL4DH feeder** - interface for data mining from digital libraries operated in the Kramerius system
- **TEI converter** - conversion and export of data and metadata from the Kramerius system to TEI format
- **Methodology** of data preparation from digital libraries for digital humanities research
- **Use case** - example of use of data from digital libraries in field of religionistics

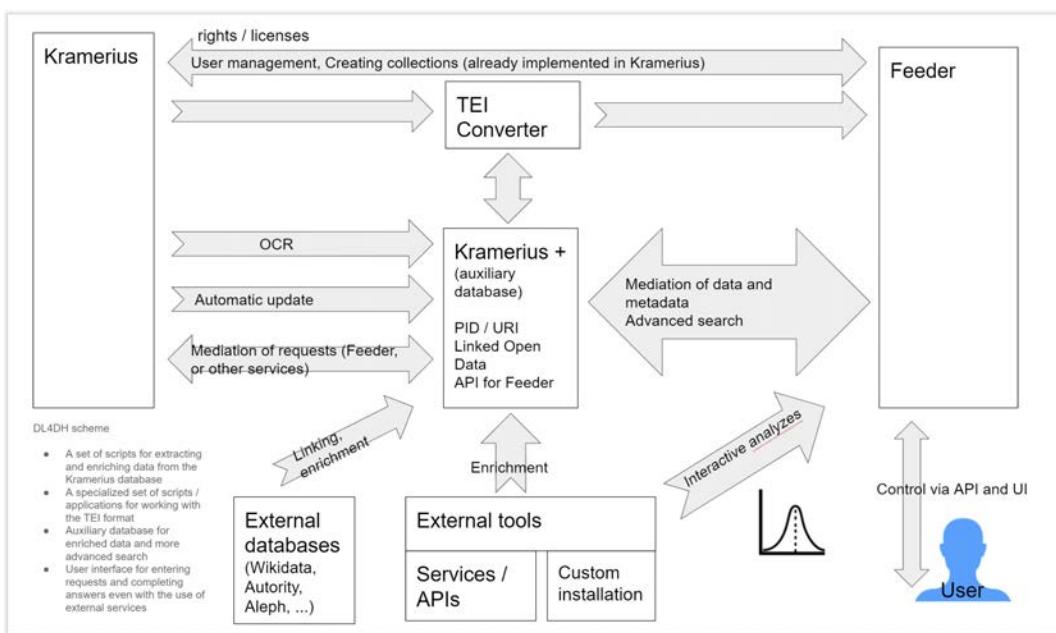
DL4DH - team

- Libraries – National Lib., Moravian Lib., Cz. Acad. of Sci. Lib.
- Researchers
 - Linguistics
 - Religious studies
 - Archeology
 - Librarianship
 - Philosophy
 - Informatics / Computer science
 - ...
 - Digital Humanities

DL4DH – basic architecture

- The main output of the developed solution is the **DL4DH Feeder** software operated as a web service which is intended for direct use by scientists.
- **Kramerius+** database serves to store metadata and enriched text data that cannot be retrieved from Kramerius digital library.
- The solution includes possibility of connecting the DL4DH Feeder web service and the Kramerius+ system with **external databases and tools for purpose of the enrichment** of metadata and text data or in the visualization of data extracted from the Kramerius system.
- **TEI Converter** will allow you to export data from the Kramerius system and the Kramerius+ database in TEI format

DL4DH – diagram



DL4DH Kramerius+ database

- The Kramerius+ extension module / database serves to **store enriched text data and metadata**, in addition to content stored in the Kramerius system (for example, lemmatization and automatically recognized entities in full texts, records from library systems, links to authorities and bibliographic records, etc.).
- Kramerius + stores selected data from **archival packages** which are not in Kramerius system
- This data enables **advanced search and filtering** of digital documents or **export of enriched content** for further use.
- Kramerius + is an **external module** of the Kramerius systém has its own **API** and communicate with both the Kramerius system and the DL4DH Feeder.

External tools for enrichment UDPipe and NameTag

UDPipe 2

<http://lindat.mff.cuni.cz/services/udpipe/>

lemmatization and morphology analysis (multilingual)
position tags and universal dependencies
web user interface + REST API, CC BY-NC-SA license

NameTag 2

<http://lindat.mff.cuni.cz/services/nametag/>

entity recognition (multilingual)
web user interface + REST API, CC BY-NC-SA license
TEI elements coverter
(<https://github.com/LIBCAS/DL4DH/issues/19#issuecomment-848394546>)

DL4DH Feeder

- User management and authentication via Shibboleth
- Searching in recognized entities (persons, geographical names, etc.)
- Information on data quality and enrichment, or on the quality of individual versions
- Export formats: TEI, TSV, CSV, JSON, plain text, PAGE, ALTO + RAW export (data in original form)
- Exporting text and image data
- Displaying list of publications in tabular form
- Visualizing results of searching in chart
- Parallel view of 2 publications
- History of queries and list of performed exports



AKADEMIE VĚD ČR

DL4DH Feeder

← → ⌂ feeder.nkp.cz/search?availability=PUBLIC

Národní knihovna České republiky
DL4DH Feeder

Search in DL4DH Feeder (word or filter the results)...

Results: 1 - 15/ 43605

Advances Search

Active filters

- ✓ Public
- Uložit filtry

Accessibility

- Public: 43605
- Private: 0

Enrichment

- Enriched only: 26
- Not enriched only: 43579

Document Type

- Book: 39381
- Map: 2254
- Newspaper: 1140
- Sheet Music: 827

Keyword

- vojenská topografie: 1984
- dějiny: 574
- piano music: 298

Page limit: 15 ▾

1 2 3 4 5 6 7 8 9 10 11 ... 2907 >

The screenshot shows a search results page for the DL4DH Feeder. The left sidebar contains various filters: Active filters (Public), Accessibility (Public, Private), Enrichment (Enriched only, Not enriched only), Document Type (Book, Map, Newspaper, Sheet Music), and Keyword (vojenská topografie, dějiny, piano music). The main area displays a grid of 15 document entries, each with a thumbnail, title, and date. The titles include "... budovali jsme republiku", "Jahres-Bericht der achtklassigen Volks-", "Jahres-Bericht des evangelischen Hilfs-", "Jahres-Bericht des Niederösterreichisc...", "Jahresbericht der fachlichen Fortbildun...", and "Jahresbericht der Königl. böhmischen L...". The dates range from 1863 to 1938.



AKADEMIE VĚD ČR

DL4DH Feeder

← → ⌂ feeder.nkp.cz/search?availability=PUBLIC

Národní knihovna České republiky
DL4DH Feeder

Search in DL4DH Feeder (word or filter the results)...

Results: 1 - 15/ 43605

Advances Search

Show more ▾

Geographical names

Media names

Personal names

Complex person names

Complex time expression

Numbers in addresses

Number expressions

Complex address expression

Complex biblio expression

Artifact names

Time expressions

The screenshot shows a search results page for the DL4DH Feeder. The left sidebar is expanded to show a detailed list of entity types: Geographical names, Media names, Personal names, Complex person names, Complex time expression, Numbers in addresses, Number expressions, Complex address expression, Complex biblio expression, Artifact names, and Time expressions. The main area displays a grid of 15 document entries, each with a thumbnail, title, and date. The titles and dates are identical to the first screenshot, showing historical reports and documents from the late 19th and early 20th centuries.



DL4DH Feeder

AKADEMIE VĚD ČR

The screenshot shows a search results page for the DL4DH Feeder. The top navigation bar includes links for 'Search in K', 'Collections', 'About', 'Go to Kramerius', 'Log in', and a language switcher. The main search bar contains the placeholder 'Search in DL4DH Feeder (word or filter the results...)'. Below the search bar, there are filters for 'Search NameTag' and 'Institutions'. The 'Geographical names' section lists various locations with their counts: Aberdeen (2), Afrika (4), Alexandria (2), Alpy (2), Altenburg (2), Amerika (4), Angle (5), Athény (3), Babylon (3), Barcelona (3), Basilej (2), Bavorško (3), Bayreuth (3). The main content area displays a grid of 15 historical documents, each with a thumbnail, title, author, and date. The documents include books, newspapers, and reports from various years and publishers. A navigation bar at the bottom shows pages 1 through 2907.



DL4DH Feeder

AKADEMIE VĚD ČR

The screenshot shows the DL4DH Feeder interface with an 'Export publications' dialog open. The dialog has a 'Count: 9' indicator. It includes a 'Cancel export' button and a 'Check' button for ALTO availability. The 'Export name' field is set to 'export-20230300-195513'. The 'Format' dropdown is set to 'TEXT'. On the left, there is a sidebar with 'Active filters' (Public, Enriched only), 'Uložit filtry', 'Accessibility' (Public, Private), 'Enrichment' (Enriched only, Not enriched only), 'Document Type' (Book, Newspaper), and 'Keyword' (svobodné zdrojátky, jazyk, dílo). The main area shows a grid of 9 selected documents, each with a checkmark and a preview. A navigation bar at the bottom shows pages 1 and 2.

DL4DH – links

- DL4DH Feeder at the National Library <https://feeder.nkp.cz/>
- DL4DH Repository on Github <https://github.com/LIBCAS/DL4DH>
- DL4DH Methodology <https://hdl.handle.net/11104/0335692>
- DL4DH Conference <https://www.lib.cas.cz/akce/odborne/konference-dl4dh-digital-libraries-for-digital-humanities/>
- DL4DH Use case: Biblical Citations in the Press during the First Republic <https://dl4dh.nkp.cz/>

Thank you for attention!

Martin Lhoták
lhotak@knav.cz



US Academic Libraries and Challenges to Fair Use

Heather Briston
University of California San Diego Library
hbriston@ucsd.edu



US Academic Libraries and Challenges to Fair Use

Heather Briston, JD, MSI Uni. of California, San Diego
Director, Scholarship Tools and Methods Program

UC San Diego
LIBRARY

US Copyright Law and Fair Use?

UC San Diego
LIBRARY

United States Copyright Law - Title 17

- Constitutional foundation
 - Article 1, Section 8
 - Title 17 - last major overhaul 1976 (1978)
 - Regularly amended
 - Berne Convention 1989

UC San Diego
LIBRARY

Fair Use - Title 17, Section 107

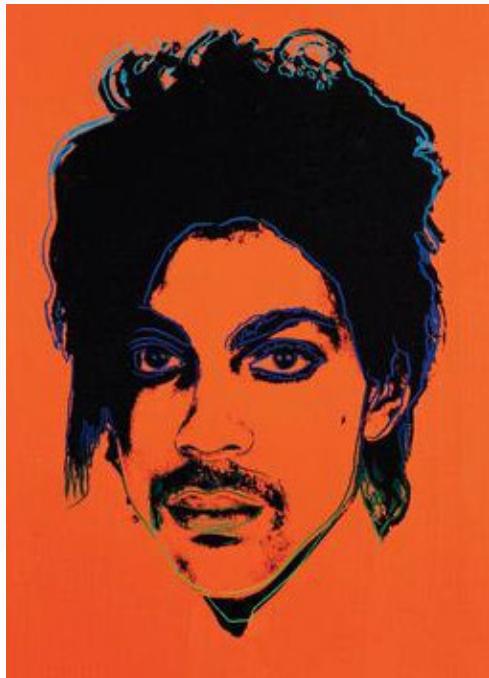
1. The purpose of the use (educational, for-profit)
2. The nature of the work (factual, creative)
3. Amount of the work being used
4. Potential impact of use on the market for original

UC San Diego
LIBRARY

Recent Challenges

UC San Diego
LIBRARY

Andy Warhol Foundation for the Visual Arts, Inc. v. Goldsmith (Docket 21–869)



What is the proper test for whether a work is “transformative” under the first factor of the Copyright Act’s fair use doctrine?

UC San Diego
LIBRARY

Hachette Book Group v. Internet Archive

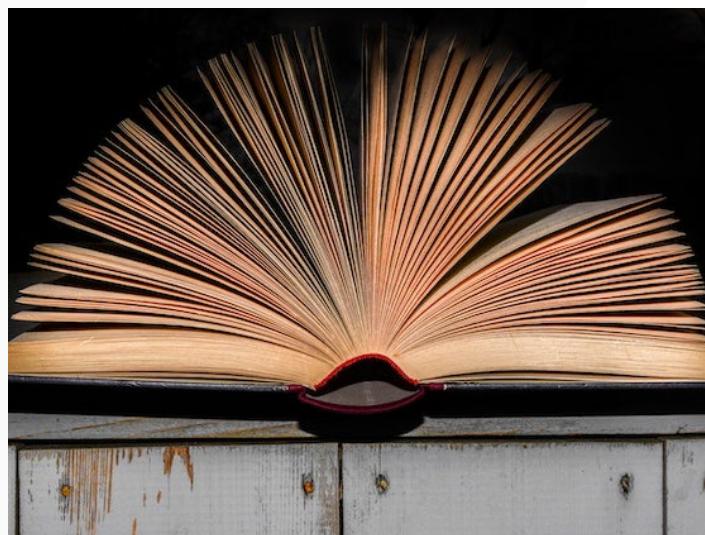


Photo by Skitterphoto

- Issue
 - Controlled digital lending - National Emergency Library
 - Fair use argument from IA
- Found not to be a fair use
- Lower court decision - (SDNY)

UC San Diego
LIBRARY

Generative Artificial Intelligence

- Copyright and human authorship
 - Burrow-Giles v. Sarony - copyright protection for photography
 - Naruto v. David John Slater
- Copyright Office - Zarya of the Dawn
 - Midjourney
 - Compilation / Rights in text
- Unanswered questions
- Training data



Public domain



Public domain

UC San Diego
LIBRARY

Fair Use and Academic Libraries

UC San Diego
LIBRARY

Students, Faculty and Researchers

- ★ Changing instructional output - contextualizing resources within argument and exposition, e.g. digital exhibits, online portfolios. Scholarly output of students
 - Publishing literacy - copyright, ethics
 - Not enough time to do a proper job - connection needs to be made between attainment of a level of literacy and successful completion of project.
 - Most take down notices are for old student digital projects.

UC San Diego
LIBRARY

Digital Collections, Exhibits and Digital Humanities

[Code of Best Practices in Fair Use for Academic and Research Libraries](#), Association of Research Libraries

[Code of Best Practices in Fair Use for the Visual Arts](#), College Art Association

[Statement of Best Practices in Fair Use for Orphan Works for Libraries, Archives, and other Memory Institutions](#), Center for Social Media and Social Impact

[Fair Use & Libraries | Fact Check - Can Librarians Exceed the 1976 Copyright Act?](#), Center for Media & Social Impact

UC San Diego
LIBRARY

Available responses for higher education

- Education - faculty, students, researchers

It takes a village . . .

- Waiting and hoping?
- No flailing and hand waving
- Never forget Section 108 and other exceptions - balancing mission with the law.
- Others?

UC San Diego
LIBRARY



UC San Diego
LIBRARY

Zborník prezentácií a príspevkov z konferencie ILIDE 2023

ILIDE 2023 conference proceedings

Zostavil: Jozef Dzivák

Vydavateľ: Fakulta chemickej a potravinárskej technológie STU v Bratislave

Počet strán: 141

Rok vydania: 2023

Elektronická publikácia

ISBN 978-80-8208-103-2

EAN 9788082081032



ISBN 978-80-8208-103-2

EAN 9788082081032