

NTK

50°6'14.083"N, 14°23'26.365"E
Národní technická knihovna
National Library of Technology

Repozitář vědeckých dat - *projekt CESNET/NTK*

Jan Pokorný

Národní technická knihovna, 20.11.2015



evropský
sociální
fond v ČR



EVROPSKÁ UNIE



OPERAČNÍ PROGRAM
LIDSKÉ ZDROJE
A ZAMĚŠTNANOST

PODPORUJEME
VAŠI BUDOUCNOST
www.esfcr.cz

Výstupy vědecké a akademické činnosti

1. Odborné články a publikace
2. Multimediální dokumenty (obrazové, zvukové, video apod.)
3. Binární soubory

1. Odborné články a publikace

- Uloženy ve standardních textových formátech (PDF, DOCX, TXT, INDD, ...)
- Vysoká standardizace metadatových schémat používaných k jejich popisu (MARC, DC, XMP, RDF)
- Široké možnosti indexování metadat i samotného fulltextu
- K dispozici množství IS pro správu a zpřístupnění (dokumentové repozitáře, e-archivy, AKS)

2. Multimediální (netextové) dokumenty

- Obrazové dokumenty, zvukové a video nahrávky, digitalizované 3D objekty nebo dokumenty se smíšeným obsah
- Uloženy ve standardních dokumentových formátech (JPG, TIFF, MP3, AVI, ...) nebo kontejnerových formátech (PDF, DOCX)
- Vysoká standardizace metadatových schémat používaných k jejich popisu (MARC, DC, XMP, RDF, EXIF, ID3, ...)
- Indexování metadat i samotného obsahu pomocí analýzy binárního objektu
- K dispozici IS pro správu a zpřístupnění (multimediální repozitáře)

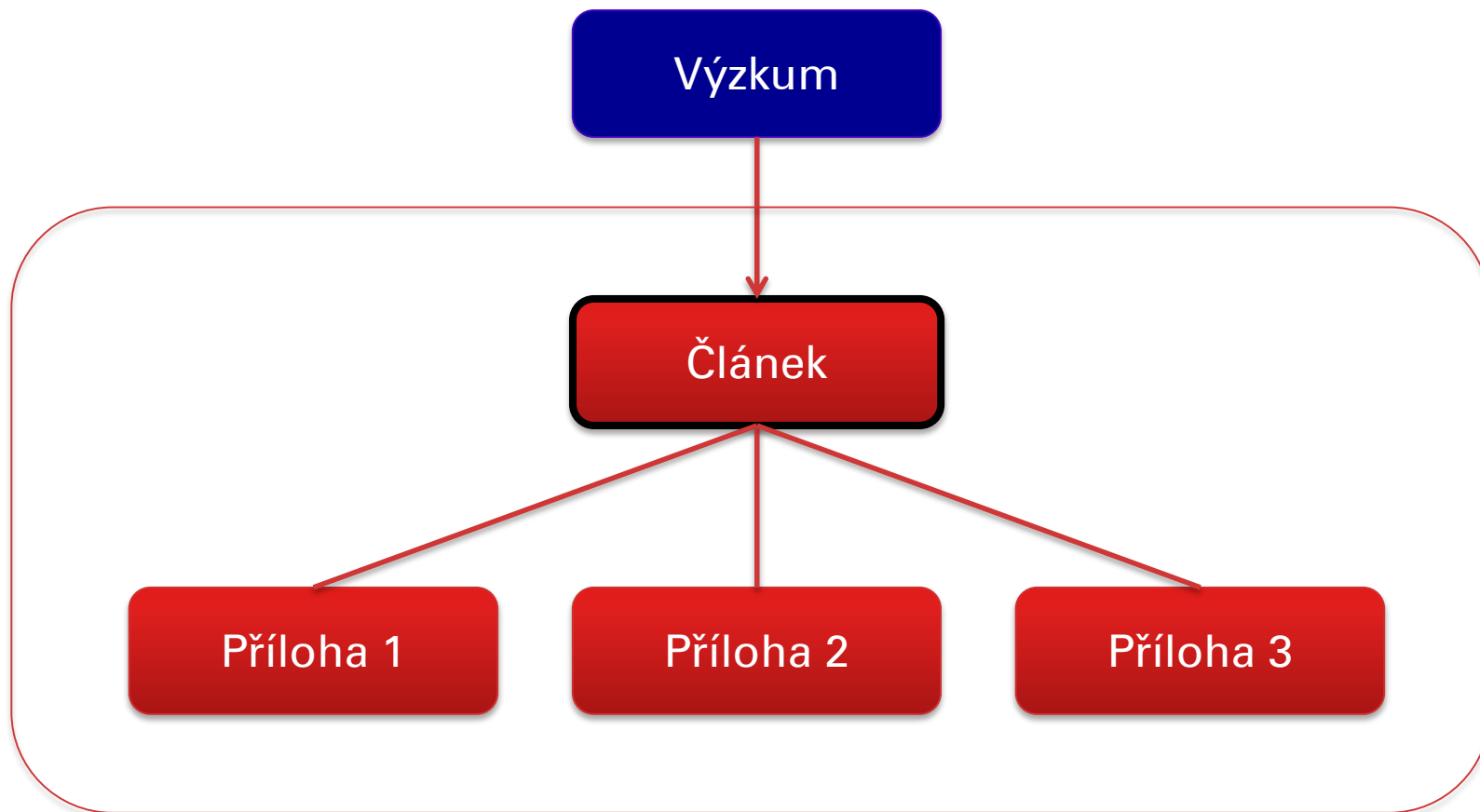
3. Binární soubory

- Soubory ve formátech, které jsou často proprietární pro danou vědeckou oblast nebo konkrétní software či přístroj
- Formátová metadata jsou proprietární nebo úplně chybí
- Lze ukládat do datových repozitářů, ale často není k dispozici dostatečný popisný aparát, který by umožnil efektivní organizaci a vyhledávání

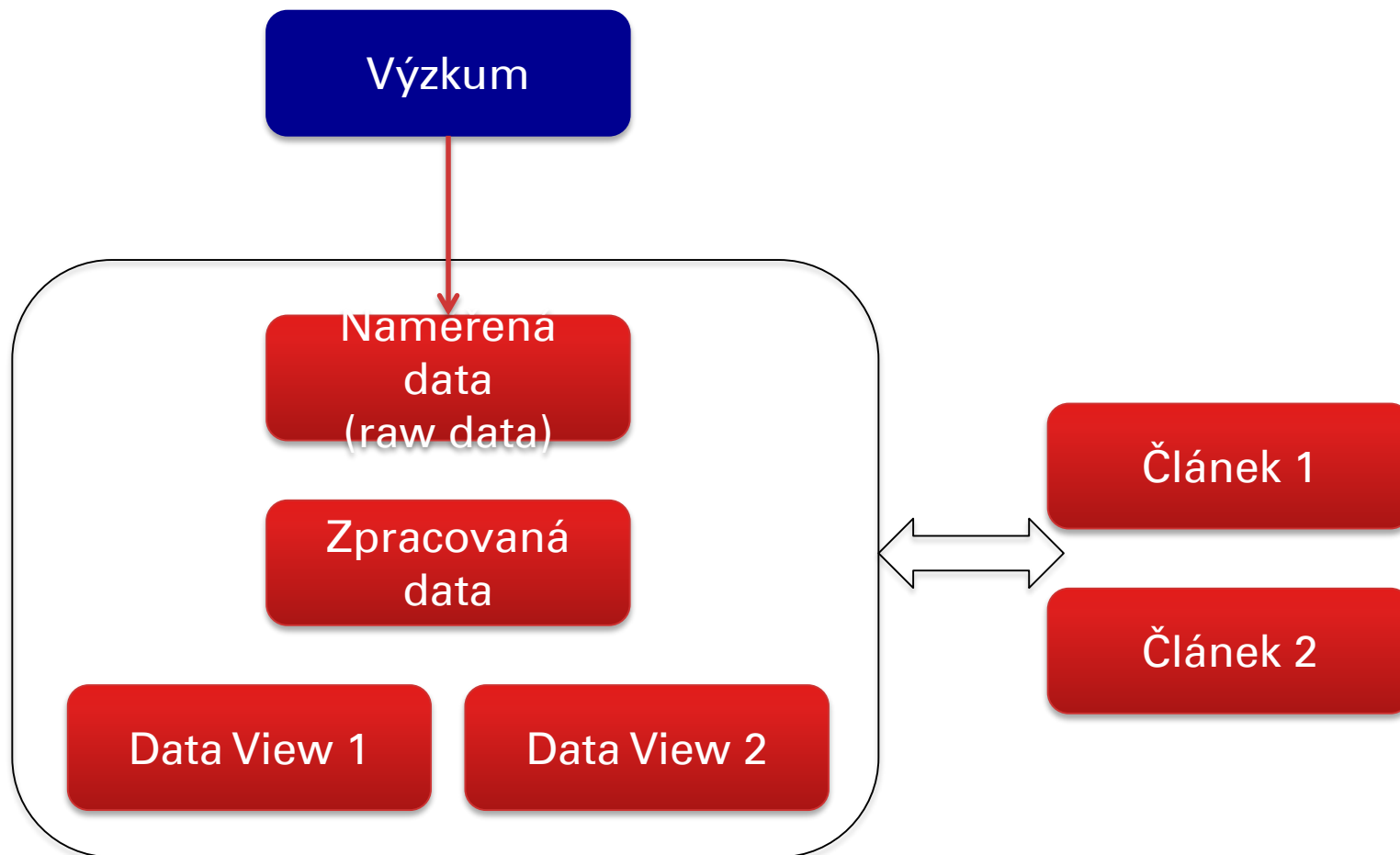
Repozitář vědeckých dat

- Univerzální repozitář, který umožňuje efektivní ukládání, správu, vyhledávání a zpřístupnění všech typů vědeckých dat
- Systém metadat umožňující efektivní popis objektů s ohledem na potřeby jednotlivých vědeckých oborů a vyjadřující vztahy mezi objekty
- API umožňující flexibilní napojení aplikací (např. editorů)
- Důvěryhodnost a bezpečnost

Možnost řešení 1



Možnost řešení 2



Možnost řešení 2

- Metadatový popis vědeckých dat specifický pro danou skupinu vědních oborů (např. jazykové, sociologické, výsledky laboratorních měření atd.)
- Možnost vazeb na publikační výstupy (vytěžování klíčových slov), ale nezávislost na nich
- Zpřístupnění dat dle oprávnění formou výstupních sestav
- Podpora verzování

Projekt RVD CESNET/NTK

- Založeno na Fedoře
- Ladění výkonu (předem nespécikovaná schémata, velké datové objemy)
- Autentizační a autorizační mechanismus (ne vše může být open)
- Základní view a edit, ale důraz na API pro nasazení zákaznických aplikací
- V první fázi bude repozitář vyladěn na několika pilotních zákaznících, následně nabídnuto všem zájemcům z vědeckotechnické oblasti v ČR
- Naplní podmínky dotačních titulů jako Horizon 2020