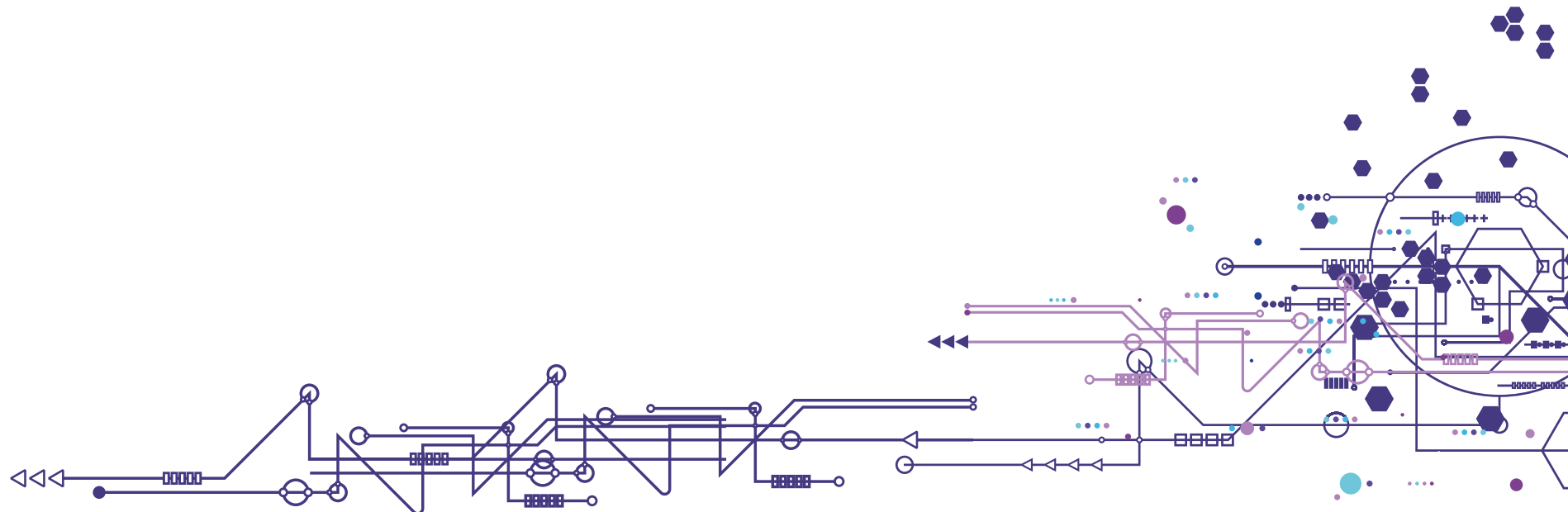


# Interakce s vědeckou komunitou

David Antoš a vedoucí tematických skupin

e-INFRA CZ/CESNET

- Jsem vědec a mám data/mám zájem o práci s daty. Případně jsem správce infrastruktury instituce/odborné skupiny. Jak se mohu zapojit?
- Různé úrovně vlivu na infrastrukturu
  - Projdeme možnosti
  - Ukázky tematických pracovních skupin



# Jak se mohu zapojit

## 1. Vůbec

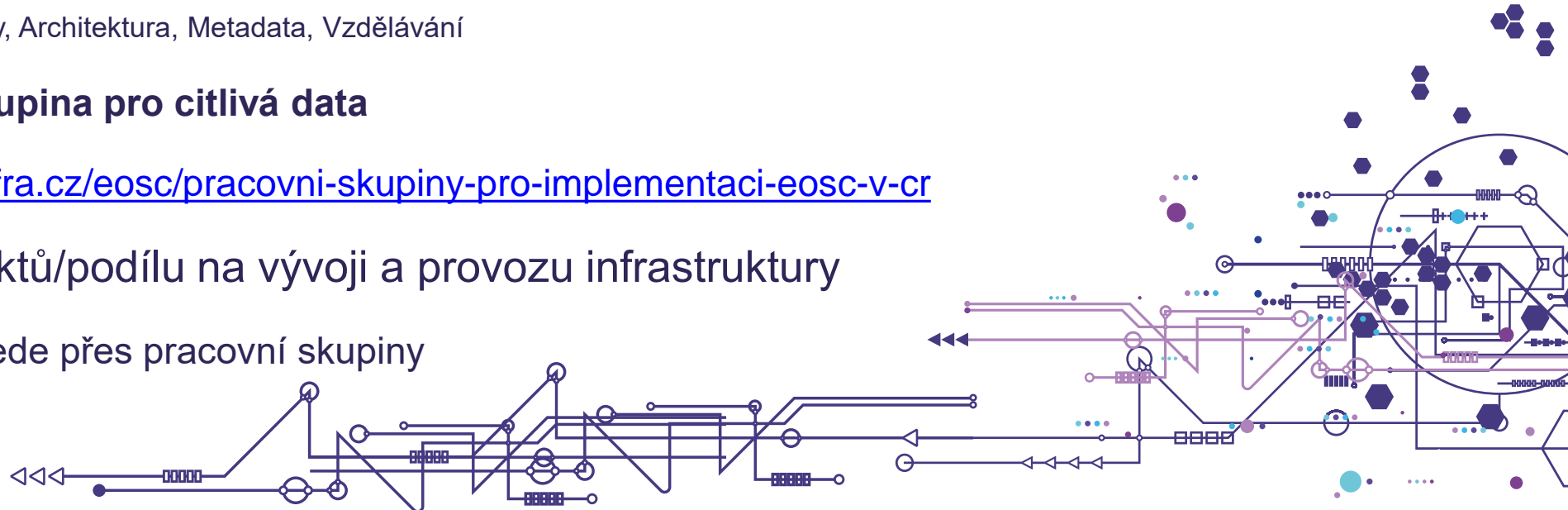
- Být „pouze uživatel“ je legitimní možnost
- Infrastruktura bude k dispozici

## 2. Do pracovních skupin: vliv na formování infrastruktury

- Průřezové
  - Základní služby, Architektura, Metadata, Vzdělávání
- **Tematické + skupina pro citlivá data**
- <https://www.e-infra.cz/eosc/pracovni-skupiny-pro-implementaci-eosc-v-cr>

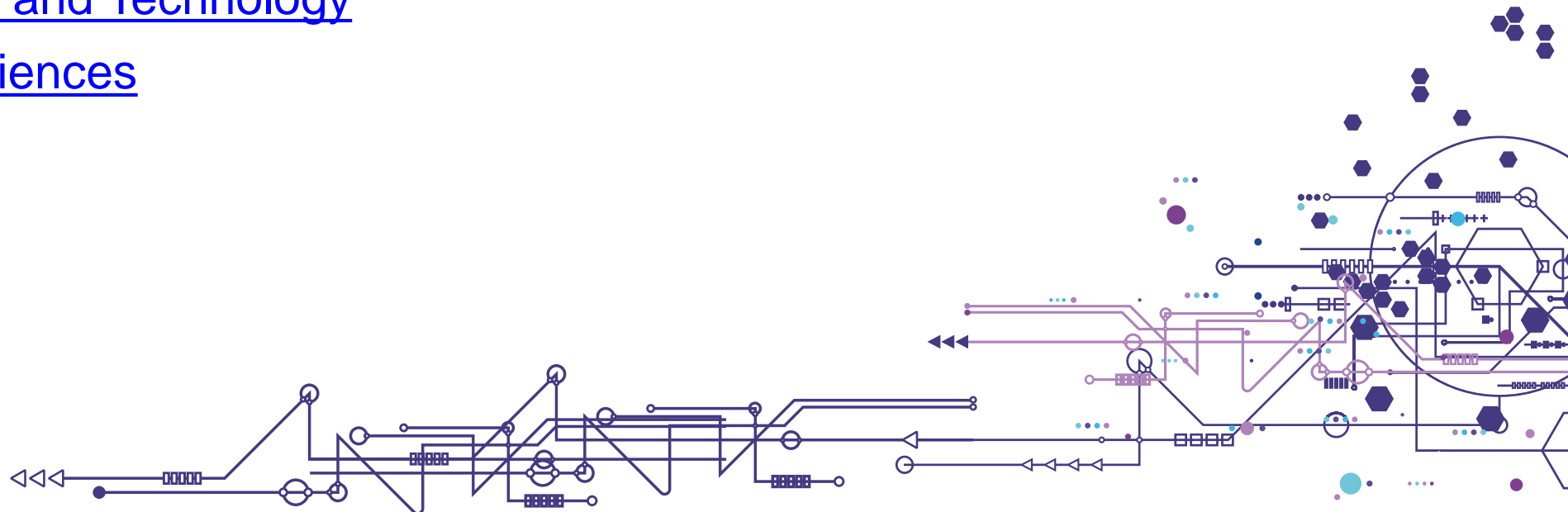
## 3. Do přípravy projektů/podílu na vývoji a provozu infrastruktury

- Vhodná cesta vede přes pracovní skupiny



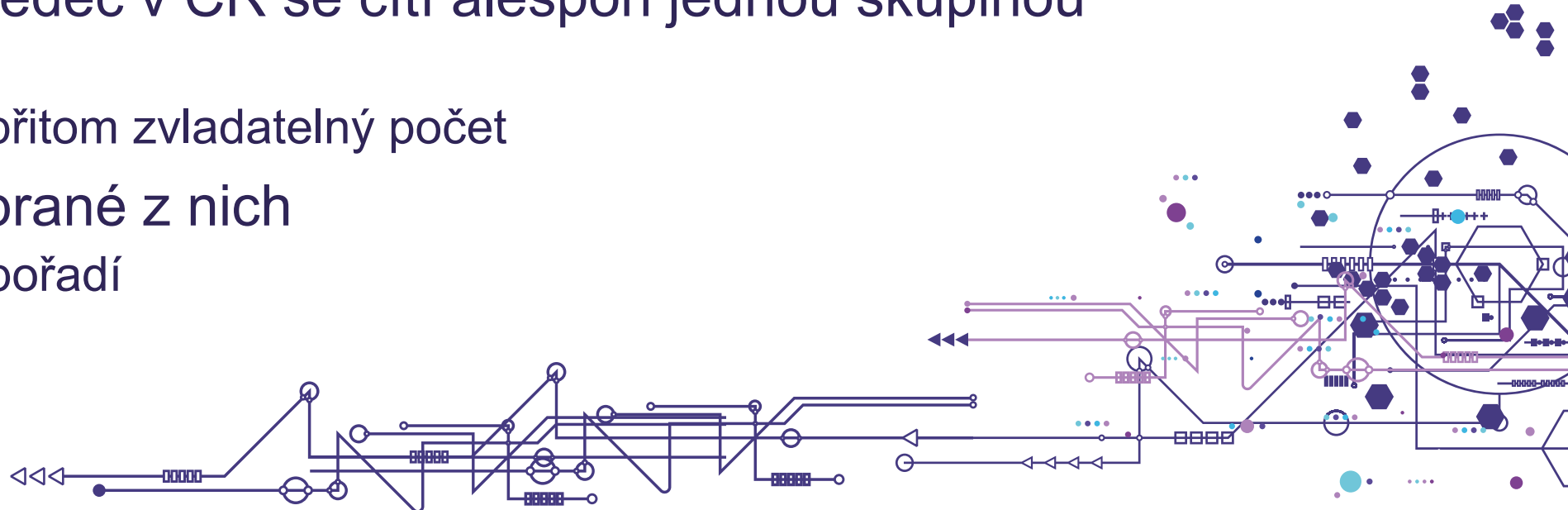
# Tematické skupiny

- [Bio/Health/Food](#)
- [Data Management for Artificial Intelligence and Machine Learning](#)
- [Social Sciences](#)
- [Physics](#)
- [Humanities and the Arts](#)
- [Material Sciences and Technology](#)
- [Environmental Sciences](#)



# Tematické skupiny – potřebují novou?

- Tematické skupiny by měly pokrývat relativně široký obor
  - Tedy ne „data pro naši katedru“
- Měly by reprezentovat obor/oborový cluster v rámci ČR
  - Optimálně s přesahem do zahraničí
- Ideál: každý vědec v ČR se cítí alespoň jednou skupinou reprezentován
  - Ale skupin je přitom zvladatelný počet
- Popíšeme vybrané z nich
  - V náhodném pořadí

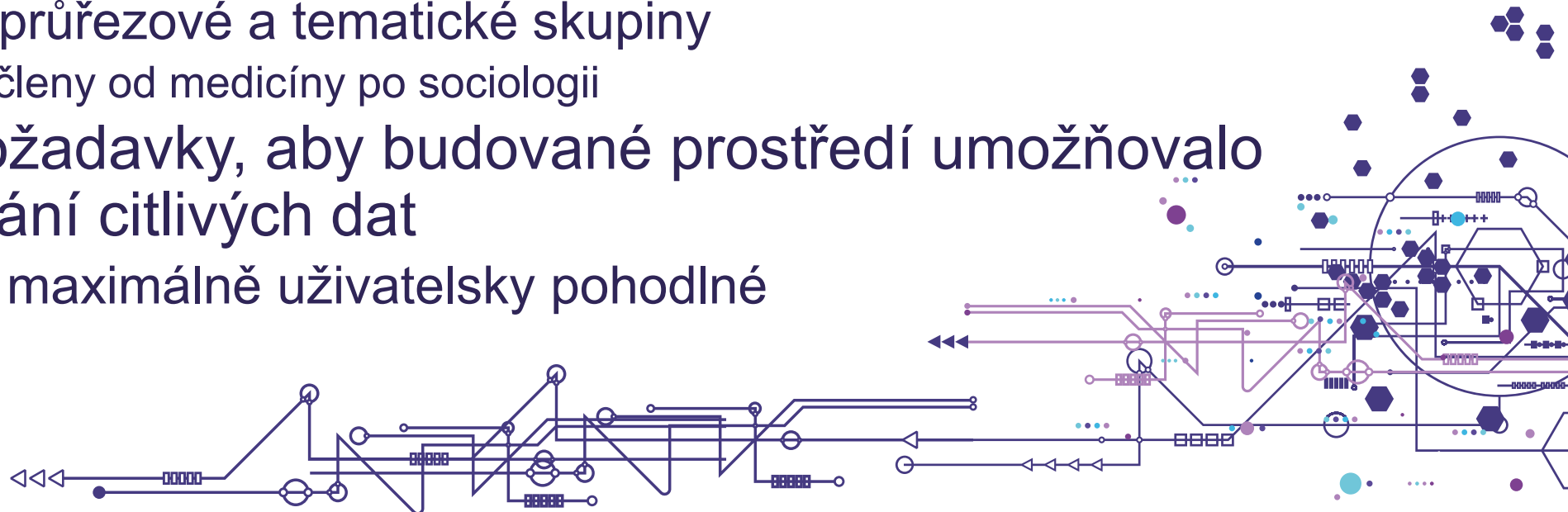


# Citlivá data

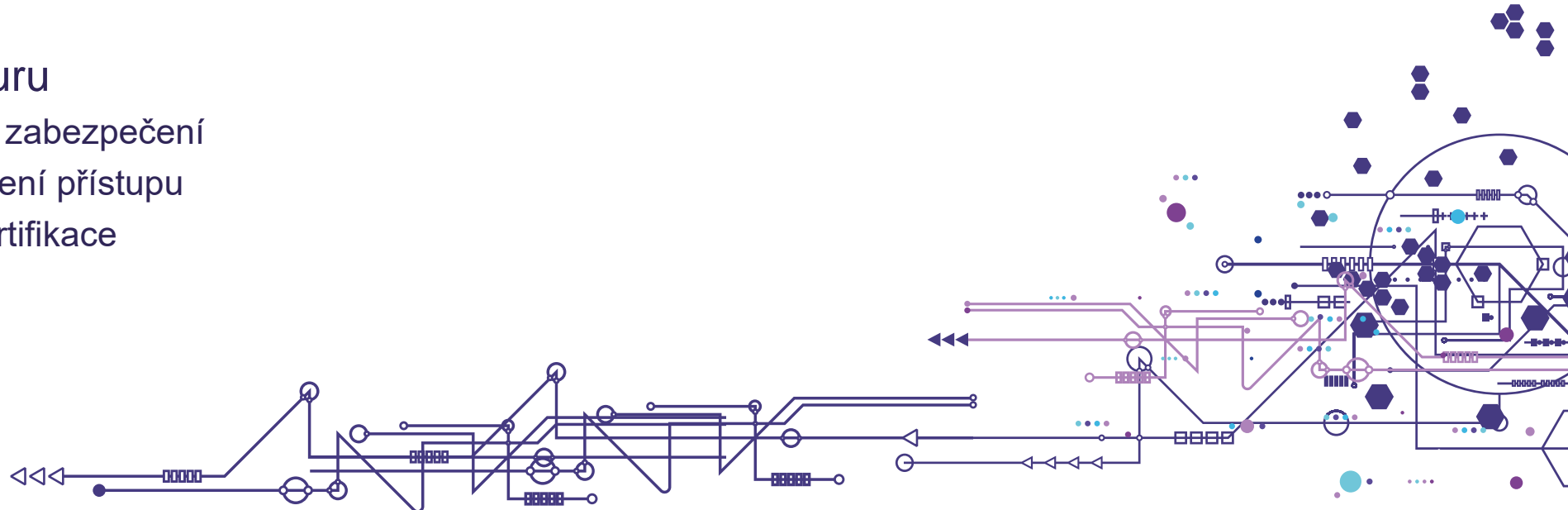
EOSC Road show

# Citlivá data

- Vedoucí: dr. Zdenka Dudová, BBMRI
- Zástupce: dr. Adam Svobodník, CZECRIN
- Citlivá data: rasa, etnikum, politika, náboženství, odbory, genetika, biometrika, zdraví a sex
- Specifické potřeby citlivých dat
  - Na pomezí průřezové a tematické skupiny
    - Zahrnuje členy od medicíny po sociologii
- Formuluje požadavky, aby budované prostředí umožňovalo legální ukládání citlivých dat
  - Přitom bylo maximálně uživatelsky pohodlné



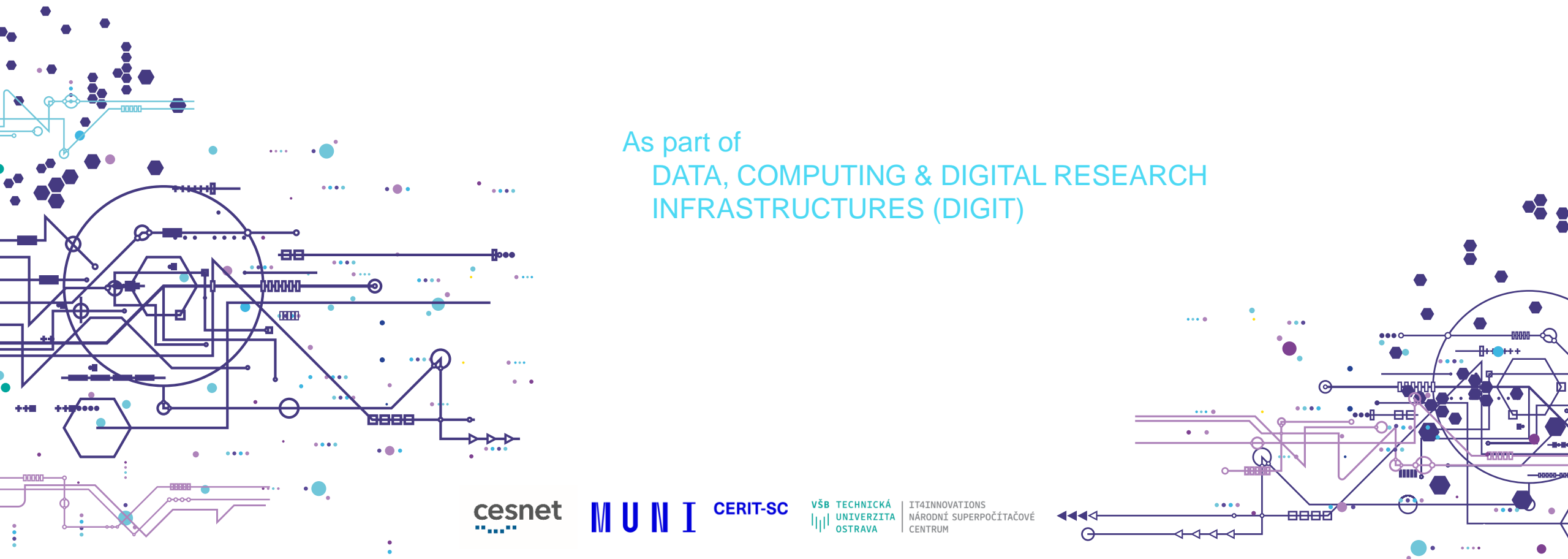
- Specifika citlivých dat
  - Jednotlivé účely zpracování
  - Z jakého právního titulu můžeme data zpracovávat?
  - Potřebujeme k tomu souhlas donora?
  - Potřebujeme k tomu speciální informovaný souhlas – např. komerční využití dat
  - Jak dlouho, kde, kdo k nim má přístup?
  - Jaká data budeme zpracovávat?
  - Jak je budeme získávat, kde uchovávat, za jakých podmínek?
  - Postavíme federovanou nebo centralizovanou infrastrukturu?
- Vliv na infrastrukturu
  - Fyzické a logické zabezpečení
  - Požadavky na řízení přístupu
  - Požadavky na certifikace





# Data Management for Artificial Intelligence and Machine Learning

As part of  
DATA, COMPUTING & DIGITAL RESEARCH  
INFRASTRUCTURES (DIGIT)



# Data Management for Artificial Intelligence and Machine Learning



**Lead:** Josef Šivic (CIIRC CVUT) <josef.sivic@cvut.cz>

**Co-leads:** Jan Martinovič (VSB), Jan Černocký/Martin Jírovec (VUT)

**Objective:** Create data infrastructure to support large-scale data together with supercomputing capabilities to develop a new generation of advanced machine learning models opening-up new data analysis applications in the exascale era.

## Tentative participants who expressed interest:

CAS (E. Pelikán, P. Jaruš), CVUT CIIRC (J. Šivic, V. Mařík), CVUT FEL (J. Matas, T. Svoboda), MFF UK (J. Hajič), MU (S. Mazurenko, J. Damborsky), VSB (J. Martinovič), VUT (P. Zemčík, J. Černocký, M. Jírovec), ZCU (J. Psutka).

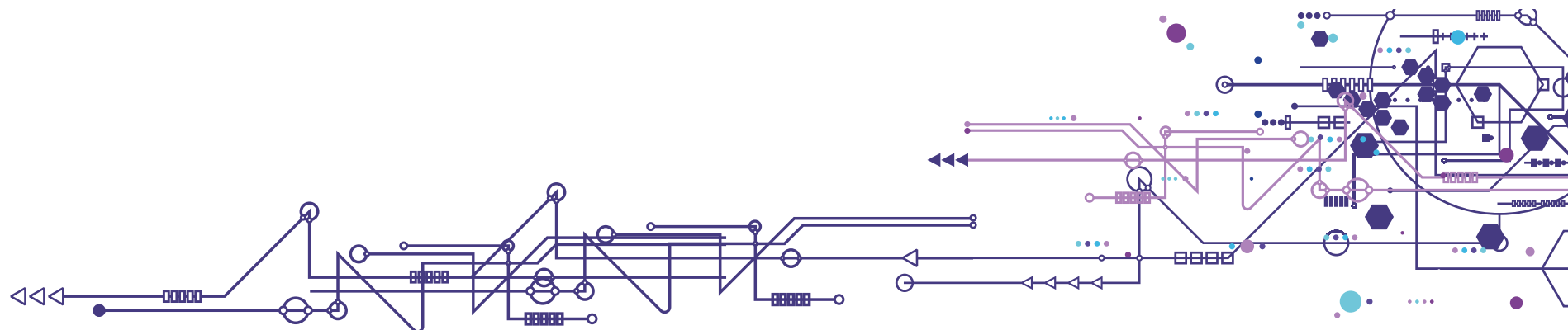


# Topics (1)

<b>Computer vision</b>	<b>3</b>
Large-scale learning for visual recognition with natural language - Josef Sivic (CVUT)	3
Perception for autonomous driving - Josef Sivic (CVUT)	4
Long-Term Visual Localization - Torsten Sattler (CVUT), Martin Cadik (VUT)	4
Gestural Language Recognition - Marek Hruz (ZCU)	5
Computer vision in traffic applications - Adam Herout (VUT)	6
OCR of historical texts combining image analysis and NLP - Michal Hradis (VUT)	6
<b>Natural language processing</b>	<b>6</b>
Pre-trained models for language transfer-learning - Jan Svec (ZCU)	6
Knowledge representations extracted from big data repositories - Pavel Smrz (VUT)	7
Very large-scale language models for EU and BRIC languages for universal use - Jan Hajic (MFF UK)	7
Large-scale machine translation models for spoken and written language - Jan Hajic (MFF UK)	8
Text analytics tools including NER and NEL - Jan Hajic (MFF UK)	8
Big data for symbolic knowledge representation - Jan Hajic (MFF UK)	8
Language data in 100+ languages for text analytics and machine translation - Jan Hajic (MFF UK)	9
<b>Automated Reasoning - Josef Urban (CVUT)</b>	<b>9</b>
<b>Speech processing</b>	<b>10</b>
Contextual representations of speech - Jan Svec (ZCU)	10
Multi-lingual speech recognition - Martin Karafiat (VUT)	11
Robust techniques for speaker characterization - Oldrich Plchot (VUT)	11
Speech data acquisition and processing for air-traffic control - Igor Szoke (VUT)	11

# Topics (2)

<b>Robotics</b>	<b>11</b>
Visual navigation using deep reinforcement learning - Jonas Kulhanek (CVUT)	11
Mobile robotics - Tomas Svoboda (CVUT)	11
<b>Simulation and analysis of complex systems</b>	<b>12</b>
Analysis of human connectome and modeling of its dynamics - Jaroslav Hlinka (CAS)	12
Geoscientific data modeling - Jaroslav Resler (CAS)	13
Network traffic analysis using machine learning techniques - Jan Korenek (VUT)	14
<b>Medical imaging</b>	<b>14</b>
2D and 3D simulation for medical imaging and ultra-sound surgery - Jiri Jaros (VUT)	14
<b>Life Sciences</b>	<b>14</b>
Analysis of protein architectures for functional annotation - Daniel Georgiev (ZCU)	14
Computational protein design - Stanislav Mazurenko, Jiri Damborsky (MU)	14
<b>Earth and Geo-Sciences</b>	<b>15</b>
Urban Change Detection - Georg Zitzlsberger (VSB)	15



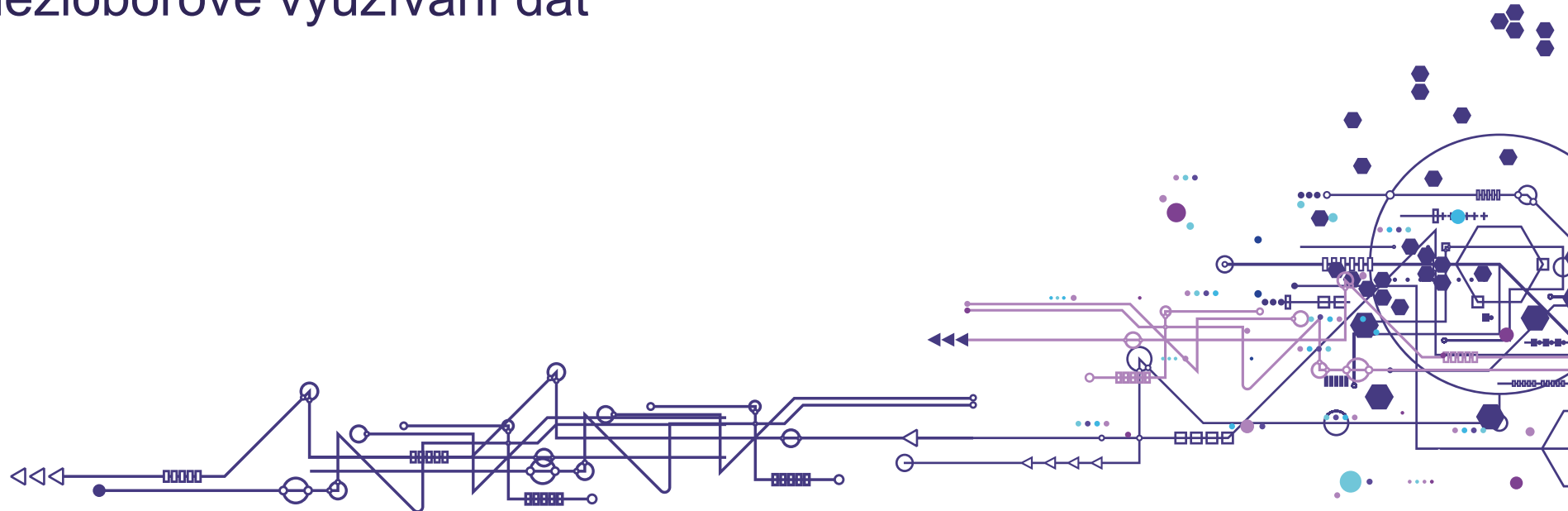
# Social Sciences

Vedoucí: Mgr. Jindřich Krejčí, Ph.D., SOÚ AV ČR

Zástupce: Mgr. Ilona Trtíková, Ph.D.

# Vymezení

- TS se zaměřuje na výzkumná data relevantní pro sociální vědy, bez ohledu na jejich disciplinární původ a metodologii
- obory: sociologie, ekonomie, politologie, psychologie a dalších disciplín sociálních věd
- významné je mezioborové využívání dat



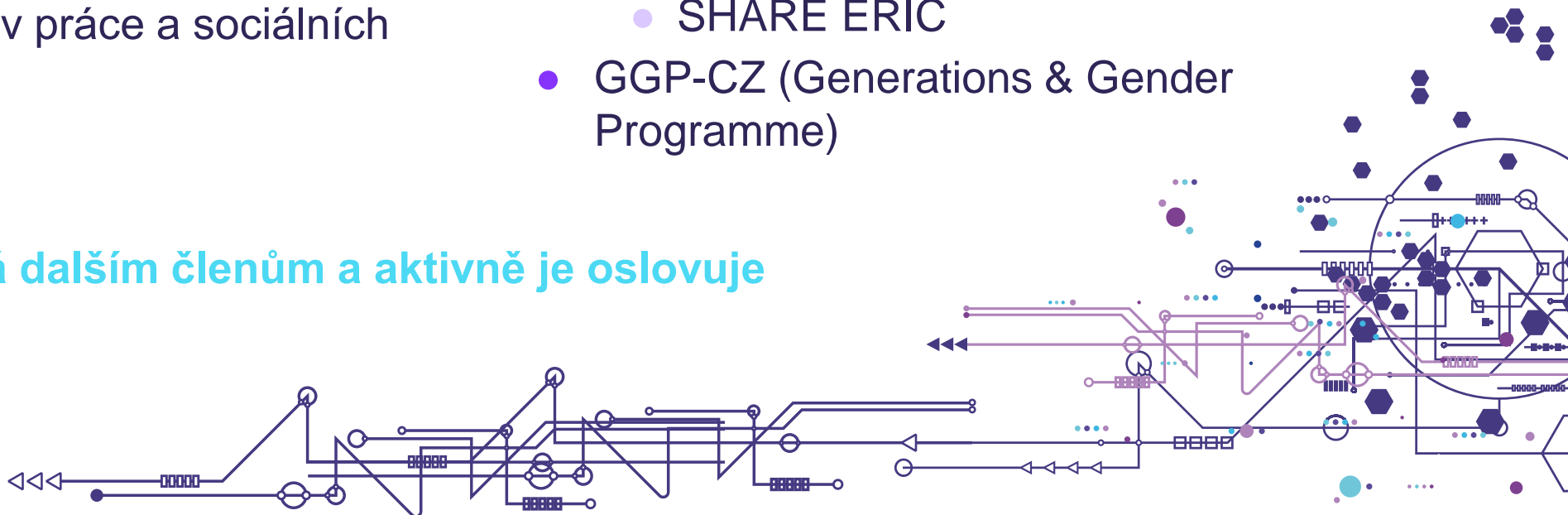
## ● Instituce

- Sociologický ústav AV ČR
- Fakulta sociálních studií MU
- Fakulta sociálních věd UK
- Katedra sociologie FF ZČU
- Provozně ekonomická fakulta ČZU
- Psychologický ústav AV ČR
- Výzkumný ústav práce a sociálních věcí
- ...

## ● Velké výzkumné infrastruktury

- ČSDA – Český sociálněvědní datový archiv
  - CESSDA ERIC
- ESS-CZ (European Social Survey)
  - ESS ERIC
- SHARE-CZ (Survey of Health, Ageing and Retirement in Europe)
  - SHARE ERIC
- GGP-CZ (Generations & Gender Programme)

Skupina je otevřená dalším členům a aktivně je oslovuje



# Cíle tematické skupiny

- prozkoumat krajinu datových zdrojů relevantních pro sociálněvědní výzkum v ČR,
- nastavit postupy a procesy implementace relevantních datových zdrojů do budovaného národního prostředí s ohledem na specifika sociálněvědních výzkumů,
- identifikovat významná mezioborová propojení v oblasti datových zdrojů a souvisejících standardů a technologických otázek práce s daty a sdílení dat relevantní pro sociální vědy,
- definovat možnosti a principy podpory datového managementu ve výzkumné praxi a odborného výcviku práce s daty s ohledem na sdílení datových zdrojů v sociálněvědním výzkumu.





# Oborový cluster fyzika

EOSC Road show

Jiří Chudoba

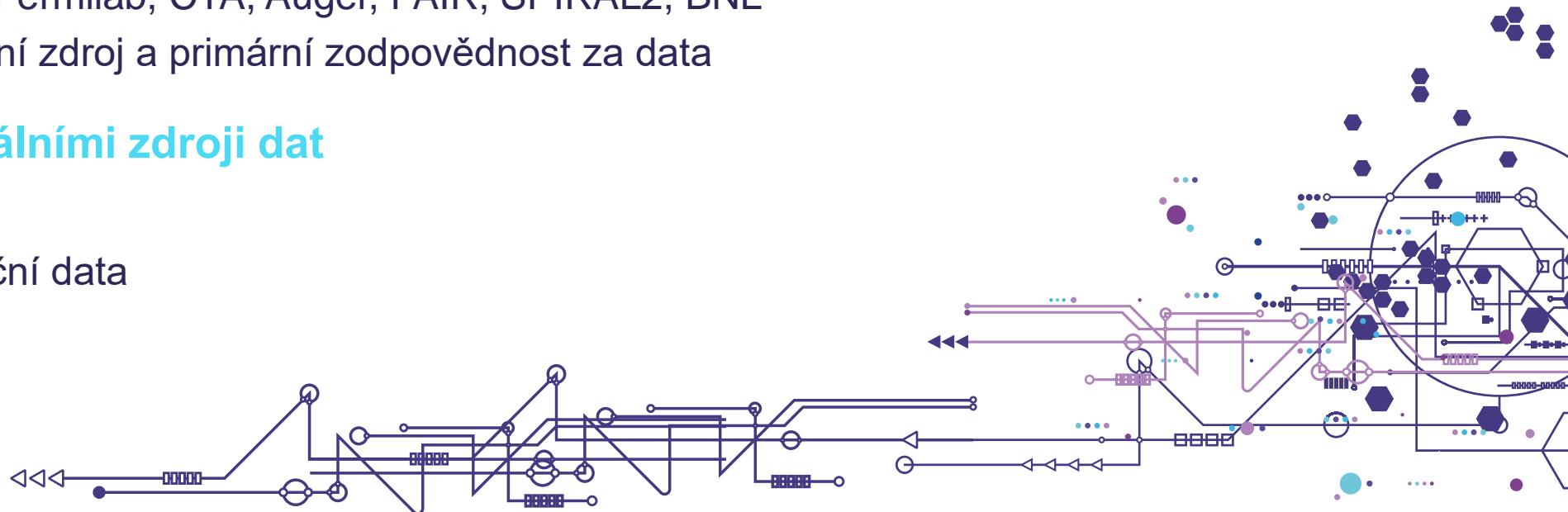
# Cluster pro VVI i pro malé skupiny

## Velké výzkumné infrastruktury ve fyzikálních oborech

- Mnohé společné znaky
  - zapojení ve velkých mezinárodních kolaboracích
  - zdroje dat zpravidla mimo ČR
  - péče o data sdílená
  - objemy dat i stovky PB (LHC experiment ATLAS – 300 PB na discích již nyní)
  - částicová, astročásticová, jaderná, laserová fyzika
    - VI CERN, Fermilab, CTA, Auger, FAIR, SPIRAL2, BNL
    - ELI – lokální zdroj a primární zodpovědnost za data

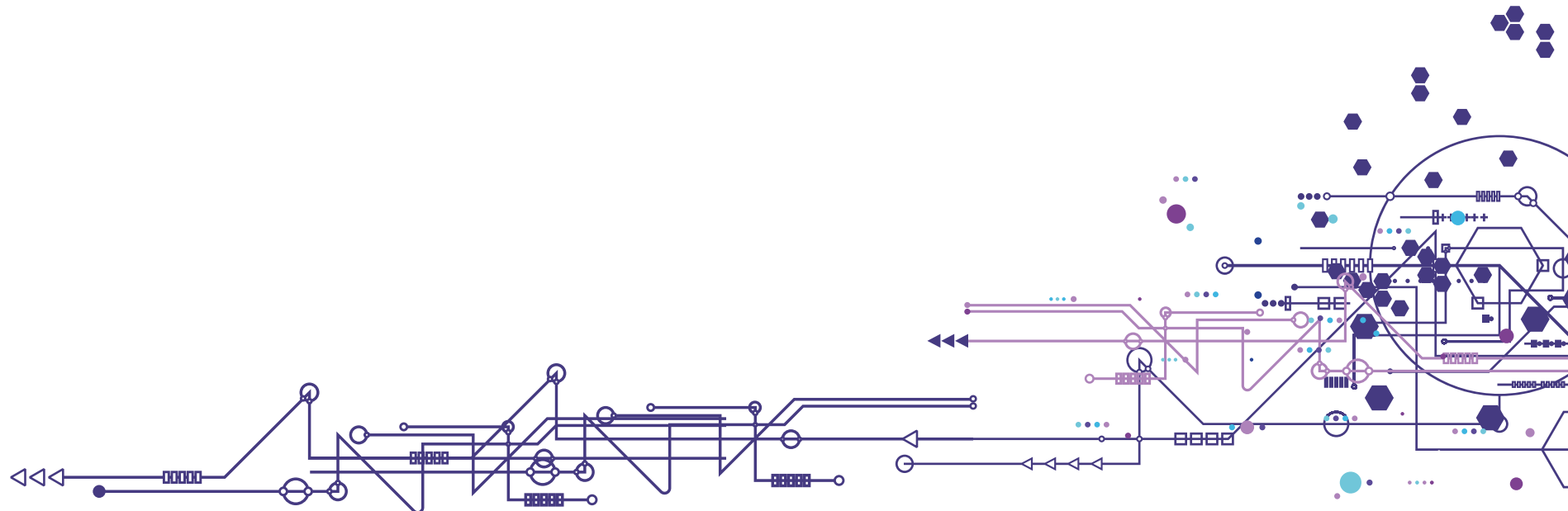
## Menší skupiny s lokálními zdroji dat

- Krystalografie
- Astronomie – sluneční data
- ...



# Cluster pro VVI i pro malé skupiny

- Cíle
  - aktivní účast na data managementu v existujících mezinárodních projektech
  - podpora vytváření datových plánů pro lokální zdroje dat
  - přenos zkušeností (nejen) mezi velkými a malými skupinami



# Národní datová infrastruktura (NDI)

## Pracovní skupina Bio/Med/Health (Health and Food)

Jiří Vondrášek

UOCHB AV ČR, ELIXIR CZ

# Koncept Bio/Med/Health (Health and Food ) WG

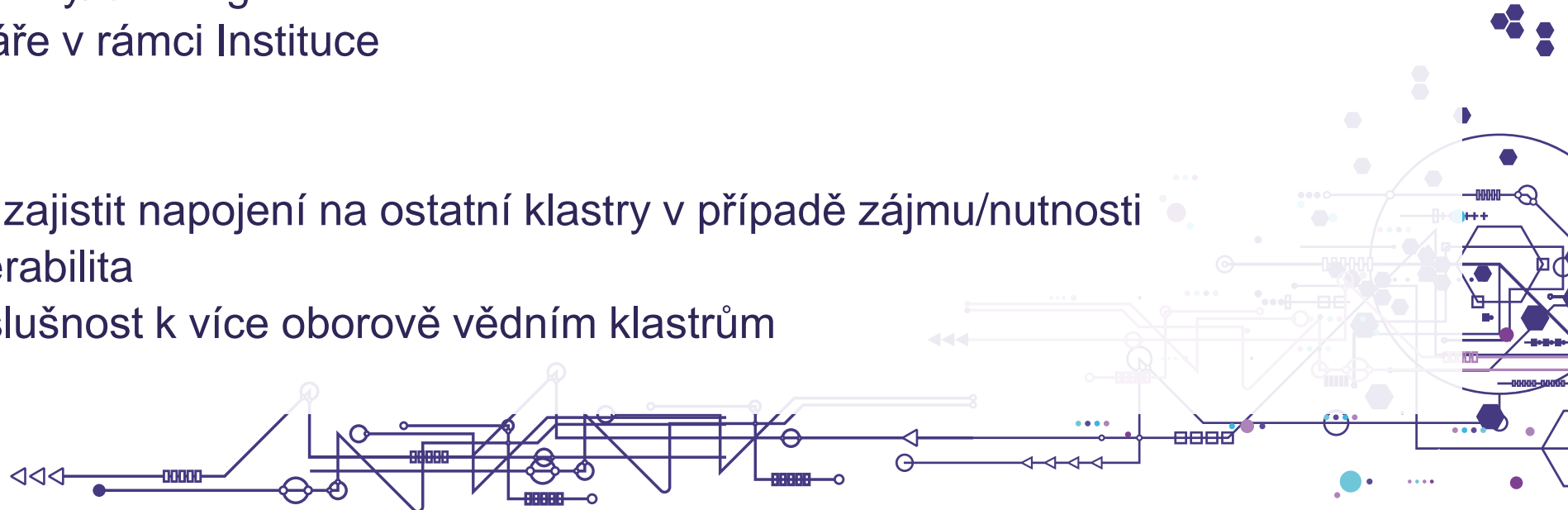
Existující repozitáře (instituce/infrastruktury) klastru  
Pouze data, která nejsou charakterizována jako **citlivá**

## Analýza

- Typ, Objem, Dostupnost dat
- Samostatné versus Součásti CZ/EU/World datových repozitářů
- Metadatové standardy/Ontologie
- Začlenění repozitáře v rámci Instituce

## Issues:

- Jakým způsobem zajistit napojení na ostatní klastry v případě zájmu/nutnosti
- Klastrová interoperabilita
- Institucionální příslušnost k více oborově vědním klastrům



# Metadatové standardy - Biologie



Animal pathology Animal physiology **Biochemistry** Biodiversity Bioengineering  
Biogeography Bioinformatics Biology Biomaterials Biomechanics **Botany** Cartography  
**Cell Biology** Chemistry Climatology Clinical Medicine Crystallography **Ecology** Entomology  
Environmental Science **Genetics** **Genomics** Geography Geology Geoscience Glaciology Hydrogeology  
Hydrography Hydrology Marine Biology **Marine Science** Marine Zoology Maritime Geography Metabolic biochemistry  
Meteorology **Molecular biology** Multi-disciplinary Neuroscience Oceanography Palaeontology Physiology  
Plant pathology Plant physiology Proteomics Soil Science Topography Toxicology **Zoology**

[ABCD - Access to Biological Collection Data](#)

[Darwin Core](#)

[EML - Ecological Metadata Language](#)

[Genome Metadata](#)

[ISA-Tab](#)

[MIBBI - Minimum Information for Biological and Biomedical Investigations](#)

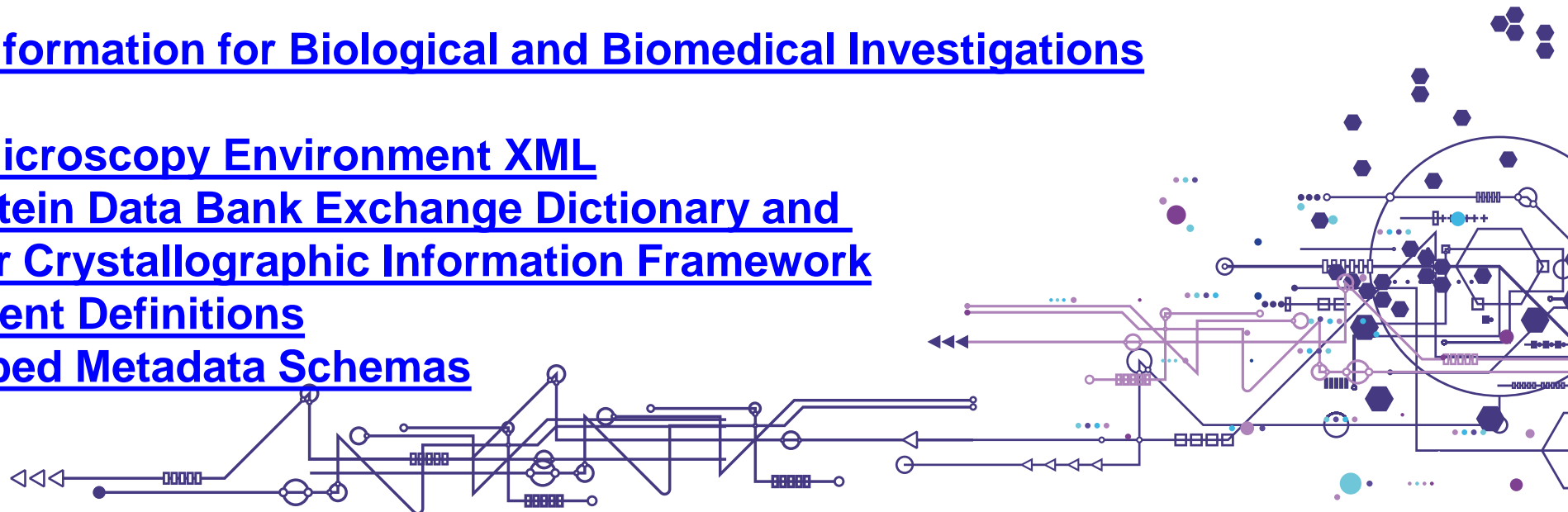
[Observ-OM](#)

[OME-XML - Open Microscopy Environment XML](#)

[PDBx/mmCIF – Protein Data Bank Exchange Dictionary and the Macromolecular Crystallographic Information Framework](#)

[Protocol Data Element Definitions](#)

[Repository-Developed Metadata Schemas](#)



## Spolupráce s WG Metadata

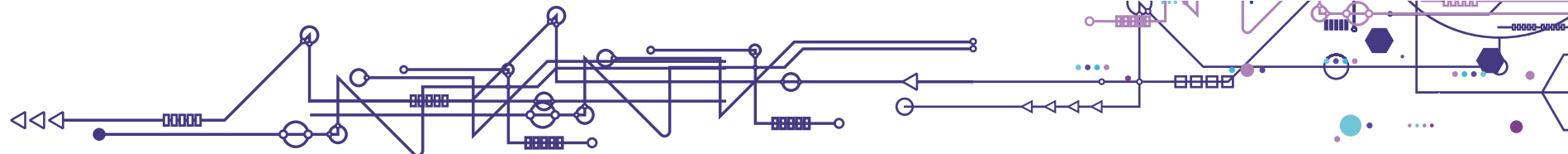
- Mapování metadatové standardy pro Bio/Med/Health
- Výstupy z ostatních WG relevantní pro Bio/Med/Health
- Diferenciace/Integrace repozitářů v rámci Bio/Med/Health klastru
- Klastrová interoperabilita



## Členské instituce/representatives

Jiří Vondrášek, UOCHB AV ČR, ELIXIR CZ  
Anna Strachotová, UOCHB AV ČR, ELIXIR CZ  
Zdenka Dudová, BBMRI.cz, MOÚ, ÚVT MU  
Roman Hrstka, MOÚ, BBMRI.cz  
Radka Svobodová, CEITEC/MUNI, ELIXIR CZ  
Regina Demlova, CZECRIN  
Petra Zemankova, CZECANCA  
Michal Kozubek, FI MU, Czech-Biolmaging  
David Bednář (FNUSA)

Karel Berka, UPOL, ELIXIR CZ  
Marian Novotný, UK, Elixir CZ  
Jan Procházka, IMG, CCP  
Ctibor Škuta, UMG AV CR, CZ-OPENSSCREEN  
Marian Hajduch, UPOL, EATRIS  
Matěj, Metrofood  
Lenka Kouřimská, METROFOOD  
Petra Zemánková, (ÚLBLD 1.LF UK a VFN)  
Petr Nehasil (ÚLBLD 1.LF UK a VFN)  
Bohdan Schneider (BTU AV ČR)





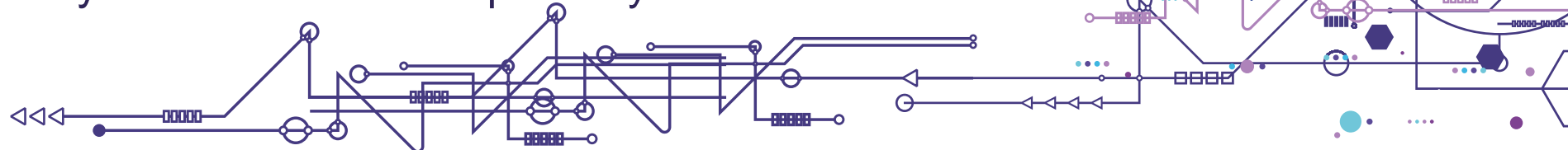
# Tematický cluster pro materiálové vědy a technologie

EOSC Road show

Marek Cebecauer

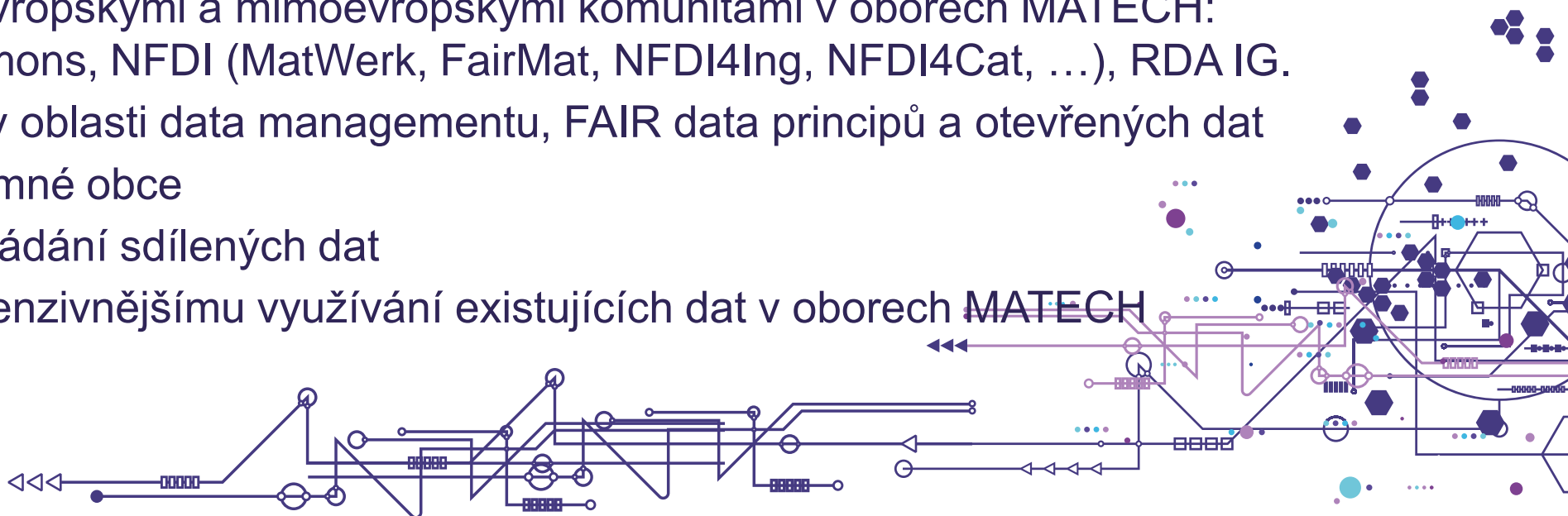
## Organizace PS pro materiálové vědy a technologie (MATECH)

- Tematická pracovní skupina pro realizaci EOSC CZ v materiálových vědách a technologiích
- Vedoucí: Marek Cebecauer (ÚFCHJH)
- Organizace zapojené v PS MATECH (vyjádřily zájem do května 2022):
  - Ústav fyzikální chemie J. Heyrovského AV ČR, Fyzikální ústav AV ČR, Ústav přístrojové techniky AV ČR, Ústav fyziky materiálů AV ČR, Ústav fotoniky a elektroniky AV ČR, Vysoká škola chemicko-technologická v Praze, Vysoké učení technické v Brně, Univerzita Karlova, Technická Univerzita v Liberci, VŠB-Technická Univerzita v Ostravě, Univerzita Palackého v Olomouci, Západočeská Univerzita v Plzni
- Zapojené infrastruktury:
  - CzechNanoLab
  - NanoEnvi
  - ...
- Spolupráce s PS pro fyziku a zahraničními partnery.



## Sdílená data v technologických oborech

- Vybudování repozitáře pro otevřená vědecká data v oblasti MATECH
  - Hardware a software pro EOSC CZ MATECH repozitář
  - Definice pravidel pro ukládání FAIR dat do EOSC CZ MATECH repozitáře
- Vytvoření prostředí pro ukládání vědeckých dat v oblasti MATECH
  - Rozvoj metadatových modelů a sémantických knihoven pro cluster MATECH
  - Prosazování principů FAIR dat v oboru
  - Propojení s evropskými a mimoevropskými komunitami v oborech MATECH: MaterialCommons, NFDI (MatWerk, FairMat, NFDI4Ing, NFDI4Cat, ...), RDA IG.
- Podpora institucí v oblasti data managementu, FAIR data principů a otevřených dat
- Vzdělávání výzkumné obce
  - Směrem k ukládání sdílených dat
  - Směrem k intenzivnějšímu využívání existujících dat v oborech MATECH



# Shrnutí

- Možnosti zapojení
  - Jako uživatel
  - Chci mít vliv na formování infrastruktury -> pracovní skupiny
  - Chci se podílet na budování a provozu -> projekt

