

Očekávané služby Národní datové infrastruktury (NDI)

Cílený stav NDI v letech 2024/2025

Michal Růžička, e-INFRA CZ, ruzicka@ics.muni.cz

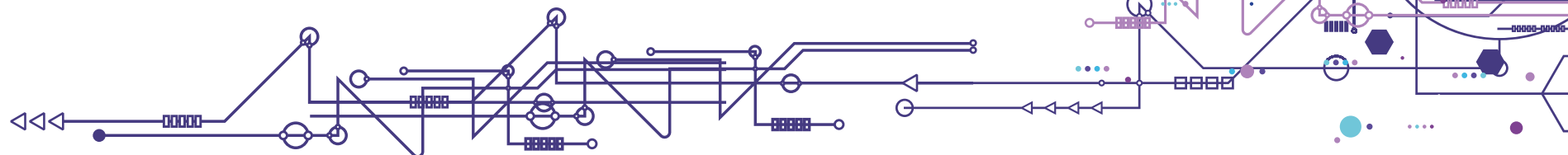
EOSC Roadshow 2022

Životní cyklus výzkumných dat

- Jaká **data (znovu) používáte**
 - včetně licencí, které vám to umožňují,
- jaká **data generujete** a jakým způsobem,
- **kde je ukládáte, zálohujete, dlouhodobě uchováváte,**
- jak je **trvale a jedinečně identifikujete,**
- jak je **zpracováváte,**
- jak je **analyzujete,**
- kde je **zveřejňujete a sdílíte,**
- kdo tohle všechno **zaplatí;**
- **o čem data skutečně jsou,**
- **k čemu jsou data (ne)vhodná,**
- **kdo může data znovu použít,**
- jaká konkrétní data podporují vaše výsledky,
- jak je použít pro **opakování** vašich **experimentů** atd.



Zdroj: ELIXIR RDMkit, <https://rdmkit.elixir-europe.org/>



Jaké služby můžete od NDI očekávat

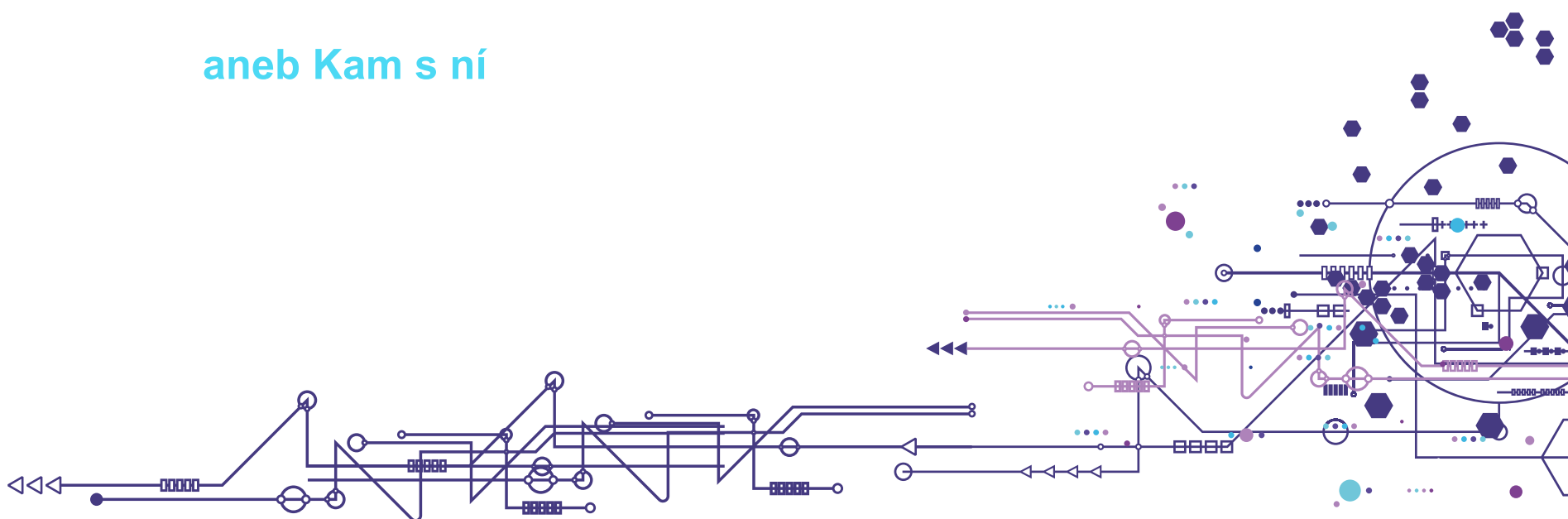
- Služby očekávané od NDI vychází z **výstupů** pracovní skupiny [Základní služby](#).
 - Pracovní skupiny pro implementaci EOSC v ČR jsou stále otevřené novým členům, viz [Pracovní skupiny implementace EOSC v ČR](#).
- Navrhované služby se dají rozdělit do několika základních kategorií:
 - **Důvěryhodné repozitáře**
 - Služby pro podporu práce s **persistními identifikátory (PID)**
 - Služby pro **řízení přístupu k datům, autentizaci a autorizaci uživatelů (AAI)** a podporu jejich **spolupráce**
 - Služby pro **podporu a plánování správy výzkumných dat**
 - **Machine actionability**
 - Služby pro **FAIRifikaci** výzkumných **dat**
 - Služby **transferu dat**
 - **Kvalita služeb**
 - **Kompatibilita s existujícími platformami**
 - **Statistiky a přehledy**
 - Národní příspěvek k **budování EOSC**





Příběh datové sady

aneb Kam s ní





Ahoj,

v prosinci jsme spolu probírali postup pro anonymizaci logů a síťového provozu pro **dataset namířený na publikaci**. Ještě jednou díky za pomoc, **dataset i článek jsou víceméně hotové**.

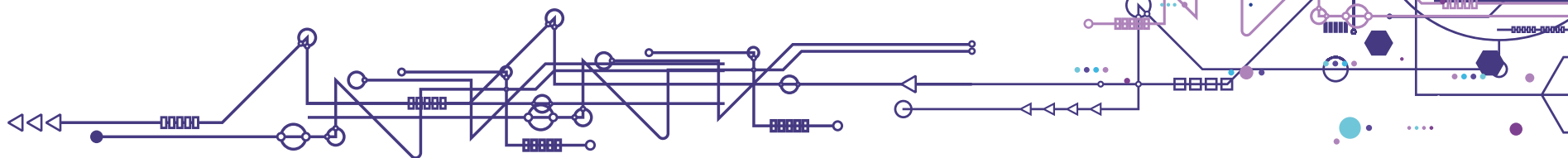
Narazil jsem ale ještě na jednu věc, se kterou bych se na vás obrátil.

Součástí datasetu jsou šifrovaná síťová data, která mají **i po komprimaci značný objem**. Celkem dataset zabírá asi 270 GB. **Úložiště Zenodo a Mendeley, která jsme už v minulosti použili pro publikaci datasetu, mě s takovým balíkem dat vcelku pochopitelně odmítla**. Přesto bychom rádi **publikovali dataset celý, bez redukce obsahu**.

Víte o nějakém prostoru, **kde by šel takhle velký dataset zveřejnit a udržovat**? Případně je nějaká možnost hostovat takový dataset přímo v rámci univerzity bez využití externích služeb?

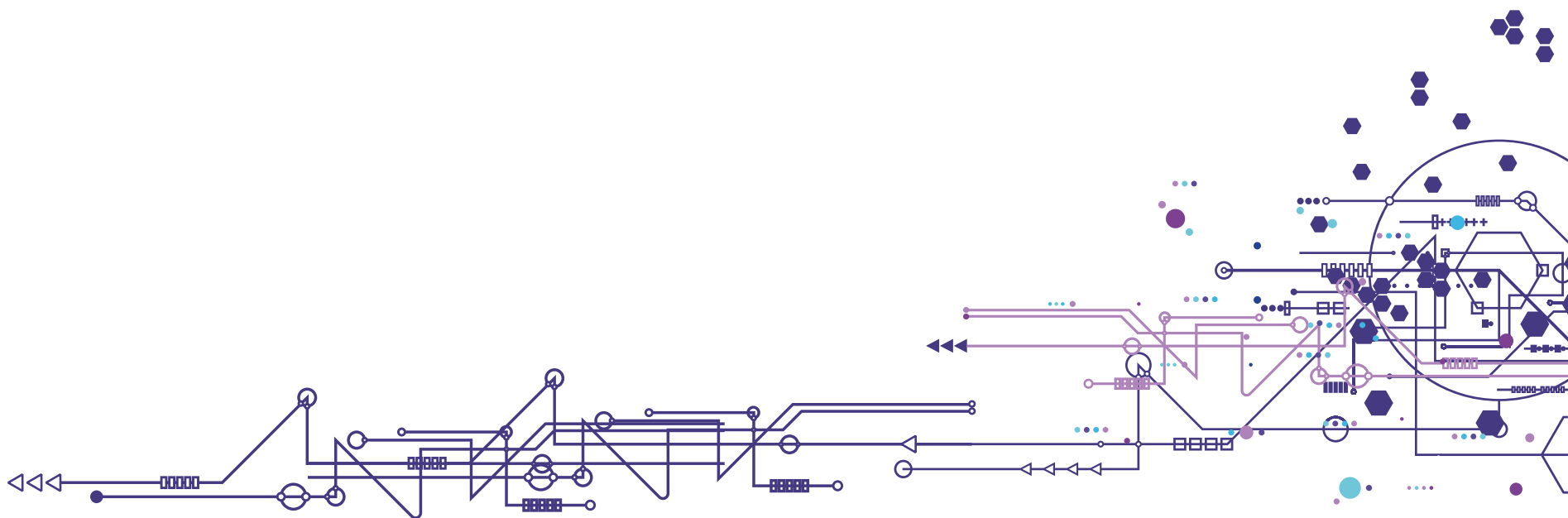
Rád bych prozkoumal všechny možnosti, než začnu dataset ořezávat.

Díky,
Standa



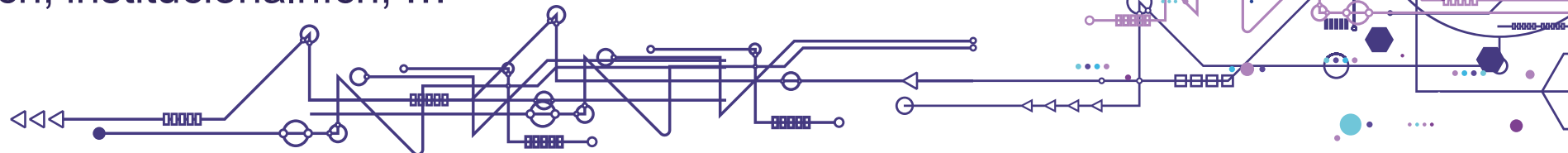


Důvěryhodné repositáře



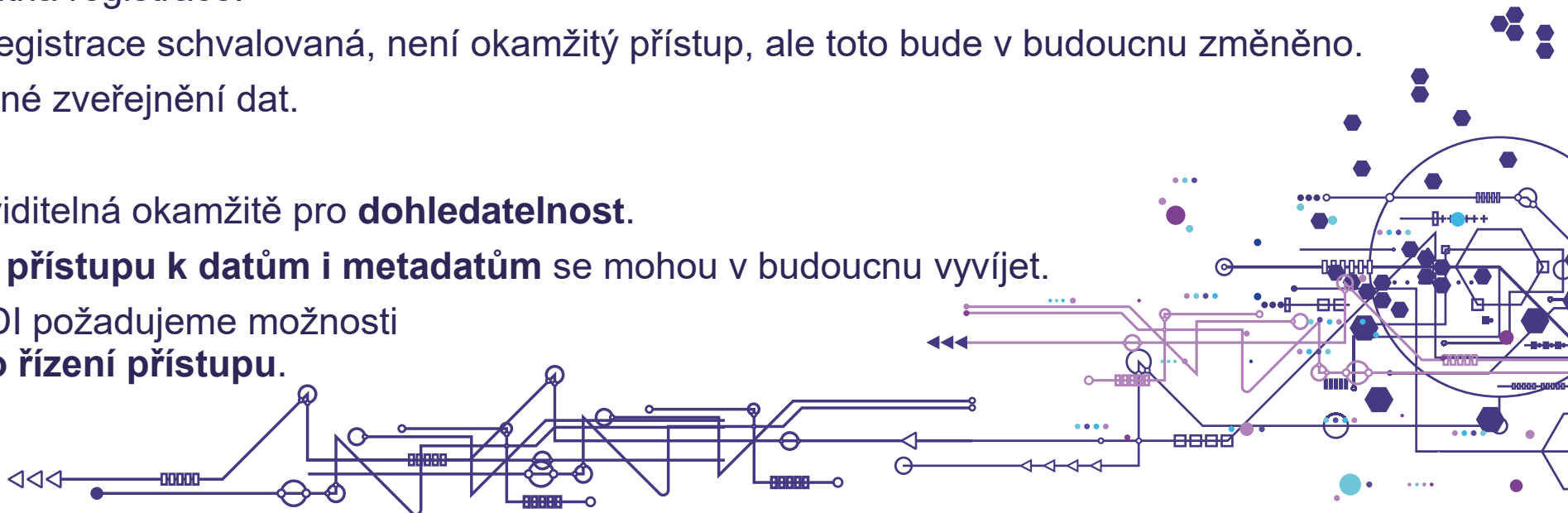
Důvěryhodné repozitáře

- Místo, kam můžete (i velká) data **nejen uložit**, ale také
 - spolehnout se, že tam budou **v následujících letech dostupná** a
 - budete moci **s daty pracovat**.
- Určeno i pro „živá“ data, tj. podpora **aktivní práce s daty**, ne jen uložení archivních dat.
- Poskytování služeb dle vhodných standardů/certifikací pro provoz **důvěryhodných repozitářů**.
 - Důvěryhodné repozitáře dle definic EU, EOSC, grantových agentur apod.
 - CoreTrustSeal?
 - Rodina standardů ISO 27 tisíc apod.?
- **Nestaráte se** o vlastní infrastrukturu, zabezpečená úložiště apod., ale
 - **využíváte služby**.
 - NDI umožní **postavení vlastních instancí repozitářů**.
 - Oborových, institucionálních, ...

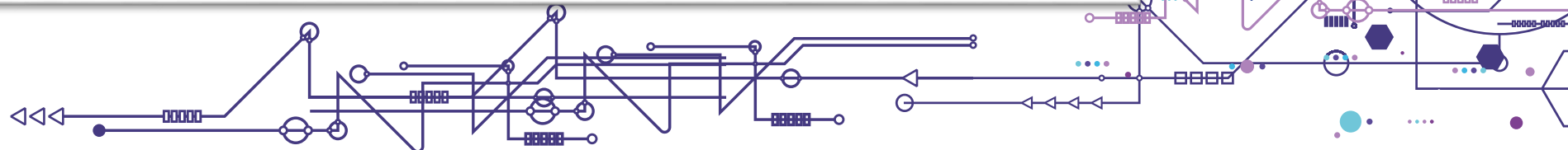
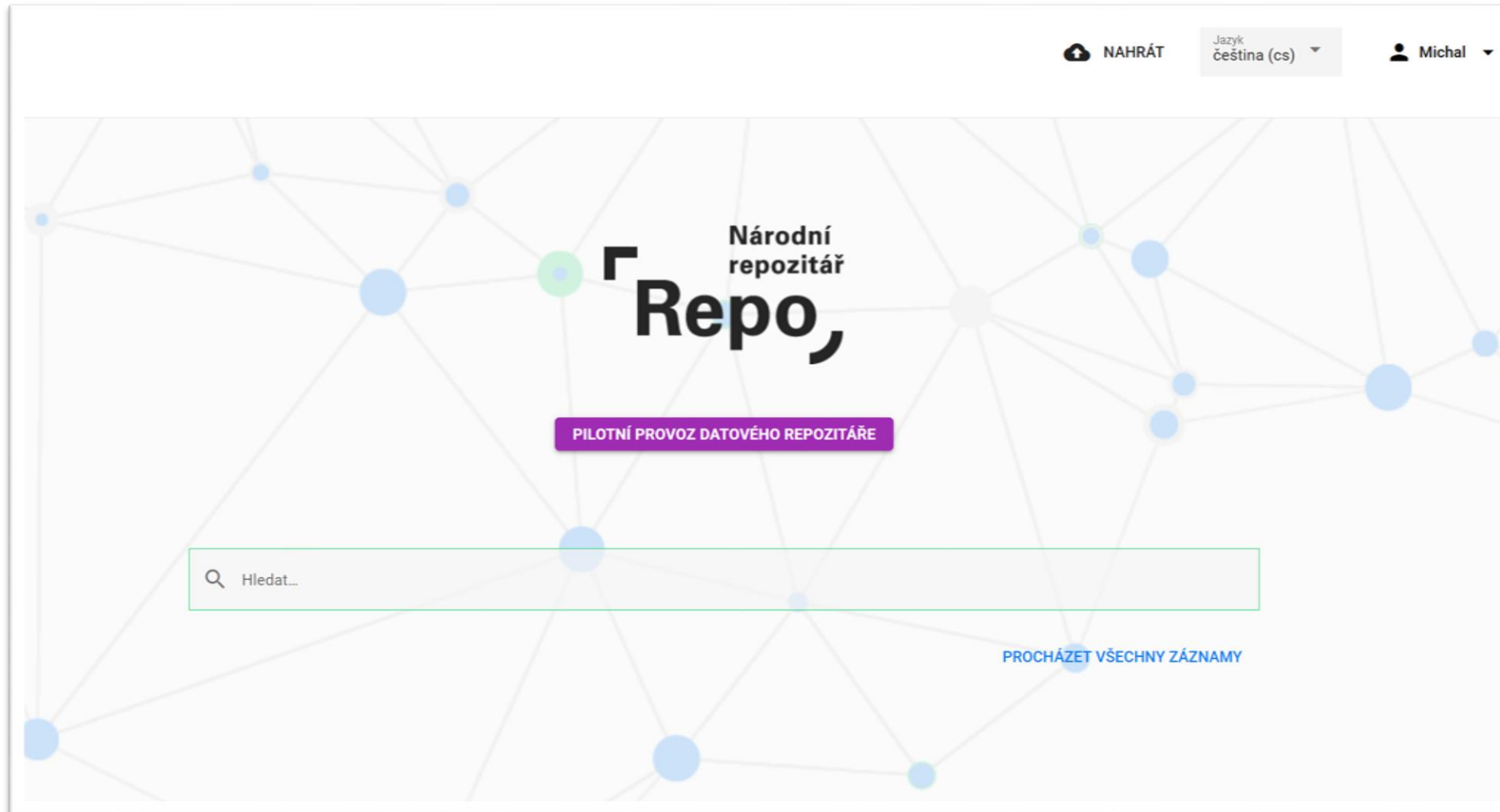


První ukázka možného budoucího vývoje: datový repozitář provozovaný CESNET

- Pilotní provoz: <https://data.narodni-repozitar.cz/>
- Návody: https://du.cesnet.cz/cs/navody/narodni_repozitar/start
- Pro **ukládání a řízené zveřejňování datových sad**.
 - Vytvoření metadatového záznamu, který popisuje datovou sadu,
 - připojení jednotlivých souborů, které tvoří vlastní datovou sadu.
 - Každý veřejný záznam **může být** na požádání **opatřen persistentním identifikátorem DOI**.
 - Možnost **vazby na publikace** založené na dané datové sadě **přes DOI** publikace v metadatech.
- Pro nahrávání dat nutná registrace.
 - V současnosti registrace schvalovaná, není okamžitý přístup, ale toto bude v budoucnu změněno.
 - Možnost odložené zveřejnění dat.
- Časové embargo.
 - Metadata jsou viditelná okamžitě pro **dohledatelnost**.
 - Možnosti **řízení přístupu k datům i metadatům** se mohou v budoucnu vyvíjet.
 - V rámci NDI požadujeme možnosti **precizního řízení přístupu**.



Datový repozitář provozovaný CESNET – titulní stránka



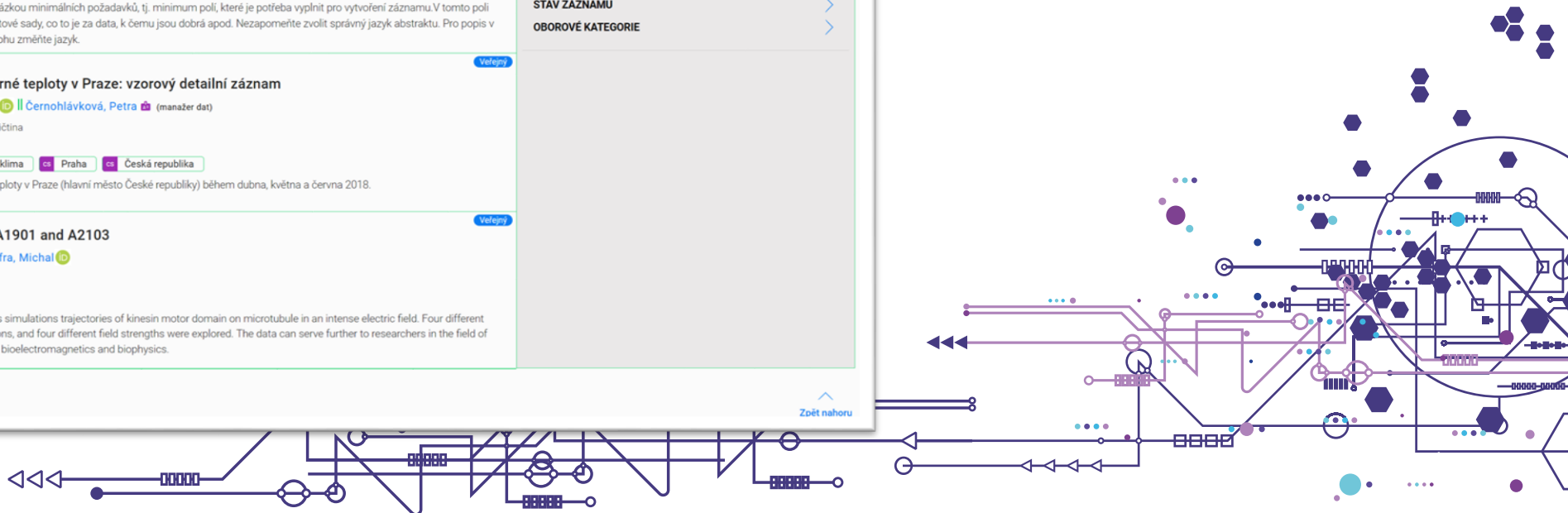
Datový repozitář provozovaný CESNET – přehled datových sad



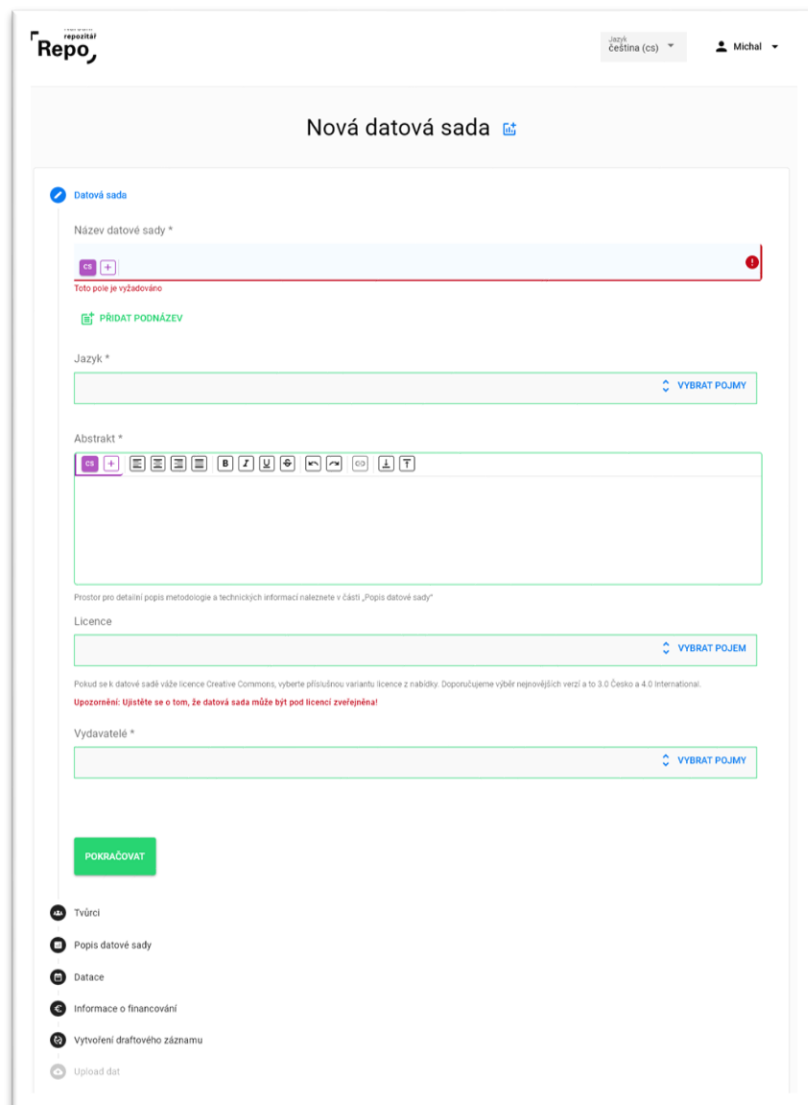
The screenshot shows the 'Repo' website interface. At the top, there is a search bar with the text 'Hledat...'. Below the search bar, it indicates 'Výsledků: 3' and 'Počet na stránku: 10'. The main content area displays three search results:

- Minimální záznam** by Horníček, Jan (2021, English). Description: 'Tento záznam je ukázkou minimálních požadavků, tj. minimum polí, které je potřeba vyplnit pro vytvoření záznamu. V tomto poli by měl být popis datové sady, co to je za data, k čemu jsou dobrá apod. Nezapomeňte zvolit správný jazyk abstraktu. Pro popis v angličtině v levém rohu změňte jazyk.' Tags: katalogizace, test.
- Měření průměrné teploty v Praze: vzorový detailní záznam** by Vyčítalová, Hana and Černošková, Petra (2022, Czech, English). Description: 'Měření průměrné teploty v Praze (hlavní město České republiky) během dubna, května a června 2018.' Tags: teplota, klima, Praha, Česká republika.
- Raw data for A1901 and A2103** by Průša, Jiří and Cifra, Michal (2022, English). Description: 'Molecular dynamics simulations trajectories of kinesin motor domain on microtubule in an intense electric field. Four different electric field directions, and four different field strengths were explored. The data can serve further to researchers in the field of bionanotechnology, bioelectromagnetics and biophysics.'

On the right side of the results, there is a sidebar with navigation links: AFILIACE, AUTOŘI, KLÍČOVÁ SLOVA, JAZYK, STAV ZÁZNAMU, and OBOROVÉ KATEGORIE. At the bottom right of the page, there is a 'Zpět nahoru' button.



Datový repozitář provozovaný CESNET – vkládání záznamu



The screenshot shows the 'Nová datová sada' (New dataset) form in the 'Repo' repository. The form is in the 'Datová sada' (Dataset) section and includes the following fields:

- Název datové sady *** (Dataset name): A text input field with a red error message 'Toto pole je vyžadováno' (This field is required) and a '+ PŘIDAT PODNÁZEV' (Add subtitle) button.
- Jazyk *** (Language): A dropdown menu with 'VYBRAT POJMY' (Select terms) as a placeholder.
- Abstrakt *** (Abstract): A rich text editor with a toolbar and a 'VYBRAT POJMY' (Select terms) button.
- Licence** (License): A dropdown menu with 'VYBRAT POJEM' (Select item) as a placeholder. A note below states: 'Pokud se k datové sadě váže licence Creative Commons, vyberte příslušnou variantu licence z nabídky. Doporučujeme výběr nejnovějších verzí a to 3.0 Česko a 4.0 International. Upozornění: Ujistěte se o tom, že datová sada může být pod licencí zveřejněna!' (If a Creative Commons license is associated with the dataset, select the appropriate license variant from the list. We recommend selecting the latest versions and 3.0 Czech Republic and 4.0 International. Warning: Make sure the dataset can be published under the license!).
- Vydavatelé *** (Publishers): A dropdown menu with 'VYBRAT POJMY' (Select terms) as a placeholder.

At the bottom left, there is a 'POKRAČOVAT' (Continue) button and a sidebar with navigation options: 'Tvůrce' (Creator), 'Popis datové sady' (Dataset description), 'Datace' (Data), 'Informace o financování' (Funding information), 'Vytvoření draftového záznamu' (Create draft record), and 'Upload dat' (Upload data).

- V rámci NDI chceme **procesy maximálně automatizovat**.
 - **NE** zbytečné **ruční vyplňování** a kontroly **formulářů**.
 - **ANO**
 - **automatické extrakci metadat z dat,**
 - **přebírání informací z existujících zdrojů** (metadatové záznamy z DataCite či ORCID, informace o datových formátech z FAIRsharing.org apod.)





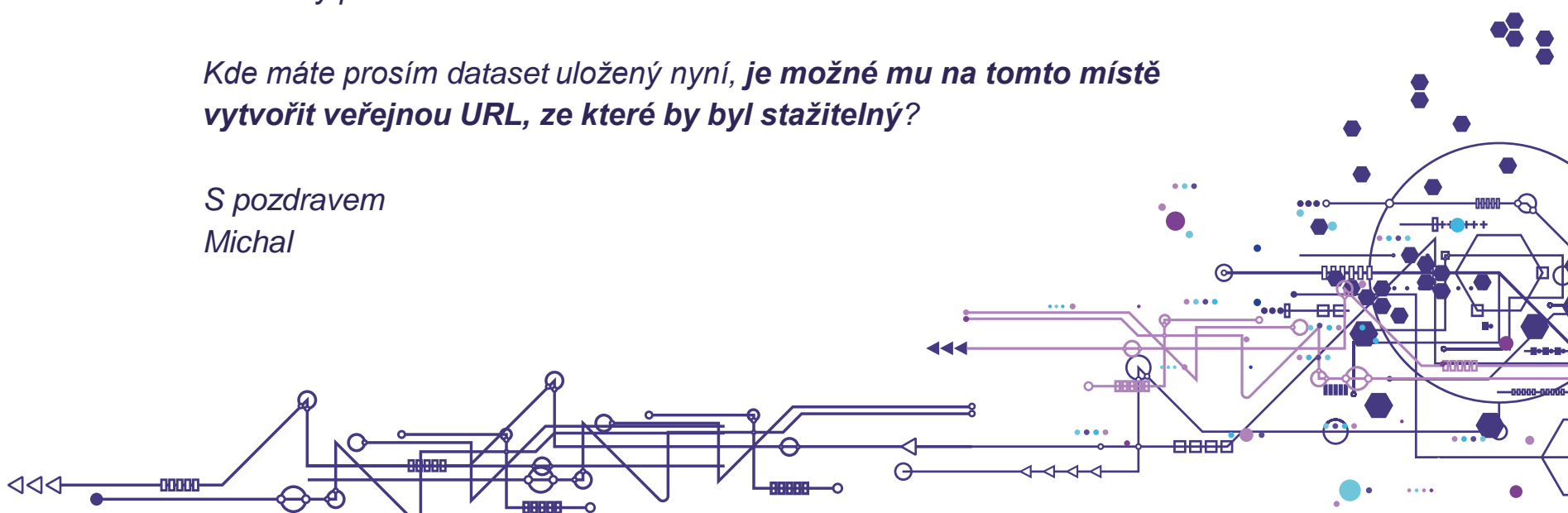
Zdravím,

*možná by bylo nejlepší **nyní** hostovat data přímo **někde** na univerzitním úložišti a **datasetu přidělit DOI**. V současné době DOI pro data za naši univerzitu umíme.*

*Pokud bychom tedy dataset hostovali na nějakém úložišti univerzity nebo CESNET, stačilo by veřejně a v publikacích **propagovat výhradně jen to DOI** a to směřovat na patřičnou URL **s aktuálním fyzickým uložením**. Nebyl by přitom problém ani s tím, že by se případně v čase fyzické **uložení měnilo**, stačilo by pak aktualizovat URL v DOI metadatach.*

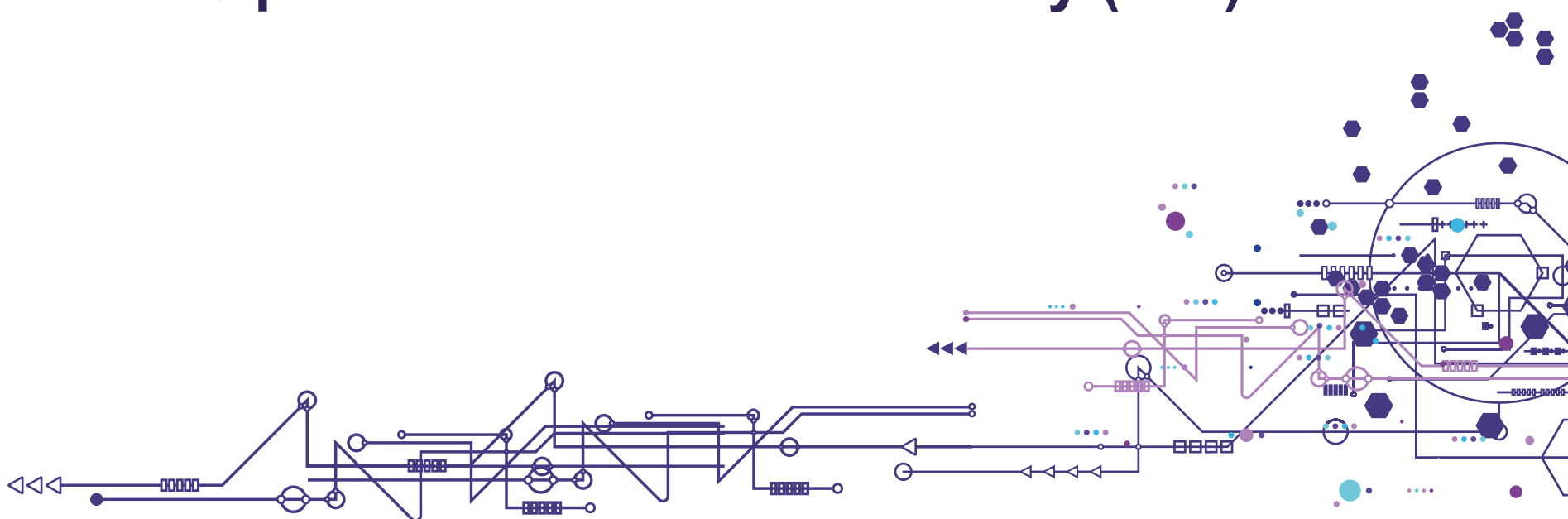
*Kde máte prosím dataset uložený nyní, **je možné mu na tomto místě vytvořit veřejnou URL, ze které by byl stažitelný?***

*S pozdravem
Michal*



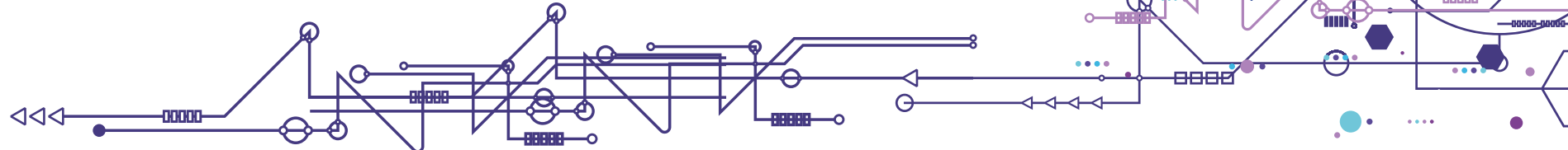


Služby pro podporu práce s persistentními identifikátory (PID)



Perzistentní identifikátory (PID)

- Mají zajistit **oddělení identifikace objektu** jako takového,
 - osoba,
 - instituce,
 - publikace,
 - dataset, ...
- **od jeho momentálního fyzického umístění.**
- **Příklad** – datová sada *https-set*
 - **Identifikátor** datové sady: <https://doi.org/10.48791/4mxx-r725>
 - Současné fyzické **umístění**: <https://ucnmuni.sharepoint.com/teams/mu-UVT-https-set/Shared%20Documents/Forms/AllItems.aspx?id=%2Fteams%2Fmu%2DUVT%2Dhttps%2Dset%2FShared%20Documents%2Fhttps%2Dset%2Dv1%2E0%2E0&p=true&ga=1>
- Fyzické umístění se **bude pravděpodobně** v budoucnu **měnit**.
 - Pilotní datový repozitář CESNET?
 - Oborový repozitář v NDI?
 - Oborový repozitář v zahraničí?
- Ale **změny nevadí**.
 - **Uživatelům** je jako odkaz na data **vždy** prezentováno [DOI 10.48791/4mxx-r725](https://doi.org/10.48791/4mxx-r725),
 - které je vždy **zavede na aktuální umístění**.

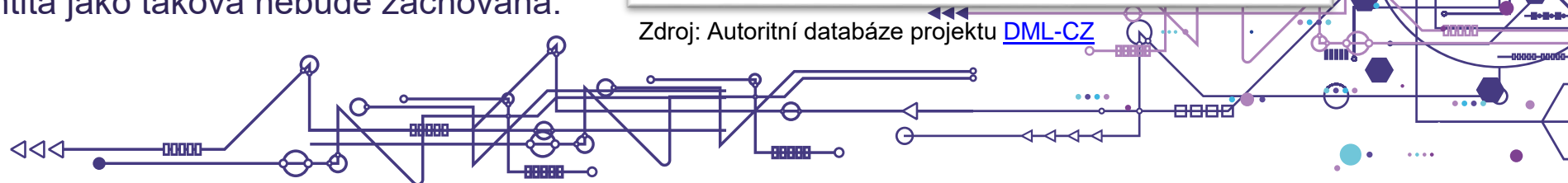


Perzistentní identifikátory (PID) (2)

- Mají zajistit **jednoznačnost**.
- **Příklad** – jména fyzických osob
 - **Více forem zápisu** jména **jedné fyzické osoby**.
 - **Více různých fyzických osob** se **stejným jménem**.
- Mají zajistit **trvalost** (perzistenci).
 - **Metadata** fyzicky umístěna **u třetí strany** nezávisle na fyzickém umístění odkazované entity.
 - Vlastník identifikátoru se stará o **aktualizaci metadat** a aktualizace **směřování** na aktuální umístění.
 - Třetí strana **pečuje o zachování** poslední verze a historie **metadat** a **existenci identifikátoru** jako takového, i pokud se vlastník identifikátoru o něj starat přestane. A dokonce i tehdy, pokud identifikovaná entita jako taková nebude zachována.

<input type="checkbox"/>	NovakD (1)	Novák, David (1)	Join	Delete
<input type="checkbox"/>	NovakE (5)	Nováková, Eva (5) Nováková, E. (0)	Join	Delete
<input type="checkbox"/>	NovakJ (52)	Novák, Josef (38) Novák, Jos. (1) Novák, J. (13)	Join	Delete
<input type="checkbox"/>	NovakJ2 (7)	Novák, Jiří (7) Novak, Jiri (0)	Join	Delete
<input type="checkbox"/>	NovakJ7 (19)	Novák, Josef (16) Novák, J. (3)	Join	Delete
<input type="checkbox"/>	NovakK (2)	Novák, Karel (2)	Join	Delete
<input type="checkbox"/>	NovakM (2)	Novák, Mirko (2) Novak, Miroslav M. (0) Novak, M. M. (0) Novák, M. (0)	Join	Delete
<input type="checkbox"/>	NovakM2 (2)	Nováková, Markéta (2)	Join	Delete
<input type="checkbox"/>	NovakM3 (1)	Novák, Miroslav (1)	Join	Delete
<input type="checkbox"/>	NovakO (2)	Novák, Ondřej (2)	Join	Delete
<input type="checkbox"/>	NovakP (1)	Novák, Petr (1)	Join	Delete
<input type="checkbox"/>	NovakS (1)	Novák, Stanislav (1)	Join	Delete
<input type="checkbox"/>	NovakV (57)	Novák, Vítězslav (55) Novák, V. (2)	Join	Delete
<input type="checkbox"/>	NovakV2 (7)	Novák, Vilém (7) Novák, V. (0)	Join	Delete
<input type="checkbox"/>	NovakV3 (53)	Novák, Vladimír (53)	Join	Delete
<input type="checkbox"/>	NovakZ (1)	Novák, Zdeněk (1)	Join	Delete

Zdroj: Autoritní databáze projektu [DML-CZ](#)



Služby pro podporu práce s persistentními identifikátory (PID)

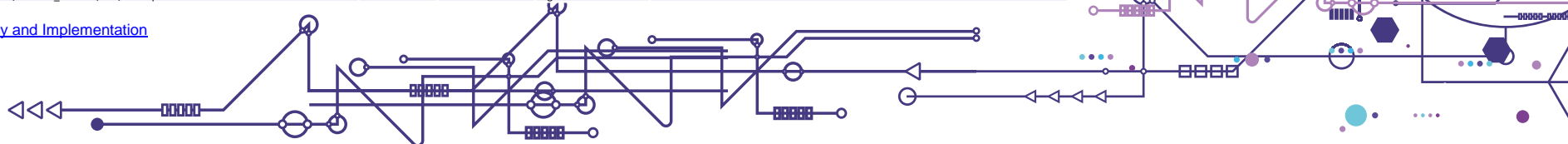
- Poskytování **PID** dle **soudobé dobré praxe**.
 - Kompatibilita s existující praxí a doporučeními,
 - např. s EOSC PID Policy and Architecture (<https://doi.org/10.2777/926037>),
 - ale také **aktivní přispívání** do vývoje **praxe PID** na úrovni **EOSC**.
 - Dialog o vhodných **PID pro identifikaci** a odkazování **výzkumných infrastruktur**.
- Podpora **přidělování PID** pro **datové sady**.
 - Jako **služba**.
 - Ne nutně jen uvnitř NDI, ale i **pro externí existující repozitáře** apod.
- Využívání **PID dalších objektů** při **metadatovém popisu** a odkazování.
 - Osoby,
 - instituce,
 - publikace,
 - ...



Řada typů PID

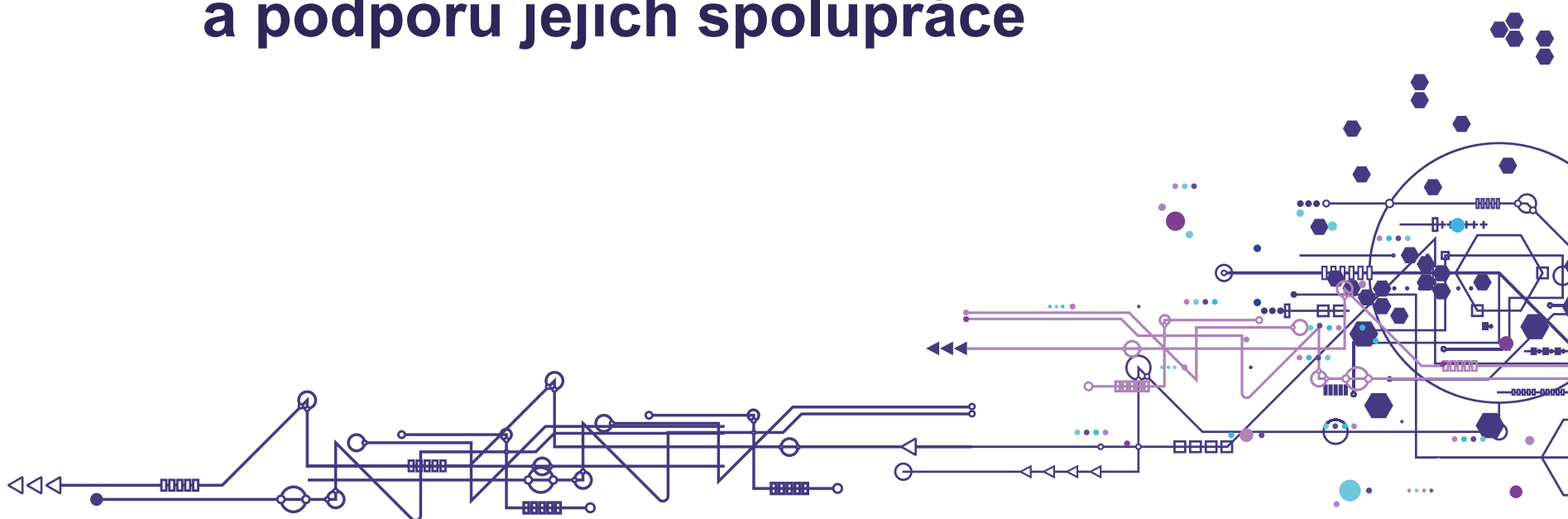
CODEVALUE	PREFLABEL_EN	HREF	Maturity	Globally resolvable	For which object type	Comments
ADSBibcode	Astrophysics Data System - Bibliographic Reference Code	https://ui.adsabs.harvard.edu/	High	Needs token	Publication	
ARK	Archival Resource Key	https://arks.org/	High	Local	Other (in comments)	Everything
arXiv	arXiv identifier scheme	https://arxiv.org/	High	Yes	Publication	
ASIN	Amazon Standard Identification Number	https://sellercentral.amazon.ca/gp/help/external/200317470?language=en-CA&ref=mpbc_200576730_cont_200317470 https://indico.cern.ch/event/780651/attachments/1776614/2888642/Conference_PIDs_and_Crossmark.pdf	High	?	Other (in comments)	Things sell by Amazon
ConfID	Conference identifier		High	Yes	Event	Not clear if it is a Crossref service
Crossref DOI			?	Yes	Publication	
Crossref_funders	Crossref Funder Registry	https://www.crossref.org/services/content-registration/grants/	?	Yes	Organisation	
Crossref_grants	Registering research grants	https://www.crossref.org/community/grants/			Other (in comments)	Grants
DataCite DOI			High	Yes	Other (in comments)	28 different resources and outputs
DOI	Digital Object Identifier	https://www.doi.org/	High	Yes	Publication	Services supporting PIDs and metadata for 40+ resource and output types
EAN13	The 13-digit International Article Number	https://www.gs1.org/standards/barcodes/ean-upc	High	?	Other (in comments)	Physical product identifier. A Whole famill of id: UPC-A, UPC-E, EAN13, EAN8
eISBN	electronic International Standard Book Number	https://www.isbn-international.org/			Publication	
eISSN	Electronic International Standard Serial Number	http://portal.issn.org/	High	https://portal.issn.org/resource/ISSN/0376-4583/	Publication	Identifies various types of serial publications (eg. journals, websites, blogs)
GRID	Global Research Identifier Database	https://www.grid.ac/	Closed	?	Organisation	Transitioned to ROR
Handle	Handle	http://www.handle.net/	High	Yes	Dataset	It is the base of DOI also
IGSN						
ISAN	International Geo Sample Number	https://www.igsn.org/	High	Yes	Other (in comments)	Physical Samples and Sampling Features
ISAN	International Standard Audiovisual Number	https://www.isan.org/		Yes?	Publication	
ISBN	International Standard Book Number	https://www.isbn-international.org/	High		Publication	
ISLI	Identifies the links between different entities	https://www.isbn-international.org/content/isli-introduction	?	Yes	Other (in comments)	Link bw. entities
ISMN	International Standard Music Number	https://www.ismn-international.org/		No	Publication	
ISNI	International Standard Name Identifier	https://isni.org/page/search-database/			Person	Contributors to creative works and their distribution and organizations
ISSN	International Standard Serial Number	http://portal.issn.org/	High	https://portal.issn.org/resource/ISSN/0376-4583/	Publication	Identifies various types of serial publications (eg. journals, websites, blogs)
ISTC	The International Standard Text Code	http://www.istc-international.org/	Defunct	No	Publication	Ceased in 2017
LSID	Life Sciences Identifier	http://www.lsid.info/	?	Yes	Other (in comments)	Metadata for life science items
ORCID	Open Researcher and Contributor ID	https://orcid.org/	High	Yes	Person	
PIC	(EC) partner identity code	https://ec.europa.eu/info/funding-tenders/opportunities/portal/screen/support/faq/1055				
PMID	PubMed ID	https://www.ncbi.nlm.nih.gov/pmc/pmctopmid/			Publication	
PURL	persistent uniform resource locator	https://archive.org/services/purl/	High	Yes	Other (in comments)	Resources on the Web
QID	Wikidata identifier	https://www.wikidata.org/wiki/Wikidata:Identifiers			Other (in comments)	Knowledge item
RAID	Persistent Identifier for research projects	https://www.raid.org.au/		Needs token	Other (in comments)	Research projects
Ringgold	Unique numerical identifier applied to organizations in the scholarly supply chain	https://www.ringgold.com/			Organisation	
ROR	Research Organization Registry	https://ror.org/	High	Yes	Organisation	
RRID	Research Resource Identifier	https://scicrunch.org/resources				
ScopusAuthorID	Scopus Author ID	https://service.elsevier.com/app/answers/detail/a_id/11212/supporthub/scopus/	?	No?	Publication	
SWHID	SoftWare Heritage persistent Identifiers	https://docs.softwareheritage.org/devrel/zhw-model/persistent-identifiers.html		Local	Source Code	
UPC	Universal Product Code				Other (in comments)	Synonym of EAN13? Product identifier
URI	Uniform Resource Identifier	https://www.gs1.org/standards/barcodes/ean-upc				
URL	Uniform Resource Locator					
URN	Uniform Resource Name					
VAT-number	VAT number	http://ec.europa.eu/taxation_customs/vies/vatRequest.html			Organisation	

Zdroj: Interní pracovní materiály [EQSC Task Force PID Policy and Implementation](#)





Služby pro řízení přístupu k datům, autentizaci a autorizaci uživatelů (AAI) a podporu jejich spolupráce



Precizní kontrola vlastníka dat nad řízením přístupu k datům i metadatům

- **Data** uložená v NDI jsou **pod** plnou **kontrolou** jejich **vlastníků**.
 - **Zpřístupnění** dalším osobám se děje **na pokyn** jejich **vlastníka**.
- Už samotná **informace o existenci** dat či metadata **může být citlivá** informace.
 - **Řízení přístupu** až na úrovni řízení **viditelnosti existence** samotných **metadata** či jejich **částí** apod.
- **Federované AAI** služby
 - Autentizace uživatelů přes **federované** služby **přihlašování** (eduID.cz apod.),
 - tj. s využitím **existující identity** z instituce, **bez** nutnosti **vytváření** dalšího **úctu**, hesla atd.



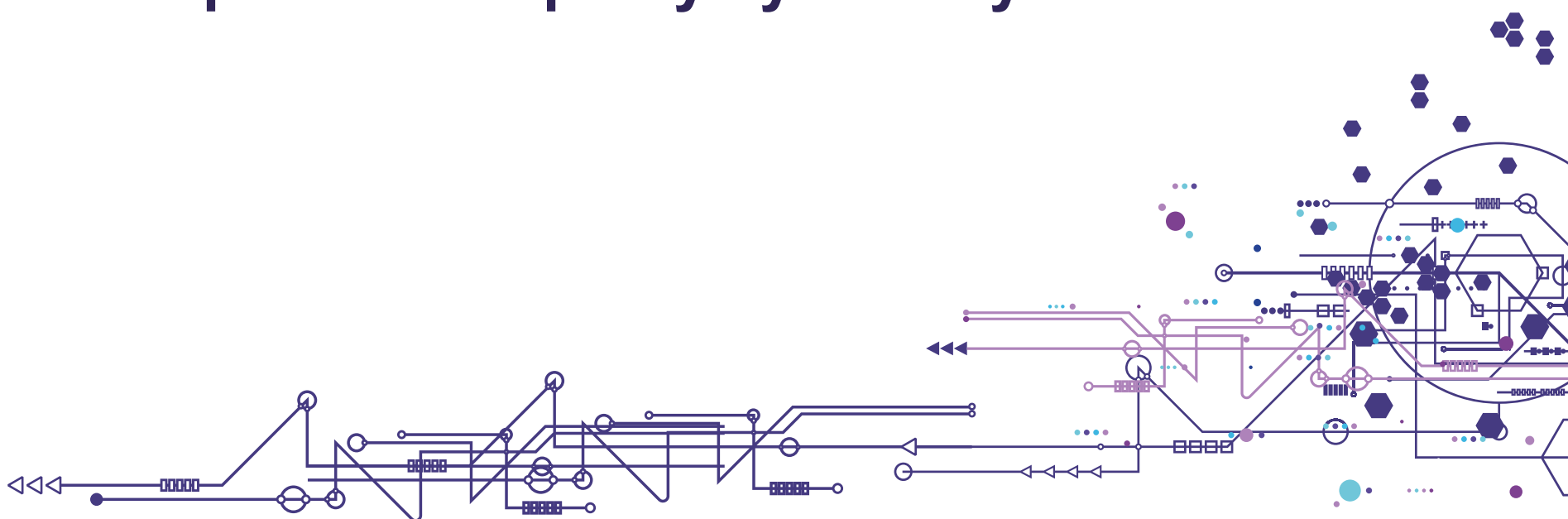
Služby pro sdílení dat s jinými uživateli

- **Sdílení dat** by mělo být možné jednak s **konkrétními osobami a skupinami**
 - výběr např. z institucionálních identit daných federovaným přihlášením apod.,
- případně **sdílením „odkazem“**.
- Může být řešeno i formou „**data visiting**“.
 - **Řízené** povolením přístupu vybraným uživatelů do „**domovského**“ úložiště s danými daty,
 - **bez** nutnosti **přenosu do jiné lokality**.
- Možnost **anonymních view-only přístupů** k datům.
 - Pro podporu předání dat **pro zaslepené recenzní řízení** apod.
 - Stále častější **povinnost sdílení podkladových dat k publikacím**.
 - Někdy již v okamžiku recenzního řízení.
 - Ukázka existující implementace např. v platformě OSF:
<https://help.osf.io/hc/en-us/articles/360042097853-Create-a-View-only-Link-for-a-Registration>





Služby pro podporu a plánování správy výzkumných dat



DMPlanning vs. DMPlan

1. Data Management Planing

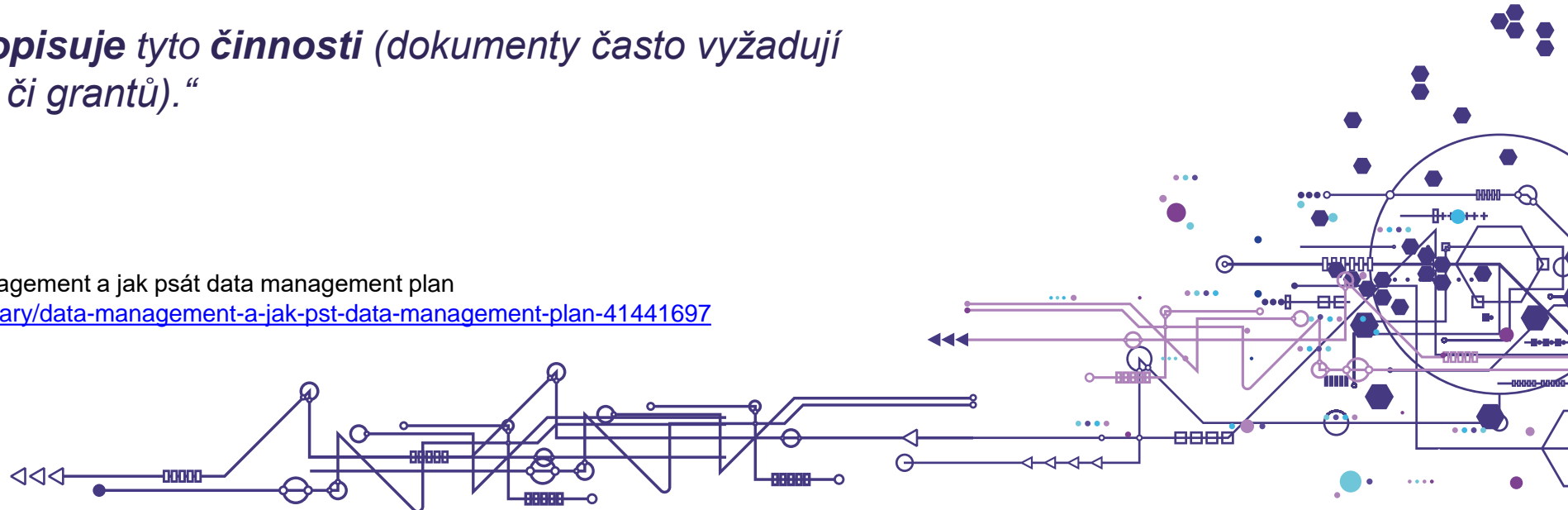
„**Proces** plánování, popisu a informování o životním cyklu dat a činností spojených s jejich správou v průběhu výzkumu.“

2. Data Management Plan (DMP)

„**Dokument**, který popisuje tyto činnosti (dokumenty často vyžadují poskytovatelé dotací či grantů).“

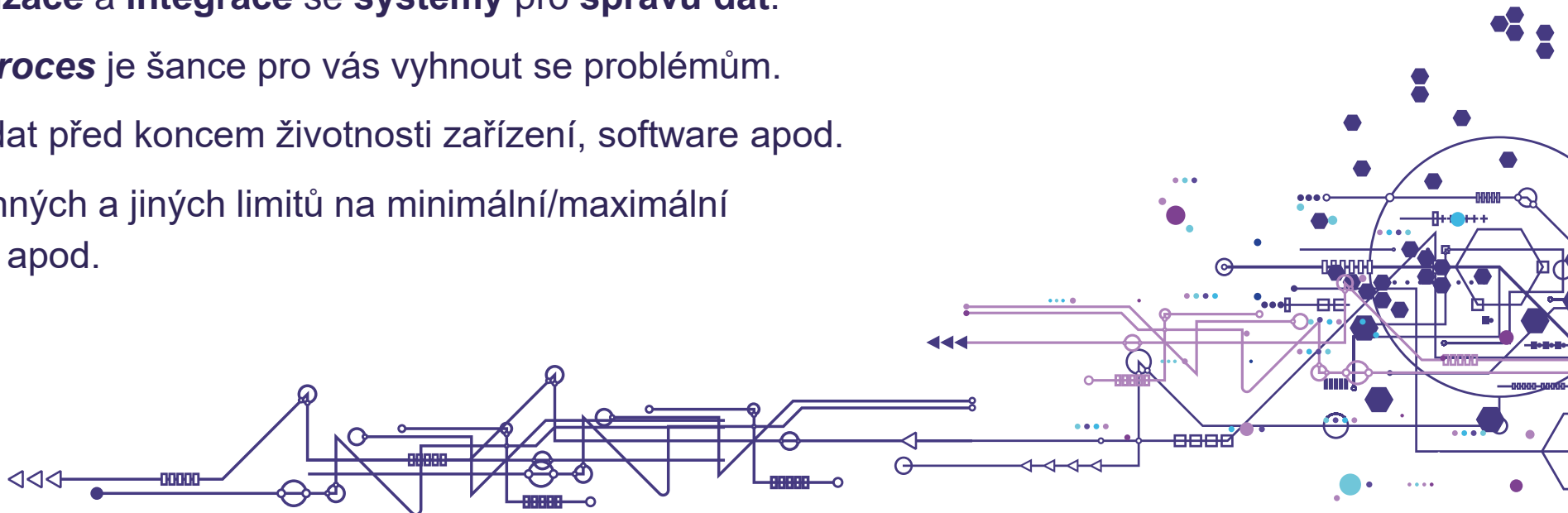
Zdroj: Petra Dědičová, Data management a jak psát data management plan

<https://www.slideshare.net/butlibrary/data-management-a-jak-psat-data-management-plan-41441697>

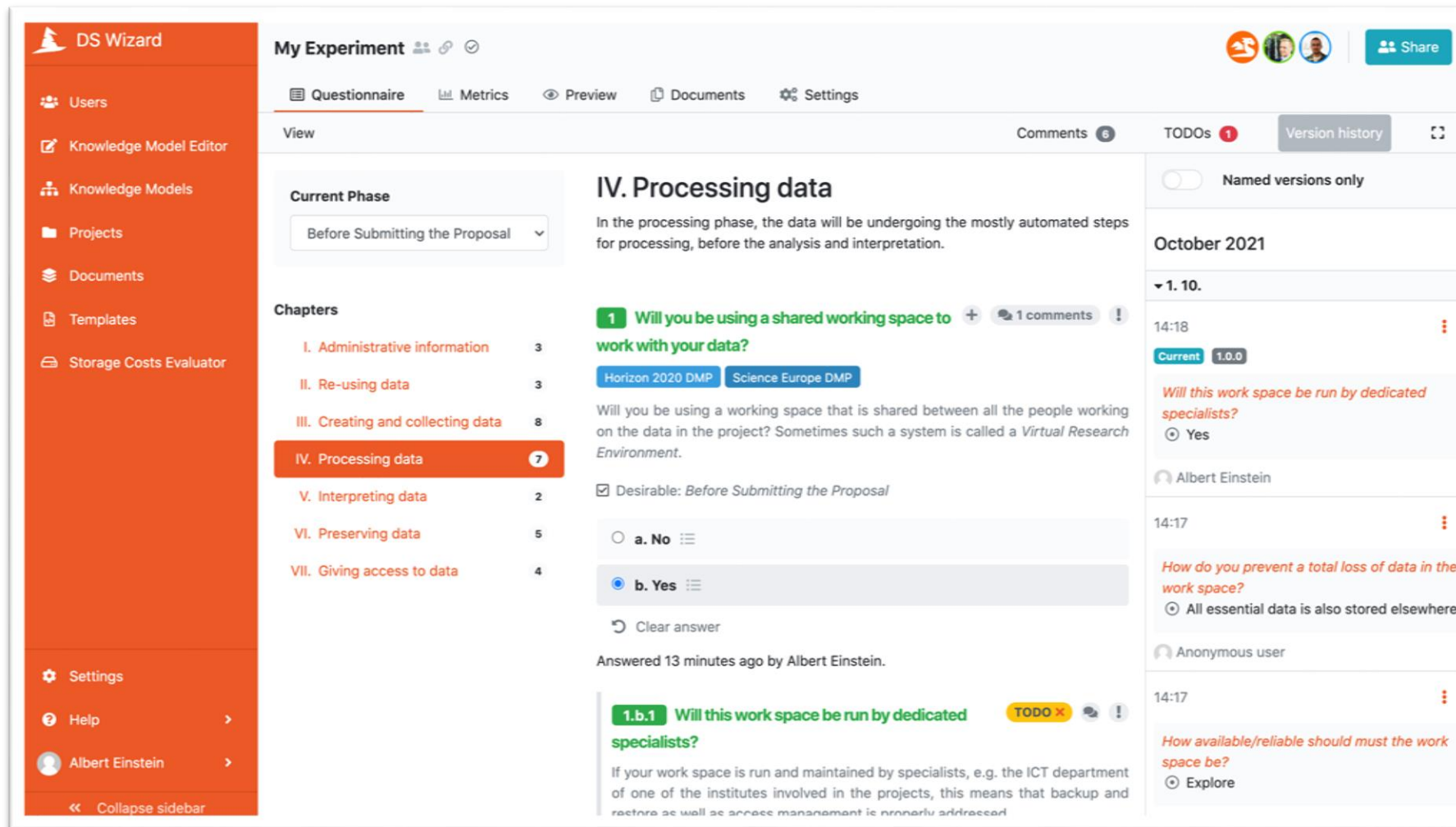


DMP jako projektová podmínka

- DMP je stále častěji **vyžadován** jako standardní **součást** podání/řešení/finalizace **projektů**.
- Např. Horizon Europe požaduje
 - iniciální DMP,
 - aktualizace DMP uprostřed projektu,
 - finální revize DMP na konci projektu.
- DMP je **živý dokument** – měl by být **průběžně aktualizován** po **celou dobu projektu**.
 - Vhodná **automatizace** a **integrace** se **systemy** pro **správu dat**.
- DMP jako **průběžný proces** je šance pro vás vyhnout se problémům.
 - Včasná migrace dat před koncem životnosti zařízení, software apod.
 - Dodržování zákonných a jiných limitů na minimální/maximální retenční doby dat apod.

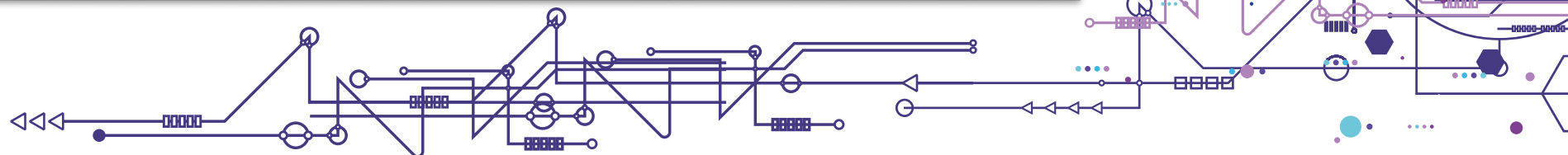


Příklad automatizovatelného DMP nástroje: Data Stewardship Wizard



The screenshot displays the 'DS Wizard' interface for a 'My Experiment'. The left sidebar contains navigation options: Users, Knowledge Model Editor, Knowledge Models, Projects, Documents, Templates, Storage Costs Evaluator, Settings, Help, and Albert Einstein. The main content area is titled 'IV. Processing data' and includes a 'Current Phase' dropdown set to 'Before Submitting the Proposal'. A list of 'Chapters' is shown, with 'IV. Processing data' selected. The questionnaire question is: '1 Will you be using a shared working space to work with your data?'. It includes a 'Desirable: Before Submitting the Proposal' checkbox and two radio button options: 'a. No' and 'b. Yes'. A comment from 'Albert Einstein' is visible, asking 'Will this work space be run by dedicated specialists?'. The interface also shows a 'Version history' panel on the right with a toggle for 'Named versions only' and a list of versions for 'October 2021', including '1.10.' with a 'Current' status and version '1.0.0'.

Zdroj: <https://ds-wizard.org/>



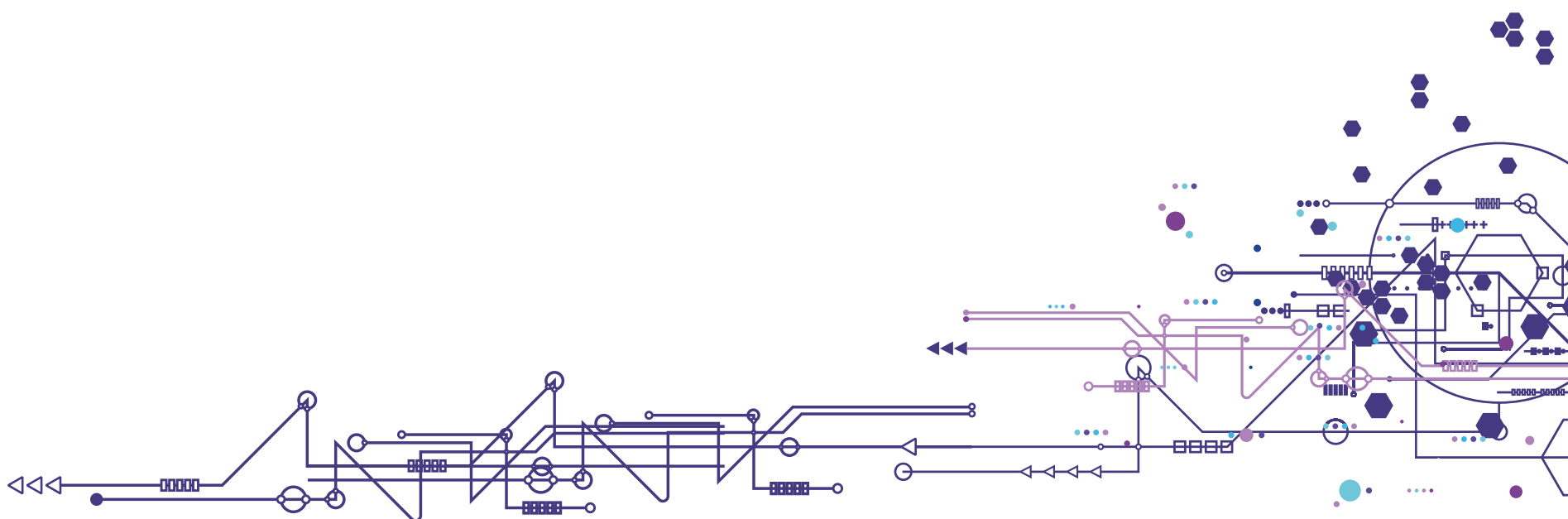
Služby pro podporu a plánování správy výzkumných dat

- **Definované** (strojově zpracovatelné) **politiky** práce s daty **v repozitářích** v NDI.
- **Podpora tvorby DMP.**
 - **Vygenerování DMP**
 - nebo vybraných částí
 - **na základě politik** použitého **repozitáře**, znalosti licencí, formátů, velikostí dat, cílených recenzí dat (maximální či minimální doba uchovávání např. na základě zákonných požadavků) apod.
 - Možnost **napojení/integrace existujících nástrojů** pro DMP.
 - Např. Data Stewardship Wizard.



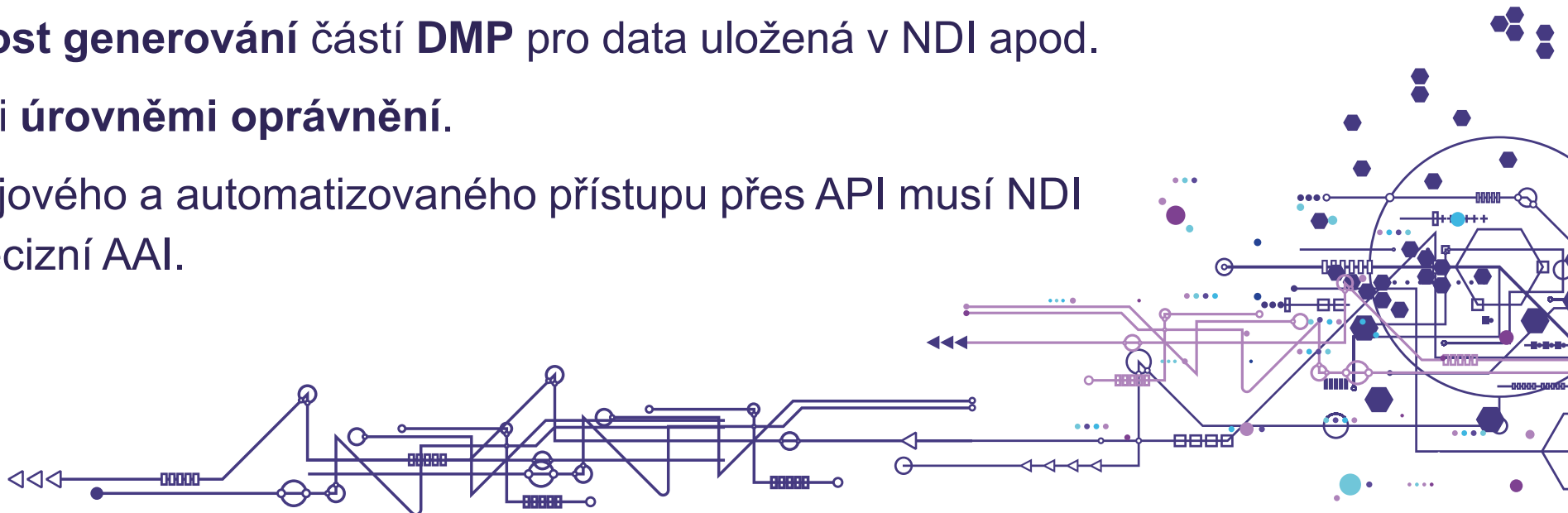


Machine actionability

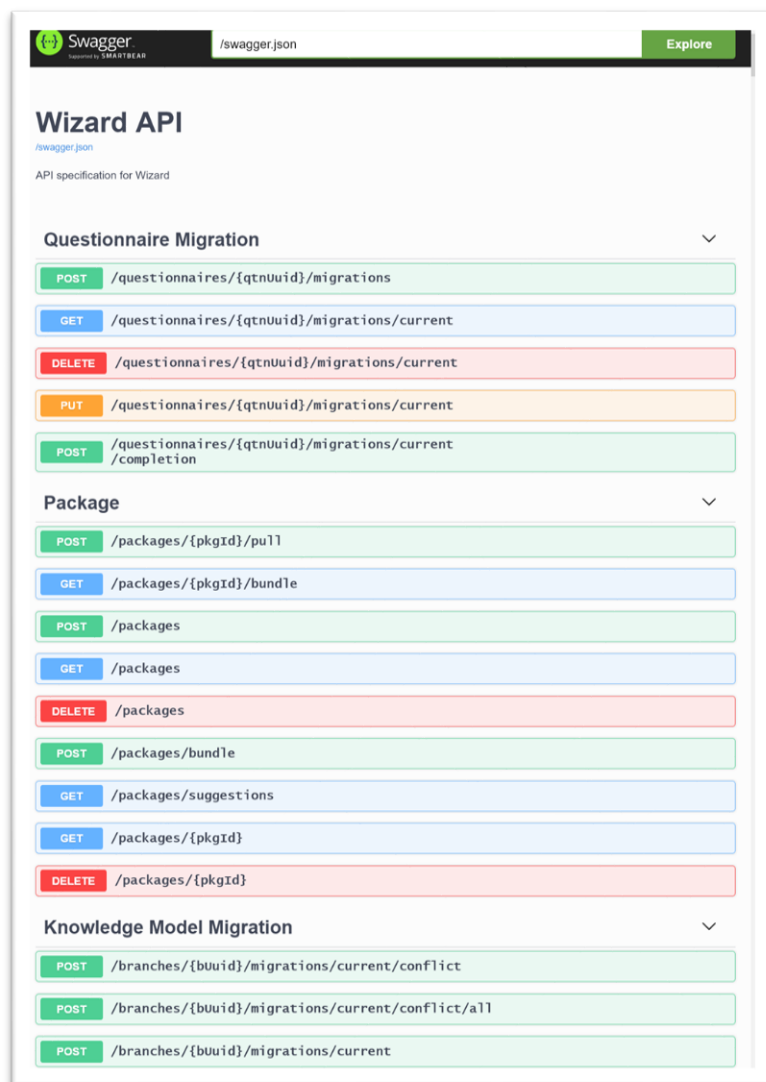


Machine actionability

- Dostupná **API** pro **všechny funkce** NDI.
 - Maximalizace možností strojového a automatizovaného použití.
 - API mohou být v odůvodněných případech autentizovaná.
- **Maximum informací** musí být uchováváno **ve strojově čitelných** a zpracovatelných **formátech**.
 - Včetně např. **politik** repozitářů v NDI apod.
 - Pro **možnost generování** částí **DMP** pro data uložená v NDI apod.
- **API klíče** s různými **úrovněmi oprávnění**.
 - I na úrovni strojového a automatizovaného přístupu přes API musí NDI podporovat precizní AAI.



Strojově použitelné API Data Stewardship Wizard



Swagger /swagger.json Explore

Wizard API

API specification for Wizard

Questionnaire Migration

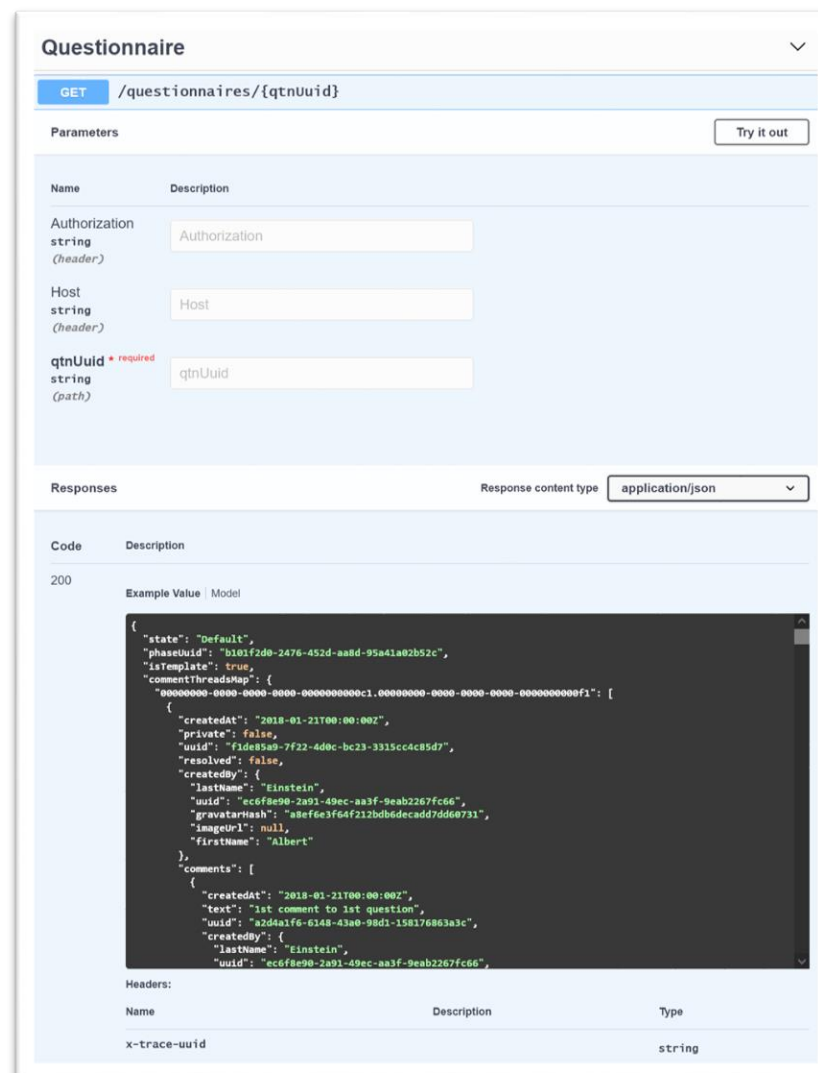
- POST /questionnaires/{qtnUuid}/migrations
- GET /questionnaires/{qtnUuid}/migrations/current
- DELETE /questionnaires/{qtnUuid}/migrations/current
- PUT /questionnaires/{qtnUuid}/migrations/current
- POST /questionnaires/{qtnUuid}/migrations/current/completion

Package

- POST /packages/{pkgId}/pull
- GET /packages/{pkgId}/bundle
- POST /packages
- GET /packages
- DELETE /packages
- POST /packages/bundle
- GET /packages/suggestions
- GET /packages/{pkgId}
- DELETE /packages/{pkgId}

Knowledge Model Migration

- POST /branches/{buid}/migrations/current/conflict
- POST /branches/{buid}/migrations/current/conflict/all
- POST /branches/{buid}/migrations/current



Questionnaire

GET /questionnaires/{qtnUuid}

Parameters Try it out

Name	Description
Authorization string (header)	Authorization
Host string (header)	Host
qtnUuid * required string (path)	qtnUuid

Responses Response content type: application/json

Code	Description
200	

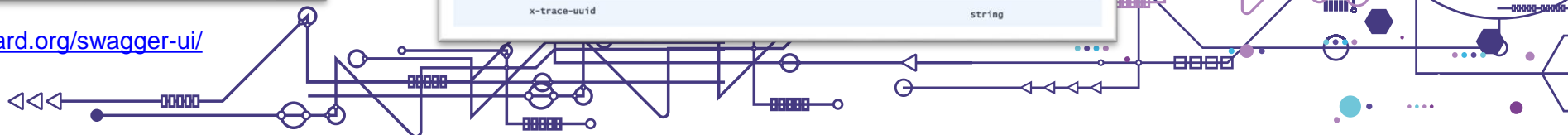
Example Value | Model

```
{
  "state": "Default",
  "phaseUuid": "b101fd0-2476-452d-aa8d-95a41a02b52c",
  "isTemplate": true,
  "commentThreadMap": {
    "00000000-0000-0000-0000-000000000001.00000000-0000-0000-0000-000000000001": [
      {
        "createdAt": "2018-01-21T00:00:00Z",
        "private": false,
        "uid": "f1de85a9-7f22-4d0c-bc23-3315cc4c85d7",
        "resolved": false,
        "createdBy": {
          "lastName": "Einstein",
          "uid": "ec6f8e90-2a91-49ec-a33f-9eab2267fc66",
          "gravatarHash": "a8ef6e3f64f212bdb6decadd7d60731",
          "imageUrl": null,
          "firstName": "Albert"
        }
      }
    ],
    "comments": [
      {
        "createdAt": "2018-01-21T00:00:00Z",
        "text": "1st comment to 1st question",
        "uid": "a2d4a1f6-6148-43a0-98d1-158176863a3c",
        "createdBy": {
          "lastName": "Einstein",
          "uid": "ec6f8e90-2a91-49ec-a33f-9eab2267fc66"
        }
      }
    ]
  }
}
```

Headers

Name	Description	Type
x-trace-uuid		string

Zdroj: <https://api-researchers.ds-wizard.org/swagger-ui/>



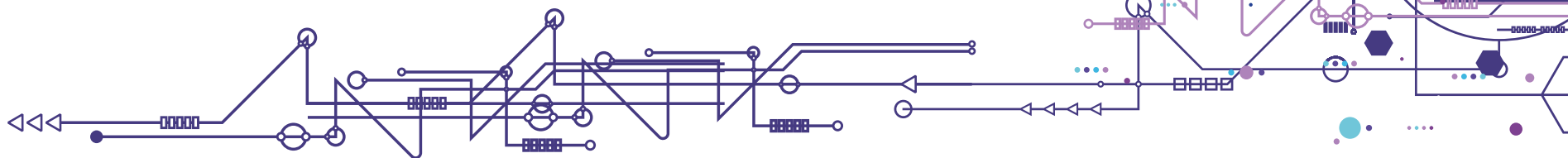


Pěkný den,
dále k jednotlivým bodům:

1. Vyrobil jsem tedy složku pro `https-set-v1.0.0`.
 - o **Jak tedy s verzováním?** Označíme toto za dataset verze 1.0.0? Pak to tak napíšeme i do DataCite metadat.
2. Díky za vyplnění **ORCID a afiliací autorů**.
3. DOI přidělím v režimu **Findable**, tj. že bude metadatový záznam datasetu **dohledatelný** podle klíčových slov atd. **na <https://search.datacite.org/>**, nebo mám DOI nechat v režimu Registered, tj. že bude fungovat, ale **bez znalosti daného ID** se člověk k metadatům **nedostane**?
 - o Pokud nemáte speciální potřebu informace o datasetu (zatím?) skrývat, určitě bych byl pro plně veřejný režim Findable.
4. **Licence CC-BY** je ok, uvedu to tak i do metadatového záznamu.
5. Co se týká vyplnění metadatového záznamu, navrhuji zvážit ještě toto:
 - o **Formats** – Technical format of the resource.
 - o **Version** – The version number of the resource.
 - Uvedl jsem tedy 1.0.0, ok?
 - o **Funding References** – Information about financial support (funding) for the resource being registered.
 - Tady je **možno** uvést **číslo grantu**, ze kterého se to platilo. Pokud číslo grantu máte, byl bych pro jej do metadat uvést, může to být dobré pro **uznávání a viditelnost.. výsledků projektu pro grantovou agenturu**.

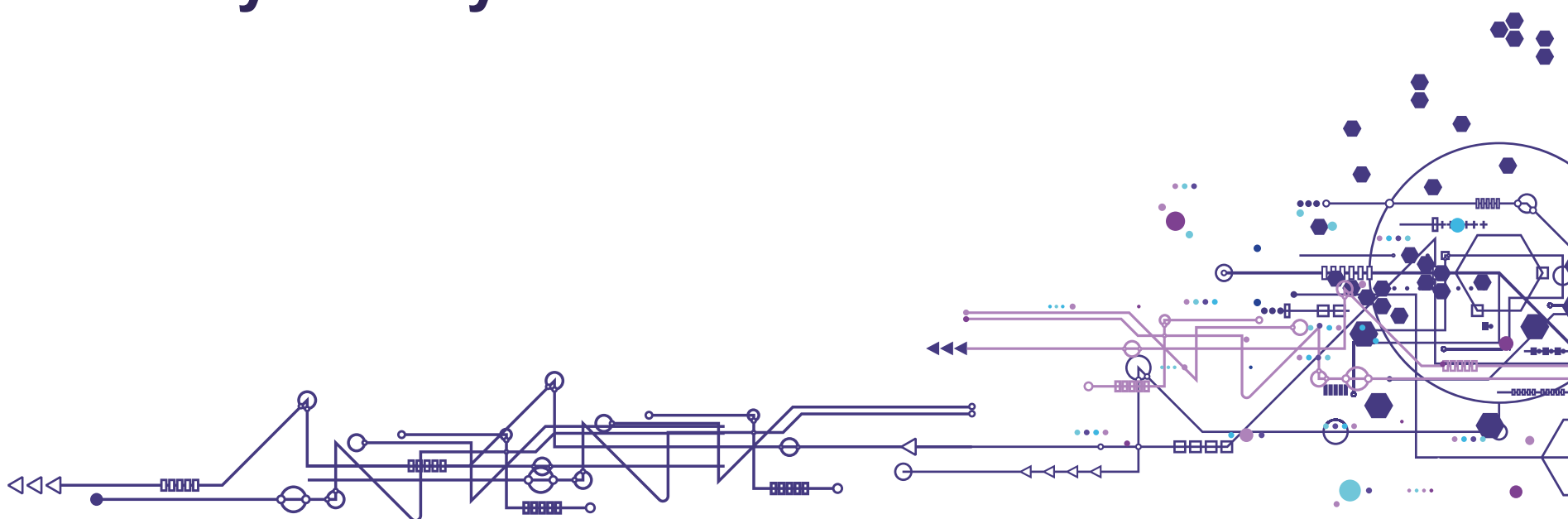
V příloze přikládám pro referenci aktuální podobu DataCite záznamu.

Michal



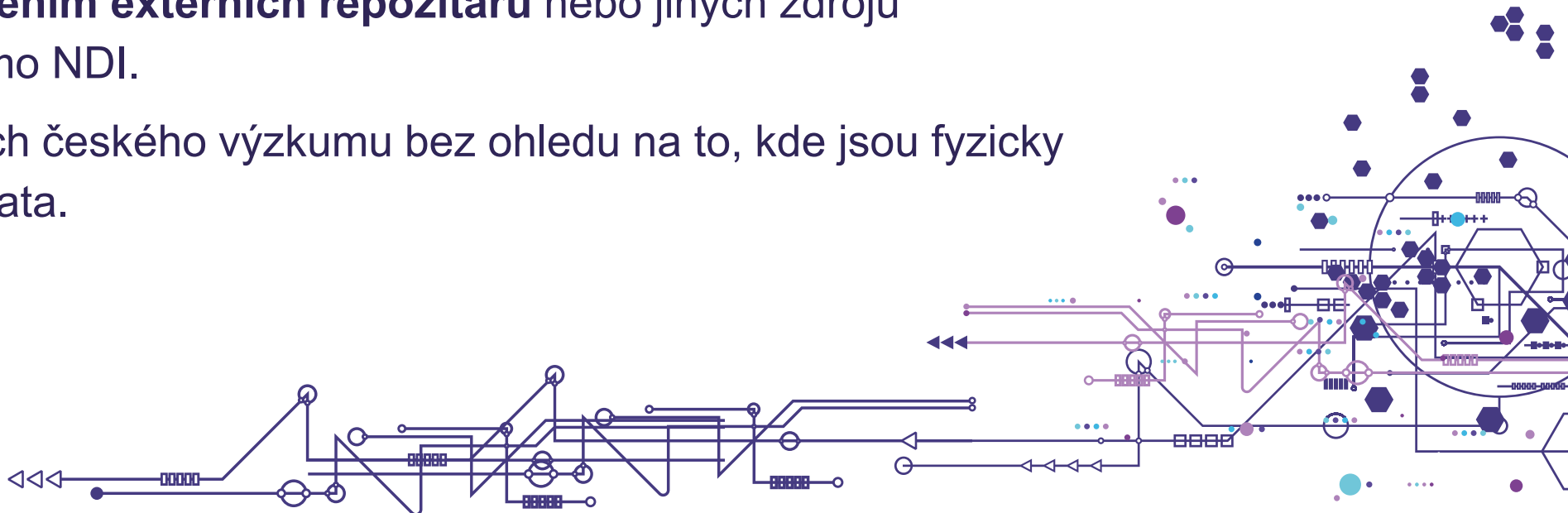


Služby pro FAIRifikaci výzkumných dat



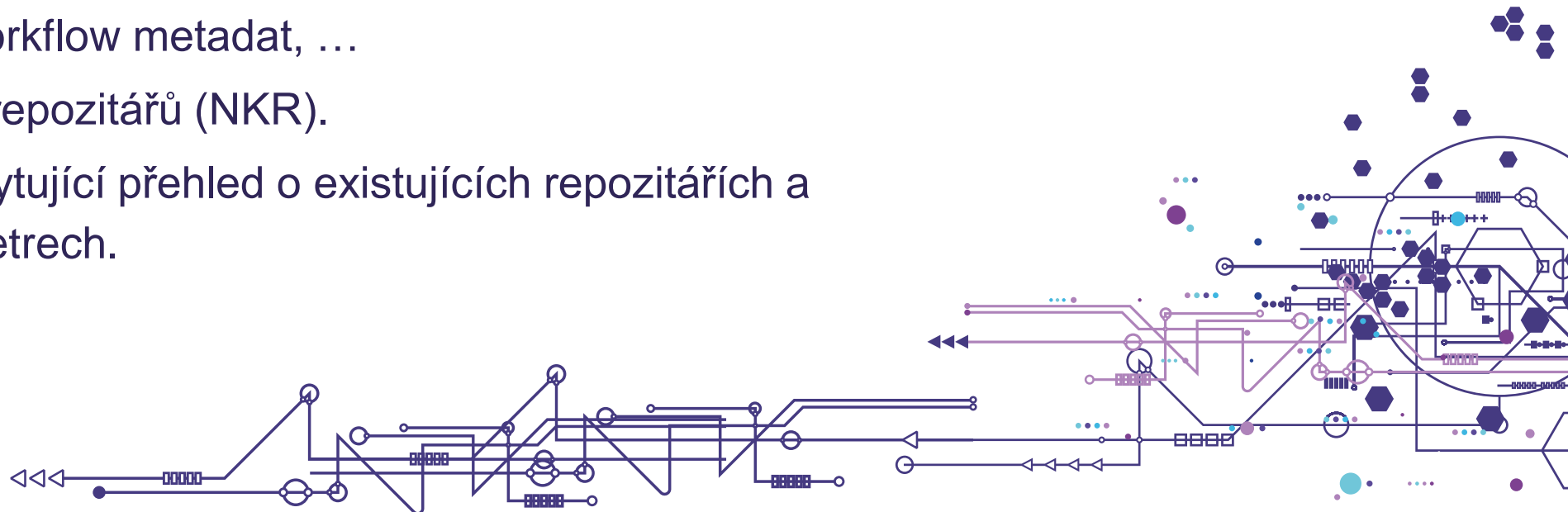
Národní metadatový adresář (NMA)

- **One-stop-shop** pro získávání **informací o výzkumných datech** produkovaných v **ČR**.
 - Důležité např. pro přehledy research funders, hodnocení výkonu a kvality výzkumných organizací apod.
 - Může to být i centralizovaný bod pro předávání metadat o českých výzkumných datech na evropskou úroveň a dále.
- Počítáme i s **napojením externích repozitářů** nebo jiných zdrojů provozovaných mimo NDI.
- Přehled o výsledcích českého výzkumu bez ohledu na to, kde jsou fyzicky uložena samotná data.

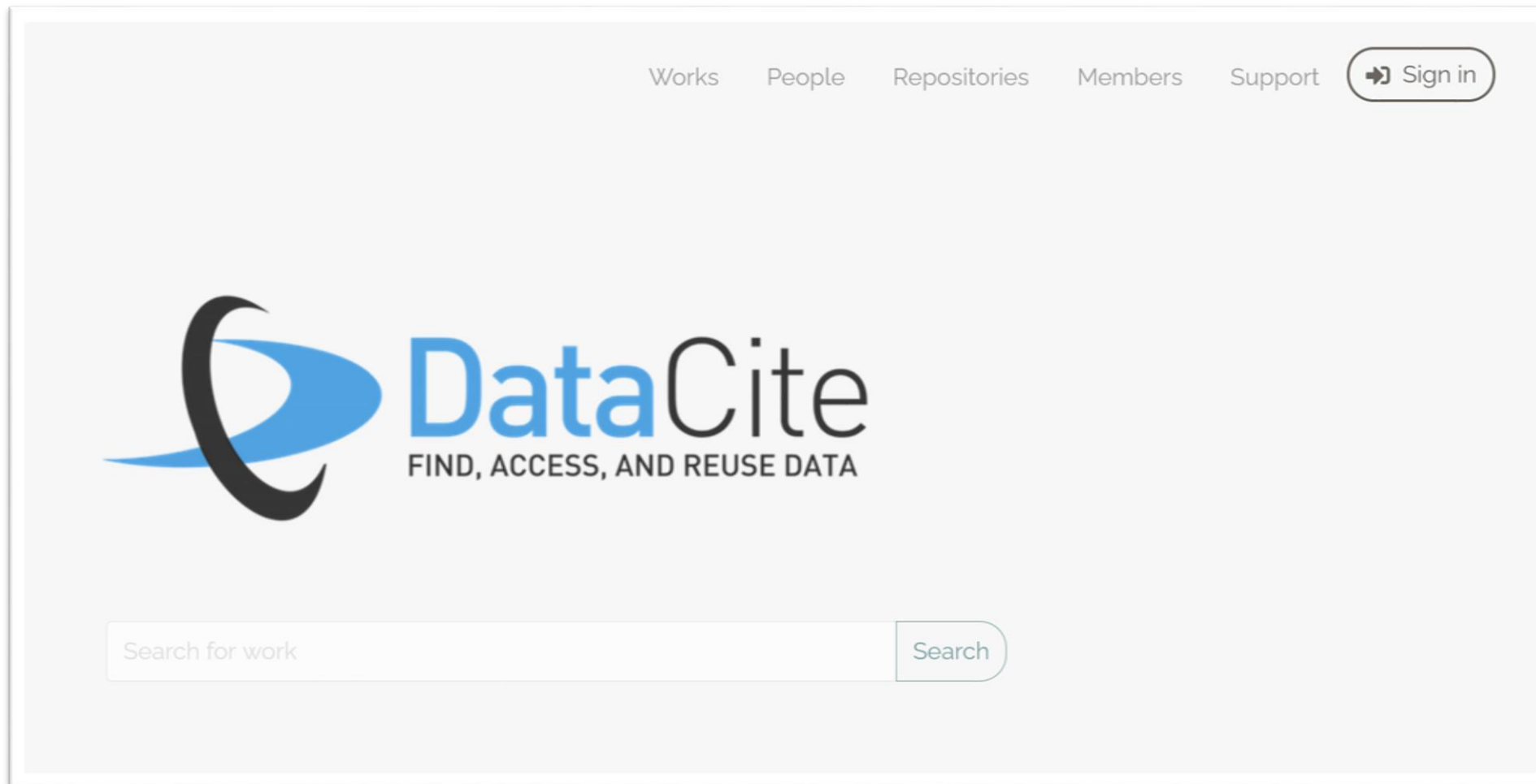


Metadatová podpora

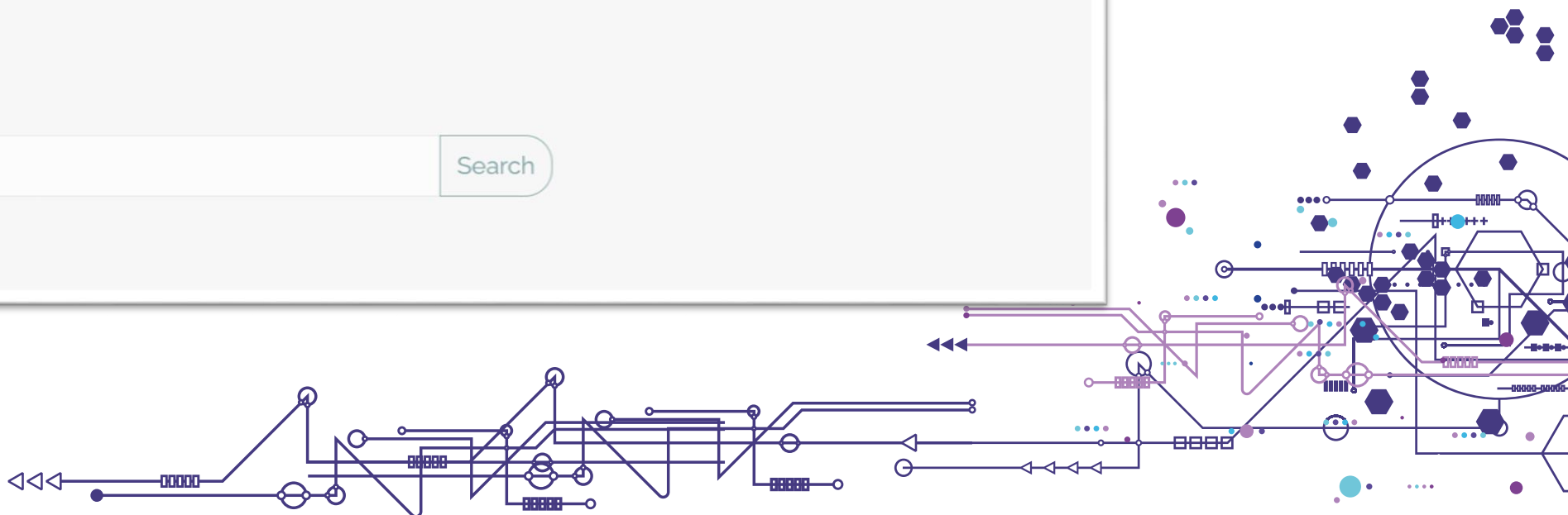
- **Součást** základních služeb **NDI** (zajišťováno z IPs CARDS).
- Primární **kontaktní bod** pro **konzultace** pro management **metadat** výzkumných dat.
- Definice **základního modelu** pro popis výzkumných dat.
 - **Minimální údaje** pro indexaci záznamů do NMA.
- Podpora pro **napojení zdrojů** na NMA.
 - Metadatové modely pro oborové clustery,
 - nastavení workflow metadat, ...
- Národní katalog repozitářů (NKR).
 - Registr poskytující přehled o existujících repozitářích a jejich parametrech.



Příklad metadatového adresáře: DataCite Search

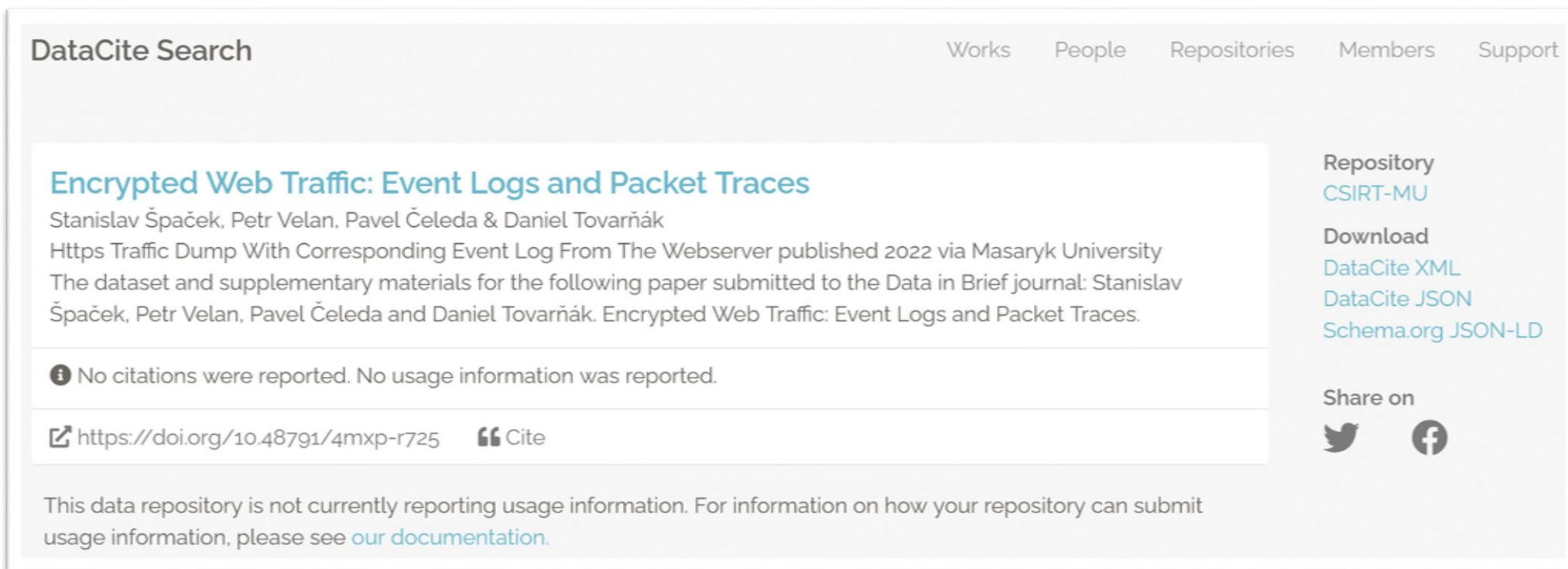


Zdroj: <https://search.datacite.org/>



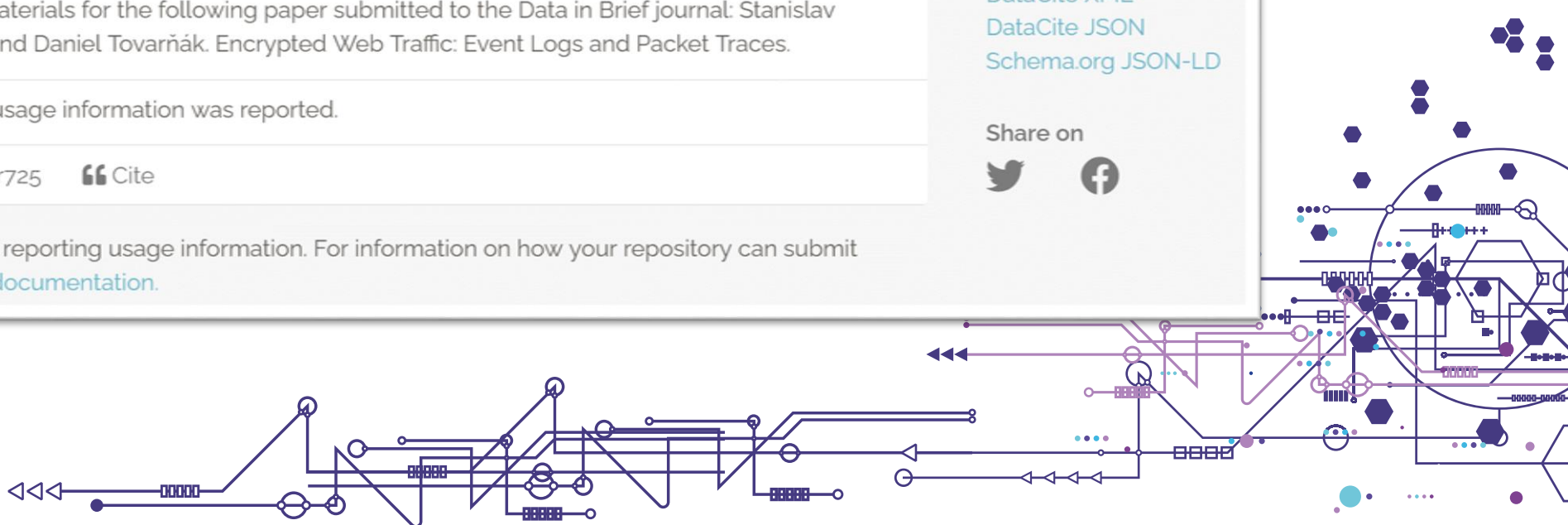
Metadata a citace dat – příklad z DataCite Search

- Přístup k metadatům – příklad *https-set*
 - <https://search.datacite.org/works/10.48791/4mxx-r725>



The screenshot shows the DataCite Search interface. At the top, there are navigation tabs: Works, People, Repositories, Members, and Support. The main content area displays the title "Encrypted Web Traffic: Event Logs and Packet Traces" in blue. Below the title, the authors are listed: Stanislav Špaček, Petr Velan, Pavel Čeleda & Daniel Tovarňák. A description follows: "Https Traffic Dump With Corresponding Event Log From The Webserver published 2022 via Masaryk University. The dataset and supplementary materials for the following paper submitted to the Data in Brief journal: Stanislav Špaček, Petr Velan, Pavel Čeleda and Daniel Tovarňák. Encrypted Web Traffic: Event Logs and Packet Traces." To the right of the main content, there is a sidebar with the following information: "Repository CSIRT-MU", "Download" options (DataCite XML, DataCite JSON, Schema.org JSON-LD), and "Share on" options (Twitter, Facebook). Below the main content, there is a message: "No citations were reported. No usage information was reported." and a "Cite" button with the DOI link: <https://doi.org/10.48791/4mxx-r725>. At the bottom of the screenshot, there is a note: "This data repository is not currently reporting usage information. For information on how your repository can submit usage information, please see [our documentation](#)."

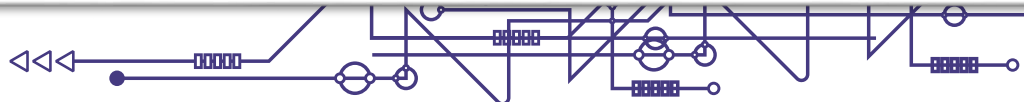
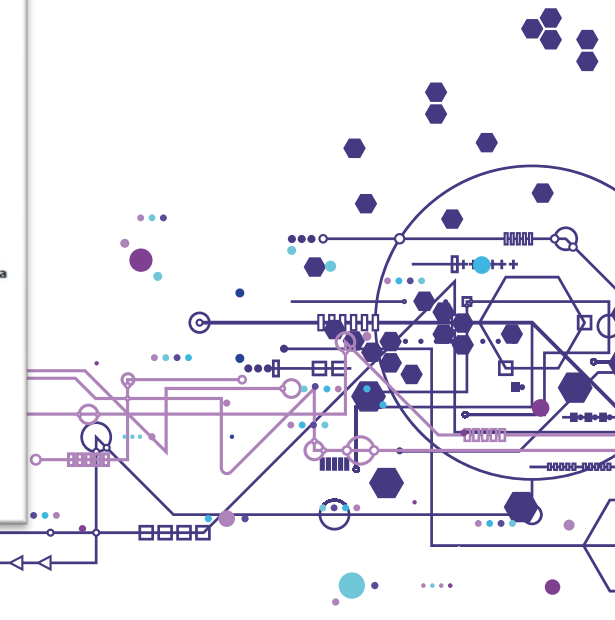
Zdroj: <https://search.datacite.org/>



Strojově čitelná metadata – příklad z DataCite Search

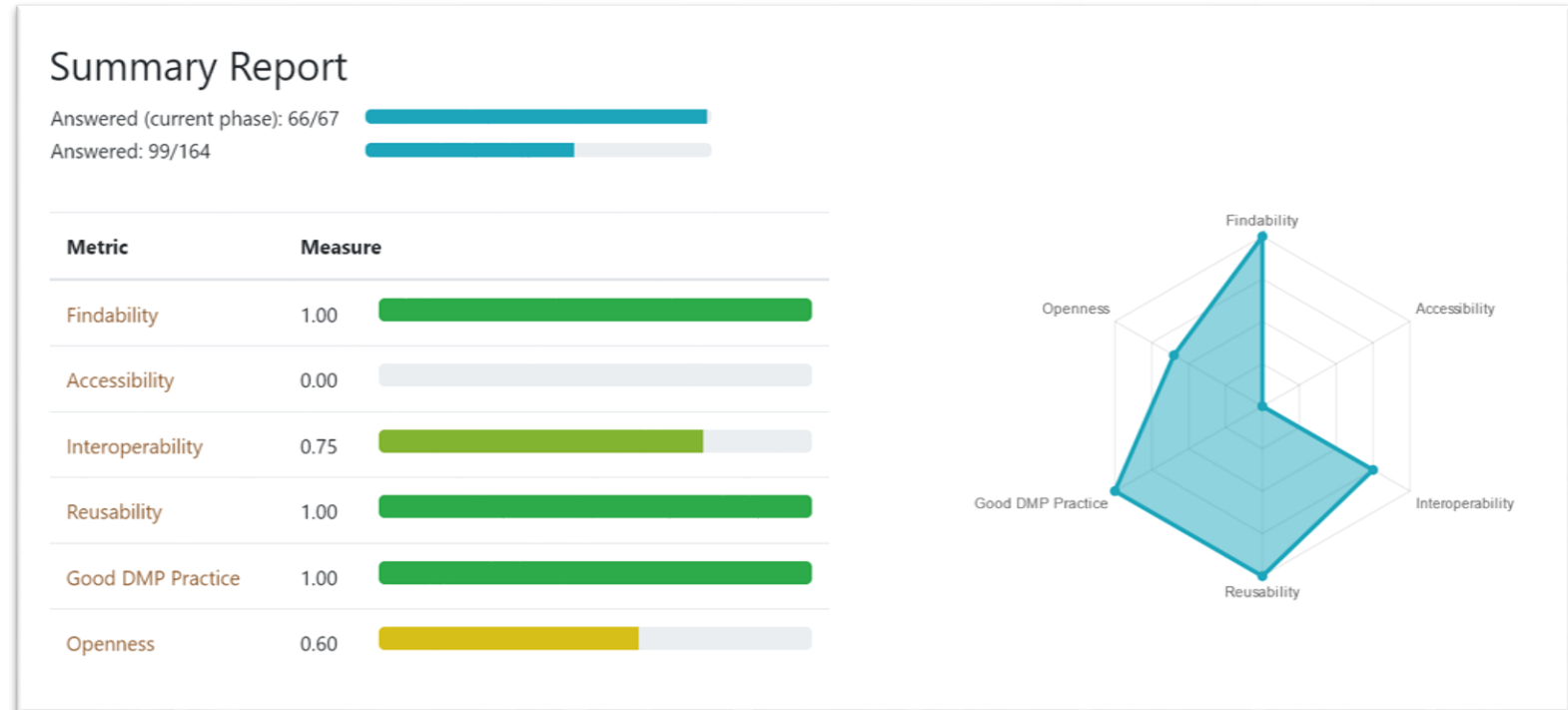


```
<?xml version="1.0" encoding="UTF-8"?>
- <resource xsi:schemaLocation="http://datacite.org/schema/kernel-4 http://schema.datacite.org/meta/kernel-4/metadata.xsd" xmlns="http://datacite.org/schema/kernel-4"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance">
  <identifier identifierType="DOI">10.48791/4MXP-R725</identifier>
  <creators>
    - <creator>
      <creatorName nameType="Personal">Špaček, Stanislav</creatorName>
      <givenName>Stanislav</givenName>
      <familyName>Špaček</familyName>
      <nameIdentifier schemeURI="" nameIdentifierScheme="Other">https://orcid.org/0000-0002-7187-5045</nameIdentifier>
    </creator>
    - <creator>
      <creatorName nameType="Personal">Velan, Petr</creatorName>
      <givenName>Petr</givenName>
      <familyName>Velan</familyName>
      <nameIdentifier schemeURI="https://orcid.org" nameIdentifierScheme="ORCID">https://orcid.org/0000-0002-2824-4299</nameIdentifier>
    </creator>
    - <creator>
      <creatorName nameType="Personal">Čeleda, Pavel</creatorName>
      <givenName>Pavel</givenName>
      <familyName>Čeleda</familyName>
      <nameIdentifier schemeURI="https://orcid.org" nameIdentifierScheme="ORCID">https://orcid.org/0000-0002-3338-2856</nameIdentifier>
    </creator>
    - <creator>
      <creatorName nameType="Personal">Tovarník, Daniel</creatorName>
      <givenName>Daniel</givenName>
      <familyName>Tovarník</familyName>
      <nameIdentifier schemeURI="https://orcid.org" nameIdentifierScheme="ORCID">https://orcid.org/0000-0002-7206-5167</nameIdentifier>
    </creator>
  </creators>
  <titles>
    <title xml:lang="en">Encrypted Web Traffic: Event Logs and Packet Traces</title>
  </titles>
  <publisher>Masaryk University</publisher>
  <publicationYear>2022</publicationYear>
  <resourceType resourceTypeGeneral="Dataset">HTTPS Traffic Dump with Corresponding Event Log from the Webserver</resourceType>
  <subjects>
    <subject schemeURI="http://www.oecd.org/science/inno" valueURI="http://www.oecd.org/science/inno/38235147.pdf" subjectScheme="Fields of Science and Technology (FOS)">FOS: Computer and information sciences</subject>
  </subjects>
  <dates>
    <date dateType="Issued">2022</date>
  </dates>
  <language>en</language>
  <sizes>
    <size>270 GiB</size>
  </sizes>
  <formats>
    <format>Network traces in Libpcap file format (see https://wiki.wireshark.org/Development/LibpcapFileFormat) and event log from IIS web servers in a newline-delimited JSON (JSON Lines) file.</format>
  </formats>
  <version>1.0.0</version>
  <rightsList>
    <rights rightsURI="https://creativecommons.org/licenses/by/4.0/legalcode">Creative Commons Attribution 4.0 International</rights>
  </rightsList>
  <descriptions>
    <description xml:lang="en" descriptionType="Abstract">The dataset and supplementary materials for the following paper submitted to the Data in Brief journal: Stanislav Špaček, Petr Velan, Pavel Čeleda and Daniel Tovarník. Encrypted Web Traffic: Event Logs and Packet Traces.</description>
    <description xml:lang="en" descriptionType="TechnicalInfo">Repository Structure - Anonymization - the anonymization folder contains the scripts and settings that were used to anonymize the data capture. - Dataset - the dataset folder contains the host-based and network parts of the dataset in two separate files. - Pcap2flow - the pcap2flow folder contains the tools, settings, and guide to convert the packet dump in the dataset into aggregated IP flows. - README.md This work is licensed under a [Creative Commons Attribution 4.0 International License: https://creativecommons.org/licenses/by/4.0/</description>
  </descriptions>
  <fundingReferences>
    - <fundingReference>
      <funderName>Horizon 2020 Framework Programme</funderName>
      <funderIdentifier funderIdentifierType="Crossref Funder ID">https://doi.org/10.13039/100010661</funderIdentifier>
      <awardNumber awardURI="https://www.concordia-h2020.eu/">830927</awardNumber>
      <awardTitle>CONCORDIA: Cyber security cOmpeteNce for Research and InnovAtion</awardTitle>
    </fundingReference>
  </fundingReferences>
</resource>
```

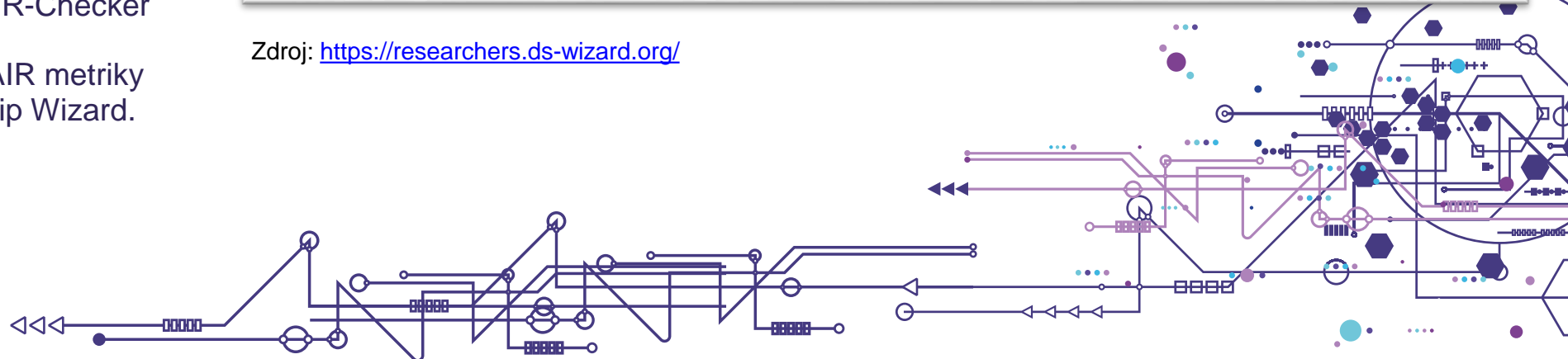


Metriky a měření plnění FAIR atributů pro výzkumná data

- Pro datasets může být zajímavé nějakým způsobem **měřit** plnění jednotlivých **FAIR atributů**.
 - NRP EOSC CZ by měla **definovat metriky** hodnocení a **poskytovat nástroje** pro **měření** plnění FAIR atributů.
- Hodnocení na základě informací o
 - datech,
 - metadatech,
 - nastavení a politikách konkrétního repozitáře apod.
- Příkladem implementace podobného nástroje může být např. FAIR-Checker (<https://fair-checker.france-bioinformatique.fr/>) nebo FAIR metriky pro DMP v Data Stewardship Wizard.



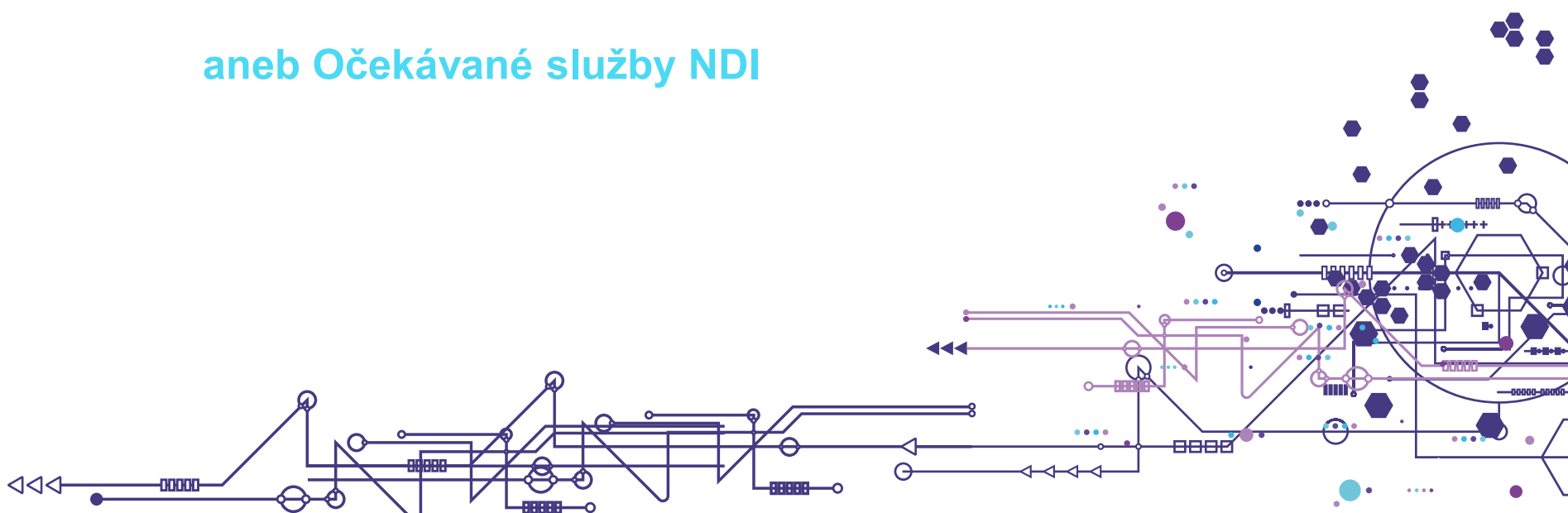
Zdroj: <https://researchers.ds-wizard.org/>





Kam s ní

aneb Očekávané služby NDI

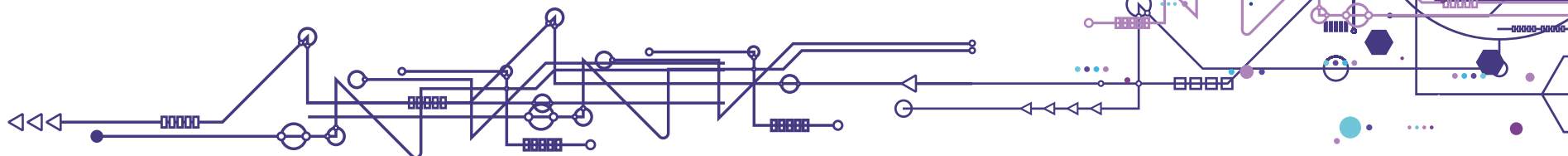


Očekávané služby NDI

- **Komplexní podpora prakticky celého životního cyklu výzkumných dat.**
 - **Uložení i velkých dat, PID, AAI, DMP, API, FAIR, automatizace, ...**
- **Uložení v ČR, řídí se zákony EU.**
- **Koncoví uživatelé využívají služeb.**
 - **Bez potřeby budování vlastního složitého technického zázemí.**
- **Součástí aktivit je také systematické vzdělávání datových odborníků.**



Zdroj: ELIXIR RDMkit, <https://rdmkit.elixir-europe.org/>





Napište nám
eosc-info@e-infra.cz

<https://www.e-infra.cz/eosc/>

A large, thin-lined circle containing the text 'e-infra.cz' in a bold, sans-serif font. The circle is partially enclosed by two curved lines on its left and right sides.

e-infra.cz