

Datový repozitář ASEP

Pracovní schůzka

16. 8. 2023

Kdy autor potřebuje uložit data do repozitáře?

- Požadavek **poskytovatele**

Podmínky poskytovatelů týkající se dat - https://openscience.lib.cas.cz/support-2/os_requirements/

- Požadavek **vydavatele**

1. Doplnující materiály k publikovanému článku (Supplementary Material) – [stránky KNAV Open Science](#)
2. Prostřednictvím datového repozitáře
3. Publikace v datovém časopise - [seznam](#)

- Požadavek **autora**

Zavedení systému v uspořádání dat: popis dat a bezpečné uložení souborů, znovupoužití

- Požadavek **pracoviště**

Proč autor ukládá data do DR ASEP, jaká data ukládat do DR ASEP

- supplementary materiál je možné ukládat jako datový záznam
- Autoři se zatím neúčastní projektů od mezinárodních poskytovatelů, kde by bylo v podmínkách ukládat data.
- Požadavky na ukládání dat jsou z iniciativy autorů.
- ústav provozuje interní uložení které neumožňuje metadatový popis.
- Ukládání dat není povinné, vedení ústavu neřeší, co bude s daty, pokud je autor má pouze u sebe a z ústavu odejde. Vědcům je někde **doporučeno** ukládat výzkumná data do ASEP.

Kam, jaká, na jak dlouho?

Podle čeho vybrat repozitář

Příprava dat

- DMP – k čemu je, proč ho vytvářet, když to nikdo nevyžaduje vytvořit?
- Je nějaký nástroj pro tvorbu DMP?
- Jaká data má smysl vůbec ukládat?
- Jak ukládat data s spoluautory?
- Jak dlouho mají být data uložena?
- Jak vybrat licenci?
- Jak pracovat s citlivými daty, mohou se vůbec uložit?



Seminář s přihlédnutím k této problematice?

Data management plány

- Data management plány vytvářejí projektoví manažeři. (Budeme je informovat o nástroji DSW wizard).
- DSW wizard nabízený knihovnou AV ČR (Inspirace pro vlastní vytvoření DMP).

Vznik repozitáře ASEP

2017

Základ – repozitář univerzity v Edinburgu

Zenodo

- Průběžné úpravy (bez praxe)

2023

Úprava repozitáře na základě praxe

Repozitář ASEP je bezpečný a důvěryhodný.

Pro data uložená v OA splňuje (+/-) FAIR principy.

Proč ukládat/neukládat data do repozitáře?

Důvody, proč autoři nechtějí ukládat data do repozitářů

Příklad mylných důvodů

<https://openscience.cuni.cz/OSCI-192.html>

Proč autoři neukládají data do repozitáře ASEP?

- **Využívají oborové repozitáře** - V určitých oborech mají autoři určené vydavateli časopisů do jakých datových repozitářů budou data ukládat.
- **Využívají repozitáře dle zvyklostí svých spolupracovníků**
- **Velké objemy dat**

Uživatelé repozitáře ASEP - role

Autor, osoba pověřená popisem a uložením dat (autor, zástupce skupiny, oddělení, zpracovatel)

Při vytvoření repozitáře se uvažovalo, že data budou ukládat primárně autoři prostřednictvím individuálních účtů.

Zpracovatel

- Potvrzuje žádost vytvoření IU autorů ústavu.
- Zajišťuje formální kontrolu uložených metadat a datových souborů.
- Zveřejnění v on-line katalogu.

KNAV

- Supervizor

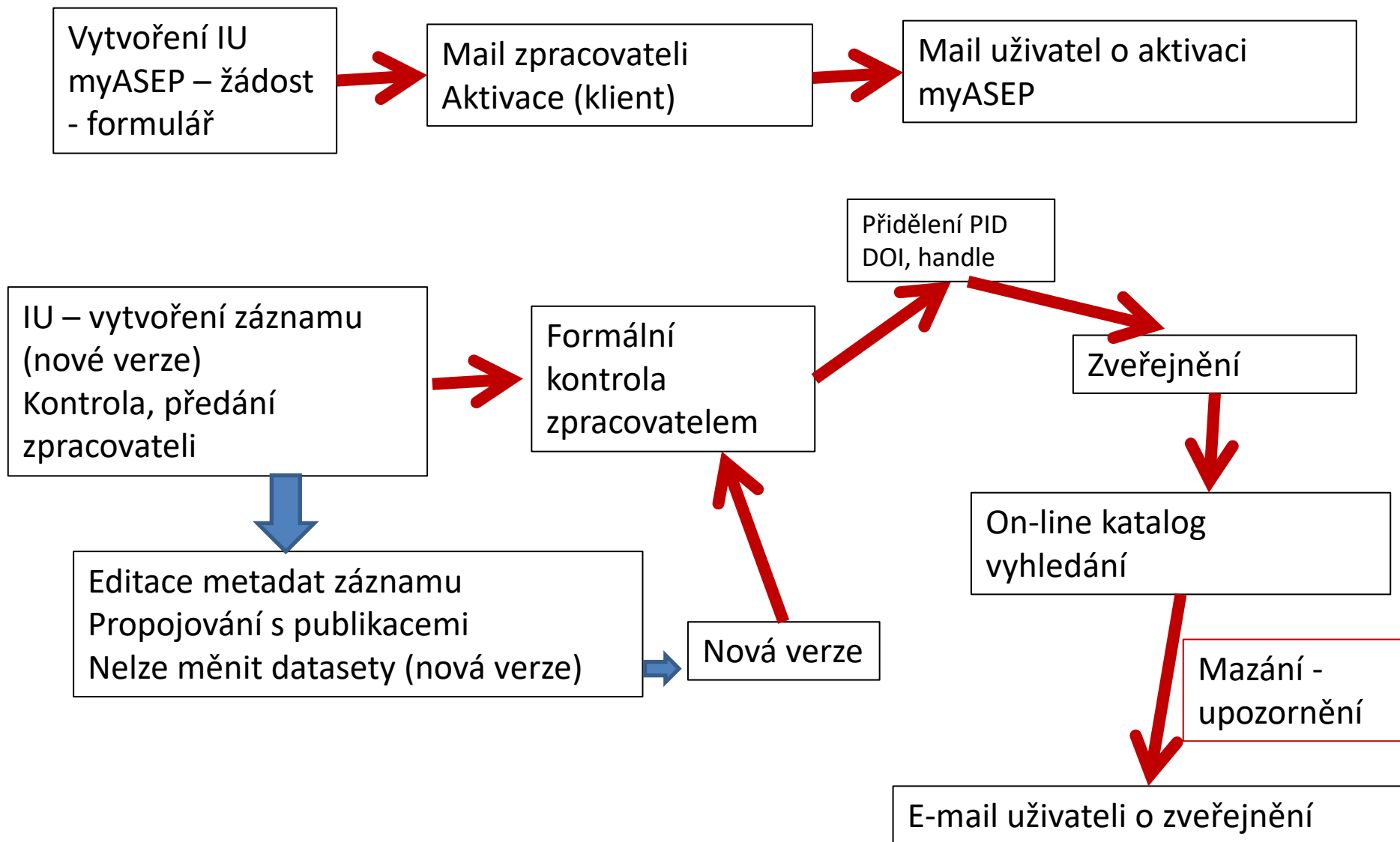
Záznamy v online katalogu

Ústav

Ústav informatiky	12
Ústav biologie obratlovců	4
Fyzikální ústav	2
Knihovna AV ČR	2
Masarykův ústav - Archiv (od r. 2006)	1
Orientální ústav	1
Ústav chemických procesů	1
Ústav fyziky plazmatu	1
— Ústav pro hydrodynamiku	1
Ústav fotoniky a elektroniky	1

Kdo inicioval uložení do repozitáře a kdo byl vkladatelem?

Workflow



1. Vytvoření individuálního uživatelského účtu

Nyní

- Webové prostředí online katalogu, uživatel vyplní formulář. **Mail by měl být ústavní.**
- Zpracovatel obdrží E-mail s žádostí o vytvoření IU, úprava záznamu IU – klient (user)
- Po úpravě – automaticky se zašle **E-mail** s informací o vytvoření IU.
- **Editace polí po odeslání:** KNAV – diskuze o tom, jaká pole je možné měnit po odeslání do IPAC v rámci editace záznamu, a které by nemělo být možné upravovat. Určitě by nemělo být možné měnit licenci a způsob přístupu.

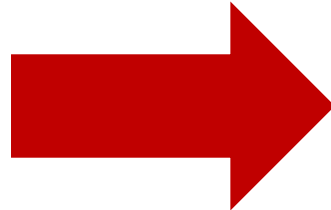
Bude

Administrace přidělení přístupu nebude na zpracovateli.
(využití institucionálního přihlášení shibboleth)

Workflow vytvoření IU



Vytvoření IU
myASEP – žádost
- formulář



E-mail **uživateli** o aktivaci myASEP
E-mail **zpracovateli** o aktivaci myASEP


Problémy s doplněním a schválením?

Co s IU, když autor přestane být zaměstnancem ústavu?

IU autora se musí smazat.

myASEP – individuální uživatel

Nápověda

Nápověda 

Datový repozitář

Nový datový záznam (metadata) 

Rozepsané datové záznamy (neodeslané)

Záznamy předané ke kontrole zpracovateli

Přehledy datových záznamů

Všechny datové záznamy v on-line katalogu

- Vytvoření nového záznamu
- Uživatel se k záznamům **může** „vracet“ dokud je nepředá ke kontrole může **editovat, mazat záznam, přidávat i mazat datasety**.
- Přehled záznamů ke kontrole, které nejsou ještě vyhledatelné v katalogu, IU **s těmito záznamy již nic nemůže dělat**.
- Vyhledatelné záznamy v online katalogu. IU může **editovat metadata, vytvořit novou verzi dat**.

Pole, která nesmí IU editovat.


myASEP - zpracovatel

Záznamy ke kontrole – aktivace odkazu, pokud se tam objeví záznam k formální kontrole

- Authority
- Klíčová slova v ČJ
- Uložení licenční smlouvy

V případě nesrovnalostí osloví vkladatele.

Datový repozitář

Nový datový záznam (metadata) 

Rozpracované datové záznamy zpracovatele

Datové záznamy autorů ke kontrole

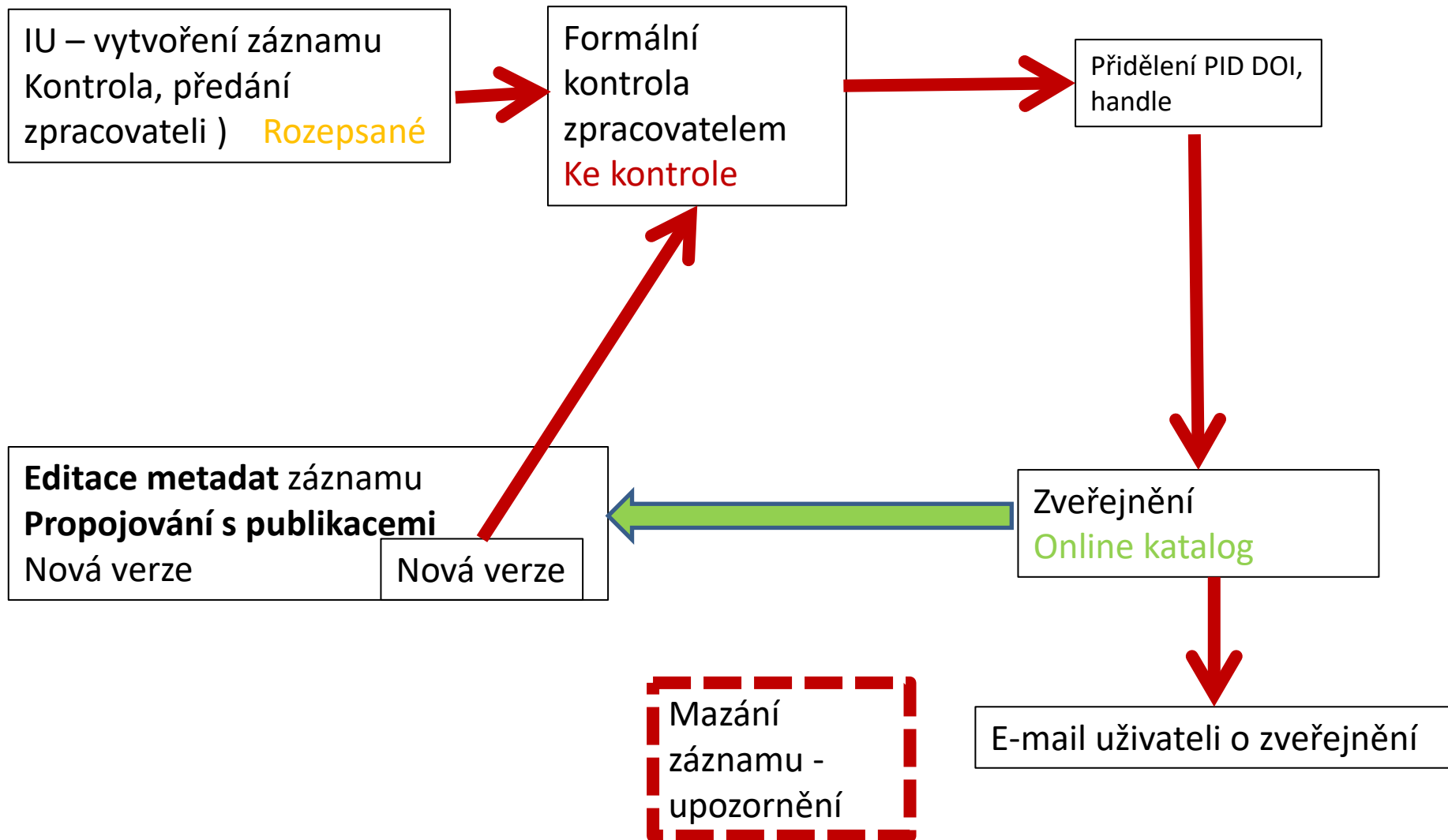
Přehledy datových záznamů

Odeslané datové záznamy v online katalogu

Zpracovatel - vkladatel

Práva stejná jako uživatel individuálního účtu.

Workflow - IU



2. Vyplnění formuláře, odeslání zpracovateli

- Kontrola záznamu není automatická.
- Změna logiky tlačítka odeslání.
- Při odeslání je automatická kontrola vyplnění povinných polí a uložení datasetu resp. odkazu na dataset.
- Záznam je předán až jsou odstraněny všechny chyby.

3. Datová struktura, pole

Povinná pole – změna zpracovatel

Nepovinná pole – změna vkladatel

Pole, která by neměl **nikdo** měnit:

Rok, PID, způsob zveřejnění, licence

Mazání – supervizor KNAV

Bude

Nová pole (vkladatel a jeho role, rozšíření rolet výběru)

Kontrola vyplnění - bude doplňovaná o kontroly autorit

4. Licence

Pokud sdílíte výzkumná data, je dobré [opatřit je vhodnou licenci](#), aby uživatelé věděli, **jak s nimi mohou nakládat**.

Licence [Creative Commons](#)

Jiná licence – pro OA i data přístupná jen na vyžádání

Žádná licence



Licence [Public Domain Mark](#) - Označení volného díla 1.0

Public Domain Mark je nástroj, který organizace Creative Commons vyvinula za účelem označení autorských děl, které již nejsou předmětem ochrany autorských práv. Může se jednat o díla, u kterých uplynula lhůta trvání autorských práv (70 let po smrti autora - „volná díla“), dále díla, která nejsou předmětem ochrany autorských práv ze zákona (např. úřední dílo) nebo díla, které nejsou výsledkem jedinečné tvůrčí činnosti.

Možnost poskytovat stejná data pod více licencemi? (například zdarma pro nekomerční využití a pro komerční využití za úplaty).

Embargo U sdílení dat platí princip „as open as possible, as closed as necessary“, tedy co nejotevřenější, uzavřená dle potřeby. Pokud chcete výzkum patentovat nebo data jinak komerčně využít, pak můžete sdílení odložit – využít embargo.

Výzkumná data se obvykle sdílí současně s publikací, ke které se vážou, pokud chcete, v licenci uveďte, že data mají být **citována**.

Dotaz Licence

- Pokud není licence žádná - licence se nebude uvádět

Při uvedení licence CC musí být záznam veřejně přístupný.

Licence

Nyní

Jedna licence ke všem datasetům

Možná úprava:

Výběr **různých licencí** k souborům

Výběr **více licencí k jednomu** souboru (např. veřejně přístupné v případě nekomerčního užití, licence pro komerční využití)

Softwarové licence MIT a GNU

Nahrát dataset


Nahrát dataset k záznamu: 0566374


VESELÁ, Lenka. (2023). Catalogus Bibliothecae Templi Cathedralis Stregnesensis - skeny. [Dataset]. ASEP.

<https://hdl.handle.net/11104/0337714>

Datové soubory v ASEP: Katalog.zip

 DOI DOI: 10.57680/ASEP.0566374

 Přidat soubory

 Spustit hromadné nahrávání

 Odstranit nenahrané soubory

Soubory s různou licenci

5. Uložené datasey na **jiném úložišti, datasey s otevřeným přístupem**

Použití:

Úložiště nemá možnost popisu, vytvoření metadat v ASEP, odkaz na záznamy.

PID jen handle

6. Uložené datasety na jiném úložišti, datasety na vyžádání

Úložiště nemá možnost popisu, data nejsou veřejně ke stažení. – Úložiště UI.

Licence jiná, předat při vyžádání.

PID - handle

Dotaz k přístupu

Bylo by možné dataset s jpeg soubory a jednoduchým popisem toho, co je v datasetu uloženo nahrát tak, aby zpřístupnění bylo na vyžádání? (z hlediska autorského zákona není možné zpřístupnit, ale autoři chtějí mít k souborům přístup pro badatelské účely) **ano**

Nebo by bylo možné nastavit určitý okruh osob, komu by to šlo zpřístupnit rovnou? Z hlediska zákona pro nás není přípustná varianta zpřístupnit pro ústav, chceme zpřístupnit jen výzkumnému týmu.

Zpřístupnění pro tým

Ve formuláři by byli vyjmenováni autoři, pro které by to mělo být zpřístupněno.

Soubor:

+ Vyber soubor

Doporučený formát: *.pdf, *.jpg.

Zakázaný formát: *.zip, *.arj, *.dll, *.com, *.bat.

Přístup:

Veřejně nepřístupný

Datum zpřístupnění:

Veřejně nepřístupný

Přístupný pro ústav

Veřejně přístupný s embargem

Veřejně přístupný

Archív

Verze:

Přístupný pro skupinu

Klíč:

...

..

Verzování

- Nová verze se vytváří v případě kdy dochází ke změně v uložených souborech ne pouze v metadatech.
- Nové verzi je přiděleno DOI a handle.

7. Verzování - dotaz

- **Verzování:**
- Zajímá nás, jestli funguje provázanost mezi jednotlivými verzemi datasetů, a pokud ano, tak jakým způsobem. Např. V repozitáři ZENODO má každá verze vlastní DOI, zároveň ale lze přes jedno DOI odkazovat na všechny verze současně – např. zde: <https://doi.org/10.5281/zenodo.4391095>.
- Vracím se tedy ještě k diskusi o výhodách takové verze DOI, které odkazuje na všechny verze a zejména vždy vede na tu poslední aktuální a nejnovější.

Smysl společného DOI spočívá v tom, že uživatelé mají různé perspektivy. Někteří odkazem na naše databáze především stvrzují to, že naše služby využívají. Společné DOI bude dobré v situaci, kdy není nutné citovat konkrétní dataset, ale spíš chceme prostřednictvím citace zdůraznit hledisko vývoje našich dat. Také budeme chtít z našeho webu prolinkovat přes DOI do repozitáře a pro tento účel by byl lepší aktuální DOI link než každý půlrok tvar DOI na našem webu (či jinde) aktualizovat.

8. Zobrazení záznamu a indexace v IPAC, vyhledání záznamu

Dotazy

- **Indexace google:**

Jak je zajištěna indexace záznamů v repozitáři službou GOOGLE? Tento dotaz vztahuji nejen na Datový modul, ale na ASEP obecně, protože když zkouším jen náhodně vyhledávat některé publikace, které mají záznam v ASEPU, tak mě Google nikdy spojení do ASEP nenabídne. Respektive nějaký záznam z r. 2020 to našlo, novější ale nic.

- V řešení

- **Implementace do ASEP:citace dat**

Citace dat - možnost využít DOI Citation Formatter

<https://citation.crosscite.org/>

(zadá se DOI a vybere citační styl)