

MUNI

DataRaptors – aktivity skupiny

Tomáš Rebok

*Ústav výpočetní techniky
Masarykova univerzita*



Aplikačně-výzkumné spolupráce v oblasti datové analytiky

Bezpečnostní výzkum – spolupráce s Policií ČR

Environmentální výzkum – spolupráce s CzechGlobe a ČZU

Další oblasti dílčích výzkumů

- fyzika materiálů – ÚFM AV ČR
- energetika – MycroftMind

Spolupráce s Policií ČR



Současné spolupráce – datové infrastruktury

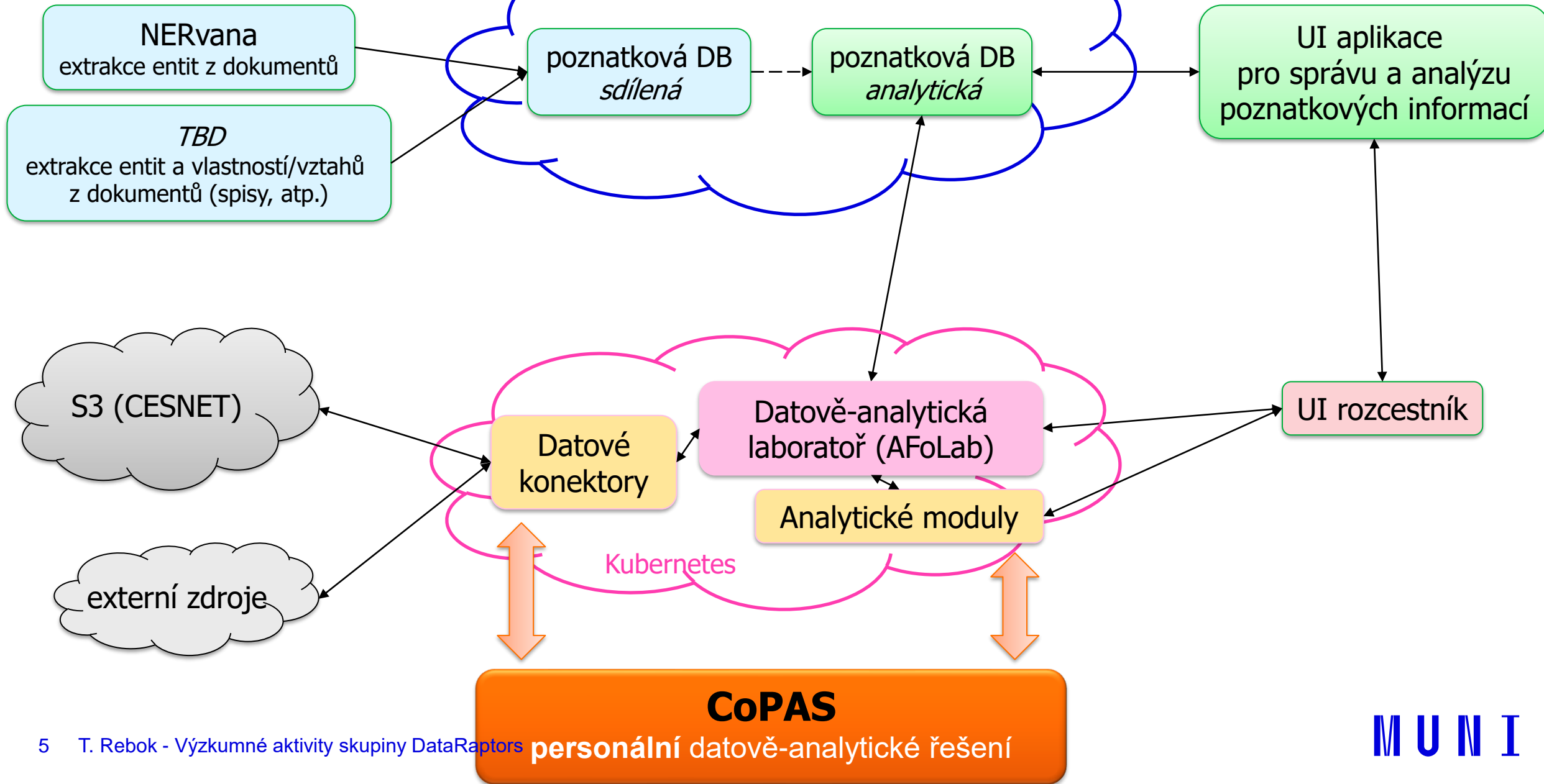
Projekt AFoLab (2023-2026) – BV MV ČR

- *Automatizovaná Forenzní Laboratoř pro odhalování komplexní kriminální činnosti*
- cloudová (týmová) infrastruktura pro zpracování a analýzu dat v prostředí PČR
 - *vlastnosti*: zabezpečení, flexibilita a modularita, identity a podpora týmové spolupráce, auditovatelnost a replikovatelnost, ...
 - zajištění celého životního cyklu dat a „stopy“ budování důkazních materiálů
 - výstupem také sada pokročilých datových analýz

„Projekt“ CoPAS

- *Cop's Personal Analytic System*
- personální platforma pro privátní, zabezpečené zpracování a analýzy dat
- inovativní myšlenka inspirovaná prostředím PČR
 - předmět připravovaného Horizon EU (PČR bude partnerem)
- cílem mj. konsolidace roztříštěných nástrojů do jednotného prostředí

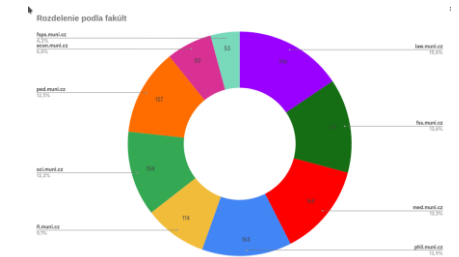
AFoLab



Současné spolupráce – výzkumná témata I.

LikPik – zpracování a analýza dat datových záchytů a komunikace

- desítky miliard semi-strukturovaných záznamů
- finalizujeme nasazení v infrastruktuře PČR
 - připravena tisková zpráva (čekáme na odsouhlasení na straně PČR)



NERvana – dolování a analýza informací z dokumentů

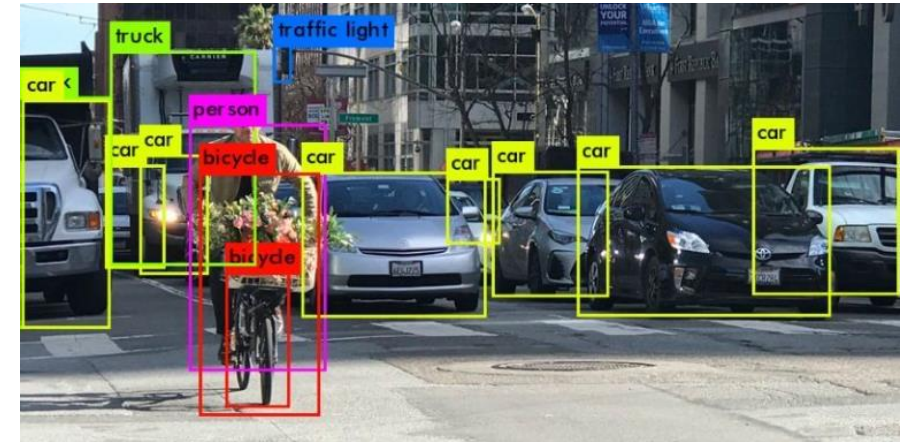
- identifikace jmen, názvů, čísel účtů, adres, emailů, datumů, ...

POZNATKY – vysoce zabezpečený systém pro správu, analýzu a prohledávání poznatkových informací

- udržování informací o subjektech v řešených případech a jejich vazbách
 - entity, vazby, atributy
 - jak v rámci případu, tak vazby na zajištěná data/dokumenty/...
- **Ize použít i jako zabezpečené úložiště entit a souvisejících metadat**
 - viz CEITEC včera

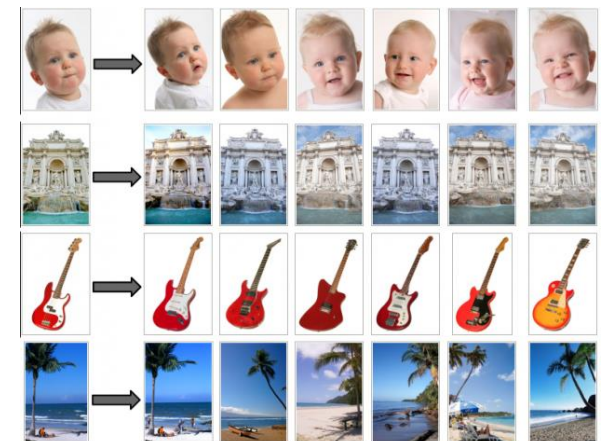
Současné spolupráce – výzkumná témata II.

Identifikace objektů zachycených na obrázcích/videích



Prohledávání obrázků a videí na základě podobnosti

- podobnost nejen dle popředí, ale také dle pozadí
- *omezení*: nutná **existence vstupu** (vzorového obrázku)



Současné spolupráce – výzkumná témata III.

Prohledávání obrázků a videí na základě textového popisu

- aplikace jazykových modelů
- vyhledávání obrázků/scén zadaných přirozeným jazykem (EN, CZ)
 - bez nutnosti vzorového vstupu
 - umožňuje jeho zobecnění, větší důraz na sémantiku obrázku

Pro ilustraci

- představte si svou domácí videotéku ...
 - 30000 videosouborů, celková délka 4000 hodin
- a v té potřebujete něco vyhledat
 - ale vůbec netušíte, kde to hledat ...

V3C

MVK

Sim Reorder



Desc: e.g.: A tennis player serving a ball on the court

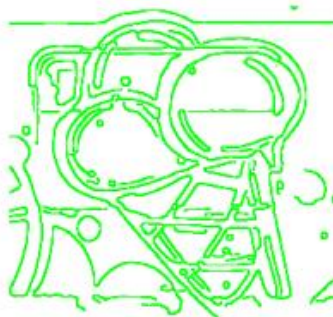
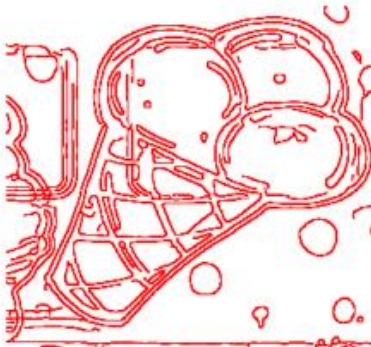
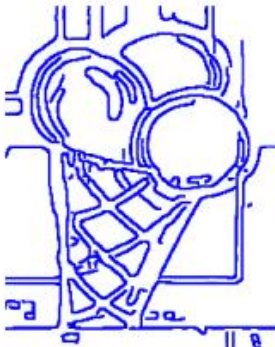
I

Současné spolupráce – výzkumná témata IV.

Analýza graffiti (pro PČR nyní velmi prestižní)

- identifikace podobných kreseb dle vizuální podoby (vč. tahů, barev)
- *pokračující cíl*: podpora při prokazování podobnosti kreseb





AI jazykové modely – další výzkumné tipy

Textový popis scén dle zachycených fotografií

- jen pomůcka vyšetřovateli (předchystání textu)
 - finalizace textu vždy v jeho odpovědnosti

Generování identikitu dle textové výpovědi (vůči databázi)

- a generování textového popisu dle fotografie
- součást podávaného OP JAK (mezisektorová spolupráce)
 - obecně zde budeme rozvíjet různé aplikace jazykových modelů

Generování datových sad pro potřeby školení

- injektáž kriminálního činu do „čistých“ dat
 - kriminální čin popsán přirozeným jazykem

Strategické směry bezpečnostního výzkumu skupiny

Datové zpracování a analýza – lokální i cloudová

- **inovace celého vyšetřovacího postupu a metod předávání důkazů**
 - připravovaný CoPAS Forensics HEU
- **konsolidace roztržitých nástrojů** do snadno-využitelného uniformního (zabezpečeného) prostředí CoPAS
 - včetně podpory uplatnění (akademických) výstupů vědy a výzkumu v praxi
- **uplatnění CoPAS pro potřeby vzdělávání a praktických školení PČR**
 - významná podpora ÚPVSP, snaha o podání projektu ERP III (zatím bez podpory)

AI a jazykové modely

- **virtuální vyšetřovatel – datově-analytický part'ák**
 - AI nejen dílčím nástrojem, ale partnerem vyšetřovatele
 - důraz nejen na výstupy dotazů, ale zejména jejich prokazatelnost
 - jednoduchá aplikace jazykových modelů má své limity
- **vize nového způsobu budoucí interakce vyšetřovatele s analytickým systémem**
 - jazykové modely nad datově-analytickou „mlátičkou“ (např. AFoLab)
 - v případě potřeby si jazykový model doptává konkrétní analýzy

Projekty

Automatizovaná forenzní laboratoř (AFoLab) – 2023-2026

- běžící, MV ČR

CoPAS Forensics – připravovaný Horizon EU

- již předjednané zapojení PČR, NSZ, policie SR
- zapojení dalších složek: Německo, Chorvatsko, Slovinsko, Polsko, Švýcarsko, ...

OP JAK

- jeden z připravovaných MU projektů zaměřen na víceoborový forenzní výzkum
- jsme součástí, navázána spolupráce s doc. Urbanovou (vedoucí Ústavu antropologie na PřírF MU)

CoPAS Education – interní rozvojový projekt PČR (ERP III)

- nevyšel, ale čekali jsme to

Probíhající příprava memoranda o spolupráci mezi PČR a MU

- policejní prezident – rektor
- zapojení do vznikajících skupin pro uplatnění AI v PČR a aplikace inovací v prostředí PČR

Environmentální výzkum

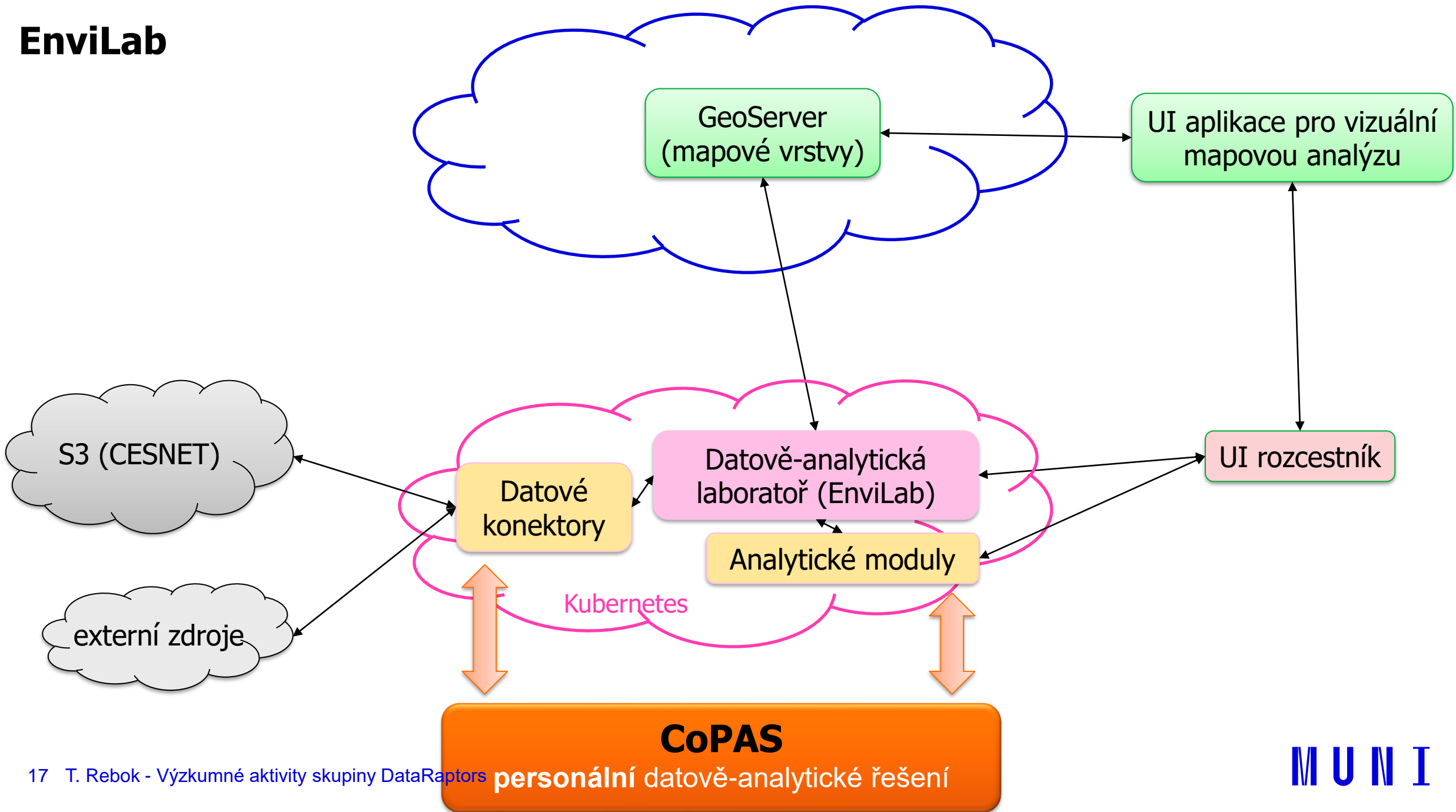


Spolupráce s CzechGlobe: projekt EnviLab

Národní datově-analytická infrastruktura pro zpracování, sdílení a analýzu dat o přírodě a krajině

Environmentální komunitě a veřejnosti (*citizen-science*) přineseme prototyp řešení, které:

- poskytne jednotný prostor pro přístup k vhodným datovým sadám
- poskytne sadu zajímavých a potřebných datových analýz
- umožní integraci a konsolidaci výsledků jiných projektů a platform



Témata EnviLab

Bezoblačné mozaiky

- vyvinuto již dříve

Analýza vegetačních parametrů

- dříve jednoduchý model, nyní snaha o zlepšení s využitím AI (Terka)

Kůrovcová kalkulačka

- součást i podaného projektu

Dendronetwork

- evropsky unikátní síť měřících stanic monitorujících „dýchání“ stromů

Radarová mapa ze snímků družice Sentinel-1 (s ČZU)

- předzpracování aktuálně nedostupné datové sady
- otevírá možnosti jiných aplikací (zájem má i Ministerstvo dopravy)

EnviLab – projekty

Přijat projekt FR CESNET na vývoj EnviLab

- hlavní nositel CzechGlobe
 - nicméně 2/3 financí pro nás

Konsolidace výstupů jiných projektů

- **Dendrometry (CzechGlobe)** – měření „dýchání“ stromů v různých lokalitách
 - smluvní výzkum – aktuálně podpis smlouvy
- **Šíření kůrovce (CzechGlobe)** – podaný projekt TAČR Sigma
 - na výsledcích už ale začínáme pracovat („kůrovcová kalkulačka“)
- **Platforma EnviroHrozeb (ČZU)** – nepřijatý projekt TAČR Sigma
 - vyústilo alespoň v první propojení s Ministerstvem dopravy
- **Hodnocení mechanické stability stromů** – nepřijatý projekt MZem (ZEMĚ II)
 - administrativní chyba
 - *naše téma*: tvorba digitálních dvojčat lesů – stacionární LiDAR, dron, letadlo

EnviLab – projekty „na střeše“

eLTER – nabídka realizace datově-analytických core-facilities vznikající evropské výzkumné infrastruktury

- jednotné EU místo pro data produkovaná sítí měřících stanic
- první reakce na naši nabídku pozitivní
 - jednání však teprve probíhají
- může shořet na financování
 - představa eLTER je 50% spoluúčast z národních zdrojů
 - MŠMT zájem o realizaci v ČR má, ale peníze zatím slíbit nedokázalo

„Zdravá města“

- úvaha většího společného projektu s CzechGlobe
- zatím bez podrobností, výzva teprve koncem roku

M A S A R Y K O V A
U N I V E R Z I T A