

Západočeská univerzita v Plzni
Fakulta aplikovaných věd

DISERTAČNÍ PRÁCE

2020

Ing. Petr Včelák

Západočeská univerzita v Plzni
Fakulta aplikovaných věd

Rozsáhlé informační systémy – standardizace metadat

Ing. Petr Včelák

disertační práce
k získání akademického titulu doktor
v oboru Informatika a výpočetní technika

Školitel: Doc. Dr. Ing. Jana Klečková
Katedra: Informatiky a výpočetní techniky

Plzeň 2020

University of West Bohemia
Faculty of Applied Sciences

Large-scale Information Systems – Metadata Standardisation

Ing. Petr Včelák

doctoral thesis
submitted in partial fulfillment of requirements
for the degree of Doctor of Philosophy
in Computer Science and Engineering

Supervisor: Doc. Dr. Ing. Jana Klečková
Department: Computer Science and Engineering

Pilsen 2020

Prohlašuji, že disertační práce je původní a vypracoval jsem ji samostatně a výhradně s použitím citovaných zdrojů uvedených v seznamu literatury.

V Plzni dne

Podpis autora

Název práce: Rozsáhlé informační systémy – standardizace metadat

Autor: Ing. Petr Včelák

Katedra: Katedra informatiky a výpočetní techniky

Vedoucí disertační práce: Doc. Dr. Ing. Jana Klečková, KIV FAV ZČU

Abstrakt: Disertační práce se zabývá standardizací metadat ve spojení informačních systémů a multidisciplinárního datově orientovaného výzkumu v oblasti medicíny. Standardizace metadat v práci nespočívá pouze ve vhodném, jednoznačném a srozumitelném popisu založeném na ontologiích v jazyce OWL. Součástí je také standardizace procesů, která je realizována prostřednictvím vytvořené výzkumné platformy poskytující jednotné řešení často opakujících se úloh: ochrana osobních údajů, tvorba metadat a ETL, ontologický popis, archivace a manipulace s daty, ostatní praktické použití dat, metadat a jejich popisu.

Klíčová slova: informační systém, medicínský výzkum, metadata, ontologie, propojená data

Title: Large-scale Information Systems – Metadata Standardisation

Author: Ing. Petr Včelák

Department: Department of Computer Science and Engineering

Supervisor: Doc. Dr. Ing. Jana Klečková, KIV FAV ZČU

Abstract: The dissertation thesis deals with the standardization of metadata in the combination of information systems and multidisciplinary data-oriented medical research. Metadata standardization in the thesis does not only consist of a suitable, unambiguous, and comprehensible description based on ontologies in the OWL language. It also includes the standardization of processes, which is implemented through a research platform providing a unified solution to frequently recurring tasks: personal data protection, metadata creation, ETL, ontological description, data archiving and manipulation, other practical uses of data, metadata, and their description.

Keywords: information system, medical research, metadata, ontology, linked data

Obsah

1	Úvod	1
1.1	Motivace a zdůvodnění	1
1.2	Formulace problému a cílů	2
1.3	Struktura práce	2
2	Informační systémy	4
3	Metadata	7
3.1	Typy metadat	9
3.2	Účel metadat	10
3.3	Úrovně interoperability	11
3.4	Granularita metadat	11
3.5	Tvorba metadat a jejich kvalita	11
3.6	Způsoby uložení metadat	13
3.7	Standardizace metadat	14
3.7.1	Schéma a elementy metadat	14
3.7.2	Tvorba schéma a standardu	15
3.8	Konceptualizace a ontologie	16
3.8.1	Součásti ontologie	16
3.8.2	Druhy ontologií	17
3.8.3	Ontologické jazyky	18
3.9	Standardy pro metadata	18
3.9.1	Registr metadat	20
3.9.2	Dublin Core	21
3.9.3	METS	22
3.9.4	PREMIS	23
3.10	Referenční modely, rámce a protokoly	24
3.10.1	OAI-PMH	24
3.10.2	OAIS	24
3.10.3	PAIMAS	26
3.10.4	RDF	27
4	Technologie sémantického webu	28
4.1	RDF koncept	28
4.2	Schéma a slovníky	29
4.3	Formáty RDF dat	29
4.4	SPARQL	30
4.5	Web Ontology Language	31
4.6	Sémantický web	32
4.7	Propojená data	33
5	Výzkum v medicíně	34
5.1	Data ve zdravotnictví	34
5.1.1	Klinická, terapeutická a laboratorní data	36
5.1.2	Obrazová data	39

5.1.3	Signálová data	40
5.1.4	Data ostatní	41
5.1.5	Klinická událost a struktura dat	41
5.2	Ochrana osobních údajů	42
5.3	Cévní mozková příhoda	44
5.4	Idiopatické střevní záněty	45
5.5	Možnosti realizace výzkumu v medicíně	46
5.5.1	Specializované systémy a registry	47
5.5.2	Obecné systémy	48
6	Použité metody	51
7	Ontologie	55
7.1	Jmenné prostory	55
7.2	MRE Ontology	56
7.3	DASTA Ontology	60
7.4	DICOM Ontology	62
7.5	Form Ontology	63
7.6	Population Ontology	64
7.7	Stroke Ontology	65
7.8	Idiopatické střevní záněty	67
7.8.1	IBDS Ontology	67
7.8.2	IBDT Ontology	68
7.9	TDB Ontology	69
8	Standardizace procesů medicínského výzkumu	71
8.1	Výzkumná platforma	71
8.1.1	Archiv	72
8.1.2	Softwarové prostředky	73
8.1.3	Úlohy a procesy zpracování dat	74
8.1.4	Metadata	75
8.2	Proces zpracování nových dat	75
8.3	Ochrana osobních údajů	75
8.3.1	AnonMed – de-identifikace dat a metadat	76
8.3.2	PureImage – osobních údaje v obrazových datech	78
8.4	MetaMed – ETL proces	79
8.5	Interakce s okolím	80
8.6	Software vs. standardizace metadat	81
8.6.1	MRE Library	81
8.6.2	MRE Ontology	81
8.6.3	MRE Core	81
8.6.4	Sparkle	82
8.6.5	Query Builder	82
8.6.6	Web Form Generator	83
8.6.7	Databáze tkáňových vzorků	83
9	Ověření výsledků a zhodnocení	84
10	Diskuze	90

11 Přínosy práce	95
12 Doporučení	96
13 Závěr	97
Seznam literatury	98
Seznam obrázků	126
Seznam tabulek	127
Seznam výpisů	128
Seznam zkratk	129
A Ontologie a jmenné prostory	140
B Návrh ontologie v tabulkovém procesoru	141
C Tvorba RDF slovníků z existujících dat	144
D Praktické využití ontologií	145
D.1 Extrakce metadat – konfigurace z ontologie	145
D.2 Příklady aplikace Sparkle	146
D.3 Generování webových formulářů z ontologie	148
D.4 Tvorba dotazů pro temporální data	150
D.5 DataTable, ontologie a RDF data	150
E Ochrana osobních údajů v obrazových vyšetřeních	151
F Metriky ontologií	153
G Klíčové požadavky a metriky	156
H Hodnocení metrik klíčových požadavků	158
I Obsah DVD	159
J Aktivity autora	161

1 Úvod

Informační a komunikační technologie propojují moderní svět. Informační systémy, s kterými se často přímo nebo nepřímo běžně setkáváme v každodenním životě, mají různá zaměření a cíle. Používány jsou pro sběr, uchování, zpracování a sdílení informací, které jsou nutné pro plánování, rozhodování nebo řízení činností k dosažení požadovaného konkrétního cíle. Informační systémy mohou být obecné nebo doménově orientované a pomáhají opakovanému použití informací pro daný účel. Nezbytností je propojení informačních systémů a vzájemné sdílení interoperabilních dat a informací. Existují metodologie, výměnné formáty nebo protokoly s různou mírou obecnosti pro zajištění spolupráce s okolím. Interoperabilitu a znovupoužitelnost komplikují odlišná schémata dat, která odpovídají specifickým potřebám a požadavkům uživatelů nebo jsou zatížena historickými důvody obvyklými v daném místě a čase.

1.1 Motivace a zdůvodnění

Práce je zaměřena na *sjednocení metadat a procesů* v informačních systémech, které mají sloužit pro výzkumné účely v mezioborových projektech využívajících nejen zdravotnická data. Základem vědeckého výzkumu v medicíně jsou heterogenní data v množství datových formátů a z různě kvalitních zdrojů dat. Standardizace metadat spočívá v zajištění jednoznačné identifikace a popisu zdrojů včetně zápisu, který lze sdílet s okolím. Možné jsou různé formy – doporučení, vyžadované postupy nebo specifikace hodnot číselníku/slovníku. Zmínit lze proces extrakce, transformace a plnění (ETL) metadat do úložiště, dotazování nebo další způsoby zpřístupnění dat a metadat. Sjednocení postupů je významné pro automatizaci nebo zajištění dlouhodobé znovupoužitelnosti archivovaných dat.

Pro aplikovaný výzkum, vývoj a inovace jsou důležitá zdrojová data, postupy, softwarové prostředky, konfigurace a dosažené výsledky. Dlouhodobě udržitelný výzkum potřebuje data správně a jednoznačně identifikovat pro zajištění následného zpřístupnění, opakovatelnosti nebo ověřitelnosti použitých metod a publikování výsledků. Rozsáhlá reálná data poskytují široké možnosti pro analýzu, trénování, validaci nebo personalizaci a tvorbu modelů z agregovaných nebo zpracovaných dat a metadat. Zásadní je kontext a postup včetně nastavení parametrů použité metody. Algoritmy mohou mít omezení, jejichž příčinou jsou např. nedostatečná kvalita dat, anatomické anomálie nebo další diagnózy.

Disertační práce se věnuje modelům metadat, ontologiím a přínosu popisu v souvislosti s principem propojených dat. Sjednoceny jsou procesy pro společné často opakující se úlohy, jako jsou zajištění unikátní identifikace, proces akceptace dat, ochrana osobních údajů, ETL, integrace dat z více zdrojů a další. Při snaze o společnou, relativně obecnou, platformu pro odborníky a výzkumné pracovníky z množství oborů, je evidentní rostoucí význam standardizace metadat, jejich interoperability, ale i aplikačních rozhraní a procesů. Rizikem jsou kontinuální změny požadavků včetně rozšiřování sledovaných dat a metadat nebo používání nových algoritmů, postupů a zdrojů dat.

Standards vytvářejí lidé, kteří je potřebují pro sjednocení a strukturování dat nebo metadat. Obvykle vycházejí ze zájmových skupin věnujících se konkrétní do-

méně. V mnoha oblastech se lze proto setkat s existencí několika konkurenčních datových formátů nebo modelů metadat. Existuje řada situací, kde mají technologie sémantického webu své uplatnění pro svoji rozšiřitelnost, univerzálnost nebo možnost odvozovat nové znalosti.

1.2 Formulace problému a cílů

Disertační práce se zabývá návrhem vhodné reprezentace metadat a procesů rozsáhlého informačního systému pro datově orientovaný medicínský výzkum, který předpokládá využití primárních zdravotnických dat. Cíle práce jsou formulovány následujícími body.

1. Navrhnout standardizované procesy pro podporu medicínského výzkumu založeného na primárních zdravotnických datech.
2. Navrhnout a implementovat doménovou ontologii pro sekundární využití zdravotnických dat v České republice.
3. Navrhnout a implementovat doménové ontologie pro rozšířená strukturovaná metadata o anamnéze, diagnostice, terapii nebo dispenzarizaci u pacientů (i) po cévní mozkové příhodě, anebo (ii) s idiopatickými střevními záněty.

1.3 Struktura práce

Úvod, formulaci problému a cíle zahrnuje 1. kapitola. Ve 2. kapitole je text věnován informačním systémům. Metadata a standardy nebo referenční modely a rámce jsou popsány ve 3. kapitole. Technologie sémantického webu uvádí 4. kapitola. Výzkum v medicíně s typy a datovými formáty používaných zdravotnických dat včetně souvislosti s ochranou osobních údajů obsahuje 5. kapitola. Přiblížena je oblast výzkumných projektů zaměřených na (i) cévní mozkovou příhodu a (ii) idiopatické střevní záněty.

V práci použité existující i vlastní metody jsou popsány v 6. kapitole. Výsledky uvádí kapitoly 7 a 8. Kapitola 7 popisuje všechny realizované ontologie. V kapitole 8 je uvedena vytvořená výzkumná platforma včetně sjednocení přístupů, knihoven a software využívajících RDF a ontologický model popisu. Ověření a hodnocení výsledků je v 9. kapitole. Diskuzi se věnuje 10. kapitola. Kapitolou 11 navazují přínosy práce následované doporučením v 12. kapitole. Poslední je závěr ve 13. kapitole.

Praktickou část doplňují přílohy. Příloha A uvádí jmenné prostory používaných existujících a vytvořených vlastních ontologií včetně prefixů. Při návrhu doménových ontologií a RDF slovníků v multidisciplinárním výzkumu se osvědčila metoda využívající tabulkový procesor, která je v příloze B. Transformaci číselníku z XML do RDF ilustruje příloha C. V příloze D jsou příklady praktického využití výsledků zahrnující extrakci metadat (D.1), aplikaci Sparkle (D.2) a webové aplikace pro generování webových formulářů dle ontologie (D.3), tvorbu SPARQL pro temporální data (D.4), a databázi tkáňových vzorků (D.5). Výsledky vlastní metody v software PureImage pro identifikaci a odstranění osobních

údajů v obrazových vyšetřeních ukazuje příloha E. Metriky vytvořených ontologií jsou v příloze F. Požadavky a metriky pro zajištění standardizace metadat uvádí příloha G. V příloze H je hodnocení těchto požadavků pro předložený výsledek navržené výzkumné platformy MRE se souvisejícími ontologiemi a nástroji, které se silně orientují na technologie sémantického webu a propojená data. V příloze I je uveden obsah DVD. Poslední příloha uvádí aktivity autora J, kde jsou zahrnuty publikace, jejich ohlas a souhrn výukových aktivit a projektů.

V práci jsou seznamy literatury (str. 98), obrázků (str. 126), tabulek (str. 127), výpisů (str. 128) a zkratk (str. 129).

2 Informační systémy

Pojem informační systém (IS) je používán desítky let. Přesto nelze tvrdit, že by existovala jediná definice, na níž by panovala shoda. Autoři se problematice informačních systémů věnují z různých perspektiv a obvykle v kontextu svého oboru. Piccoli a Pigni (2018) definuje informační systém jako „*formální, socio-technický a organizační systém navržený pro shromažďování, zpracování, ukládání a distribuci informací*“ v [1]. Laudon & Laudon (2018) [2] vysvětlují IS jako „*soubor vzájemně provázaných komponent, které shromažďují (nebo získávají), zpracovávají, ukládají a distribuují informace pro podporu rozhodování a řízení (koordinaci a kontroly) v organizaci*“. Současně uvádí, že „*může pomáhat manažerům i pracovníkům analyzovat problémy, vizualizovat komplexní objekty nebo vytvářet nové produkty*“. [2] Valacich a Schneider (2017) ve své knize [3] informační systémy definují jako „*kombinaci hardware, software a komunikačních sítí, které lidé vytvářejí a používají za účelem sběru, tvorby a distribuce užitečných dat*“. Bulgacs (2013) uvádí, že „*každý informační systém má za cíl podporovat činnosti, řízení a rozhodování*“. [4] Dle Altera (2013) [5] je informační systém speciální případ tzv. pracovního systému (work system). Tento pohled publikoval již dříve např. v [6, 7]: „*lidé nebo stroje provádějí práci (procesy a činnosti) využívající informace, technologie a další zdroje k tvorbě produktů nebo služeb pro zákazníky*“. Veškeré aktivity, jako je pořizování/snímání, přenos, ukládání, načítání, manipulace a zobrazování, označuje za *zpracování informací*. [7] O'Hara et al. (1999) popisují v [8] informační systém jako „*čtyři komponenty: úlohy, lidé, struktury/role a technologie*“. Velmi obecná definice v knize Theoretical Analysis of Information Systems [9] publikovaná Langefors (1973) IS uvádí jako „*technologické médium pro záznam, ukládání a šíření jazykových výrazů, jakož i pro odvození závěrů*“.

Jak vidíme, existuje mnoho rozličných definic, které se vztahují k následujícím třem tematickým okruhům.

1. **Čím je IS tvořen?** Informační systém je technickým prostředkem [1, 9] vytvářeným lidmi [3], kde pracují lidé [3, 5, 8] prostřednictvím procesů [7, 8] s daty [3, 9] nebo informacemi [1, 2, 7]. Z toho lze odvodit, že informační systém je tvořen komponentami, jimiž jsou (I.) lidé [3, 5, 8, 10], (II.) procesy/úlohy [7, 8, 10], a (III.) technologie [7–10], které tvoří hardware [3], software [3] a komunikační sítě [3]. Pouze autoři Bourgeois (2014) nevnímají data jako samostatnou entitu nad komponentami, ale zařazují je do komponenty technologie [10]. Piccoli a Pigni (2018) označují IS jako způsob organizace (např. právě komponent) pro zajištění svého cíle [1].
2. **Jakou činnost a jak IS vykonává?** V informačním systému jsou činnosti prováděny lidmi [3] nebo stroji [5] a to na základě procesů [5] využívajících technické prostředky [7–10]. Všechny procesy pracují s daty [3, 9] nebo informacemi [1, 2, 7], čímž dochází k jejich manipulaci [7] – získávání/shromažďování, ukládání [1, 2, 7, 9], zobrazování [7], zpracování [1, 2, 7] a distribuci/přenosu/sdílení [1, 2, 7, 9].
3. **Jaký je cíl IS?** Z výše uvedených definic autorů lze odvodit cíle informačního systému jako (I.) podpora činností [4], (II.) řízení [2, 4], kam spadají

koordinace [2] nebo kontrola [2], a (III.) rozhodování [2, 4]. Konkrétní podporované činnosti budou závislé na typu informačního systému. Může se jednat o tvorbu produktů [2, 5] nebo služeb [5]. Se splněním cílů pomáhají data a informace, které se v informačním systému za tímto účelem tvoří/shromažďují [3], zobrazují [2], analyzují [2, 9] nebo sdílí [9].

Ve zjednodušeném pohledu je IS modelem reálného světa a uživatelů, kteří jej využívají pro dosažení svého cíle. Zatím co první okruh vysvětluje, jaké komponenty tvoří IS, druhý a třetí okruh objasňují jeho roli. Informační systém zajišťuje interakci mezi (a) lidmi, (b) procesy, (c) technologií a případně (d) daty. Právě procesy jsou způsobem, jak se s informačním systémem interaguje. V organizacích, společnostech a firmách, kde se informační systémy nasazují, jsou procesy obvykle sjednoceny a lze je považovat za jakýsi (alespoň lokální) standard, který musí všichni lidé v organizaci dodržovat. Uvedený postup a vhodné provázání procesů mezi sebou vede k přesnějším výsledkům v kratším čase. Již v roce 1984 se Ives ve svém článku [11] zaměřoval právě na výhody a konkurenceschopnost nebo včasné odhalování příležitostí ve spojení s používáním informačního systému. Velmi důkladně se problematikou úspěšnosti informačních systémů věnovali DeLone a McLean, kteří publikovali svá zjištění v [12] (1992), [13] (2002), [14] (2003) a společně s Petter (2013) také v [15].

Pro reálné použití musí informační systémy spolupracovat a k tomu je nutná interoperabilita dat a metadat. V definicích není přímo věnována pozornost interakci s okolím, resp. s dalšími informačními systémy. Alter (2013) přiznává, že procesy mohou být prováděny i strojem [5], což by se dalo na spolupráci informačních systémů vztáhnout. Také O'Brien a Marakas opakovaně (2009, 2012) v [16, 17] uvádějí IS jako kompozici pěti základních komponent tvořených (a) lidmi, (b) hardware, (c) software, (d) daty a (e) počítačovou sítí. Valacich (2017) se o komunikační síti zmiňoval v [3]. O'Brien a Marakas v [16, 17] explicitně zmiňují *komunikaci s okolím*. Zejména v případně systémů silně datově orientovaných je zásadní, aby byla aktuální, úplná a kvalitní data a metadata.

Pro tvorbu IS byly navrženy různé metodologie a techniky. Používají se přístupy orientované (a) datově, (b) procesně (DFD), (c) kombinace datově a procesně orientovaných (Yourdonova metoda) a (d) strukturované metody. Tvorba IS má fáze – úvodní studie, rozbor zadání, analytické modelování, systémový design, objektový design, implementace, zkušební provoz a nasazení. [18, 19] Integrací se zabýval například Hasselbring (2000), kde ve svém článku [20] zmiňuje možné komplikace komponent informačních systémů rozdělených na obchodní, technologickou a aplikační vrstvu. Procesy, pravidla a organizační struktura patří do obchodní vrstvy. Technologická vrstva definuje informační a komunikační infrastrukturu. V aplikační vrstvě je implementace propojující obchodní a datovou vrstvu. Dle [21, 22] se v oblasti integrace objevují přístupy zabývající se přímo integračními platformami jako služby (Integration Platform as a Service, zkráceně iPaaS).

Rozsáhlé informační systémy Pojem *rozsáhlé informační systémy* vymezuje spíše hlavní rysy, ale jinak je definice volnější. V angličtině se pro slovo *rozsáhlý* používají označení „extensive“, „large“, „large-scale“ nebo „ultra-large scale“. Pro rozsáhlá data jsou zažity pojmy velká (big) a obrovská (huge) data, mající souvislost s *velkým množstvím různých typů dat*. Z perspektivy čistě lingvistické má

slovo *rozsáhlý* významy „*mající velký rozsah, velké rozměry, rozlehlý, prostorný, rozměrný, mající velký obsah, obsáhlý, obširný*“ dle [23]. V kontextu informačních systémů, proto celé komponenty nebo jejich části mohou být řádově velkého rozsahu/rozměru. Metrikou pro rozsah by mohl být objem, množství nebo typy dat, počet uživatelů, ale i např. územní rozlehlost. Bagheri (2014) [24] charakterizuje obecný pojem „*ultra-large scale system*“ (ULS system), volně přeložený na *velmi rozsáhlý systém*. Podobně Gabriel (2006) [25] nerozlišuje konkrétní typ ani účel systému. Mezi metriky rozsahu Bagheri (2014) uvažuje např. počet uživatelů, množství dat (uložené, zpřístupňované, nebo obecně zpracovávané), ale i počet spojení/závislostí mezi komponentami systému, počet hardwarových prvků, počet řádek zdrojového kódu, apod. Právě zásadní navýšení rozsahu mění způsob a techniky týkající se požadavků na návrh, vývoj, automatickou koordinaci a řízení komplexních systémů, procesů i výměnu informací. [24, 26] S rozsahem IS se zvyšuje pravděpodobnost k selhání, jak uvádí Jormanainen (2018) [27]. Rezaei (2014) [28] doplňuje, že selhání mohou být nepředvídatelná a stávají se v ULS normou. Evropská komise (2010) [29] a Buyle (2017) [30] uvádějí zvyšování interoperability jako klíčový aspekt umožňující budoucí růst.

Vysvětlení pojmu *rozsáhlé informační systémy* v souladu s [24] podporují další publikace. Interoperabilitu považují za důležitou Jormanainen (2018) [27] a Rezaei (2014) [28]. Chen (2012) [26] doplňuje význam interoperability pro udržitelnost. Architekturu informačních systémů v souladu s ULS se věnoval Chen (2012) v [26], kde zdůrazňuje neznámé nebo konfliktní požadavky, neustálou evoluci procesů, kontextů a v čase se měnící zdroje i hranice systému.

Interoperabilitou mezi rozsáhlými informačními systémy Evropské unie (EU) se zabýval Buttarelli (2017/2018), ve funkci evropského inspektora ochrany údajů (European Data Protection Supervisor, EDPS), v [31] a česky [32] s cílem vytvořit harmonizované technické prostředí systémů pro správu vnějších hranic schengenského prostoru a posílení vnitřní bezpečnosti Evropské unie: *European Dactyloscopy or fingerprint identification* (EURODAC), *European Police Office* (EUROPOL), *Visa Information System* (VIS), *Schengen Information System* (SIS), *Customs Information System* (CIS) a *Internal Market Information System* (IMIS). Další dokumenty vztahující se k problematice interoperability informačních systémů EU byly [33, 34] v letech 2017 a 2018. K uvedenému publikovali Järvsoo et al. (2018) [35] případovou studii pro Estonsko za situace, kdy existuje množství nařízení a omezení, ale zcela chybí jednotný standard a kooperace při implementaci v členských státech.

Problematikou úspěšné implementace a nasazení rozsáhlých zdravotnických informačních technologií se zabývali Cresswell et al. (2013) v publikaci [36], kde uvádějí deset klíčových faktorů v průběhu čtyř fází životního cyklu informačních technologií. Přehled literatury zpracovali Sligo et al. (2017) ve svém článku [37] se zaměřením na plánování, implementaci a hodnocení rozsáhlých zdravotnických informačních systémů.

3 Metadata

Data jsou údaje, které používáme pro formální vyjádření nějakého jevu nebo vlastnosti obecně libovolného objektu. Údaje mohou být reprezentovány libovolnými řetězci znaků na informačním nosiči. Data se získávají měřením nebo pozorováním a lze je interpretovat, přenášet, uchovávat nebo jinak zpracovávat. Data jsou vhodný formalizovaný způsob jak reprezentovat informace. [38, 39] Dle [38] data „zpravidla nemají význam sama o sobě, ale teprve jsou-li pochopena, interpretována, komunikována a využita člověkem nebo počítačem, stávají se smysluplnými informacemi.“ *Interpretace* (interpretation) je výklad nebo objasnění, na základě kterého mohou z dat vzniknout informace. [40] *Informace* (information) je dle [41] „údaj o reálném prostředí, jeho stavu a procesech v něm probíhajících“. Cílem sdělování informací je zvýšení znalosti a tím snižování nebo úplné odstranění neurčitosti (entropie). [39, 41, 42] *Znalost* (knowledge) je schopnost uchovávat, komunikovat a zpracovávat informace do systematicky a hierarchicky uspořádaných znalostních struktur. Znalost je charakterizována schopností abstrakce a generalizace dat i informací. [39, 43]

Metadata (meta data) jako slovo pochází z řeckého *meta* (mezi, za) a latinského *data*. Přeloženo označuje *to, co je dáno*. Metadata jsou **strukturovaná data obsahující informaci o primárních/popisovaných datech**. [44, 45] Často používané zjednodušené vysvětlení uvádí *data o datech*. V [46] je definována *meta informace* (meta information) jako „*Informace, která je v transformačním vztahu k jiné informaci. Je nástrojem popisu vztažné informace. Slouží jako prostředek vybavení obsahu popisované informace ve vyhledávacích systémech.*“

Interoperabilita (interoperability) je schopnost systémů *vzájemně spolupracovat*. Dochází k součinnosti aplikací nebo ke vzájemné výměně dat a informací mezi aplikacemi nebo celými systémy. [40, 44]

Zdroj (resource) je označení pro obecně libovolnou jednoznačně identifikovatelnou věc a to bez rozdílu, zda je fyzická, digitální/virtuální, anebo s jakoukoliv mírou abstrakce. Může se jednat o osobu, událost, dokument, soubor nebo pouhou část objektu, který je naším zájmem. *Jednoznačná identifikace zdroje* (Uniform Resource Identifier, URI) je základním prvkem pro všechny popisy, protože je nutné nezaměnitelným způsobem přesně identifikovat popisovaný zdroj a dva různé zdroje od sebe odlišit (validace). Pro identifikaci zdrojů se používá **Uniform Resource Identifier** (URI), který byl definován v RFC 3986 (2005) [47] Primárně se jedná o identifikátor, ale současně může plnit lokalizační funkci. Rozlišení na dvě podmnožiny je dle definice v [47] nepřesné a může působit zmatek [48]. Obě podmnožiny identifikátorů vyžadují označení použitého schématu na počátku identifikátoru. [47] Příklady URI jsou uvedeny ve výpisu 3.1. První podmnožina se označuje *Uniform Resource Locator* (URL) a zajišťuje jednoznačnou identifikaci a lokalizaci zdroje, protože poskytuje primární přístupový mechanismus a cestu ke zdroji. [47] Výhoda URL spočívá v možnosti lokalizace zdroje, ale je současně i nevýhodou. Jestliže zdroj přemístíme, je nutné změnit identifikátor, jehož je lokalizace součástí. Nevýhodu odstraňuje *trvalé URL* (Persistent URL, PURL), které se nemění v čase. Využívá se *směrovací služba*, která spravuje seznam souvisejících URL a jejich metadat pro cílový zdroj. Funkce PURL je identifikační i lokalizační, která míří pouze na zmiňovanou *směrovací službu* sprava-

vující seznam souvisejících URL (nebo i metadat) pro cílový zdroj a zprostředkuje poskytnutí aktuální platné URL požadovaného zdroje. [49] Na identifikaci a bez možnosti lokalizovat zdroj, je zaměřena druhá podmnožina URI. Označuje se *Uniform Resource Name* (URN) viz výpis 3.1. Identifikátor URN vždy začíná textovým řetězcem *urn:* a obvykle následuje další označení typu použité metody identifikace zdroje. [47] Identifikátor může být odvozen od již běžně používaných způsobů pojmenování nebo značení jako jsou Digital Object Identifier (DOI), International Standard Book Number (ISBN), International Standard Serial Number (ISSN), Universally Unique Identifiers (UUID, ISO/IEC 9834-8:2014, [50]), Organization for the Advancement of Structured Information Standards (OASIS) [51] a řady dalších. Pro identifikaci lze používat i hodnoty z čárových kódů, identifikace na rádiové frekvenci (Radio Frequency Identification, RFID), poštovní směrovací čísla, apod. V případě potřeby lze zajistit chybějící funkčnost lokalizace zdroje s URN prostřednictvím [52] např. HTTP proxy server¹. *Internationalizovaný identifikátor zdroje* (Internationalized Resource Identifiers, IRI) dle [53] může na rozdíl od URI obsahovat i mezinárodní znaky Universal Character Set (Unicode/ISO 10646).

Výpis 3.1: Příklady URI, URL a URN. V prvním sloupci je příklad identifikátoru. Klasifikace, zda se jedná o URI, URL nebo URN, je uvedena ve druhém sloupci.

<code>file:///etc/fstab</code>	URI, URL
<code>file://localhost/etc/fstab</code>	URI, URL
<code>https://tools.ietf.org/rfc/rfc3986.txt</code>	URI, URL
<code>mailto:vcelak@kiv.zcu.cz</code>	URI, URL
<code>doi:10.1016/j.ijmedinf.2019.02.011</code>	URI
<code>urn:doi:10.1016/j.ijmedinf.2019.02.011</code>	URI, URN
<code>https://doi.org/10.1016/j.ijmedinf.2019.02.011</code>	URI, URL
<code>issn:1386-5056</code>	URI
<code>urn:issn:1386-5056</code>	URI, URN
<code>uuid:00112233-4455-6677-8899-aabbccddeeff</code>	URI
<code>urn:uuid:00112233-4455-6677-8899-aabbccddeeff</code>	URI, URN
<code>urn:oasis:names:specification:docbook:dtd:xml:4.1.2</code>	URI, URN

Nejnovější publikaci [54] věnující se porozumění metadatům vytvořil Riley (2017) pod záštitou National Information Standards Organization (NISO). Současně s knihou [55], jejímž autorem je Lambe (2014), se jedná o aktuální prameny. Autor v [54] objasňuje, jaké jsou typy metadat, jak je ukládat i sdílet a nechybí zmínka o standardizaci metadat: (a) prostřednictvím kontrolovaných slovníků nebo (b) formou pravidel, postupů a zásad jejich tvorby. [54]

Definice pojmů data, informace a metadata existuje, ale přesto už v roce 2000 autoři Bargmeyer a Gillman v publikaci [56] o standardizaci metadat přišli s přístupem *nerozlišovat mezi pojmy data a metadata*. Přístup to není ojedinělý a K. Hüner, B. Otto a H. Österle ve svém článku *Collaborative management of business metadata* [57] používají shodný přístup. Důvodem je nejasnost hranice mezi daty a metadaty, která jsou daty také. Data i metadata lze uložit shodným způsobem do libovolného úložiště. Pouhým pohledem na uložená data nelze rozlišit, zda se jedná v tom kterém případě o data nebo metadata. Pro jednu aplikaci se jedná o metadata, ale pro jinou to mohou být data. [58]

¹DOI pro lokalizaci zdroje používá HTTP proxy server `https://doi.org/<DOI>`, kde se za `<DOI>` doplní identifikátor zdroje. [52]

3.1 Typy metadat

Ani rozdělení typů metadat není mezi autory jednotné. Záleží na úhlu jejich pohledu. Riley (2017) [54] rozděluje metadata na následující čtyři skupiny.

Popisná metadata (descriptive) slouží pro lokalizaci, nalezení (discovery), zobrazení (display) či porozumění (interoperability) zdroji. Popisná metadata jsou o individuálních instancích aplikačních dat a jejich obsahu (data content). Alternativní označení *data o obsahu dat* (data about data contents) proto dává smysl. Příkladem jsou informace o autorovi, datu publikování, název a klíčová slova dokumentu.

Strukturální metadata (structural) popisují relace a vztahy mezi zdroji včetně jejich organizace. Při zobrazení umožňuje tento typ metadat navigaci v datech. Může se jednat např. o místo v hierarchii tříd nebo obsah s odkazy na jednotlivé kapitoly. Mezi strukturální metadata patří *data o kontejneru pro data*, protože při návrhu nebo specifikaci datových struktur nelze o datech mluvit, neboť ještě žádná neexistují.

Administrativní metadata (administrative) jsou potřeba pro správu zdrojů a souvisí s jejich vznikem nebo přístupem k nim. Lze je členit na technická, archivační a právní. *Technická* (technical) metadata jsou nutná pro dekódování nebo zobrazení záznamů (např. typ a velikost souboru, použitý typ komprese). *Archivační* (preservation) metadata slouží pro dlouhodobé zachování zdroje, archivaci, ale i správu zdrojů (např. kontrolní součet). *Právní* (rights management) metadata se vztahují k ochraně práv souvisejících s použitím zdrojů (např. copyright, licence, práva držitele) nebo jejich důvěrnosti.

Značkovací jazyk integruje strukturální nebo jiné sémantické vlastnosti (metadata) přímo v původním obsahu. Použití může být pro navigaci i pro porozumění (např. obsahu nebo struktury textu). Příkladem mohou být nadpis, odstavec, seznam, ale i formátovací značky pro zvýraznění části textu.

Při porovnání publikace NISO [54] (2017) s její o 13 let starší verzí vydané [59] (2004) je zřejmý postupný vývoj oblasti. Typ administrativních dat byl v [59] původně pod pojmem *ochrana zdrojů*, který obsahoval archivační a právní podtypy. Vedle této spíše kosmetické změny však nová verze přinesla rozšíření typů metadat, neboť v [54] je nově uváděn také *značkovací jazyk* jako typ metadat.

Typy metadat se v minulosti zabývalo více autorů. Již v roce 1994 mezi ně patřili Bretheron & Singley s publikací [60] zaměřenou na vědecké databázové systémy, kde z pohledu uživatele členili metadata do dvou skupin: (1) strukturální/řídící a (2) klíčová metadata. Strukturální byla určena pro popis struktury počítačového systému, jako byly tabulky, sloupce a indexy. Klíčová metadata měla za cíl pomoci lidem najít konkrétní položky za využití klíčových slov v přirozeném jazyce. Dnes bychom je označili jako strukturální a popisná metadata. Publikace přišla relativně krátce po vzniku pojmu *World Wide Web* (1990) [61, 62], který masové rozšíření teprve čekalo. World Wide Web Consortium (W3C) bylo založeno až v říjnu 1994. [63] Autoři Kimball et al. (2008) v knize [64], věnující se životnímu cyklu datových skladů, dělí metadata na typy (1) interní/technická, (2) externí/obchodní a (3) procesní.

3.2 Účel metadat

Účel metadat je výstižně popsán v publikaci [59] ve znění: „*Metadata is key to ensuring that resources will survive and continue to be accessible into the future.*“ Volně přeloženo „*Metadata jsou klíčová, aby zdroje přežily a byly přístupné i do budoucna.*“ Pozitivní přínos [54] je možné spatřovat s ohledem na jednotlivé případy užití. Rámcový výčet nejčastějších účelů použití metadat je uveden v abecedním pořadí: *nalezení* (discovery), *navigace* (navigation), *porozumění* (interoperability), *zachování* (preservation) a *zobrazení* (display).

Nalezení zdroje a navigace Metadata mohou pomoci získat zdroje efektivně, zkrátit čas potřebný k jejich nalezení, anebo snížit nároky na výpočetní výkon nebo vstupně/výstupní operace. Množství datových formátů obsahuje data i metadata. Nereálná je představa datového úložiště s obrovským množstvím dat a nutnosti pro vyhledávání pracovat pokaždé s každým jednotlivým souborem. Užitečné je mít metadata k dispozici v jednoduché formě vně původních dat. Vhodná je existence aplikačního nebo uživatelského rozhraní, které je umožní přímo používat a vyhledávat v nich. Nalezení je možné přímo na základě znalosti URI. Pro ostatní případy potřebujeme použít metadata. Popisnými metadaty se realizují katalogy, rejstříky nebo je možné zdroje různě klasifikovat a třídít do skupin zdrojů. Prostřednictvím strukturálních metadat jsou propojovány různé zdroje a popisujeme reálný svět, jeho objekty a vztahy mezi nimi (např. hierarchie). Všechna tato metadata pomáhají s nalezením požadovaného zdroje, zdrojů nějakým způsobem příbuzných, ale mohou být použity i pro navigaci mezi zdroji.

Porozumění datům, metadatům a interoperabilita Správné a jednoznačné pochopení významu dat nebo metadat je nezbytnou součástí při zajištění jejich interoperability. Pro správnou interpretaci je potřeba znát definici vlastností, kterými zdroj popisujeme. Pokud tato konkrétní vlastnost neupřesňuje, je vhodné doplnit další metadata (anotace). Strany účastníci se výměny by měly definice a anotace sdílet, aby nebylo pochyb a nedošlo k nedorozumění nebo dokonce chybám.

Zachování zdroje Metadata mají význam při evidenci změn digitálních objektů a mohou být využita pro archivaci a ochranu zdrojů – prevence poškození nebo ztráty. Ochrana je nutná zejména vůči (1) uživateli a (2) poškozenému hardware nebo software. Ochrana před *uživatelem* je obvykle na úrovni administrativních metadat, kde lze udělit povolení přístupu pro čtení (zobrazení) nebo zápis dat (zachování). V ideálním případě by nemělo docházet ke změnám původních dat. V reálném světě může vzniknout oprávněný zájem původní data upravit, protože jsou neúplná, anebo chybná. Pak je vhodné uchovávat metadata o provedených změnách, identitu uživatele, datum, čas a další potřebné údaje. Ke ztrátě nebo poškození může dojít vlivem hardwarové nebo softwarové chyby. Metadata mohou poskytnout dostatečné podklady pro kontrolu zdrojů (kontrolní součet), opětovné vytvoření (postup) nebo získání (původce). [54, 65] Pro tento účel existuje např. slovník a doporučení (PREMIS Data Dictionary for Preservation Metadata) s datovým modelem definujícím pět entit – objekty, prostředí, události, agenti a práva. [65]

Zobrazení zdroje Účel poměrně úzce navazuje na nutnost rozumět datům a metadatům. Pro zobrazení zdroje mohou být v některých případech potřeba metadata, která určují jak data interpretovat, aby byla prezentována koncovému uživateli správně. Příkladem mohou být metadata, jaký audio/video kodek je nutný pro přehrávání, jaká barevná paleta má být použita, apod. Metadata mohou ovlivňovat vhodný způsob zobrazení nebo výběr jednoho z několika možných.

3.3 Úrovně interoperability

Rozlišovat lze mezi interoperabilitou strukturální, syntaktickou a sémantickou. Dublin Core Metadata Initiative (DCMI) vytvořila [66] v roce 2009 stupnici o čtyřech úrovních interoperability pro metadata v rámci Dublin Core. Tolk (2003) nejprve v [67] uváděl pět úrovní konceptuální interoperability. Později (2007) v [68] model rozšířil na celkem sedm úrovní skládajících se z (0) žádná, (1) technická, (2) syntaktická, (3) sémantická, (4) pragmatická, (5) dynamická, a (6) konceptuální. Tolk u systémů rozlišuje integrovatelnost (integratability), interoperabilitu a možnost kompozice (composability). Nejvyšší úroveň konceptuální interoperability odpovídá systému, který je „*plně specifikovaný, ale implementačně nezávislý model*“ dle [68].

3.4 Granularita metadat

Úroveň strukturování metadat se označuje termínem *granularita* a určuje ji ten, kdo navrhuje schéma metadat. Menší granularita znamená méně detailní strukturu informací a menší náklady na tvorbu i následnou údržbu metadat. Vyšší míra granularity umožňuje detailnější popis, ale následující technická manipulace může být náročnější. Granularita se projeví, když metadata zastarají nebo je nutná transformace do jiné podoby. Pro ilustraci si lze představit měření teploty. Co vše budeme potřebovat, aby pro nás měla data o měřených teplotách význam a poskytovala relevantní informace? Mohou to být použité jednotky veličiny, datum, čas, místo, přístroj a jeho přesnost nebo další okolnosti.

3.5 Tvorba metadat a jejich kvalita

Přidáním metadat zvyšujeme kvalitu dat nebo popisovaných zdrojů. [54] Potíže mohou vzniknout v případech, kdy metadata nejsou tvořena na základě řízených slovníků, číselníků, systematických nebo datových slovníků a thesaurů nebo ve shodě s pravidly pro indexování a katalogizaci zdrojů. Přes respektování pravidel může být příčinou i způsob tvorby metadat.

National Information Standards Organization v [59] uvádí, že se lze setkat s problémy: (a) chybějící povinné elementy, (b) chybně použité elementy, (c) chyby v syntaxi schématu a (d) nekonzistentní terminologie. Při syntaktických chybách ve schématu není možné metadata správně zpracovat. Nekonzistentní terminologie komplikuje nalezení relevantních nebo podobných či souvisejících informací a i přes jejich existenci mohou být kvůli odlišné terminologii nevyhledatelné. [59] NISO v dokumentu *The Framework of Guidance for Building Good*

Digital Collections [69] poskytuje následujících šest principů pro tvorbu dobrých metadat.

1. Metadata by měla odpovídat materiálům v kolekci, jejím uživatelům a určenému využití zdroje.
2. Podporují interoperabilitu.
3. Ve vztahu k obsahu zdroje používají standardizované řízené slovníky pro vyjádření co, kde, kdy a kdo.
4. Jasně určení okolností a podmínek užití zdroje.
5. Metadata jsou objekty, které jsou autoritativní a ověřitelné. Mohou mít atributy archivovatelnosti, perzistence, unikátní identifikátor, apod.
6. Podporují dlouhodobou správu objektů ve sbírce.

Kvalitu metadat ovlivňuje způsob jejich vytváření nebo získávání. Vytvářet a získávat metadata lze (1) manuálně, (2) automatizovaně nebo (3) kombinací obou přístupů. Každý způsob má výhody i nevýhody.

Manuální tvorba nebude vhodná pro velký objem dat a i pro menší množství zdrojů může často docházet k překlepům nebo chybám v důsledku nepozornosti. Překlepy a chyby budou časté, zejména pokud se bude jednat o rutinní, nudné a dlouhé vytváření. Kvalitu může snižovat různorodost uživatelů, kteří metadata vytváří, např. v důsledku jejich různé odbornosti nebo znalosti problematiky.

Automatizovaný postup vyžaduje existenci softwarového nástroje. Pro snazší tvorbu metadat existuje řada nástrojů, které jsou vyvíjeny s podporou konkrétního metadatového schématu nebo sady používaných elementů. Dle [59] lze rozlišit následující čtyři typy nástrojů pro automatickou tvorbu metadat.

Šablony poskytují předdefinovaná pole, z nichž je následně vygenerována sada dvojic atributu a hodnoty.

Značkovací umožňují strukturovat atributy s hodnotami dle konkrétního schématu a většina z nich generuje XML dokumenty.

Extraktory jsou nástroje pro automatické získání metadat na základě zpracování nebo analýzy digitálních zdrojů. Kvalita takto získaných metadat se může lišit v závislosti na použitých algoritmech a obsahu nebo struktuře zdrojových dat. Je vhodné je brát jako pomůcku pro tvorbu metadat a následně provést manuální validaci, revizi a případnou úpravu.

Konverzní provádí převod mezi odlišnými formáty metadat. V závislosti na podobnosti převáděných formátů je potřeba dalších manuálních úprav.

Při kombinaci manuálního a automatizovaného způsobu může být základem automatická forma získávání. Následně uživatel provede úpravy/korekce nebo doplnění dalších metadat, která nelze automaticky zjistit v požadované kvalitě.

Většina základních strukturálních a administrativních metadat je zajišťována původcem při tvorbě dat, ale mohou být generovány automaticky. Je-li to možné, je vhodné použít kombinaci obou přístupů. Do takové kategorie zcela jistě patří

výzkumná data nebo dosažené výsledky. Část metadat lze získat z uložených souborů automatizovaně. Přesto obvykle znalost způsobu jejich vzniku má pouze autor samotný, a proto je vhodné, aby je vytvářel rovnou původce dat. Vytváření metadat původcem dat by mělo být v souladu s určitými pravidly, aby byla data popisována shodně dalšími původci. Metadata nemusí vždy vytvářet pouze původce, ale může se jednat o specialistu v daném oboru (např. radiolog, lékař). Přesto může docházet k odlišnostem v kvalitě metadat, pokud daná osoba nemá dostatek času nebo zkušeností pro tvorbu správných a kvalitních metadat. Řešením může být revize takto vytvořených metadat provedená osobami s potřebnou odborností. Ověří správnost, konzistenci a shodu s používaným schématem a používanými metodickými pokyny.

Negativní vliv může mít i původce nebo jeho zařízení. Pokud již datový formát podporuje uchování metadat a byla původcem zapsána – jejich kvalita záleží na původci. Například u fotografie, kde je v metadatech zapsán datum a čas pořízení snímku, může chybný údaj pocházet z nesprávně nastaveného data a času nebo časové zóny ve fotoaparátu. S novým snímkem vzniknou metadata, která budou chybná. Tento typ zavlečené chyby může být velmi obtížné odhalit a následně opravit.

3.6 Způsoby uložení metadat

Data existují v množství podob, kódování a formátů. Podporuje-li datový formát metadata, pak přímo ovlivňuje způsob uchování – bude se jednat o interní nebo vložená metadata. Rozlišují se tři přístupy k uložení metadat: (1) interní, (2) vložená a (3) externí. *Interní* metadata spočívají v uložení ve stejném datovém souboru společně s daty (speciální část, hlavička). *Vložená* (embedded) metadata jsou také ve stejném datovém souboru jako metadata interní, ale jsou přímo součástí dat. *Externí* metadata jsou uchovávána odděleně. Netvoří součást zdroje nebo datového souboru. Některé datové formáty neumožňují vložení metadat nebo nejsou vhodné pro všechny jejich typy. [59]

Každý z uvedených přístupů k uložení metadat má své klady i zápory. *Interní a vložená* metadata, poskytují snazší přenos metadat, protože jsou přenesena společně s daty a jsou s nimi vždy k dispozici. Nevýhodou je, že metadata nelze vyměňovat samostatně bez dat a je vysoká redundance metadat, neboť shodné atributy se musí opakovat v každém souboru. Při více kopiích totožných dat je problematické udržet konzistenci metadat po změně v jednom ze souborů. Aktualizace je jednoduchá, jen nejedná-li se o velký soubor dat. Extrakce metadat je nutná, aby mohly být následně používány. Pro získání metadat z datového souboru je nutné konkrétní datový formát znát. *Externí* uchování metadat umožňuje jejich seskupování, odstranění redundance. Mají výhody v případě hromadné správy metadat (např. v databázi), kdy může být vyhledávání a získávání metadat i následně dat jednodušší a rychlejší. Metadata lze poslat odděleně od dat samotných. Rizikem je možnost ztráty vztahu mezi metadaty a původními daty. Proto je nutné důkladně řešit identifikaci metadat a vztah k datům. Při změně původních dat nemusí dojít k přímé aktualizaci externích metadat. Popisovaný zdroj reálného světa nemusí mít žádný identifikátor a pro externí zápis je nutné jej uměle vytvořit. Pro sdílení externích metadat se používají k tomu určené obecné výměnné formáty nejčastěji vycházející z formátů XML nebo JSON. [54]

3.7 Standardizace metadat

Standardy vytvářejí lidé, kteří je potřebují pro sjednocení a strukturování dat nebo metadat. Obvykle vycházejí ze zájmových skupin věnujících se konkrétní doméně. Standardizace, aby měla význam a skutečný přínos, musí přesahovat hranice jednotlivých autorů nebo vydavatelů norem. Pro standardizaci metadat je potřeba, aby všechny zúčastněné strany byly ve shodě. Systémy pak mohou efektivně spolupracovat a vzájemně si poskytovat služby. Za tímto účelem existují nebo stále vznikají zájmové skupiny, asociace, spolky, konsorcia, komise, různé iniciativy a výzvy na národní nebo mezinárodní úrovni. Standardy pro metadata mohou být v různé míře závaznosti, od prostých pravidel a doporučení až po de-facto standard nebo formu závazné mezinárodní normy. V České republice se vytvářením, vydáváním a zveřejňováním Českých technických norem (ČSN) až do konce roku 2008 zabýval Český normalizační institut na základě rozhodnutí Ministerstva průmyslu a obchodu č. 237/1997 Sb. a provádějícího zákona č. 22/1997 Sb. Od roku 2009 má úlohu vydávání i tvorby norem *Úřad pro technickou normalizaci, metrologii a státní zkušebnictví*² (ÚNMZ). Od roku 2018 se zabývá technickými normami *Česká agentura pro standardizaci*³ (ČAS), která je zřízena jako příspěvková organizace ÚNMZ dle zákona č. 265/2017 Sb. Ve Spojených státech amerických (United States of America, USA) je obdobou nezisková *National Information Standards Organization* (NISO⁴), jež je akreditována *American National Standards Institute* (ANSI). Právě pod záštitou NISO se metadataům věnuje Riley (2017) ve své publikaci [54]. Na mezinárodní úrovni má hlavní roli *International Organization for Standardization* (ISO). Více doménově orientované jsou obvykle právě různé zájmové skupiny. V oblasti webu a webových technologií je nejznámější *World Wide Web Consortium*⁵ (W3C). [63]

3.7.1 Schéma a elementy metadat

Při standardizaci metadat se používá pojem *schéma metadat*, které definuje celkovou podobu popisu řešené domény, typu dat nebo obecně informačního zdroje. Schéma metadat se označuje také jako model. Ve schématu je definována množina *elementů*, které se používají k popisu. Element je označení pro koncept nebo pojem/termín přidružený k určitému významu (sémantika), který je jednoznačně identifikovatelný. Mezi elementy mohou existovat *vztahy*. Při popisu zdrojů obsahuje element *hodnotu*. Každý z elementů musí mít definován *název*, *význam* (sémantiku) a volitelně další *pravidla pro hodnotu elementu* jako jsou např. vyžadované formáty, povolené hodnoty nebo slovníky. Schémata mohou být abstraktní nebo mohou vyžadovat konkrétní *syntaxi elementů* specifikující kódování nebo obecně způsob vyjádření/zápisu hodnoty. Syntaxe elementů se často z důvodu obecnosti schématu uvádí v tzv. *aplikačním profilu*, který určuje konkrétní implementační pravidla. Aplikační profil může využívat elementy z více schémat současně a určovat další vlastnosti elementu (volitelnost nebo opakovatelnost). Aplikační profil nebo implementace schématu staví velmi často na Standard Generalized Markup Language (SGML) nebo eXtensible Markup Language

²<http://www.unmz.cz>

³<http://www.agentura-cas.cz>

⁴<https://www.niso.org/>

⁵Výčet standardů a návrhů W3C je dostupný na <https://www.w3.org/TR/?tag=data>.

(XML). Příkladem může být Resource Description Framework (RDF, viz 3.10.4) nad XML. Reálně se pro popis používá více schémat současně.

Existující schémata lze dělit na (1) hierarchické, (2) lineární (jednorozměrné) nebo (3) rovinné (dvourozměrné). *Hierarchické* schéma je, pokud existují vzájemné vztahy mezi elementy metadat s vazbou typu rodič a potomek. Příkladem hierarchického schéma je Learning Object Metadata [70], kde elementy metadat mohou náležet do rodičovského elementu. [71] *Lineární* schéma má každý element zcela nezávislý na ostatních elementech a klasifikován je pouze v jednom rozměru (Dublin Core, viz kapitola 3.9.2). [71] *Rovinná* schémata mají každý element zcela nezávislý od ostatních elementů, ale klasifikace je ve dvou vzájemně ortogonálních rozměrech. U rovinného schéma je nutný i typ mapování, který umožní zobrazení na metadata podle zvoleného aspektu. Často se jedná o vrstvení geografických a geologických informací. [71–73]

Schéma v sobě zahrnuje standardy a modely metadat. Konkrétní podoba zápisu vychází z pravidel vytvořených pro strukturování polí a elementů metadat. [74] Schéma pro metadata může být vyjádřeno v množství značkovacích nebo programovacích jazyků s odlišným způsobem zápisu. Metadata dle Dublin Core [75] mohou být zapsána v čistém textu, HTML, XML a RDF. Schéma metadat lze formalizovat a zapsat ve formě ontologie (viz kapitola 3.8).

3.7.2 Tvorba schéma a standardu

Výbor ISO/TC 46/SC 11⁶ zodpovědný za tvorbu norem pro správu archivů/záznamů doporučuje provést analýzu dostupných schémat a vyhodnotit jejich vhodnost pro daný účel před vlastní implementací. Výbor vydal normu *ISO 15489*⁷ (2016) [76] definující pojmy, postupy a principy pro vytváření, příjem a správu záznamů (dokumentů). Norma je základem řady dalších mezinárodních norem a technických zpráv. Součástí druhé části normy jsou specifikace metadat vycházejících ze základní sady dle ISO 15836 (Dublin Core, viz 3.9.2) pro dosažení souladu s legislativními požadavky uchování záznamů. Další vydanou normou je *ISO 23081* tvořená třemi částmi, které rozšiřují ISO 15489 a určují obecný návod popisu záznamů s ohledem na možnosti budoucích inovací. První část *ISO 23081-1* (2017) [77] spojuje požadavky se základní sadou metadat dle ISO 15489-1 pro všechny záznamy, jejich metadata, procesy, systémy, ale i celé organizace, které za ně zodpovídají. Druhá část *ISO 23081-2* (2009) [78] určuje rámec pro definování elementů metadat a je praktickým přístupem k implementaci nebo jejím variantám. Řeší správu metadat a konceptuální model metadatových elementů pro záznamy. Poslední, třetí část *ISO/TR 23081-3* (2011) [79] je kontrolním seznamem, který umožní posoudit silné/slabé stránky navrženého metadatového schéma ve shodě s normou.

Ve shodě s výše uvedeným dle ISO/TC 46/SC 11, Australian National Data Service (2018) [80], University of North Carolina (2019) [81], ale i řadou dalších autorů se doporučuje prozkoumat možnost existence schématu před tvorbou vlastního. V minulosti mohlo vzniknout vhodné řešení pro dosažení požadovaného cíle a není nutné vymýšlet duplicitní schéma. Pokud existující schéma požadav-

⁶<https://committee.iso.org/tc46sc11>

⁷ISO 15489-1 Information and documentation – Records management – Part 1: Concepts and principles. ČSN ISO 15489-1 je v platnosti od června 2018.

kům vyhovuje, může být snazší jej **adoptovat** a vzít za své. V případě větších odlišností nebo jiné úrovně granularity metadat je alternativou **adaptace**, kdy lze existující schéma použít jako základ. Upravit jej můžeme zredukováním na potřebnou podmnožinu elementů nebo naopak o chybějící elementy rozšířit. Oba postupy pomohou dosáhnout interoperability a předejít budoucím problémům.

Při návrhu i tvorbě vlastního schématu University of North Carolina (2019) [81] explicitně doporučuje, aby minimální elementy určené pro popis zdroje poskytovaly (a) unikátní *identifikátor*, (b) člověku srozumitelný *název*, (c) *popis*, (d) *formát* souboru, médium, rozměry, přístupová metoda, (e) *metadata o metadataových elementech* jako specifikaci a význam jednotlivých používaných elementů, (f) *držitel práv*, (g) *práva* ke zdroji, (h) *kontakt* na původce/správce dat a (i) vše ostatní, co je potřeba uchovávat z perspektivy budoucího uživatele metadat.

3.8 Konceptualizace a ontologie

Konceptualizace je sémantické (významové) vymezování konceptů – pojmů, kategorií, taktik, strategií, scénářů, námětů a vztahů mezi nimi. Konceptualizací je myšlen abstraktní a zjednodušený model pohledu na reálný svět. Cílem konceptualizace je systematicky reprezentovat určitou doménu ve smyslu části světa, věcné oblasti, skutečnosti nebo jevů. Konceptualizace je základem pro formální reprezentaci znalostí – objekty, koncepty, entity a vztahy existující v dané oblasti zájmu. [82] Libovolnou znalostní bázi nebo na znalosti orientovaný systém lze považovat za implicitní nebo explicitní konceptualizaci reálného světa a jeho části. [82]

Explicitní konceptualizací reálného světa je *ontologie*. Termín ontologie pochází z filozofie, ale používán je v mnoha významech také v dalších odvětvích. V informatice je ontologie způsobem jak modelovat reálný svět a znalosti domény s velkým důrazem na jasně a pevně dané pojmy v podobě řízených slovníků. Množina znalostí, objektů i popisu vztahů je označována jako popis světa (universe of disclosure), který je reprezentován slovníkem. Definice ontologie je spojená s pojmenováním entit při popisu světa, jako jsou třídy, relace, funkce a další objekty s člověku čitelným popisem významu a formálních principů (axiomů) omezujících interpretaci i správné použití termínů. [83]

Ontologie definuje závazný slovník sdílený mezi spolupracujícími stranami (agenty) a je používán shodným a konzistentním způsobem pro dotazy i další uplatnění. [84] Důležité je zmínit způsob použití, kdy není potřeba, aby spolupracující agenti (strany) mezi sebou sdíleli znalostní bázi, tj. každý z agentů má jen některé znalosti. [85] Agenti nemusí dát odpověď na všechny dotazy, které lze pomocí sdíleného závazného slovníku formulovat. Ontologie *zajišťuje a garantuje konzistenci*, nikoliv však kompletnost. Odpověď nemusí být známa, což znamená, že neexistuje (Open World Assumption). Neodpoví-li ani jediný agent na položený dotaz – odpověď neznáme. [86]

3.8.1 Součásti ontologie

Současné ontologie mají podobné vyjadřovací struktury bez ohledu na jazyk, v němž jsou vyjádřeny. Součástmi ontologií jsou:

třída představuje typy objektů nebo věcí;

individuum je *instance* nebo základní/výchozí objekt;

atribut představuje aspekty, vlastnosti (ve významu properties i features), charakteristiky nebo parametry objektů;

relace je vztah, jaký mohou mít třídy nebo individua mezi sebou;

funkční člen je komplexní struktura složená z určitých relací, které mohou být použity jako individuum v tvrzení;

omezení jsou popisy vstupních hodnot, které musí být splněny, pro použití v tvrzení;

odvozovací pravidla jsou tvrzení zapsaná v podobě *if-then* (antecedent-consequent) popisující logické dedukce, které mohou být utvořeny z tvrzení v konkrétních případech;

axiomy tvrzení a pravidla;

událost popisuje změnu atributů nebo relací.

Uvedený výčet součástí je použit pro konceptualizaci, ale formální zápis se liší v závislosti na konkrétním ontologickém jazyce a jeho syntaxi.

3.8.2 Druhy ontologií

Nejčastější dělení ontologií je dle (a) oboru a (b) předmětu formalizace. Rozdělení podle *oboru* představuje skupiny (1) *terminologické* – tezaury a strukturování termínů, (2) *informační* – nadstavba nad datovými zdroji, podobnost databázovým schémátům a (3) *znalostní* – reprezentace znalostí, logické odvozování. Podle *předmětu formalizace* je možné dělení ontologií na (1) *generické* ontologie zaměřující se na obecné a vyšší koncepty, (2) *doménové* ontologie zaměřené na konkrétní domény znalostí a (3) *úlohové* pro postup řešení úlohy. Existují ještě (4) *aplikační*, které jsou vázány na specifické aplikace a často spojují doménové a úlohové ontologie.

Generické ontologie, někdy označované jako vyšší (upper ontology), jsou určeny k modelování obecných objektů. Lze je aplikovat napříč širokým spektrem doménových ontologií. Zajišťují obecný základní slovník termínů a s nimi souvisejících popisů objektů, které se používají v různých doménách.

Doménové ontologie (domain; domain-specific) mají za cíl modelování konkrétní domény, tedy určité omezené části reálného světa. Význam pojmů z domény je popsán právě v ontologii. S doménovými ontologiemi pro shodnou doménu bývá problém s jejich vzájemnou kompatibilitou, kterou ve své doméně mají zajišťovat. Mezi možné nedostatky patří různá úroveň zaměření na detaily. Rozdílné ontologie ve stejné doméně mohou vznikat z různých úhlů pohledu nebo jsou závislé na kulturních odlišnostech, vzdělání, ideologii a dalších faktorech, mezi něž patří např. volba odlišných ontologických jazyků. Všechny tyto rozdíly je těžké překonat při spojování ontologií, které jsou vytvořeny nad odlišnou základní ontologií (např. rozdílné způsoby popisu konceptů). Takový proces je časově velmi náročný a nelze automatizovat. Je nutný zásah odborníka s důkladnou znalostí domény a spojovaných ontologií.

3.8.3 Ontologické jazyky

Pro formální zápis ontologie se používá *ontologický jazyk*. Vybrané existující ontologické jazyky jsou v následujících bodech seřazených abecedně.

Common Logic je standardizován v ISO 24707 pro skupinu ontologických jazyků, které lze mezi sebou překládat.

CycL je vlastní ontologický jazyk projektu Cyc⁸ založený na Lispu. Existuje podpora konverze do OWL i práce z jazyka Java. [87]

DOGMA⁹ poskytuje na fakta orientované modelování s vyšší úrovní sémantické stability.

KIF syntaxe predikátové logiky prvního řádu založená na S-výrazech.

OBO jazyk určený pro biologické a biomedicínské ontologie. [88] Mapováním do OWL se zabývala publikace [89], kde autoři vytvořili metodologii pro konverzi z OBO do OWL včetně implementace v jazyce Java.

OWL je zkratka pro Web Ontology Language, který je široce používaným ontologickým jazykem. Vychází z RDF a starších ontologických jazyků OIL, DAML a DAML+OIL. Modelovat může libovolnou oblast (viz kapitola 4.5).

RIF kombinuje ontologie a pravidla (Rule Interchange Format a F-logic).

SADL je zkratka pro Semantic Application Design Language¹⁰, který je podmnožinou OWL, ale používá anglický jazyk pro tvorbu sémantických modelů a výukových pravidel.

3.9 Standardy pro metadata

V minulosti byla vytvořena řada metadatových schémat obecných nebo orientovaných na konkrétní doménu s různou granularitou metadat. Digital Curation Centre (DCC) se zabývá správou výzkumných dat a poskytuje i poradenství výzkumným organizacím. Na webu DCC [90] jsou dostupné seznamy metadatových standardů, aplikačních profilů, případových studií a nástrojů. Další seznam členěný dle typu je dostupný na [91]. V oborech archivace a bibliografie/knihovnictví byly pravděpodobně jedny z prvních snah na vytvoření standardizovaného metadatového schéma, právě s cílem efektivnější katalogizace a vyhledávání ve spravovaném fondu. Ostatně Hamill ve své nedávné (2017) knize [92] uvádí, že lze mnohá pravidla a zásady z oborů archivnictví uplatnit v elektronické oblasti. Archivací se zabývají Encoded Archival Context (EAC) [93], Encoded Archival Description (EAD) [94], General International Standard Archival Description (ISAD-G) [95] a PREMIS Data Dictionary for Preservation Metadata (PREMIS) [96]. V oblasti knihovnictví a bibliografie jsou používány: konceptuální model Functional Requirements for Bibliographic Records (FRBR) [97], MACHine-Readable Cataloging

⁸<http://www.opencyc.org/>

⁹Developing Ontology-Grounded Methods and Applications

¹⁰<http://sabl.sourceforge.net/>

(MARC) [98], Metadata Encoding and Transmission Standard (METS) [99], Metadata Object Description Schema (MODS) [100], ONline Information eXchange (ONIX) [101], Resource Description and Access (RDA) [102]. Existuje množství schémat pro metadata z oborů jako je biologie, ekologie, geografie, umění a vzdělávání. Darwin Core (DWC) pro sdílení informací o biologické rozmanitosti [103]. Ecological Metadata Language (EML) [104] pro ekologii. Z geografie uvádí Federal Geographic Data Committee (FGDC) seznam standardů [72] a to včetně ISO 19115-1:2014 [73]. Z oblasti umění se jedná o Visual Resources Association Foundation (VRAF) se schématem VRA Core [105] Pro vzdělávání existuje Learning Object Metadata (LOM) [70].

Existují obecná schémata jako je *Dublin Core* dle ISO 15836-1:2017 [75], které je pravděpodobně nejznámějším ([106]) schématem metadat. Vznikly standardy pro popis kolekcí dat Data Catalog Vocabulary (DCAT) [107], nebo organizačních informací a taxonomií ISO/IEC 11179 Information technology – Metadata registries (MDR) [108–113], a Simple Knowledge Organization System (SKOS) [114]. Autoři DCC (2019) [90] a Gonzalez-Perez (2017) [106] zmiňují referenční modely nebo rámce orientované na archivaci, skladování, sdílení nebo výměnu metadat, jako jsou Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) [115], Open Archival Information System (OAIS, ISO 14721:2012) [116, 117] a Producer-Archive Interface Methodology Abstract Standard (PAIMAS) [118, 119]. Dlouhodobému uchování elektronických dokumentů se věnuje norma ISO/TR 18492 (2005) [120], která byla v roce 2013 potvrzena a je platná.

Při bližším zkoumání [106] jsou schémata většinou založena na XML. Polovina schémat v [106] je přímo na XML založena a druhá polovina používá Resource Description Framework (RDF, viz 3.10.4), kde bylo XML základním stavebním prvkem. Dříve se používalo pro zápis schéma XML dokumentu tvořeného elementy a atributy *Document Type Definition* (DTD). Později vzniklo XML Schema (XSD), které významně rozšiřuje (oproti DTD) možnosti a zavádí datové typy a jmenné prostory. Jedná se o strukturovaný formát a právě prostřednictvím XSD lze definovat strukturu, kterou může příjemce dat využít. Výhodou XML dokumentu je možnost použití více schémat současně. [121]

Pro porozumění je důležitá konzistence kódování/zápisu konkrétních hodnot metadat základních datových typů. Proto existují normy pro číselníky (např. ISO 3166 pro kódy států [122] a ISO 639 pro kódy jazyků [123]), slovníky, ale i pro zápis základních datových typů jako je např. datum/čas (ISO 8601 [124]) nebo sjednocení, jak zapisovat číselky, jména/příjmení osob nebo názvy a zkratky. Lze použít Metadata Authority Description Schema (MAD/MADS) [125] nebo Resource Description and Access (RDA) [102].

Rozsáhlé informační systémy na úrovni Evropy, resp. Evropské unie (EU), a nutná interoperabilita je řešena zejména ve vztahu k veřejné správě a poskytovaným službám. [33, 34] Na základě rozhodnutí EU č. 2015/2240 byl schválen (2015) program *Interoperability solutions for public administrations, businesses and citizens*¹¹ (ISA²) na období let 2016–2020. [126] V roce 2017 byla vydána brožura *New European Interoperability Framework* (New EIF) [127] věnující se službám a toku dat evropské veřejné správy. Evropská komise (EC) založila platformu *Joinup*¹² pro dosažení interoperability veřejné správy, podniků a občanů.

¹¹https://ec.europa.eu/isa2/home_en

¹²<https://joinup.ec.europa.eu/>

Joinup provozuje otevřený registr *Open Metadata Registry*¹³ (původně National Science Digital Library Registry) za účelem standardizace a s využitím sémantických technologií (viz kapitola 4). Uživatelům umožňuje registraci schématu, slovníků a aplikačních profilů. Poskytuje seznam a přístup k existujícím schématům¹⁴ a slovníkům¹⁵ včetně jejich dokumentace. Evropská unie přispěla na vznik dokumentu s definicí *Europeana Data Model* (EDM, 2017) [128] a příručku s pokyny, jak metadata správně mapovat [129].

3.9.1 Registr metadat

Mezinárodní norma ISO/IEC 11179 vydaná ISO se vztahuje na tzv. *registry metadat*. Norma je určena pro výměnu přesně definovaných dat řízenou na základě metadat v heterogenním prostředí. Princip MDR spočívá v uchování datových prvků (data elements), které obsahují (1) *sémantiku* i (2) *reprezentaci* v chráněném prostředí, kde mohou změny provádět pouze autorizovaní jedinci. Obě dvě části musí být od sebe jasně odděleny. V sémantické části MDR je uchována sémantika (význam) přesně definovaných datových elementů. Reprezenční část MDR (např. v databázi nebo strukturovaném formátu XML) je pro konkrétní formát a obsahuje metadata vztahující se k omezením¹⁶. Norma má celkem šest částí. [108]

První část ISO/IEC 11179-1 (2015) je základem a spojovacím prostředkem všech ostatních částí normy a je nutná pro porozumění pojmů vztahujících se k metadatům a MDR celkově. Popisuje základní myšlenky související s datovými elementy, doménami hodnot, pojmy datových elementů, jejich doménami a klasifikačními schématy. [108]

Druhá část ISO/IEC 11179-2 (2005) definuje soubor principů, metod, procedur a technik pro sdružování datových elementů s klasifikačními schématy. Všechny typy spravovaných položek lze klasifikovat – objektové třídy, vlastnosti, zobrazení, domény hodnot, pojmy datová hodnota i datové elementy. Norma používá klasifikační schémata proměnlivé diskriminační síly, jako jsou (1) klíčová slova, (2) slovníky, (3) taxonomie a (4) ontologie. Všechna uvedená schémata jsou užitečná pro dokumentaci objektů v reálném světě. [109]

Třetí část normy ISO/IEC 11179-3 (2013) má největší význam. Definuje model metadatového registru, který se používá k uchování informací o datových elementech a přidružených pojmech (pojmové domény a domény hodnot). Význam spočívá v definování charakteristik dat pro jejich správné využití (např. interpretaci a zobrazení) – *metadata*. Všechna metadata norma označuje jako *metadatové položky*, které jsou nutné (1) k jasnému popisu, (2) záznamu, (3) analyzování, (4) klasifikaci a (5) spravování dat. [110]

Čtvrtá část normy ISO/IEC 11179-4 (2004) specifikuje požadavky a doporučení konstrukce definic dat a metadat z pohledu sémantiky. Jedná se o definice datových elementů a konstruktů – (1) typy entit, (2) entity, (3) vztahy, (4) atributy, (5) objektové typy/třídy, (6) objekty, (7) složené elementy, (8) kódové výstupy a (9) data. [111]

¹³<http://metadataregistry.org/>

¹⁴<http://metadataregistry.org/schema/list.html>

¹⁵<http://metadataregistry.org/vocabulary/list.html>

¹⁶Omezeními mohou být např. minimální nebo maximální délka textového řetězce, přesnost a konkrétní reprezentace desetinného čísla.

Pátá část normy ISO/IEC 11179-5 (2015) se zabývá tvorbou názvů a identifikátorů, jejich částmi a strukturou s cílem zajistit jednoznačnou identifikaci všech položek v MDR. Pravidla se vztahují na všechny čtyři skupiny názvů: (1) pojmy datových elementů, (2) pojmových domén, (3) datových domén a (4) domén hodnot. [112]

Poslední, šestá část normy ISO/IEC 11179-6 (2015) popisuje postup pro registraci spravovaných položek v MDR. Specifikuje, jak zajistit unikátnost identifikátorů na mezinárodní úrovni. Mezinárodně unikátní identifikace je složená z (1) identifikátoru registrační autority, (2) unikátního identifikátoru položky v rámci registrační autority a (3) verze položky. [113]

V České republice byly jednotlivé výše uvedené části ISO/IEC 11179 převzaty pod označením ČSN ISO/IEC 11179. V platnosti byly během období 1996–2015. Následně byly jednotlivé části ČSN ISO/IEC 11179 bez náhrady zrušeny v letech 2015¹⁷ 18¹⁸ 19¹⁹ a 2016²⁰. Naposledy, v červnu 2019, byla zrušena druhá část normy ČSN ISO/IEC 11179-2²¹. Poslední platnou částí je ČSN ISO/IEC 11179-4²².

V souvislosti s MDR vznikla v roce 2008 iniciativa *eXtended Metadata Registry* (XMDR). Základem bylo použití ontologií pro obsah MDR s cílem poskytovat rozsáhlejší sémantický rámec než lze dosáhnout pouze konvencemi lexikálního a syntaktického pojmenování. Vývoj se týkal vylepšení o podporu sémantiky datových elementů, terminologií a struktur konceptů v MDR. Pro účely XMDR bylo používáno RDF (viz kapitola 3.10.4), OWL (viz kapitola 4.5) a SPARQL (viz kapitola 4.4) k prokázání správnosti konceptu. Výsledkem iniciativy byla třetí edice ISO/IEC 11179. [56] Další rozšíření označené XMDR+ publikoval Kim et al. (2011) v [130].

3.9.2 Dublin Core

Dublin Core (DC) je označením pro normu ISO 15836-1 (2017) [75] definující schéma metadat. První zmínka byla v roce 1995 prostřednictvím Dublin Core Metadata Initiative (DCMI). Dublinské jádro definuje celkem patnáct deskriptorů určených výhradně pro popis zdroje informace. Všechny prvky obsahují návěští, definici a možnost poznámky. Co má být zdrojem, není nijak omezeno a schéma lze používat obecně. Norma nedefinuje konkrétní způsob implementace. Všechny elementy se však používají v kontextu aplikačního profilu, který určuje shodu a soulad s požadavky. [75] Druhá část normy s množstvím dalších vlastností a tříd vznikala od roku 1999 a vydána byla teprve v prosinci 2019. [131] Ve shodě s normou ISO 15836-1 (2017) [75] je definováno následujících patnáct elementů:

Contributor (příspěvatel) určuje, kdo je zodpovědný nebo má podíl na zdroji, např. osoba, organizace nebo služba. Uvádí se jméno/název příspěvatele.

Coverage (pokrytí) prostorové nebo časové téma, použitelnost nebo jurisdikce zdroje. Může se jednat o pojmenování místa, které je preferováno, nebo jeho

¹⁷ČSN ISO/IEC 11179-3 (platnost v období 04/1996–11/2006, 12/2006–10/2015).

¹⁸ČSN ISO/IEC 11179-5 (platnost v období 06/1997–08/2008, 09/2008–05/2015).

¹⁹ČSN ISO/IEC 11179-6 (platnost v období 06/1998–04/2009, 05/2009–10/2015).

²⁰ČSN ISO/IEC 11179-1 (platnost v období 02/2002–11/2006, 12/2006–07/2016).

²¹ČSN ISO/IEC 11179-2 (platnost v období 05/2001–08/2008, 09/2008–06/2019).

²²ČSN ISO/IEC 11179-4 (platnost v období 04/1997–11/2006, 12/2006–nyní).

geografické souřadnice. Doporučeno je použití tezauru geografických jmen (Thesaurus of Geographic Names²³, TGN).

Creator (tvůrce) primární původce zodpovědný za popisovaný zdroj. Uvádí se jméno/název tvůrce.

Date (období) nebo okamžik události životního cyklu ve vztahu ke zdroji. Doporučeno je uvedení v souladu s ISO 8601 [124].

Description (popis) zdroje např. formou abstraktu, obsahu nebo volného textu zprávy, který vhodně zdroj popisuje.

Format (formát) zahrnuje popis souborového formátu, typ média, rozměry, rozsah nebo další podobné vlastnosti zdroje. Doporučeno je uvedení v souladu s Internet Media Types²⁴ (MIME) [132].

Identifier (identifikátor) je jednoznačným identifikátorem zdroje v daném kontextu.

Language (jazyk) použitý ve zdroji. Doporučeno zapisovat dle RFC 4646.

Publisher (vydavatel) zodpovědný za zpřístupnění zdroje.

Relation (vztah) k souvisejícímu zdroji.

Rights (práva) majetková nebo práva duševního vlastnictví vůči zdroji.

Source (zdroj) je související zdroj, z něhož byl popisovaný zdroj odvozen.

Subject (téma) je hlavním předmětem zdroje. Doporučuje se uvádět klíčová slova nebo klasifikační kódy na základě řízeného slovníku.

Title (název) je jméno dané zdroji. Doporučuje se použít název, pod nímž je formálně známý.

Type (žánr) určuje povahu zdroje. Doporučuje se použít řízený slovník (např. DCMI Type Vocabulary²⁵). Formát souboru se popisuje v samostatném elementu *Format*.

3.9.3 METS

*Metadata Encoding & Transmission Standard*²⁶ (METS) je výměnný formát zaměřený na kódování administrativních, popisných, strukturálních, ale i technických metadat. Vznikl v roce 2001 pod záštitou Digital Library Federation a nyní jeho údržbu zajišťuje Kongresová knihovna (Library Of Congress, Washington, USA). Formát METS se využívá ve spojení s úložišti dat, jako jsou digitální archivy a knihovny. [99]

²³<http://www.getty.edu/research/tools/vocabularies/tgn/>

²⁴<https://www.iana.org/assignments/media-types/media-types.xhtml>

²⁵<http://dublincore.org/specifications/dublin-core/dcmi-type-vocabulary/>

²⁶<http://www.loc.gov/standards/mets/>

Formát je založen na široce rozšířeném XML, kde je kořenovým elementem `<mets>`, a v němž jsou všechna metadata umístěna. Přenáší se celý XML soubor. Cílem METS je sdílení mezi různými systémy prostřednictvím jednotného formátu. METS je složen ze sedmi sekcí, jimiž jsou: hlavička (tvůrce, datum/čas vytvoření a modifikace), popisná metadata (vhodné pro vyhledání, možnost použít existující schémata a jejich elementy), administrativní metadata (popis původního zdroje, práva pro zpřístupnění a šíření), seznam souborů (všechny popisované fyzické soubory včetně umístění), strukturální mapy (zachycení hierarchie a vazeb mezi soubory, objekty a metadaty), strukturální odkazy (odkazování mezi uzly strukturální mapy), pravidla chování (definice akcí nebo události v případě manipulace s dokumentem). [99, 133]

Profily jsou cestou jak flexibilně přizpůsobit poměrně obecný a komplexní formát METS lokálním potřebám v praxi pro účel dané instituce. Profil METS se skládá ze třinácti částí, kterými jsou unikátní identifikátor (přidělené URI), název, abstrakt, datum a čas vytvoření, kontaktní informace, související profily, rozšiřující schémata METS dokumentu, pravidla popisu, řízené slovníky, strukturální požadavky (omezení struktury METS), technické požadavky, nástroje (např. software je vhodné použít s METS dokumentem) a ukázkový METS dokument (jedna nebo více příloh s příkladem METS dokumentu ilustrující jeho použití). Vždy je doporučeno nejprve ověřit existenci²⁷ vhodného profilu a teprve následně vytvářet vlastní. Vytvořený profil je doporučeno zaregistrovat, aby byl viditelný i pro ostatní. [133]

3.9.4 PREMIS

*PREservation Metadata: Implementation Strategies*²⁸, zkráceně PREMIS, je snahou přinést strategie pro kódování, uchování a správu metadat digitálních objektů. Vznikly na základě podpory Online Computer Library Center (OCLC) a Research Libraries Group (RLG) s cílem zajištění dlouhodobé digitální udržitelnosti informací. [96] Strategie vychází a jsou kompatibilní s referenčním modelem ISO 14721 [116] známým pod označením Open Archival Information System (OAIS), který je blíže popsán v [117] a kapitole 3.10.2. Kompatibilita s OAIS má vliv na použití PREMIS v systémech, které z OAIS vycházejí. Standard je složen z datového slovníku, XML schématu a nezbytné dokumentace. Datový slovník ve verzi 3.0 definuje čtyři entity (původně pět), kterými jsou objekty (objects; předmět dlouhodobé ochrany), události (events; akce zahrnující objekt a agenta), práva (rights; práva nebo oprávnění) a agenti (agents; osoba, organizace, software/slужba). Dále jsou definovány vlastnosti entit a vztahy mezi nimi. Ve verzi PREMIS 3.0 se objevuje prostředí (environment) jako nedílná součást objektu a specifikuje technologii nutnou pro jeho správnou interpretaci (např. spuštění nebo zobrazení). [134]

Od verze PREMIS 2.2 je vedle XML schéma k dispozici současně ontologie pojmenovaná PREMIS OWL. Poslední verze 3.0 (v období 2016–2018) spojuje osvědčené postupy tzv. propojených dat (viz kapitola 4.7) a využití vhodných existujících ontologií a slovníků (Dublin Core, Provenance ontology²⁹ a další slov-

²⁷<http://www.loc.gov/standards/mets/mets-registered-profiles.html>

²⁸<https://www.loc.gov/standards/premis/>

²⁹<https://www.w3.org/TR/prov-o/>

níky³⁰). Metadata v souladu s PREMIS je možné vyjádřit prostřednictvím RDF, což je dle [134] vhodnější v případech jako dotazování, publikování nebo sdílení.

3.10 Referenční modely, rámce a protokoly

Referenčních modelů, abstraktních rámců nebo jejich následných rozšíření vzniklo v minulosti pro správu metadat a archivaci informací nebo digitálních zdrojů více. V kapitole 3.10.1 je protokol pro jednosměrné získávání metadat od poskytovatele dat k poskytovateli služeb jako je např. vyhledávání. Kapitola 3.10.2 popisuje abstraktní referenční model *Open Archival Information System*, který definuje obecný rámec, ale nezabývá se přímo konkrétní implementací. Detailům mezi tvůrcem dat a archivem se věnuje kapitola 3.10.3. V kapitole 3.10.4 je popsán obecný rámec pro popis zdrojů, z něhož vychází princip propojených dat (4.7).

3.10.1 OAI-PMH

The Open Archives Initiative³¹ (OAI) měl za cíl usnadnit efektivní šíření obsahu a zajištění interoperability. *Open Archives Initiative Protocol for Metadata Harvesting*³² (OAI-PMH) není přímo datový výměnný formát, ale spíše aplikační rozhraní. [115] Původci obsahu nebo datových repositářů vystavili prostřednictvím OAI-PMH metadata. Poskytovatelé (vyhledávacích) služeb využívali protokol k získání vystavených metadat prostřednictvím protokolu HTTP. Použitím XSLT (eXtensible Stylesheet Language Transformations) lze získaná metadata transformovat do RDF. [135] Existuje veřejně dostupný seznam registrovaných poskytovatelů³³, kde jsou mezi zdroji zejména odborné časopisy, univerzitní nebo fakultní knihovny a další zdroje odborných publikací (např. CiteSeerX) stejně jako instance DSpace. Seznam není zcela aktuální, protože větší množství poskytovatelů ze seznamu bylo nedostupných (v čase zkoušení). Firma Google ukončila podporu OAI-PMH v roce 2008. Nástupcem je soubor XML Sitemap, který je používán doposud a poskytuje mapu webu včetně data modifikace. [136]

3.10.2 OAIS

Aktuální verze doporučení pro praxi [117] definuje abstraktní referenční model *Open Archival Information System* (2012), jehož autorství má Consultative Committee for Space Data Systems (CCSDS). OAIS je konceptuálním modelem pro dlouhodobou archivaci. Model určuje, jaká metadata mají být uložena společně s archivovanými daty, aby bylo dosaženo odpovědné dlouhodobé udržitelnosti a ochrany digitálních objektů. Záměrem je použitelnost určené cílové skupině. Referenční model OAIS byl vydán jako mezinárodní norma ISO 14721 [116] a pro svoji obecnost a otevřenost se stal velmi rozšířeným a často implementovaným.

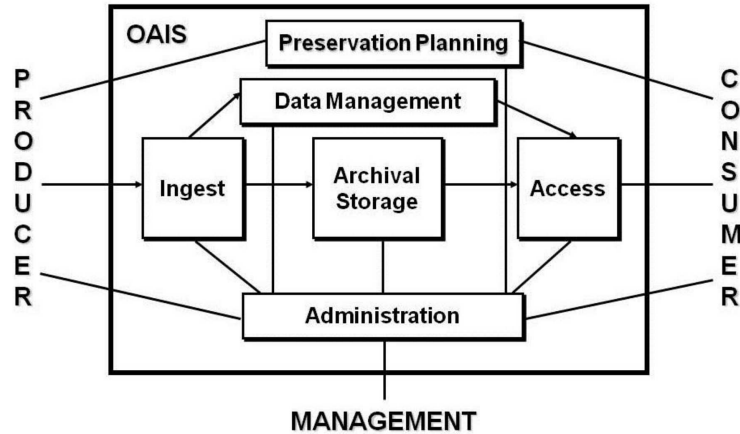
Přehlednou příručku připravil Lavoie (2014) [137], v níž popisuje jednotlivá specifika referenčního modelu. Hlavní částí je samotný archiv OAIS, dále jen *Archiv*, poskytující funkce Producentům, Konzumentům a Řízení. *Producentem* je

³⁰<http://id.loc.gov/vocabulary/preservation.html>

³¹<https://www.openarchives.org/>

³²<https://www.openarchives.org/pmh/>

³³<https://www.openarchives.org/Register/BrowseSites>



Obrázek 3.1: Funkční model OAIS dle Lavoie (2014) [137].

původce nebo tvůrce dat, který předává data do Archivu na základě jejich dohody. Dohoda o dodání dat (Submission Agreement) upravuje typ dat, potřebná metadata, práva a oprávnění k datům. Pouze schválená data se následně automaticky zpracují a umístí do Archivu. *Konzumentem* se označuje cílová skupina uživatelů Archivu, který jej používá např. pro vyhledávání nebo získání uložených dat. *Řízení* má autorita, která rozhoduje, plánuje nebo obecně určuje a vynucuje pravidla Archivu. Ve své obecnosti je potřeba předpokládat Producenta i Konzumenta jako případné uživatele Archivu. Může se jednat o zcela jiné osoby, celé skupiny uživatelů, organizace, ale i různé instance Archivu, které si předávají data. [137]

Samotný Archiv má šest funkčních celků, které jsou uvedeny na obrázku 3.1 a tvoří *funkční model*, který odpovídá procesům v systému Archivu: (1) příjem (ingest), (2) perzistentní uchování (archival storage), (3) správa dat (data management), (4) administrace (administration), (5) zpřístupnění (access) a (6) plánování uchovávání (preservation planning). [137]

Nad uvedenými šesti funkčními celky existuje *informační model*, který je založen na předávání *informačních balíčků* mezi sebou. *Vstupní informační balíček* (submission information package, SIP) je předáván Producentem do Archivu. Podoba balíčku je na dohodě původce dat se zodpovědnou autoritou Archivu. Pro vlastní archivaci je důležitý *archivační informační balíček* (archival information package, AIP), který může být typu *archivační informační jednotka* (archival information unit, AIU) a *archivační informační kolekce* (archival information collection, AIC). Jednotkou se rozumí konkrétní digitální objekt. Kolekce je seskupením více jednotek. Konceptuální model OAIS je skutečně obecný, proto AIU a AIC nejsou zcela přesně vymezeny. Jednotkou může být kniha, ale mohou to být i jednotlivé kapitoly (např. každá kapitola v samostatném souboru). *Výstupní informační balíček* (dissemination information package, DIP) je odpovědí na požadavek zpřístupnění dat z Archivu a to ve formátu jakému Konzument rozumí. [137]

Balíček AIP je pro archivaci právě tím nejdůležitějším, protože je trvale uložen v Archivu. Samotná existence SIP na vstupu a DIP na výstupu má nezastupitelnou roli právě pro interoperabilitu Archivu. Umožňuje přijímat data od Producenta v jednom formátu a transformovat je pro účely trvalého uchování. Stejně

tak je pro Konzumenta možné provést transformaci AIP do výstupního formátu, který Konzument požaduje a ty mu odeslat v DIP. [137]

Celková obecnost a vysoká úroveň abstrakce modelu OAIS je výhodná pro svoji nezávislost na konkrétní implementaci. Splnit určený rámec modelu lze více způsoby, což ve svém důsledku může způsobovat nekonzistentnosti, pokud více autorů implementuje své řešení a třeba i se svým schématem. Z obecnosti plyne problematika kontroly a ověření, zda implementace OAIS vyhovuje. Problematikou certifikace a určení souladu s OAIS zejména ve vztahu k důvěryhodnosti Archivu pro Producenty a Konzumenty se proto věnuje např. ISO 16363:2012, který byl v roce 2017 revidován a potvrzen za platný, a ISO 16919:2014. [137, 138] Vedle norem existují certifikace např. Data Seal of Approval³⁴ (DSA), které jsou založeny na recenzním hodnocení šestnácti zásad a úrovně jejich splnění dle [139] (starší verze v češtině [140]). Také existují kontrolní seznamy a další způsoby pro kontrolu shody s OAIS viz [137, 138].

Autor Paul Arthur Laughton (2011) se v [141] zabýval organizací World Data Centre³⁵ (WDC) patřící International Science Council³⁶ (ISC; původně International Council for Science), která poskytuje přístup k vědeckým datům různých vědních disciplín. Dotazníkovou metodou Laughton analyzoval 52 členů z 12 zemí s cílem určení, zda je možné vytvořit společný framework založený na OAIS, protože členové WDC mají jiné požadavky na ochranu a správu dat, rozsah nebo četnost dotazů kladených na data. Mezi nejznámější implementace OAIS patří např. Archivematica³⁷ a DSpace³⁸. *Archivematica* je sada open source nástrojů, které uživatelům umožňují uchování digitálních objektů v souladu s OAIS. Při implementaci vycházející z OAIS se často využívají schémata metadat jako jsou Dublin Core (3.9.2), METS (3.9.3), PREMIS (3.9.4), BagIt [142] a další. V České republice se zabývá (2016–2020) vytvořením komplexního řešení na bázi OAIS pro dlouhodobé uchování digitálních knihovnických sbírek projekt ARCLib³⁹.

3.10.3 PAIMAS

Huc et al. (2004) publikovali *Producer-Archive Interface Methodology Abstract Standard* (PAIMAS) [118, 119]. Norma s označením ISO 20652:2006 [143] byla vydána v roce 2006, platnost této normy byla potvrzena v roce 2014. PAIMAS navazuje na referenční model OAIS (viz kapitola 3.10.2), ale zaměřuje se pouze na části Producenta a Archivu v otázce příjmu dat. Doporučení se vztahuje na aspekty týkající se dohody nutné pro předání dat, vlastní přijetí dat i zajištění kvality. Dle PAIMAS je vyžadováno, aby vždy (1) bylo definováno, co bude archivováno, (2) byl vytvořen formální souhlas, (3) došlo k provedení přenosu dat a (4) byla provedena validace získaných dat. Tím lze dosáhnout toho, že se do Archivu dostanou pouze data, která jsou schválena a očekávána. Po přenesení jsou zkontrolována. [143]

³⁴<https://www.datasealofapproval.org/> a česky na <https://dsa.cuni.cz/>

³⁵<https://www.icsu-wds.org/>

³⁶<https://council.science/about-us>

³⁷<https://www.archivematica.org/>

³⁸<https://duraspace.org/dspace/>

³⁹<https://arclib.cz/>

3.10.4 RDF

Resource Description Framework (RDF) je způsobem nebo *rámcem pro popis zdrojů*, který byl od počátku navrhován pro metadata v prostředí webu. První specifikace vznikla roku 1997 a v roce 1999 bylo vydáno doporučení pod záštitou World Wide Web Consortium (W3C). Specifikace konceptu RDF 1.0 (2004) byla vydána v [144] a RDF 1.1 o deset let později (2014) [145]. RDF je abstraktní datový model pro metadata a obecnou metodou pro konceptuální popis nebo modelování informací zejména webových zdrojů. Existuje podobnost s konceptuálním modelováním ve formě diagramu entity-relationship nebo tříd. Princip je založen na tvorbě tvrzení o popisovaném zdroji. RDF pochází z období vzniku dalších metadatových schémat jako bylo např. Dublin Core (viz [75] a kapitola 3.9.2). Velké množství takových schémat [106] lze proto ve spojení s RDF přímo používat k popisu zdrojů. Koncept RDF modelu měl za cíl řešení následujících pěti úloh v souladu s [144].

- Metadata pro web. Poskytnutí informací o webových zdrojích a systémy, které je budou využívat (např. hodnocení obsahu).
- Aplikace vyžadující otevřené informační modely místo omezených (plánování, popis organizačních procesů, anotace webových zdrojů).
- Strojově zpracovatelné informace (data aplikací) jako obdoba WWW a hypertextu. Umožnit zpracování dat mimo konkrétní místo jejich vzniku.
- Spolupráce aplikací a využití jejich dat pro získání informace nové.
- Automatizované zpracování informací z webu softwarovými agenty. Snaha změny od člověkem čitelných informací k celosvětové síti spolupracujících procesů.

Abstraktní syntaxe [144] kladla velký důraz na šest cílů: (1) jednoduchost modelu; (2) možnost odvozování/usuzování; (3) rozšiřitelnost slovníku; (4) syntaxi založenou na XML; (5) podpora datových typů XML schéma a (6) umožnit každému tvořit tvrzení o jakémkoliv zdroji.

Při návrhu aplikací založených na RDF je nutné počítat s tím, že informace mohou být neúplné nebo nekonzistentní. Nekonzistentnost může nastat právě, pokud kdokoliv vytvoří tvrzení (nesmyslné) v rozporu s ostatními tvrzeními. [144] Podstatou RDF je vytěžit maximum z možnosti sdílení informací. Ostatně W3C (2004) [144] uvádí: „*význam a hodnota informace roste s její přístupností a dosažitelností více aplikacemi napříč Internetem*“. RDF bylo navrženo maximálně flexibilně nad XML a XSD. Přes původní inspiraci v XML je možné použít více způsobů zápisu (viz kapitola 4.3), prostřednictvím nichž lze metadata sdílet mezi systémy. Reprezentace v abstraktním modelu RDF má minimum omezení, aby mohly být informace snadno sdíleny. RDF data nemusí být vždy veřejně dostupná. Mohou vzniknout i samostatné a uzavřené aplikace s vlastními slovníky, které nemají vazbu na okolní svět a internet. [144]

4 Technologie sémantického webu

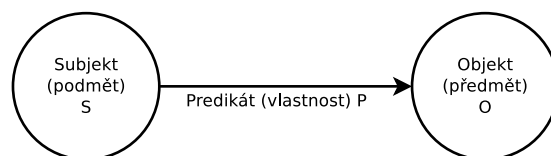
4.1 RDF koncept

První zmínka o Resource Description Framework (RDF) byla v kapitole 3.10.4. Současná verze RDF konceptu byla publikována W3C (2014) [145] a uvádí podstatu datového modelu RDF spočívající ve vytváření *tvrzení* (statements) o zdrojích, kterými může být zcela obecně libovolný popisovaný objekt. Odlišnosti oproti předchozí verzi z roku 2004 [144] uvádí [146]. Každé tvrzení má podobu *trojice* (triple) skládající se ze *subjektu* (subject, podmět, S), *predikátu* (predicate, vlastnost, P) a *objektu* (object, předmět, O). Subjektem je vždy popisovaný zdroj. Predikát představuje aspekty, vlastnosti nebo rysy zdroje a někdy se také označuje jako vlastnost (property) zdroje. Predikát vyjadřuje vazbu mezi subjektem a objektem. [147]

Trojice je zobrazena na obrázku 4.1 uzly (N) a orientovanou hranou (E), kde trojice S-P-O odpovídá uspořádání N-E-N. Orientace hrany v grafu je důležitá, protože označuje uzly, které jsou subjektem a objektem pro dané tvrzení. Množina trojic subjekt-predikát-objekt (S-P-O, SPO) je označuje jako *RDF graf*, který má orientované hrany. RDF graf je tvořen všemi trojicemi z množiny RDF trojic současně. [145, 147, 148]

Datový model RDF je velmi flexibilní. Disponuje schopností modelovat a reprezentovat různorodé druhy znalostí. Je na uživateli jakou granularitu metadat a schéma zvolí. Právě abstraktnost RDF umožňuje vytvářet tvrzení také o predikátech a objektech stejným způsobem, jakým jsou popisovány subjekty (zdroje). Pokud chceme popsat nějaký predikát (doplnit k němu metadata), musí existovat tvrzení, kde je identifikátor predikátu subjektem. Model disponuje formální sémantikou s přísně definovanými pravidly pro odvozování a dedukování z RDF dat. [145, 149, 150]

K pojmenování zdrojů a jejich jednoznačné identifikaci se používá URI [47, 144] nebo IRI [53, 145]. Identifikátor slouží k přesné identifikaci zdroje. Lokalizační funkce identifikátoru je doporučována. Popisovaný zdroj může být objektem z reálného světa a elektronická podoba nemusí existovat. V prostředí webu na webových aplikacích je časté, že IRI/URI se používá k lokalizaci s protokoly HTTP nebo HTTPS. Záleží na účelu identifikátoru, protože lze použít prakticky jakoukoliv variantu splňující platnou normu dle [47, 53]. Doporučení označené *Cool URI* (2008) [151] má zajistit jednoduchost na zapamatování, dlouhodobou stabilitu a výhodu při migraci schéma. Dokument [151] je zaměřen na URI, ale zřejmě lze aplikovat také na IRI. Subjekt je identifikován IRI nebo se používá *prázdný uzel* (blank node) – existuje nepojmenovaný anonymní zdroj. S anonymním zdrojem nelze přímo pracovat, protože není identifikovatelný. Predikát musí



Obrázek 4.1: RDF trojice obsahující subjekt, predikát a objekt.

být vždy pojmenován IRI a nesmí být prázdným uzlem. Objekt může být IRI, prázdný uzel nebo textový literál v Unicode. [145, 148]

4.2 Schéma a slovníky

Všechny strany účastníci se výměny RDF tvrzení musí mít shodu použitých identifikátorů a jejich sémantiky pro zajištění interoperability. Shoda mezi zúčastněnými stranami nezávisí přímo na RDF samotném, který je pouze obecný rámec pro popis zdrojů metadaty. Může být použito libovolných identifikátorů, ale existují vyhrazené jmenné prostory a URI pro základní slovníky RDF nebo ontologie. Úplná volnost není výhodná pro vzájemné porozumění. Základní stavební prvky jsou *RDF Vocabulary* (označení RDF) a *RDF Schema* (označován RDFS) [152]. Pro zkrácení zápisu jmenného prostoru mají slovníky vyhrazené prefixy `rdf` a `rdfs` (viz příloha A a výpis A.1). Definovány jsou základní třídy a vlastnosti pro popis zdrojů. Třídy jsou: `rdfs:Resource` pro vše popisované prostřednictvím RDF, `rdfs:Class` definuje RDF třídu, `rdf:Property` je rodičovská třída pro vlastnosti, `rdfs:Datatype` pro deklarace datových typů, `rdfs:Literal` zastupuje základní datové typy (literály), `rdf:langString` pro textové řetězce s definovanou jazykovou variantou, `rdf:HTML` pro HTML obsah a `rdf:XMLLiteral` pro literály XML. [152]

Definované jsou vlastnosti `rdfs:comment` pro komentáře, `rdfs:domain` označuje příslušnost zdroje s vlastností do uvedených tříd, `rdfs:isDefinedBy` uvádí, kde je zdroj definován, `rdfs:label` je člověkem čitelný název zdroje, `rdfs:range` určuje obor hodnot, `rdfs:seeAlso` označuje zdroj s dalšími doplňujícími informacemi, `rdfs:subClassOf` pro hierarchie tříd, `rdfs:subPropertyOf` pro hierarchie vlastností, `rdf:type` určuje, že instance je dané třídy (typu). [152]

Kontejnery jsou podporovány prostřednictvím třídy `rdfs:Container`, která je rodičem pro `rdf:Bag` (množina, neuspořádané prvky), `rdf:Seq` (sekvenčně uspořádané prvky), `rdf:Alt` (alternativní prvky, první je výchozí), `rdfs:ContainerMembershipProperty` (určena pro prvky kontejneru) a vlastnost `rdfs:member`. Pro kolekce je určena třída `rdf:List` a vlastnosti `rdf:first`, `rdf:rest` a `rdf:nil`. [152]

4.3 Formáty RDF dat

Pro sdílení a výměnu metadat je nutné, aby existoval syntaktický zápis abstraktního modelu. Primární je XML syntaxe [153], neboť RDF bylo od počátku XML inspirováno včetně podpory jmenných prostorů. Zápis RDF prostřednictvím XML se označuje RDF/XML a má MIME typ `application/rdf+xml` dle RFC 3870. Přesto existuje množství dalších způsobů zápisu abstraktního RDF modelu jako je N-Quads [154], N-Triple [155], TriG [156], Turtle [157] a JSON [158]. RDF popis je možné prostřednictvím RDFa¹ [159] vložit přímo do HTML dokumentu (atributy). Příklad dat v RDF/XML je na výpisu 4.1 a stejná data zobrazuje výpis 4.2, kde bylo použito syntaxe Turtle. Na příloženém DVD jsou reálné příklady v adresáři `data/` viz příloha I.

¹Resource Description Framework in Attributes

Výpis 4.1: Syntax RDF/XML

```
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:ds="https://mre.zcu.cz/ontology/dasta.owl#"

  <rdf:Description rdf:about="https://mre.zcu.cz/id/PX01">
    <ds:age>75</ds:age>
    <ds:name>PX01</ds:name>
  </rdf:Description>
</rdf:RDF>
```

Výpis 4.2: Syntax Notation 3

```
@prefix ds: <https://mre.zcu.cz/ontology/dasta.owl#>.

<https://mre.zcu.cz/id/PX01>
  ds:age "75";
  ds:name "PX01".
```

4.4 SPARQL

SPARQL představuje jazyk a protokol vyvinuté W3C pro dotazování a manipulaci s RDF grafem na webu nebo v RDF úložišti. Verzi SPARQL 1.1 [160] z roku 2013 předcházela SPARQL 1.0 (2008) [161]. Další existující jazyky pro dotazování² v RDF úložištích nemívají velkou podporu a neposkytují možnosti srovnatelné se SPARQL. Aktuální specifikace sestává z částí zabývajících se dotazovacím jazykem [162], modifikacemi grafu [163], popisem SPARQL služeb [164], federativními dotazy [165], sémantickou implikací (entailment) [166] a protokolem HTTP [167] nebo manipulací s RDF grafy v úložišti [168]. Nechybí formáty pro vrácení výsledků (JSON, CSV, TSV, XML).

Záleží na implementaci a možnostech používaného RDF úložiště, zda podporuje ODBC, JDBC nebo SPARQL Endpoint. *SPARQL Endpoint* umožňuje přijímat a zpracovávat SPARQL požadavky prostřednictvím URL a protokolu HTTP dle [167, 168].

Dotaz ve SPARQL má na počátku definovány prefixy, aby mohly být identifikátory zkráceny. Za deklarací prefixů následuje vlastní dotaz, který začíná klíčovým slovem označujícím typ dotazu (ASK, CONSTRUCT, DESCRIBE, a SELECT). Proměnné se označují prefixem znaku otazníku (?id) nebo dolaru (\$id) před názvem. V části WHERE jsou jednotlivé podmínky tvořeny trojicí, jejíž složky odpovídají subjektu, predikátu a objektu datového modelu RDF. Všechny podmínky jsou na RDF graf aplikovány současně a musí být splněny jako by bylo použito logické funkce AND. Podmínky omezují obsah RDF grafu. Při dotazu vzniká podgraf, který podmínkám vyhovuje, a je vrácen jako výsledek. Lze používat komplexnějších struktur, jako jsou skupiny podmínek (uvnitř složených závorek), vnořované dotazy, filtry nebo volitelné podmínky (klíčové slovo OPTIONAL). Při práci s více grafy je možné ještě před částí WHERE, uvést jejich označení formou FROM <...> nebo FROM NAMED <...>. [162]

Na výpisu 4.3 je jednoduchý SPARQL dotaz typu SELECT pro získání seznamu všech pacientů, kde bude uveden jejich identifikátor (?id), příjmení (?prijmeni)

²Existují také například RDQL (<https://www.w3.org/Submission/RDQL/>), Versa, RQL, SeRQL a XUL.

pohlaví (?pohlavi), datum narození (?narozeni) a úmrtí (?umrti). Bez klíčového slova OPTIONAL by byl vrácen pouze seznam zemřelých pacientů, protože mají neprázdný datum úmrtí.

Výpis 4.3: SPARQL

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
PREFIX ds: <https://mre.zcu.cz/ontology/dasta.owl#> .
SELECT ?id ?prijmeni ?pohlavi ?narozeni ?umrti
WHERE {
  ?patient rdf:type ds:Patient .
  ?patient ds:patientID ?id .
  ?patient ds:lastName ?prijmeni .
  ?patient ds:sex ?pohlavi .
  ?patient ds:datetimeBirth ?narozeni .
  OPTIONAL {?patient ds:datetimeDeath ?umrti .}
}
ORDER BY ASC(?id)
```

4.5 Web Ontology Language

Web Ontology Language (OWL) [169] je široce používaným ontologickým jazykem a využívá RDF včetně slovníků RDF a RDFS. Inspirován byl staršími ontologickými jazyky³. V roce 2012 byl vydán OWL 2 [169], který nahradil OWL [171] z roku 2004. Všechny původní platné ontologie OWL jsou kompatibilní s OWL 2. Novinky OWL 2 popisuje publikace [172]. Součástí, jako jsou třídy, vlastnosti a individua, jsou definovány v podobě RDF zdrojů a pro identifikaci se používá IRI/URI. Zápis OWL je nejčastěji ve formátu RDF/XML (viz kapitola 4.3). Tvorba ontologií v jazyce OWL je podporována např. editorem Protégé⁴ a pro odvozování nových tvrzení existují nástroje (reasoners) jako jsou Pellet, RacerPro, FaCT++ nebo HermiT. [171] Obě verze OWL podporují distribuované zpracování běžné v prostředí sémantického webu. Informace mohou pocházet z distribuovaných zdrojů dostupných na webu. Souvisí s tím import dalších potřebných ontologií, požadavek na odvozování v prostředí tzv. Open World Assumption (OWA).

OWL 2 umožňuje vysokou míru expresivity založené na deskripční logice (Description Logic, DL) SROIQ [173], která vychází z predikátové logiky prvního řádu. OWL 2 definuje profily [174], které jsou vhodné pro různé účely. Volba profilu závisí na struktuře ontologie a cílech odvozování. Profil *OWL2-EL* je založen na EL++ [175, 176] a škálovatelném uvažování v tzv. TBoxu (terminologická komponenta). Tento profil je pouze fragmentem OWL 2, kde většina inferenčních úloh včetně klasifikace je možná v polynomiálním čase. Profil *OWL2-QL* má základem DL-Lite [177] a škálovatelný dotaz v ABoxu (tvrzení, assertion component), kde je mnoho instancí dat, ale relativně jednoduchý TBox. Profil OWL2-QL je vhodný pro snazší přístup a dotazování v databázi uchovávaných dat. Profil *OWL2-RL* vychází z deskriptivní logiky programů (Description Logic

³OIL (Ontology Inference Layer), DAML (DARPA Agent Markup Language) a jejich kombinace DAML+OIL. [170]

⁴<https://protege.stanford.edu/>

Programs, DLP) [178, 179]. Výrazové schopnosti (expresivita) jsou podmnožinou OWL2 DL.

OWL 1 definuje profily označené (1) *OWL Lite*, (2) *OWL DL* a (3) *OWL Full*. V uvedeném pořadí roste míra vyjadřovacích schopností. Každá z variant je rozšířením předchozí jednodušší varianty. [171] Profil *OWL1-Lite* poskytoval odlehčenou variantu možností jazyka OWL 1. Vhodný byl pro uživatele, kteří potřebují pouze hierarchii pro klasifikaci nebo tvorbu systematického slovníku s některými základními omezeními. Ekvivalence je možná pouze mezi třídami. Vazba rodič-potomek je možná pouze u pojmenovaných tříd a nikoliv u libovolných výrazů. Omezení (restrictions) lze definovat jen u pojmenovaných tříd. Kardinalita může mít pouze hodnoty 0 a 1. [171] Varianta profilu *OWL1-DL* byla navržena poskytnout maximální možnou výřečnost při současném zachování výpočetní kompletnosti. Lze vypočítat (skončí v konečném čase) všechny logické důsledky a rozhodnutelnosti. OWL1 DL se shoduje s deskriptivní logikou a lze využívat i nástroje určené pro výpočetní odvozování. [171] Profil *OWL1-Full* je nerozhodnutelná ontologie (na rozdíl od předchozích OWL1-Lite a OWL1-DL). Žádný software pro odvozování není schopen provést její kompletní rozhodování. OWL1-Full respektuje kompatibilitu s RDF Schema, které v některých případech koliduje s OWL1-DL. Je podporována možnost rozšířit význam předdefinovaných slovníků RDF i OWL. [171]

4.6 Sémantický web

Sémantický web (Semantic Web) definoval [180] Tim Berners-Lee, autor WWW, v roce 2001. Sémantický web měl být evolučním krokem webu, neboť mělo jít o převedení stávajícího stavu na *web tvořený daty* (web of data). Prostý obsah na webu byl doplněn a provázán hypertextem. Stále byl obsah nestrukturovaný případně pouze semi-strukturovaný. Web tvořený daty měl být obohacen o strojově využitelná metadata. Sémantické informace mělo být možné doplnit prostřednictvím RDF do HTML dokumentu dle [159]. Zdroje na webu by byly srozumitelné člověku i čitelné a zpracovatelné strojově. Cílem bylo zvýšení efektivnosti a přesnosti uživateli poskytovaných služeb a zejména informací. Tim Berners-Lee (2001) definoval web dat slovy: „*A web of data that can be processed directly and indirectly by machines.*“ Volně přeloženo jako „*web dat, který může být přímo i nepřímo zpracováván stroji.*“ [180] Uvedený směr sémantického webu se v praxi příliš neprosadil. Již v roce 2006 sám Berners-Lee v [181] uvádí, že „*Tato jednoduchá myšlenka . . . zůstává z velké části nerealizovaná.*“ Běžný obsah byl autory na webu obohacen maximálně o několik málo metadat jako jsou autor, titulek, popis, klíčová slova nebo datum vytvoření. Přínos těchto metadat byl ryze praktický, dosažení vyšší pozice ve webových vyhledávačích. Skutečný sémantický význam pro webový obsah a jeho síla pro praxi významně využita nebyla. Určitý posun je znát s odstupem času, kdy W3C (2009) v [182] uvádí sémantický web jako „*společný rámec umožňující sdílení a opakované použití dat napříč aplikacemi, podniky a komunitami*“⁵ V roce 2013 byl sémantický web, s odstupem 12 let od

⁵W3C (2009) v [182]: „*The Semantic Web provides a common framework that allows data to be shared and reused across application, enterprise, and community boundaries. It is a collaborative effort led by W3C with participation from a large number of researchers and industrial partners.*“

svého vzniku, stále spíše okrajovou záležitostí. Rozhodně nešlo mluvit o velkém rozšíření a využití, které bylo předpovídáno. V prosinci 2013 byla původní aktivita *W3C Semantic Web Activity*⁶ zahrnuta jako součást *W3C Data Activity*. [183, 184] V minulosti vznikly další snahy jak RDF datový model a princip popisu využít nebo posunout dále i mimo prostředí webu. Prvním z nich byl sémantický desktop [185], po kterém následoval princip *propojených dat* (linked data).

4.7 Propojená data

Propojená data (linked data) využívají webové technologie a RDF pro vytvoření vazeb na související data. Prakticky se jedná o zobecnění hypertextu, kde RDF nahrazuje HTML odkaz. Bez existence propojení ve formě odkazů by nemohli objevit lidé ani stroje žádný obsah – web dat neboli sémantický web. [180, 181] Podstatou [181] je především doplnění informace o vztahu mezi daty (dokumenty, zdroji), ale nikoliv pouze lokalizační funkce. Index kompetencí (2017) propojených dat (Linked Data Competency Index, LDCI) přehledně shrnuje [186]. Propojená data poskytují metodu vystavení, sdílení, propojování dat, informací nebo znalostí a jejich částí, za použití standardizovaných postupů, kterými jsou HTTP, RDF a URI/IRI. Technologie a postupy lze aplikovat pro použití s libovolnou aplikací nebo informačním systémem. Pro svoji obecnost a rozšiřitelnost je možné princip založený na propojených datech použít ve všech aplikacích, kde je potřeba interoperability strukturovaných dat ve strojově zpracovatelné podobě. Na data lze odkazovat, prostřednictvím HTTP je jednoduchým způsobem propojovat, ale také distribuovat napříč webem a aplikacemi. [186]

Podle Berners-Lee v [181] platí pro růst webu čtyři jednoduchá pravidla: (1) pro pojmenování používat URI; (2) použít HTTP URI, aby mohli lidé pojmenování vyhledat; (3) jestliže někdo vyhledává URI, poskytnout užitečné informace s využitím standardů RDF a SPARQL; a (4) zahrnout odkazy na další URI, aby mohli lidé více objevovat. V [181] Berners-Lee uvádí doporučení, jak pravidla splnit. O čtyři roky později, v roce 2010, Berners-Lee přidává k pojmu propojených dat požadavek na otevřenost a vznikají *propojená otevřená data* (Linked Open Data, LOD). Prvním a nejdůležitějším požadavkem je, aby byla data publikovaná pod otevřenou licencí. Druhým požadavkem je, aby data byla strukturovaná a strojově čitelná. Třetím je použití neproprietárního datového formátu pro data. Čtvrtým požadavkem je používání otevřených standardů W3C pro identifikaci, aby mohli uživatelé na taková data odkazovat. Poslední pátý požadavek je přidávat odkazy na data ostatních, aby byl poskytován kontext. Je-li splněno všech pět bodů, pak se jedná o maximální splnění podmínek kladených na LOD. [181]

Téma propojených dat je stále aktuální. *W3C Data Activity*⁷, kam původní sémantický web přešel, uvádí existenci projektů s podporou Evropské komise v roce 2019 jako jsou např. Big Data Europe⁸ [184], nástroj pro škálovatelné zpracování rozsáhlých RDF dat SANSA⁹ nebo SPECIAL¹⁰ [187].

⁶<https://www.w3.org/2001/sw/>

⁷<https://www.w3.org/2013/data/>

⁸<https://www.big-data-europe.eu/>

⁹Semantic Analytics Stack; <http://sansa-stack.net/>

¹⁰Scalable Policy-aware Linked Data Architecture For Privacy, Transparency and Compliance; <https://www.specialprivacy.eu/>

5 Výzkum v medicíně

Projekty v oblasti zdraví populace a medicínského výzkumu mají množství podob. Jednotliví lékaři mají i osobní motivaci, protože přicházejí s pacienty přímo do kontaktu. Vhodná řešení přesahují jednotlivce. Existují a vznikají celé mezioborové týmy, kde s lékaři spolupracují další odborníci z jiných vědních oborů. Cíle jsou dosažení vyšší kvality života, snížení incidence, prevalence, mortality, morbidity, invalidizace, příp. celkové zrychlení diagnostiky nebo léčebného procesu pro danou diagnózu nebo skupinu populace. Významnou roli má sběr dat ve formě databází, registrů nebo repositářů, odkud mohou být následně používány pro *retrospektivní* nebo *prospektivní* analýzy a zpracování dat. Využívány jsou statistické metody a testování stanovených hypotéz. Je důležitý popis datového souboru, aby nedocházelo k chybné interpretaci výsledků.

Klíčová jsou data – diagnostická a léčebná, jejich aktuálnost, dostupnost, kvalita, přesnost, objem a úroveň podrobnosti. Existuje více způsobů, jak si taková data opatřit. Velmi záleží na konkrétní doméně¹, ale lze se setkat (veřejně nebo po registraci) s daty reálných pacientů v anonymizované podobě odvozenými od více skutečných osob současně (např. zprůměrováním) nebo různě vytvořenými idealizovanými modely a vzorky. Důležitou roli má motivace vedoucí ke vzniku a publikování dat. Existují archivy nebo registry poskytující rozličnou kvalitu datových sad z jedné nebo více domén, modalit, apod. Za všechny alespoň *The Cancer Imaging Archive*² (TCIA) [188] s rozsáhlým archivem anonymizovaných obrazových vyšetření s výskytem rakoviny, které lze volně stahovat a citovat. Může jít o publikaci dat za účelem porovnávání různých algoritmů pro řešení určité problematiky. Příkladem je web Grand Challenge³ uvádějící množství výzev zabývajících se analýzou medicínských snímků. Mezi prvními výzvami byla *Segmentation of the Liver Competition 2007* (SLIVER07) [189, 190], která je stále aktivní. Nabízí se ještě možnost spolupráce přímo se zdravotnickým zařízením a lékaři. V takovém případě je nutné zajistit ochranu osobních údajů. [VK10, VKKK19]

Každý typ dat má výhody i nevýhody. Veřejně dostupné datové sady mají výhodu v anonymitě. Současně mívají nevýhodu v omezenosti vzorku dat (maximálně) na řádově desítky případů. Výsledek nemusí být statisticky významný s ohledem na kontext dalších vedlejších diagnóz, pohlaví, věku, apod. Zcela jistě mohou být problémy s daty reálných pacientů i daty od nich odvozených, protože každý člověk je svým způsobem originál. [191] Připravený algoritmus zpracování dat nemusí správně fungovat např. pro zdravé osoby, protože ty se na některá vyšetření mohou dostat jen výjimečně.

5.1 Data ve zdravotnictví

Zdravotnictví zahrnuje doporučení, postupy a zdravotnická zařízení pro zajišťování zdravotní péče (health care) a veřejného zdraví obyvatelstva. Základem jsou zdravotnická zařízení. Zdravotní péče je poskytována v závislosti na typu takového zařízení. Jedná se o komplexní problematiku sestávající z množství disciplín

¹Rozcestník například https://resources.dfg.de/interdisz_is_en.html.

²<https://www.cancerimagingarchive.net/>

³<https://grand-challenge.org/challenges/>

a vědních oborů zahrnující vše od prevence (předcházení nemoci), přes diagnostiku, léčbu nemocí (patologického stavu těla nebo mysli), následnou péči, až po rekonvalescenci a návrat do běžného života. [192, 193]

Pro léčbu pacienta je nejdůležitější správné a včasné určení (hlavní) diagnózy. Roli a vliv mívají i ostatní souběžné diagnózy. Ke stanovení slouží provedení anamnézy, fyzikální vyšetření, vyšetření specialistou, laboratorní vyšetření, zobrazovací diagnostické metody a další (např. audiometrie, spirometrie a zátěžové vyšetření). Při zjišťování diagnózy pacienta vzniká relativně velký objem heterogenních dat, z nichž lékař vychází. Taková data označujeme pojmem primární data a jejich účel je primárně pro určení diagnózy a zvolení správného způsobu léčby nebo její kontroly. [192] Uvedené způsoby získání přehledu o stavu pacienta odpovídající vytvářeným datům, které lze klasifikovat na data (1) klinická, terapeutická a laboratorní, (2) obrazová, (3) signálová a (4) ostatní.

Ve zdravotnických zařízeních s agendou pomáhají informační systémy. Poskytují sběr, uchování a zpřístupnění dat pacienta. Mezi další funkce patří nepochybně možnost výměny dat s ostatními zdravotnickými zařízeními. Na zdravotní péči navazují agendy jako vykazování úkonů zdravotním pojišťovnám, předávání dat Ústavu zdravotnických informací a statistiky ČR (ÚZIS), České správě sociálního zabezpečení ČR (ČSSZ), i elektronické předepisování receptů⁴.

Legislativně je nejdůležitější zákon č. 372/2011 Sb. *Zákon o zdravotních službách a podmínkách jejich poskytování* (zákon o zdravotních službách). Zdravotnickou dokumentaci a její náležitosti stanovuje vyhláška č. 98/2012, o zdravotnické dokumentaci, ve znění pozdějších předpisů. Novelizována byla vyhláškou č. 137/2018 Sb. Elektronická zdravotní dokumentace (EZD) je kolekce dat nebo informací o pacientovi v digitální podobě. V angličtině je EZD označována *electronic health record* (EHR) nebo *electronic medical record* (EMR). Ani novelizovaná vyhláška č. 137/2018 Sb. neuvádí žádný strukturovaný a strojově zpracovatelný výměnný formát pro data ve vztahu k elektronické zdravotní dokumentaci a pro účely EZD je preferován formát PDF/A s požadavky na použití uznávaného elektronického podpisu a kvalifikovaného časového razítka. Transformace z klasické listinné podoby je možná dle zákona č. 300/2008 Sb., o elektronických úkonech a autorizované konverzi dokumentů.

Doporučení Evropské komise ze dne 2. července 2008 o přeshraniční interoperabilitě systémů elektronické zdravotní dokumentace [193] definuje EHR: „*elektronická podoba komplexního lékařského záznamu nebo podobné dokumentace o minulém a současném fyzickém a duševním stavu jedince, která zajišťuje okamžitou dostupnost těchto údajů pro lékařské ošetření a další úzce související účely*“⁵. Ministerstvo zdravotnictví ČR zaštiťuje *Národní strategii elektronického zdravotnictví*⁶. Navazuje na aktivity v prostředí Evropské unie a snahy interoperability zdravotních záznamů ve formě *eHealth Digital Service Infrastructure*⁷ (eHDSI) – zejména ve vztahu na služby *ePrescription* (eRecept) a *Patient Sum-*

⁴Zákon č. 378/2007 Sb., o léčivech. Web eReceptu je <https://www.epreskripce.cz/e-recept>.

⁵EHR dle [193]: „*a comprehensive medical record or similar documentation of the past and present physical and mental state of health of an individual in electronic form, and providing for ready availability of these data for medical treatment and other closely related purposes*“

⁶<http://www.nsez.cz/>

⁷https://ec.europa.eu/health/ehealth/electronic_crossborder_healthservices_en

mary (pacientský souhrn)⁸. [194, 195] Zavedením přeshraničních služeb eHealth v České republice se zabývá NIX-ZD⁹. V prezentaci [196] je uveden stav v dubnu 2019 a zmíněny jsou legislativní problémy v České republice (např. chybějící novelu zákonů).

5.1.1 Klinická, terapeutická a laboratorní data

Datový standard Ministerstva zdravotnictví ČR

*Datový standard*¹⁰ (DASTA, DS) vydává Ministerstvo zdravotnictví České republiky (MZČR). DASTA je národní formát určený pro výměnu dat nejen mezi informačními systémy ve zdravotnických zařízeních. Slovenská republika oficiální podporu DASTA nemá, ale přesto je standard využíván v informačních systémech dodávaných firmami z Česka. V ostatních zemích není formát DASTA využíván. [197] Oficiální verze DS1 byla vydána v roce 1997 a jednalo se o čistě textový formát. Verze DS2 (2002) byla již ve formátu XML s DTD. O rok později (2003) byla vydána DS3. V roce 2007 byla vydána poslední a doposud aktuální platná a vyvíjená verze DS4, která je založená na XSD. [198] Od července 2016 byla ukončena podpora DS3. Jediným platným a podporovaným formátem je DS4. [197] Nezbytnou součástí je rozsáhlý Národní číselník laboratorních položek (NČLP), který vychází z mezinárodně uznávané nomenklatury International Federation of Clinical Chemistry and Laboratory Medicine¹¹ (IFCC). Každá položka v NČLP má unikátní klíč a je definována pěticí hodnot: (1) systém, (2) komponenta, (3) procedura, (4) druh veličiny a (5) jednotka. NČLP je průběžně doplňován a vychází pravidelně jednou za čtvrt roku. [199]

Aktuální verze jsou dostupné na webu DASTA: DS 04.20.03 a NČLP 2.70.01 (1. červenec 2020). Číselníky jsou dostupné prostřednictvím webových služeb¹². V datových blocích DASTA je využíváno množství interních a externích číselníků, které lze rozlišit dle původu – přímo číselníky Datového standardu (DS), Národní číselník laboratorních položek (NČLP) a Ústavu zdravotnických informací a statistiky ČR (ÚZIS; Národní zdravotní registry). [200, 201]

Komunikační standard DASTA podporuje sdílení dat mezi informačními systémy. Umožňuje výměnu dat zahrnující identifikaci pacienta, informace základní, urgentní, anamnézu, aktuální i trvalé diagnózy, léky, očkování a zejména tzv. klinické události. Klinickou událostí může být vyšetření laboratorní, radiologické, konzilium, ambulantní zpráva a další typy. Klinická událost může být v množství fází (objednávka, potvrzení, zpracovávání zakázky, předání dat třetí straně). [200, 201]

Formát DASTA je využíván také pro výměnu informací, které se nevztahují přímo k léčbě a vyšetření pacienta. DASTA slouží k vykazování zdravotním pojišťovnám, do Národního zdravotnického informačního systému (NZIS), Ústavu zdravotnických informací a statistiky (ÚZIS) a podávání hlášení o pracovní ne-

⁸Zejména pro urgentní případy mimo zemi pacienta má pomoci odstranit jazykovou bariéru. Pacientský souhrn obsahuje nejdůležitější informace, jako jsou např. alergie nebo aktuální medikace a to v jazyce ošetřujícího lékaře.

⁹<https://www.nixzd.cz/>

¹⁰<https://www.dastacr.cz/>

¹¹<https://www.ifcc.org/>

¹²<http://ciselniky.dasta.mzcr.cz/>

schopnosti na Českou správu sociálního zabezpečení (ČSSZ). Mimo to DASTA počítá se speciálními datovými bloky, které mohou využít informační systémy lokálně. [200, 201]

Rozšíření formátu DASTA mimo Českou a Slovenskou republiku je nepravděpodobné. Zásadní komplikací bude zcela jistě jazyková bariéra. Dokumentace je pouze v českém jazyce. XML elementy a atributy datového formátu jsou v češtině nebo z českých slov odvozeny. Obecně nejsou číselníky vícejazyčné. Kladem je rozsáhlý číselník laboratorních položek. [200]

Health Level Seven

*Health Level Seven International*¹³ (HL7) je nezisková organizace akreditovaná American National Standards Institute (ANSI) od roku 1987. Zaměřuje se na vývoj standardů z oblasti zdravotnictví. Doménou standardu HL7 je výměna dat o zdravotní péči pacientů a související administrativě. Zavádí standardy, doporučení a metodologie, které mají sloužit ke zjednodušení komunikace mezi různými zdravotnickými systémy nejen na národní úrovni. Uvedená organizace Health Level Seven International vytváří, spravuje a vydává stejnojmenný mezinárodní standard *Health Level Seven* umožňující sdílení a interoperabilitu informací o zdravotní péči. Slovní spojení „Level Seven“ odkazuje na sedmou vrstvu síťové architektury ISO/OSI¹⁴ modelu označovanou aplikační vrstva. [202, 203] Vlastní zprávy HL7 mohou být ve vzájemně nekompatibilních verzích v2 a v3.

HL7v2 Druhá řada HL7 je úzce zaměřena na komunikaci v klinické praxi (registrace pacienta, objednávky). Formát zpráv je čistě textový s použitím speciálních jednoznakových oddělovačů lišících se dle toho, co oddělují (roura, stříška, apod.). Zpráva je členěna do segmentů tvořených jednou řádkou. Segment obsahuje vždy jednu specifickou kategorii informace. Každý segment má tříznakové pojmenování a může být ve zprávě zopakován. V rámci segmentu existují pole (composites/-fields) a jejich podpole (sub-composites/subcomponents). V podpoli může být vloženo další podpole. První segment je pojmenovaný MSH s polem identifikujícím typ zprávy a dle toho jsou očekávány další segmenty ve zprávě. [204]

HL7v3 Třetí řada HL7 má modelovat veškeré informace, vztahy a procesy ve zdravotnictví. Zaměřuje se na související činnosti jako je řízení, vykazování, apod. Zásadním rozdílem třetí řady je vývoj a modelování založený na formální metodologii *HL7 Version 3 Development Framework* (HDF; ISO/HL7 27931:2009) a objektově orientovaných principech s použitím UML. Pro přenos je používán XML. [205] Základem jsou definice slovníků a HL7 Reference Information Model (RIM) dle ISO/HL7 21731:2014. RIM vyjadřuje data nutná pro specifický klinický nebo administrativní kontext a poskytuje explicitní reprezentaci sémantických a lexikálních vazeb existujících mezi informacemi v jednotlivých polích HL7 zpráv. Slovníky a používané kódy mohou pocházet z různých zdrojů, ale je vyžadováno, aby systémy implementující standard měly sémantickou informaci (jednoznačné porozumění) o všech používaných kódových systémech a jejich hodnotách. [204, 205]

¹³<https://www.hl7.org/>

¹⁴International Organization for Standardization (ISO), Open Systems Interconnection (OSI)

Clinical Document Architecture (CDA nebo HL7 CDA) je standard založený na XML a určuje způsob kódování, struktury a sémantiky klinických dokumentů pro jejich sdílení a výměnu. CDA nedefinuje způsob serializace. Tím umožňuje použití čistě textového formátu dle HL7v2 stejně jako XML dle HL7v3. První vydání bylo v roce 2000 odvozené z referenčního informačního modelu pro HL7. Od roku 2005 existuje druhé vydání, které se zaměřuje na sémantickou reprezentaci klinických událostí. V roce 2009 bylo druhé vydání CDA uvedeno pod označením ISO/HL7 27932:2009¹⁵ s aktualizací v roce 2012 a revizí roku 2015, kdy byla konstatována a potvrzena platnost normy. Důraz byl kladen na sdílení a výměnu dat. Dokument CDA se zapisuje prostřednictvím XML a jeho struktura je odvozena od RIM s použitím terminologických slovníků. Velkým posunem je možnost postupného zvyšování sémantické interoperability. Dokument umožňuje předávat informace s prakticky nulovou sémantickou informací, ale také dokumenty s bohatou strukturovanou hlavičkou a sémantické informace v maximální podobě. Povinné textové části dokument obsahuje kvůli čitelnosti člověkem. Strukturované části jsou nepovinné a mohou využívat různých kódových systémů, jako jsou např. SNOMED CT¹⁶ a LOINC¹⁷. Strukturované části jsou určeny pro snadnou strojovou interpretaci dokumentu. [202, 204, 205]

DASTA vs. HL7

O volbě mezi DASTA a HL7 se v České republice vede diskuze mnoho let. Na podporu HL7 vzniklo v roce 2001 občanské sdružení *HL7 Česká republika*¹⁸. [206] Sdružení mělo cíle jako navázat oficiální vazbu na mezinárodní úrovni s HL7, implementaci, zohlednění národních potřeb a sdílení zkušeností. Proběhlo několik seminářů, ale žádné výstupy nebyly publikovány. To potvrzuje v roce 2013 [207] jeden z autorů DASTA. Stav webové prezentace sdružení to potvrzuje (2019), kde byla poslední aktualizace v roce 2015 po volbě nového představenstva.

V minulosti bylo provedeno několik průzkumů mezi tvůrci informačních systémů ohledně používání formátu DASTA, jeho budoucnosti a vztahu k HL7. Odhalují skutečný stav standardu, silné a slabé stránky, i rizika při použití. V provozu byl dokonce software používající více než 15–20 let neudržované verze DASTA (DS1 a DS2). DS4 byl v době průzkumu (duben 2016) implementován ve více než polovině systémů. Podpora DS3 přitom končila v červenci 2016. Přehledová tabulka [201] prezentuje používané verze a stáří podporovaných verzí DASTA v roce 2011. Firmy nemají zájem ani snahu o přechod na vyšší verze DASTA. Neimplementovaly ani kritické opravy umožňující chybné sdělování laboratorních výsledků (změny z května 2010 zapracovalo pouze 16 % tj. 4 tvůrci z 24). Firmy nejsou motivovány ani legislativně. Zdůvodnění firem je zejména finanční dopad – náklady na nutný vývoj, změny a následné nasazení u zákazníků. [201, 208] Starší průzkum z roku 2010, publikovaný v [201, 209] uváděl, že 56,3 % vývojářů (9 z 16 odpovědí) je i proti pouhým úvahám o HL7. Ochotných o HL7 uvažovat bylo 37,5 % (6 z 16 odpovědí). [201] Firmy provozují software v České republice (příp. Slovenské republice) a expandovat do zahraničí se nechystají. Dle [201, 209] pouze

¹⁵Data Exchange Standards – HL7 Clinical Document Architecture, Release 2

¹⁶Systematized Nomenclature of Medicine–Clinical Terms

¹⁷Logical Observation Identifiers Names and Codes

¹⁸<https://www.hl7cr.eu/>

jediný IS z 26 podporoval HL7 v roce 2011. Podporu HL7 mají především subjekty poskytující služby v zahraničí. [201, 208] Firmy nemají zájem a vlastně ani důvod interoperabilitu s HL7 podporovat. Velmi specifický národní formát pomáhá odfiltrovat mezinárodní konkurenci, která by s HL7 mohla na trh přijít.

V listopadu 2018 proběhlo setkání tvůrců a správců standardů s dodavateli informačních technologií v souvislosti přípravou zákona o eHealth a budoucnosti nebo dalším rozvoji DASTA. [200] Na setkání byl také zástupce HL7 ČR, který uvedl stav a praktické příklady HL7 z některých zemí Evropské unie. [206] V období 2018–2019 byla verze DS4 rozšířena o patientský souhrn (Patient Summary) a některé další potřebné číselníky v souladu s HL7. Důvodem bylo zajištění eHDSI, které bude založeno na transformaci potřebné části z národního formátu DASTA¹⁹ do HL7. [196] V únoru 2019 bylo vydáno doporučení [211] týkající se evropského výměnného formátu EHR. Odstavec 11. uvádí patientský souhrn a eRecept jako u eHDSI, ale nově rozšiřuje záběr na laboratorní výsledky, obrazová vyšetření a zprávy. V případě hospitalizace se vztahuje i na propouštěcí zprávy. Doporučení v [211] uvádí konkrétní datové formáty, které mají být pro interoperabilitu zdravotních záznamů používány. Uvedeny jsou formáty DICOM (viz 5.1.2) a *HL7 CDA Release 2* (5.1.1 v úrovni 3 a/nebo 1 (PDF/A)). [211]

Zcela jistě bude potřeba do budoucna v informačních systémech počítat s podporou obou formátů a to i v jejich různých verzích. Může dojít na velmi dlouhý a náročný proces změny hluboce zakořeněného formátu DASTA, kterému bude ještě předcházet (možná nejistá) fáze přizpůsobení HL7 do národního prostředí. Možnou cestou je další rozvoj a rozšiřování stávajícího DS4 a provádění obousměrné transformace DASTA a HL7 na uzlech přeshraniční výměny dat. Takový směr je aktuálně realizován při zajištění eHDSI. Dle [207] se v NČLP objevují už i popisy v anglickém jazyce. Vynutit přechod na DS4 (tam, kde dosud není podporován) může být jednodušší, než od základů změnit systémy na HL7. Nelze jednoznačně tvrdit, že by byl formát DASTA ideálním. Množství problémů včetně těch organizačních v případě vývoje a implementace DASTA by bylo možné označit za alarmující. Z vnějšího pohledu přenos a sdílení dat prostřednictvím DASTA mezi informačními systémy ve zdravotnických zařízeních funguje. Roztříštěnost implementovaných verzí DASTA je a do budoucna může být rizikem pro interoperabilitu.

5.1.2 Obrazová data

Digital Imaging and Communications in Medicine (DICOM, ISO 12052:2006²⁰) je standard vyvinutý ve spolupráci mezi American College of Radiology (ACR) a National Electrical Manufacturers Association (NEMA). Aktuální je třetí verze (označovaná PS 3) publikovaná poprvé v roce 1993. Od první publikace došlo k množství úprav, a proto se pro rozlišení používá rok vydání změny. Standard je určen pro distribuci, výměnu a zobrazování medicínské obrazové dokumentace nezávisle na jejich původu. Umožňuje přímou výměnu mezi počítači i technickým vybavením, např. počítačová tomografie a magnetická rezonance. Poskytuje specifikaci datového formátu a řadu služeb souvisejících s přenosem dat po síti,

¹⁹Část obsahující Patient Summary přidána do DS4 od verze DS04.16.00. [200, 210]

²⁰Health informatics – Digital imaging and communication in medicine (DICOM) including workflow and data management

např. potvrzování přenosu (uložení na perzistentní médium), dotazování a získávání z PACS (Picture Archiving and Communication System), získávání informací o pacientech na zařízení (modalitě) nebo výpis provozních informací o využití modality, možnost tisku na DICOM tiskárně. [212]

Ve formátu DICOM může být uchováváno množství různých obrazových i neobrazových dat. Informace ve formátu DICOM jsou seskupovány (data sets) – obrazová data jsou uložena společně s jejich popisem. Identifikace pacienta a další informace jsou vždy přítomny v souboru s obrazovými daty. Je-li vyšetření složeno z více snímků, pak bývá každý snímek uložen v samostatném DICOM souboru a všechny soubory obsahují kompletní a v řadě atributů redundantní popis. Tím je zabráněno možné chybě s přiřazením dat pacientovi, kdyby byly informace a obrazová dokumentace od sebe odděleny. [212]

Datový formát DICOM představuje binární soubor. Organizačně je vyšetření pojmenováno *Studie* (Study), která obsahuje *Série* (Serie) složené z jednotlivých *Snímků* (Image). Existují i objekty *Pacient* (Patient), *Návštěva* (Visit), *Zařízení* (Equipment) a některé další. V hlavičce DICOM souboru jsou uložena metadata ze všech úrovní této hierarchie. Metadata jsou identifikátory (pacienta, studie, série, snímku), informace o pacientovi (např. jméno, příjmení, datum narození, pohlaví, výška/váha), studii (např. typ použité modality, datum/čas, název/popis), sérii (např. název/popis, datum/čas, protokol) a snímku (např. datum/čas, způsob zobrazení, tloušťka řezu, rozměry). Metadata jsou zapsána ve formě atributů identifikovaných hexadecimálním číselným tagem. Každý atribut má své jméno, popis, datový typ a příznak, zda je atribut povinný. Obrazová data mohou být ve třech nebo dokonce čtyřech rozměrech při použití více snímků v jednom souboru (cine-loop a frames). Obrazová data i metadata mohou být komprimována. [212]

5.1.3 Signálová data

European Data Format (EDF) je formát pro uchování a výměnu vícekanálových biologických a fyziologických signálů, který byl publikován v [213] (1992). EDF se stal de-facto standardem pro elektroencefalogram (EEG) a polysomnographii (PSG) z komerčních zařízení stejně jako pro výzkumné projekty mezi spolupracujícími centry. Formát EDF je nezávislý na hardware a software. Umožňuje použití libovolného vybavení, jako jsou převodníky, předfiltrování nebo vzorkovací frekvence. Vznikl pro spolupráci sedmi laboratoří, výměnu jejich dat, vybudování společné databáze záznamů a přinesl možnost porovnání algoritmů na shodných datech. Formát přispěl ke spolupráci při vyhodnocování ručních i automatických metod analýzy dat. [213] O deset let později byl vytvořen s EDF kompatibilní EDF+ [214] (2002), který umožnil uchovat doplňující textové anotace s časovou značkou, podněty (stimuly), průměrné hodnoty signálů, parametry elektrokardiografu, dočasná zástava dýchání (apnoea) a další. EDF+ podporuje uchování nejen anotací, ale také elektromyografie, evokovaných potenciálů, elektroneurografie, elektrokardiografie a řadu dalších vyšetření. V EDF+ je možné uchovat výsledky automatické i ruční analýzy. [214] Specifikace obou formátů je v publikacích [213, 214], které jsou volně k dispozici na webových stránkách EDF²¹.

²¹<https://www.edfplus.info/>

Walter Graphtek je původní označení německé firmy vyvíjející EEG systémy a nástroje pro neurofyziologickou diagnostiku. Vlastní a veřejně nedokumentovaný binární formát pro signálová data je WG1 a obsahuje data pro jednotlivé kanály. Identifikaci pacienta navíc poskytuje až WG2. Konverzi do EDF/EDF+ může nabízet software dodávaný s EEG systémy. Vhodný nástroj pro konverzi zmiňuje [215]. Firmu převzala *inomed Medizintechnik GmbH* v roce 2012. [216]

5.1.4 Data ostatní

Pod ostatní data náleží jakýkoliv výše neuvedený formát. Mohou to být formáty strukturované, semi-strukturované, ale i nestrukturované. Mohou ad hoc vznikat nebo měnit se včetně použitého formátu nebo jejich struktury (schéma). Obecně heterogenní data mohou být specifická pro konkrétní doménu, situaci, zařízení nebo jiný zdroj dat. Vlastní specifický formát mívají často výrobci zařízení, jako jsou EEG systémy, glukometry [KKK⁺16a, KKK⁺16b] nebo moderní nositelná elektronika, která se stává součástí běžného života. Může poskytovat cenné informace o stavu a chování populace. Množství informací o léčbě pacienta je, nebo může být, doposud pouze v papírové podobě a není možné ani reálné tato data hromadně digitalizovat a propojovat se stávajícími informacemi o léčbě pacienta. Příkladem mohou být hodnoty teploty a krevního tlaku, který nelékařský zdravotnický personál sleduje a zaznamenává u hospitalizovaných pacientů do tzv. *teplotní tabulky*. Po operacích se tyto údaje mohou zapisovat do tzv. *akutní karty* společně s informacemi, kterými jsou např. stav vědomí, příjem a výdej tekutin a dalšími. Informace nebývají nebo nemusí být vždy v elektronické podobě. Pro tento příklad, by mohl být list s hodnotami naskenován, ale jiný přínos, než možnou elektronickou archivaci, by pravděpodobně neměl.

Současně je nutné uvažovat budoucí vývoj, kdy mohou vzniknout nové verze používaného formátu nebo být proveden přechod na formát zcela nový. V případě dlouhodobého sběru dat může průběžný vývoj oblasti přinést změnu struktury sbíraných dat, jejich objemu, ale i modifikace číselníků. Předem neznámá data (v době návrhu) se budou sbírat nebo vznikat při výzkumné činnosti.

5.1.5 Klinická událost a struktura dat

Z formátu DASTA je nejdůležitější pojem *klinická událost* (KU) představující jakoukoliv interakci mezi pacientem a zdravotnickým zařízením. Klinická událost může být vyžádaná (odeslání pacienta na CT vyšetření) nebo nevyžádaná (ad hoc příchod pacienta). Rozlišují se čtyři stavy KU (a) zadáno, (b) objednáno, (c) čekající a (d) provedeno. Klinickou událostí může být příjem/propuštění pacienta (hospitalizace), anamnéza, dekurz nebo provedené vyšetření (laboratorní, obrazové, apod.). Každá KU má povinné údaje odkazující na pacienta, datum, čas a typ události, pracoviště/oddělení, lékař/tým, použité přístroje, provedené výkony, použitý materiál a zejména dokumentaci. Dokumentace je nejčastěji ve formě prosté textové lékařské zprávy (nestrukturovaný text), laboratorních výsledků (strukturované), ale mohou být připojeny přílohy (např. obrazová dokumentace).

Formáty DASTA a HL7 popsané v kapitole 5.1.1, umožňují formálně zapsat vše výše uvedené. Jedna klinická událost odpovídá právě jednomu souboru v ně-

kterém z uvedených formátů. Pro obrazová vyšetření je používán formát DICOM (viz 5.1.2). Celé obrazové vyšetření může být zapsáno více způsoby. Vkládat DICOM soubory do XML (DASTA, HL7v3) není praktické ani vhodné a nemusí to být ani možné (HL7v2). Propojení se v DASTA realizuje uvedením názvu souboru ve speciálním elementu označujícím přílohy. Současně se zapisuje element s číselným označením DICOM studie. Obecně může být připojen soubor jakéhokoliv formátu (např. EDF+, PDF, skenovaný dokument), pokud to podporuje informační systém. Formát DASTA (resp. HL7v3) funguje ve významu pomyslné obálky. Pro obrazová data je radiologem provedený nálezný zapsán v části textové zprávy. Na přiloženém DVD jsou příklady v adresářích `data/data-*/` viz příloha I.

5.2 Ochrana osobních údajů

Zajištění ochrany osobních údajů je nezbytnou součástí, pokud mají být použita primární data zdravotní péče pro výzkumné účely, kde není *identita* osob důležitá. Ani personalizovaná medicína nevyžaduje znalost skutečné *identity* osob. Na mezioborovém výzkumu se podílí množství dalších osob. Pojem *lékařský výzkum* neznamena výzkum jediného lékaře. Identita může být odhalena prostřednictvím atributů, které jsou samy o sobě anonymní. V kontextu nebo ve spojení s další informací mohou být unikátní a stanou se *nepřímým* identifikátorem. V případě obrazové dokumentace mohou být těmito atributy anatomické struktury a tvary. [191] Kostí nebo tvar lebky v obrazových datech mohou poskytovat biometrické informace. Obrazová vyšetření jsou opatřena metadaty, kterých se lze relativně jednoduše zbavit [VK11c], ale modalit mohou zapisovat (vypalovat) textovou informaci do snímků vyšetření. Na některých modalitách je to dáno historicky jako prevence záměny pacientů (ECHO/SONO), anebo jsou tímto způsobem doplněny různé formy anotací. [VKKK19]

V České republice byl od června 2000 v platnosti Zákon č. 101/2000 Sb.²² až do 24. dubna 2019, kdy byl zrušen zákonem č. 110/2019 Sb. o zpracování osobních údajů. Nový zákon č. 110/2019 Sb. upravuje zpracování osobních údajů podle nařízení Evropského parlamentu a Rady (EU) 2016/679²³ známé pod označením GDPR, které vešlo v platnost 25. května 2018. K velkým změnám nedošlo, což dokazuje dokument [217] zpracovaný Úřadem pro ochranu osobních údajů²⁴ (UOOU). V zákoně č. 110/2019 Sb. se druhý díl věnuje *Zpracování osobních údajů prováděné pro novinářské účely nebo pro účely akademického, uměleckého nebo literárního projevu*, kde § 17 odst. 1 umožňuje *zpracování osobních údajů přiměřeným způsobem také pro účely akademického, uměleckého nebo literárního projevu*. Podmínku přiměřenosti dle GDPR ovlivňuje, zda osobní údaje zahrnují tzv. zvláštní kategorie osobních údajů (dle GDPR čl. 9 odst. 1), kam informace o zdravotním stavu patří. Výjimku dle GDPR čl. 9, odst. 2., písm. (j) tvoří *zpracování pro účely vědeckého výzkumu nebo pro statistické účely*, které má být přiměřené sledovanému cíli, dodržovat podstatu práva na ochranu osobních údajů,

²²o ochraně osobních údajů a o změně některých zákonů

²³Nařízení Evropského parlamentu a Rady (EU) 2016/679 ze dne 27. dubna 2016 o ochraně fyzických osob v souvislosti se zpracováním osobních údajů a o volném pohybu těchto údajů a o zrušení směrnice 95/46/ES (obecné nařízení o ochraně osobních údajů).

²⁴<https://www.uouu.cz/>

poskytovat vhodné a konkrétní záruky pro ochranu základních práv a zájmů subjektu údajů.

V souvislosti s GDPR vznikla novela č. 137/2018 Sb., jejíž podstatou je rozšíření zdravotní dokumentace o další informace. Nově budou informace o ostatních kontaktních údajích pacienta (telefonní číslo, adresa elektronické pošty, adresa pacientova bydliště) vedeny na základě právního předpisu. Poskytovatel zdravotních služeb může nově žádat o informace a kontaktní údaje třetích osob (zákonný zástupce pacienta, opatrovník, další osoby oprávněné udělit za něj souhlas).

I před zavedením GDPR se ochraně osobních údajů v souvislosti s EHR věnovala poměrně velká pozornost. S rozvojem výpočetní techniky a síťové infrastruktury vznikla snaha tuto oblast chránit ve Spojených Státech Amerických v podobě *Health Insurance Portability and Accountability Act* známý pod označením HIPAA od roku 1996. Souhrn pravidel uvádí množství publikací a odkazují na regulaci jako novou éru ochrany zdravotních záznamů [218, 219], ale později se objevují i další snahy [220] (2012), [221] (2013) a [222] (2014). Důsledky GDPR na vědecký výzkum zmiňuje publikace [223]. Další detaily ve vztahu k ochraně dat zdravotní péče a výzkumným účelům popisuje příručka [224] (2018) zejména v kapitolách 2.1.1, 9.3 a 9.4. Pro vědecké účely je zmíněna možnost pseudonymizace dat, jako vhodný způsob skrytí skutečné identity osob. V souvislosti s eHealth, lze zmínit zjištění případové studie [35] (2018) o bezpečnosti mezinárodních informačních systémů (EURODAC, SIS II, VIS) v EU.

k-anonymita [225, 226] má za cíl zamezit případu, kdy by pro některé specifické dotazy měl být vrácen jediný výsledek. Ochrana osobních údajů provedená přímo na úrovni zdrojových souborů sama o sobě nemusí stačit. Provedení deidentifikace není totéž jako anonymizace a pro skutečné zajištění ochrany osobních údajů je nutné počítat s obezřetností, protože v datech mohou být nepřímé identifikátory, které společně s další informací povedou k odhalení identity osob. Zejména při publikaci výsledků je nutné uvažovat *k-anonymitu*. Identifikujeme atributy jednoznačně ukazující na konkrétní osobu dle následující definice.

Definice 1 (*k-anonymita*). *Mějme k celé číslo ostře větší než 1. Podmínka k -anonymity je splněna právě tehdy, když v neprázdném souboru dat \mathcal{D} vyhledávání $f(\mathcal{D})$ vrátí výsledky \mathcal{X} , jejichž počet je větší nebo roven číslu k .*

$$\mathcal{D} \notin \emptyset, k \in \mathcal{N}, k > 1,$$

$$\mathcal{X} = f(\mathcal{D}), \mathcal{X} \subseteq \mathcal{D},$$

$$|\mathcal{X}| > k$$

Číslo k musí být větší než 1, protože existoval-li by pouze jediný výsledek vyhledávání, již by se jednalo o konkrétní osobu. Je nutné zajistit **nejvyšší hodnotu k -anonymity**, která je možná. Pro výzkumné účely může být vhodné některé informace ponechat nebo je pseudonymizovat. Při publikování dat je nutné dbát na hodnotu k -anonymity, která může ovlivnit proces ochrany osobních údajů.

Důsledek definice ilustruje příklad, kdy hledáme v souboru čtyř žen a šesti mužů podle pohlaví. Vyhledávací podmínka na ženy vrátí výsledek o velikosti čtyř záznamů, tj. soubor dat poskytuje *4-anonymitu* ($k = 4$). Zpřesníme vyhledávání doplněním další podmínky na věk osoby. Dokážeme identifikovat konkrétní osobu, pokud obdržíme ve výsledcích pro některou hodnotu věku pouze jediný záznam. Identita osoby může být odhalena při znalosti reálného prostředí (např. PSC malé obce nebo oblasti, v níž došlo k jedinému případu onemocnění).

5.3 Cévní mozková příhoda

Cévní mozková příhoda (CMP) je vždy urgentní stav ohrožující život v důsledku poruchy krevního oběhu v mozku pacienta. Může se jednat o poškození funkcí celkové nebo v části mozku. Na úspěch léčby pacienta a zotavení má zásadní význam včasné rozpoznání příznaků iktu, rychlá přednemocniční neodkladná péče i následná nemocniční péče. [227, 228] Typy CMP se rozlišují podle způsobu vzniku, kterému odpovídá způsob léčby pacienta. *Ischemické* CMP představují asi 80 % všech CMP. Nejčastěji jsou způsobeny postupným trombotickým nebo náhlým tromboembolickým uzávěrem přívodné tepny. *Hemoragické* CMP se vyskytuje asi v 15 % případů CMP. Jedná se o krvácení do mozku vyvolané rupturou některé z mozkových tepen. *Subarachnoidální* CMP se vyskytuje asi v 5 % případů CMP. Jde o krvácení, které je vyvoláno výronem krve do subarachnoidálních prostor. [229, 230]

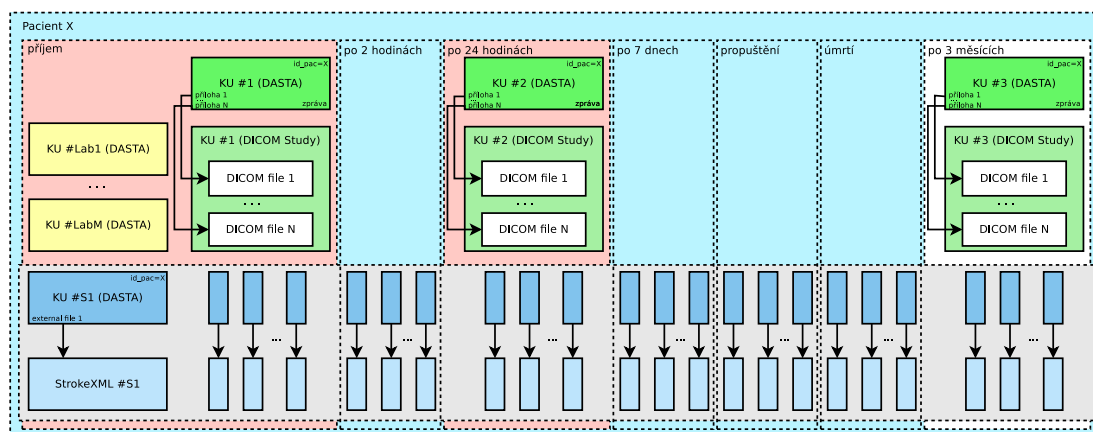
Diagnostika Bezprostředně po příjmu pacienta s vnějšími příznaky nebo podezřením na CMP je provedeno urgentní (statim) CT vyšetření a všeobecná sestra provede hodnocení stavu pacienta dle *National Institutes of Health Stroke Scale*²⁵ (NIHSS), standardizovaného neurologického vyšetření²⁶ pro popis deficitu u pacientů s iktem. Typ CMP je odvozován z vnějších příznaků pacienta, nálezu klinické neurologické symptomatologie (NIHSS) i zobrazení s pomocí zobrazovacích diagnostických metod (např. CT, MR, sonografie karotid).

Léčba Léčba CMP je vždy individuální a založena na příčině, typu CMP a dalších faktorech. Mezi další faktory patří především věk a zdravotní stav. Po provedení léčby následuje sledování stavu pacienta v délce až tří měsíců (dispenzarizace).

Data Významné okamžiky jsou příjem a následně stav pacienta po dvou a čtyřiaadvaceti hodinách, sedmi dnech a třech měsících (výstupní kontrola). Také se zaznamenává stav v případě propuštění nebo úmrtí. Nejdůležitější data vznikají bezprostředně při příjmu. Zjišťují se rizikové faktory, ale i diagnózy nebo provedená léčba. Zaznamenává se použitý způsob léčby včetně dávkování. Hodnocení NIHSS je prováděno ve všech klíčových bodech léčby stejně jako krevní tlak a teplota. Kontrolní CT vyšetření se provádí po čtyřiaadvaceti hodinách a vyhodnocuje se úspěšnost léčby. Po třech měsících může být volitelně provedeno také obrazové vyšetření. Vizualizace podoby dat je uvedena na obrázku 5.1. Jedná se o dvě až tři klinické události (KU 1 až 3) – lékařské zprávy ve formě nestrukturovaného textu (DASTA), které odpovídají souvisejícím obrazovým vyšetřením (CT, výjimečně MR) při příjmu, po 24 hodinách a v některých případech také po třech měsících (DICOM). Příklad (bez obrazových) dat je v adresáři `data/data-imaging/` viz příloha I. Klinická událost pro laboratorní výsledky může být jedna, ale může jich být i více (KU Lab1–LabM) viz příklad v příloze I a adresáři `data/data-laboratory/`

²⁵(https://www.stroke.nih.gov/documents/NIH_Stroke_Scale_508C.pdf)

²⁶Jde o testování, zda pacient může mluvit, porozumí otázkám, orientuje se v okolí, má motorické schopnosti, apod. NIHSS skóre udává úroveň postižení. Čím blíže je hodnota celkového skóre k 0, tím je u pacienta menší postižení.

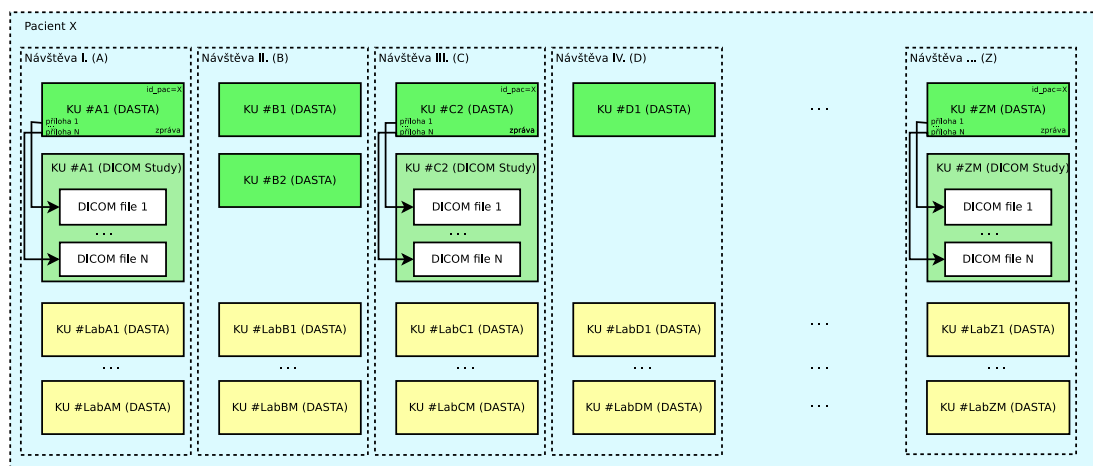


Obrázek 5.1: Data a jejich vztahy pro léčbu a sledování pacienta po cévní mozkové příhodě.

Registr dat Data související s léčbou CMP jsou v anonymní strukturované podobě zadávána do mezinárodních registrů. V minulosti, do roku 2019, to byl registr SITS. Od počátku roku 2019 je využíván RES-Q. Oba registry (blíže popsané v kapitole 5.5.1) mají pouze formulářové webové uživatelské rozhraní pro zadávání dat. Potřebné údaje byly lékaři opisovány nebo kopírovány z elektronické zdravotní dokumentace (klinických událostí) nebo poznámek v listinné podobě. Od poloviny roku 2015 jsou strukturovaná data vytvářena s použitím řízených číselníků vycházejících ze zvyklostí v oboru přímo v informačním systému zdravotnického zařízení. Data z elektronické zdravotní dokumentace v nemocničním informačním systému (NIS) jsou částečně předvyplňována automaticky. Data je možné sdílet automatizovaně ve formátu *StrokeXML* (XML s DTD) S1, který je při exportu přílohou klinické události v DASTA označené KU S1, jak je uvedeno na obrázku 5.1. Sdílení je možné automaticky po změně hodnoty, jak naznačuje diagram na obrázku 5.1 ve spodní části. Soubor formátu StrokeXML neobsahuje přímé identifikátory. Datum a čas události může být považován za nepřímý identifikátor. Příklady souborů DASTA a StrokeXML jsou v adresáři `data/data-stroke/` viz příloha I.

5.4 Idiopatické střevní záněty

Mezi zánětlivá střevní onemocnění označovaná IBD (Inflammatory Bowel Disease) patří Crohnova choroba (CD) a Ulcerózní kolitida (UC). Jsou to onemocnění trávicího traktu, která bývají provázána množstvím příznaků také mimo střeva. *Crohnova choroba* je chronický nespecifický zánět, který může postihnout prakticky jakoukoliv část trávicího traktu, kde je zánět v jednom nebo více segmentech. *Ulcerózní kolitida* je nespecifický zánět postihující nejčastěji konečník a přilehlé části. *Neurčitá kolitida* má rysy obou uvedených, které brání jednoznačnému určení, zda se jedná o CD nebo UC. Oba typy onemocnění začínají v dětství, příp. během dospívání. Příčina není jasná, ale incidence a prevalence IBD se v posledním období zvyšují. Dispenzarizaci vyžaduje chronická nemoc po celou dobu života, proto není časově omezena a dochází ke kontinuálnímu sledování pacientů. [SSC⁺17]



Obrázek 5.2: Data a jejich vztahy pro diagnostiku idiopatických střevních zánětů

Diagnostika Diagnostika se provádí na základě laboratorních výsledků (FW, CRP, hemoglobin, trombocyty), kolonoskopie s případnou biopsií, MRI enterografie, enteroskopie a další zobrazovacích diagnostických metod (SONO, RTG).

Data Podobu dat ilustruje obrázek 5.2. V případě laboratorních výsledků získaných z biologického materiálu (např. biopsie), jsou data ve formátu DASTA (KU LabA1 až LabAM). Příklad DASTA s laboratorními výsledky je v adresáři data/data-laboratory/ viz příloha I. Pro zobrazovací metody je KU také ve formátu DASTA (A1) k níž jsou připojeny snímky v DICOM souborech. Četnost návštěv pacienta je závislá na jeho momentálním zdravotním stavu – tomu odpovídá množství a typy dat. Vyšetření prostřednictvím zobrazovacích metod nemusí být provedeno vždy, jak ukazují II. a IV. návštěva pacienta. Další data mají být sbírána přímo ve strukturované podobě k jejich použití pro výzkumné účely.

5.5 Možnosti realizace výzkumu v medicíně

Pro výzkumné účely, dlouhodobou podporu nebo udržitelnost výzkumu se objevují a vznikají specifické informační systémy, které jsou velmi často úzce zaměřeny na konkrétní doménu. Příkladem jsou systémy uvedené v kapitole 5.5.1. Častým případem doménově orientovaných systémů jsou registry, jejichž hlavním cílem je sběr dat pro následné analýzy a vyhodnocování. Obecným řešením, které mají nebo mohou mít některé z uvedených rysů, nebo jsou v některém z nich univerzální, je věnována kapitola 5.5.2. Doménově orientované systémy se pro své úzké zaměření zabývají řešenou problematikou detailněji. Projekty pro CMP (5.3) a IBD (5.4) jsou vhodnými kandidáty pro vytvoření registru. Právě pro sledování stavu pacientů po CMP existují registry SITS a RES-Q (5.5.1). Problematické je přizpůsobení silně doménově orientovaných registrů pro jinou oblast.

5.5.1 Specializované systémy a registry

CARMEN Virtual Laboratory

*CARMEN*²⁷ *Virtual Laboratory* byl projekt zaměřený na neurofyziologii. Platforma umožňující nejen sdílení dat, ale přímo analýzu a vlastní zpracování dat zaregistrovanými autory. Po registraci bylo možné využívat zveřejněná data nebo služby od ostatních uživatelů. Zpracování probíhalo výhradně na hardware portálu. Uživatel mohl nahrát vlastní data a k nim povolit/zakázat přístup ostatním. Podobně při využití sdílených služeb pro zpracování dat. [231–233] Portál byl odstaven a od roku 2019 je mimo provoz.

Registry of Stroke Care Quality

*REgistry of Stroke Care Quality*²⁸ (RES-Q) je registr pro sběr dat a pomoc se sledováním léčby pacientů po CMP. Cílem je zkvalitňování léčby zahrnující odstranění rozdílů v léčbě napříč nemocnicemi a zeměmi. RES-Q byl vytvořen jako prospektivní mezinárodní registr v roce 2016 pod záštitou *European Stroke Organisation*²⁹ (ESO), resp. část *Enhancing and Accelerating Stroke Treatment* (ESO-EAST). [234, 235] Model komplexních cerebrovaskulárních center byl Ministerstvem zdravotnictví České republiky založen v letech 2010–2012 ve spolupráci s Cerebrovaskulární sekci České neurologické společnosti (CSCNS). Do RES-Q jsou předávána data z cerebrovaskulárních center prostřednictvím formulářů ve webovém prohlížeči. [236] Základem jsou metriky kvality dle Norrving et. al (2015) [237], kde je sledováno 24 hodnot, z nichž je většina odpovědí pouze ano/ne.

Safe Implementation of Treatments in Stroke

*Safe Implementation of Treatments in Stroke*³⁰ (SITS) je mezinárodní registr zaměřený na akutní léčbu a prevenci CMP. [238, 239] SITS zastřešuje Karolinska Institute ve Švédsku jako akademickou, neziskovou a mezinárodní iniciativu pro spolupráci zdravotnických odborníků. Registr je rozšířen v 80 zemích světa, kde je celkově na 1 600 spolupracujících klinických centrech. [240] Lékaři vyplňují data prostřednictvím interaktivního webového uživatelského rozhraní s rozsáhlými formuláři. Zadávají terapeutické záznamy, reakce léčiv, hodnocení stavu pacienta podle *National Institutes of Health Stroke Scale* (NIHSS, klasifikační škála pro mrtvice), popis nálezu CT/MR snímků, příčin úmrtí a mnoho dalších informací. Lékař má přehled pouze o jím vytvořených záznamech a je vlastníkem dat. Aktuální (2019) verze SITS poskytuje grafické přehledy založené na analýze počtu typů mrtvic a výsledků léčby. Lékař má možnost stažení velmi rozsáhlého CSV souboru se všemi svými daty. Pro zpracování dat je nutné jejich předzpracování a vyčištění, protože údaje nejsou konzistentní z důvodu v minulosti prováděných změn datového modelu nebo uživatelského rozhraní.

²⁷Code Analysis, Repository and Modelling for E-Neuroscience

²⁸<https://qualityregistry.eu/>

²⁹<https://eso-stroke.org/>

³⁰<https://www.sitsinternational.org/>

Rick Hansen Spinal Cord Injury Registry

*Rick Hansen Spinal Cord Injury Registry*³¹ (RHSCIR) představuje kanadský národní registr pro pozorování osob s traumatickým míšním poraněním. Registr je určen ke sběru, správě a analýze významných dat pacientů s uvedeným poraněním během fází akutní péče, rehabilitace i opětovné integrace do společnosti. Cílem registru je lepší pochopení problematiky, zvýšení efektivity a lepší výsledky poskytované léčby, cvičení a programů. Spolupracují pacienti, výzkumníci, lékaři, zdravotníci, ale i poskytovatelé služeb pro snížení handicapu a zvýšení kvality života pacientů. [241] Dlouhodobý sběr dat slouží pro výzkumné aktivity, predikční modely, jejich validace a další jako uvádějí např. Thibault-Halman et al. [242] (2017) nebo Phan et al. (2019) [243].

5.5.2 Obecné systémy

Pro podporu medicínského výzkumu lze použít existující obecné systémy pro správu obsahu, dokumentů, ale i archivační (informační) systémy. Možnosti pro správu dat a dokumentů poskytují i podnikové IS. Množství dalších nástrojů a software může být v průběhu výzkumu využíváno (nástroje pro návrh/vývoj software, organizaci myšlenkových map, systémy pro správu verzí, správa požadavků nebo chyb, lokalizace, dokumentace, návody).

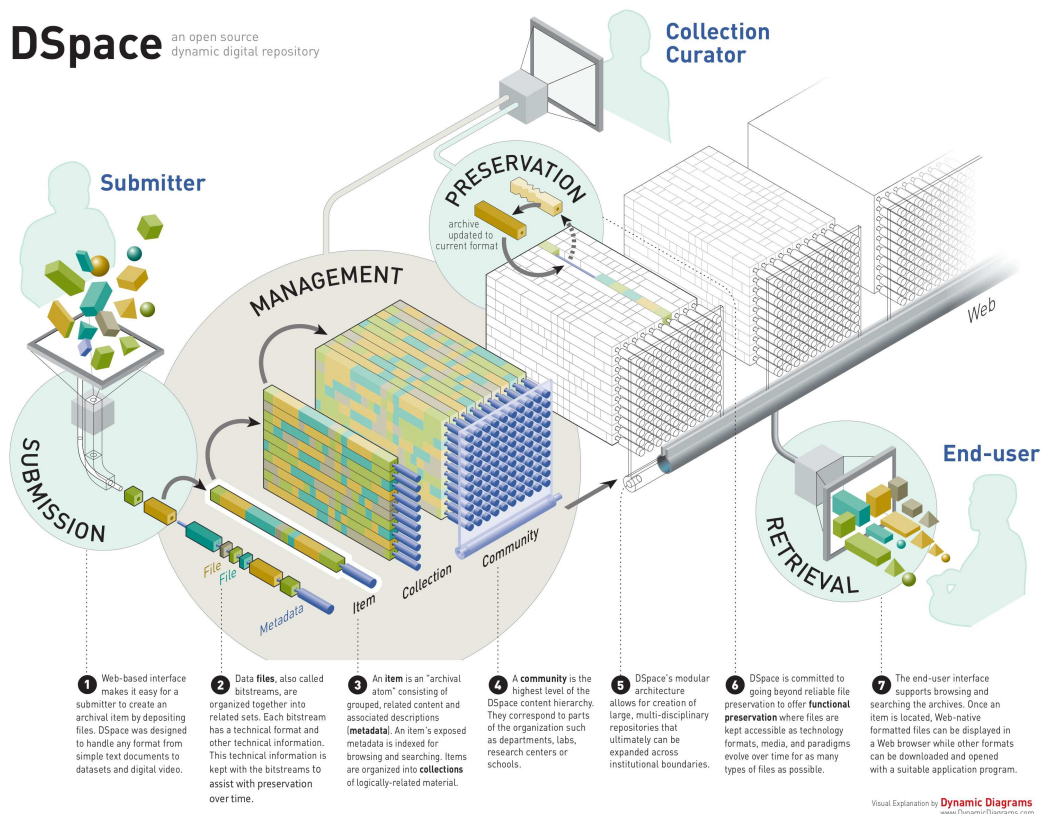
DSpace

DSpace je open source řešení pro tvorbu otevřených digitálních repositářů od roku 2002. DSpace byl publikován v [244] v kontextu požadavků OAIS (viz kapitola 3.10.2). Umožňuje pracovat s množstvím typů digitálního obsahu (text, obrázky, video, datasey). Nejčastěji je DSpace používán jako virtuální knihovna v akademickém a výzkumném prostředí s otevřeným přístupem, kde poskytuje správu vědeckých výstupů fakult a jejich studentů. Lze jej používat jako úložiště pro digitální objekty institucí, obrazové, multimédia, reporty, výukové materiály, ale i pro muzejní a kulturní dědictví v řadě institucí. [244, 245] Diagram na obrázku 5.3 je převzat od Patel (2012) [246] a ilustruje funkce DSpace zahrnující příjem, digitální ochranu, uchování, správu dat a zpřístupnění koncovému uživateli. Ve své publikaci [247] se Cortese et al. (2017) zabývali možnostmi rozšíření. O koncept propojených dat ve vztahu na DSpace se zabýval např. Becker (2016) [248], ale uvádí je i dokumentace DSpace 6.x od září 2016.

Systém správy obsahu

Výhodou systémů pro správu obsahu (Content Management System, CMS) je doménová nezávislost, protože umožní manipulaci s libovolným obsahem nebo jakýmkoliv dokumenty. Podobně systémy pro *správu podnikového obsahu* (Enterprise Content Management, ECM) jsou určeny pro manipulaci s velkým množstvím heterogenních dokumentů v prostředí podniku. Dokumentem může být textový nebo tabulkový dokument, prezentace, multimédia, ale i e-maily, kalendářová data nebo jiná strukturovaná a nestrukturovaná data. ECM má za cíl udržet přehled o dokumentech, zajistit bezpečné uložení a snadné efektivní sdílení s kolegy i mezi

³¹<https://rickhansenregistry.org>



Obrázek 5.3: Diagram DSpace ilustrující funkce příjmu (submission), digitální ochrany a uchování (preservation), správy dat (management), a zpřístupnění (retrieval) koncovým uživatelem. Zdroj diagramu dle Patel (2012) [246].

odděleními podniku. [249] Princip ECM je blízký potřebám podpory výzkumu, kde spolupracuje více uživatelů s velkým množstvím dat (dokumentů, postupů, výsledků) a jejich verzí. Nevýhody CMS a ECM mohou spočívat v omezené možnosti popisu dokumentů, obtížné vytvoření vazeb mezi dokumenty, nebo při častějším provádění změn. Pokud lze vlastní metadata doplnit (např. rozšířením datového modelu), může být problém s jejich dlouhodobou udržitelností³².

Služby pro interoperabilitu správy obsahu

Standard *Content Management Interoperability Services* (CMIS) má mezi různě implementovanými systémy zajišťovat interoperabilitu s pomocí webových služeb. Aktuální specifikace CMIS je verze 1.1 [250] (2015) a byla vydána pod záštitou *Organization for the Advancement of Structured Information Standards* (OASIS). Výhodou CMIS je oddělení služeb a obsahu, protože definuje abstraktní vrstvu pro řízení různých systémů správy obsahu a úložišť pomocí webových protokolů – doménový model a webové služby, které lze využívat v aplikacích. Dokumenty označuje pojmem *objekt*, pro které poskytuje služby pro ukládání a jejich opětovné získávání, příp. definování vztahů mezi nimi. Definuje použití široce používaných protokolů WSDL/SOAP a Representational State Transfer (REST). [250] Omezení uvádí [250] v podobě nepodporování všech konceptů plně funkč-

³²Mohou být komplikace, anebo může dojít ke ztrátě dat v případě bezpečnostní aktualizace nebo přechodu na vyšší verzi.

ního ECM. CMIS implementovali např. Marin and Brown (2015) [251], věnuje se mu Bergljung (2014) v knize [252]. Existuje framework Chemistry³³ jako součást Apache. [253] Automatizovanému propojení služeb pro digitální uchování se věnují Westerlund et al. (2019) [254].

Global Research Platform

*Global Research Platform*³⁴ (GRP) je platforma pro sběr strukturovaných dat prostřednictvím webových formulářů. Lékař si může vytvořit vlastní formulář na míru a bez potřeby větších technických znalostí. Podporováno je následné uchování dat, ale i přímo možnost výzkumu formou SaaS. Za vznikem platformy v roce 2008 byl *Rick Hansen Institute*³⁵ pod záštitou *Rick Hansen Foundation*³⁶. Mezi podporované typy patří obecně registry, klinické pokusy (clinical trials), klinický výzkum (clinical research), studie proveditelnosti (feasibility study) a zlepšení kvality (quality improvement initiatives). Spolupracovat může více center současně. Autoři mezi hlavní výhody uvádějí flexibilitu, použitelnost, škálovatelnost a vysokou míru zabezpečení a řízení přístupu ke shromažďovaným datům. Zadávaná data mohou být validována už na vstupu pro dosažení jejich vysoké kvality. Skladování dat je řízeno centrálně, výhradně v zašifrované databázi a dedikovaném serveru v Kanadě. Komunikace je možná pouze webovým prohlížečem za použití SSL. Všechna data vlastní uživatel a ostatním je může zpřístupnit přidělením různých úrovní přístupu. Export dat je podporován ve formátech jako Microsoft Access nebo Microsoft Excel, SPSS, SAS, EpInfo a Crystal Reports. [241, 255] Nabízí se podobnost a porovnání s Google Forms³⁷. Obojí nabízí pouze možnost sběru dat a prakticky jen v rozsahu základních datových typů. [255] Na použití GRP odkazují např. Jones et al. (2019) [256] a Noonan et al. (2018) [257]. GRP uvádí (2019) existenci probíhajících 13 studií/registrů zahrnující i RHSCIR (viz kapitola 5.5.1).

³³<https://chemistry.apache.org/>

³⁴<https://www.rhigrp.net/>

³⁵<https://rickhanseninstitute.org/>

³⁶<https://www.rickhansen.com/>

³⁷<https://support.google.com/docs/answer/6281888>

6 Použité metody

Pro zajištění stanovených cílů je nezbytné řešit (1) *schéma* stejně jako (2) *procesy* manipulace s heterogenními daty a metadaty při datově orientovaném výzkumu. Kvalitativní výzkum slouží pro objevení, hodnocení nebo verifikaci teorií na základě systematického popisu konceptů v řešené doméně, jejich vztahů, kontextu včetně okolí nebo účastníků. Předpokladem je iterativní postup založený na sběru dat a jejich analýze vedoucí k identifikaci nezbytných konceptů a vztahů pro modelování řešené oblasti. Předpokládaný rozsah dat a rozmanitost projektů ovlivňuje způsob a techniky týkající se požadavků na návrh, vývoj, koordinaci procesů nebo výměnu informací ve shodě s [24–26, 28, 37]. Nutnou podmínkou je zajištění interoperability dat a metadat [26–28, 30].

Použity jsou současně metody *empirické* a *teoretické*. Z empirických metod se jedná o *dotazování* a *pozorování*. Konzultace s lékaři, jejich pozorování při diagnostice a léčbě jsou v doméně medicínského výzkumu základem pro použití zdravotnických dat. Při spolupráci s lékaři je získán detailní popis řešené domény, používané postupy, škály nebo metriky pro hodnocení stavu pacientů. Vyhovují metody deskriptivního přístupu, sběru dat a požadavků nebo analýza existujících dokumentů. Pro analýzy dat a účely prospektivní databáze lze uvažovat o metodách *měření* a *experimentování*. Mezi měření mohou být obecně zahrnuty nové metody a metriky, které mohou být v průběhu medicínského výzkumu ověřeny a potvrzeny. Laboratorní hodnoty jsou zahrnuty ve zdravotnických datech. Při medicínském výzkumu mohou být laboratorně stanovené hodnoty nebo další metriky hodnocení pacientů základem pro experimentování a testování hypotéz.

Výzkumná platforma Standardizace schématu metadat a procesů potřebných pro realizaci datově orientovaného výzkumu je řešena jako komplexní platforma s označením **Medical Research & Education** (MRE), která představuje podstatu rozsáhlého výzkumného informačního systému. Inspirací byly referenční modely *Open Archival Information System* (OAIS; ISO 14721:2012, kapitola 3.10.2) [116, 117] a *Producer-Archive Interface Methodology Abstract Standard* [118, 119] (3.10.3) s rolí producenta a archivu dat.

Standardizace procesů Mezi cíli práce je *standardizace procesů* vztahující se k platformě rozsáhlého informačního systému. Vhodné jsou metody spočívající v *systémovém přístupu* s důrazem na široké možnosti přizpůsobení dle možných budoucích požadavků. Platforma plní roli informačního systému s funkcemi podpory činnosti [4] výzkumu, jeho řízení a koordinace nebo kontrolu [2, 4] a rozhodování [2, 4]. Uplatňují se metodologie a techniky [18, 19] související s projektováním IS a v souladu s [258] je platforma považována za systém, který lze jako celek členit na jednotlivé subsystémy. Celý systém i každý subsystém musí mít jasně vymezený účel, strukturu a vlastnosti včetně vazeb na bezprostřední okolí. Informační systém zajišťuje interakci mezi lidmi [3, 5, 8, 10], procesy [7, 8, 10], technologií [3, 7–10] a případně daty. Procesy pracují s daty [3] nebo informacemi [1, 2, 7], čímž dochází k jejich manipulaci [7] – získávání/shromažďování a ukládání [1, 2, 7, 9] nebo zobrazování [7], zpracování [1, 2, 7] nebo distribuci/sdílení [1, 2, 7, 9].

Schéma metadat Při *standardizaci schématu* je provedeno *modelování* pro nalezení objektů popisujících řešenou doménu. Cílem modelování je vytvoření ideálního vzoru s použitím metody *abstrakce* zahrnující analytické a syntetické postupy, které vymezují zásadní vlastnosti a vztahy – konceptualizace. Identifikovat lze významné prvky a procesy reality, které s nimi souvisí. Abstraktní modelování poskytuje objekty datových, formálních, teoretických nebo znalostních objektů dle [258]. Pro dosažení cílů disertační práce to jsou zejména datové a znalostní typy objektů abstraktního modelování. Pro komplexní objekty je vhodné použít metodu *analýzy* spočívající v rozkladu na menší části. Opakem analýzy je *syntéza*, kde dochází ke spojování částí, vlastností nebo vztahů vytvořených analýzou. Často je potřeba *generalizace*, kdy se provede vyloučení společných znaků pro nalezení obecného pojmu – ze zdrojových dat, pojmů získaných sběrem dat, požadavků, na základě dotazování nebo konzultace s lékaři. Provedení *komparace* je předpokladem generalizace a analogie, která určuje podobnost nebo pravděpodobnost shody pouze některých vlastností. V prvních fázích návrhu je nalezeno množství *anomálií*, které jsou s lékaři vyjasněny. Metoda *klasifikace* je vhodná pro používané pojmy a jejich třídění nebo řazení s použitím abecedního pořadí nebo přiřazeného skóre dle zavedených zvyklostí (např. slovníky). Další metody jako je agregace, indukce nebo dedukce mají svůj význam, ale budou vhodné v navazujících výzkumných projektech souvisejících s analýzou dat, které nesouvisí přímo s tvorbou schématu pro metadata.

Tvorba ontologií a slovníků Tvorba ontologie má fáze: (1) studie proveditelnosti, (2) analýza domény, (3) konceptualizace, (4) implementace a (5) údržba. Prvním krokem je *studie proveditelnosti* s uvedením potenciálních problémů, příležitostí a existujících řešení. Měly by být uvedeny ekonomické důsledky vývoje nové ontologie a rizika pro software založený na ontologii. Nezbytností je *analýza domény*, která má zodpovědět, co je motivací, s jakou úrovní detailu se bude pracovat a identifikovat existující vhodná nebo podobná řešení. Na základě analýzy lze realizovat *konceptualizaci*, kde je vhodné provést integraci s existujícími řešeními, které byly zjištěny ve studii proveditelnosti. Výsledkem bude ontologický model. Poslední fáze tvorby ontologie je *implementace* v ontologickém jazyce OWL. *Údržba* spočívá v průběžném ověření souladu a přizpůsobování budoucím požadavkům nebo doplnění mapování na jiné ontologie. Pravidla tvorby ontologií a slovníků jsou následující.

1. Vybrat vhodný jmenný prostor a prefix.
2. Používat URI/IRI pro všechny zdroje (třídy, vlastnosti i instance).
3. Názvy tříd a vlastností v URI/IRI jsou preferovány v angličtině. Textové popisy a anotace doplnit dle preferovaného jazyka. Doporučena je angličtina a čeština.
4. Popis ontologie (`owl:Ontology`) v rozsahu: název `dc:title`; textový popis `dc:description`; vydavatel `dc:publisher`; přispěvatel `dc:contributor`; datum vydání `dc:issued`; poslední změny `dc:modified`; práva `dc:rights`; verze `owl:versionInfo`, `owl:versionIRI`; a komentář `rdfs:comment`.

5. Popis třídy (class): povinně název `rdfs:label`; volitelně popis a komentář (`dc:description`, `rdfs:comment`).
6. Popis vlastnosti (property): neopomenout třídu subjektu `rdfs:domain` a objektu `rdfs:range`, je-li to vhodné. Povinně název `rdfs:label` a volitelně popis `dc:description`. Pro datové vlastnosti stanovit obor hodnot `rdfs:range` a popsat detaily jako jsou jednotka, přesnost nebo způsob kódování hodnoty (např. desetinná tečka vs. čárka, oddělovač tisíců, datum a čas) prostřednictvím komentáře `rdfs:comment`. Preferovat mezinárodní přenositelnost a hodnoty uvádět ve standardizovaných jednotkách (SI), případně dle jednotek a přesnosti běžných v daném oboru. Vhodné je doplnit odkaz na jiný online zdroj s informacemi o vlastnosti.
7. Pro hodnoty je důležitá konzistence a integrita. Dodržovat obor hodnot a správnou reprezentaci dat.
8. Pro slovníky preferovat třídu existující nebo definovat vlastní. Instancím (individua) vždy doplnit srozumitelný název prostřednictvím `rdfs:label` (alternativně `dc:title` nebo `skos:prefLabel`). Doporučeno je přidat anotace dle *Form Ontology* ke třídám, vlastnostem a instancím slovníků.
9. Dokumentovat všechny části ontologie.
 - (a) Atributy `rdfs:comment` nebo `rdfs:seeAlso` jsou doporučeny pro lepší pochopení a doplnění kontextu.
 - (b) Doporučeny jsou vícejazyčné popisy. Předpokladem je použití v uživatelském rozhraní nebo jako dokumentace.
 - (c) Doporučeno je používat výchozí texty (bez specifikovaného `xml:lang`) pouze v angličtině. Pro ostatní jazykové varianty doplnit informaci o jazyce `xml:lang`.
 - (d) Anotovat zastaralé části ontologického popisu, např. prostřednictvím anotace `vs:term_status` s hodnotou `archaic` jako obdoba hodnoty `deprecated`.

Specifika multidisciplinární spolupráce V případě multidisciplinárního výzkumu je nutné komunikovat a konzultovat návrh ontologického modelu s množstvím dalších členů týmu. Existuje řada nástrojů pro tvorbu ontologií, ale množství z nich může být komplikací pro běžné uživatele, kteří nemohou volně instalovat software nebo neznají ontologický jazyk. Při spolupráci na návrhu ontologie bylo pro fázi konceptualizace použito tabulkového procesoru. Listy sešitu poskytovaly (a) anotace ontologie, (b) třídy, (c) vlastnosti a (d) individua vztahující se nejčastěji k číselníkům. V příloze B je na obrázcích B.1 a B.2 ukázka dvou tabulek realizovaných v LibreOffice. Prostřednictvím [259] bylo možné provést automatickou konverzi na OWL.

Technologie Pro metadata jsou použity *technologie sémantického webu* (viz kapitola 4) [150, 180], jejichž základem jsou Resource Description Framework (RDF, 4.1) [145, 147, 148], Web Ontology Language (OWL, 4.5) [169] a dotazovací jazyk SPARQL (4.4) [160]. Slovníky RDF/RDF Schema [152] a OWL

[169, 174] poskytují formální možnosti popisu. Jsou to především obecné definice pro třídy/typy a vlastnosti. OWL je pokročilejším jazykem a rozšiřuje možnosti popisu například o omezení. Při realizaci vlastního schématu bylo adaptováno nebo adoptováno několik existujících slovníků a ontologií. Jejich jmenné prostory jsou uvedeny v příloze A ve výpisu A.1. Technologie sémantického webu poskytují flexibilní obecný rámec, který je určen k definici, publikaci a sdílení ontologií. Integrace různých datových sad je provedena na principu propojených dat (4.7) [181, 186]. Pro zpřístupnění a manipulaci s metadaty slouží ODBC/JDBC nebo SPARQL Endpointy (4.4) [167, 168].

Vytvořený model a schéma pro metadata mohou mít pozitivní i negativní dopady, které vychází z možných rizik nebo v souvislosti s úrovní interoperability (3.3) [66, 68], granularitou (3.4), kvalitou (3.5) nebo dalšími aspekty v souladu s [54]. Požadavky na změnu jsou očekávány a předpokládány, protože dlouhodobý sběr dat s navazujícím výzkumem bude nepochybně v čase rozšiřován, zpřesňován, ale mohou být modifikovány i stanovené koncepty, vlastnosti a jejich vazby nebo okolí. Podpora RDF, OWL a SPARQL je v množství programovacích jazyků a RDF úložišť. Sdílet je možné data společně se sémantikou schématu. Integraci řeší doplnění mapování mezi schématy stejně jako data. Schéma i data mohou být vícejazyčná. V souladu s [21, 22] je vhodné využít webové služby pro navazující řešení výzkumných projektů a integraci nad výzkumnou platformou. SPARQL Endpointy umožňují využívat RDF data z libovolné aplikace prostřednictvím protokolu HTTP(S).

7 Ontologie

Schéma metadat na platformě MRE je tvořeno celkem třinácti různě rozsáhlými ontologiemi, jak ukazuje diagram na obrázku 7.1. Deset z nich je aktivně používáno. *MRE Ontology* je základní ontologií výzkumné platformy viz kapitola 7.2. Pro klinická a terapeutická zdravotnická data jsou používány *DASTA Ontology* a pro číselníky *DASTA Code-lists* (7.3). Vyšetření zobrazovacích diagnostických metod popisuje adoptovaná *DICOM Ontology* (7.4). *Form Ontology* (7.5) definuje anotace pro metadata potřebná k realizaci zobrazení a umožnění editace RDF dat prostřednictvím HTML formulářů. *Population Ontology* (7.6) je určena pro údaje o populaci vždy s referenčním rokem na daném území dle pohlaví a věku. Ostatní ontologie jsou úzce orientovány na výzkum v dané oblasti. Pro diagnostiku a sledování průběhu léčby a stavu pacientů po cévní mozkové příhodě slouží *Stroke Ontology* (7.7), která nahradila původně používanou dvojici *SITS Ontology* a *NIHSS Ontology*. Pro idiopatické střevní záněty (IBD) je používána *IBDS Ontology* (7.8.1). *IBDT Ontology* (7.8.2) poskytuje vyšší míru granularity metadat a množství dalších konceptů a řízených slovníků pro klasifikaci a hodnocení pacientů s IBD. *TDB Ontology* (7.9) je zaměřena na evidenci vzorků tkáně a jejich zařazení v klinických studiích. Pro metadata obrázků ve formátech JPG, PNG nebo TIFF se používá *Image Ontology* dle [260] (viz 8.4). Dokumentace ontologií¹ je na adrese <https://mre.zcu.cz/ontology/<nazev>>, kde je část <nazev> skutečné označení ontologie. Používané jmenné prostory a prefixy jsou uvedeny ve výpisu A.2 v příloze A. Metriky ontologií jsou uvedeny v příloze F. Zjednodušené modely jsou v textu následujících podkapitol pro každou ontologii. Na příloženém DVD jsou uvedeny kompletní modely v adresáři `mre-ontology-visual/` viz Příloha I.

7.1 Jmenné prostory

Jmenný prostor využívá protokol HTTPS a doménu `mre.zcu.cz` dle výpisu 7.1. První řádek představuje doporučené URL ontologie, kde část <nazev> je její označení, resp. název souboru. Druhý řádek je pro popis verze ontologie ve vlastnosti `owl:versionIRI`, ale volitelně může mít i význam URL pro získání ontologie. Třetí řádek výpisu zahrnuje označení identity <identita> do jmenného prostoru, které může být vhodné pro specifické potřeby uživatelů nebo projektů. Používané jmenné prostory v rámci MRE zobrazuje příloha A. Pro identifikaci RDF zdrojů je používán zápis dle výpisu 7.2, kde je část <id> tvořena libovolným unikátním identifikátorem platným pro URI/IRI, včetně Cool URI [151].

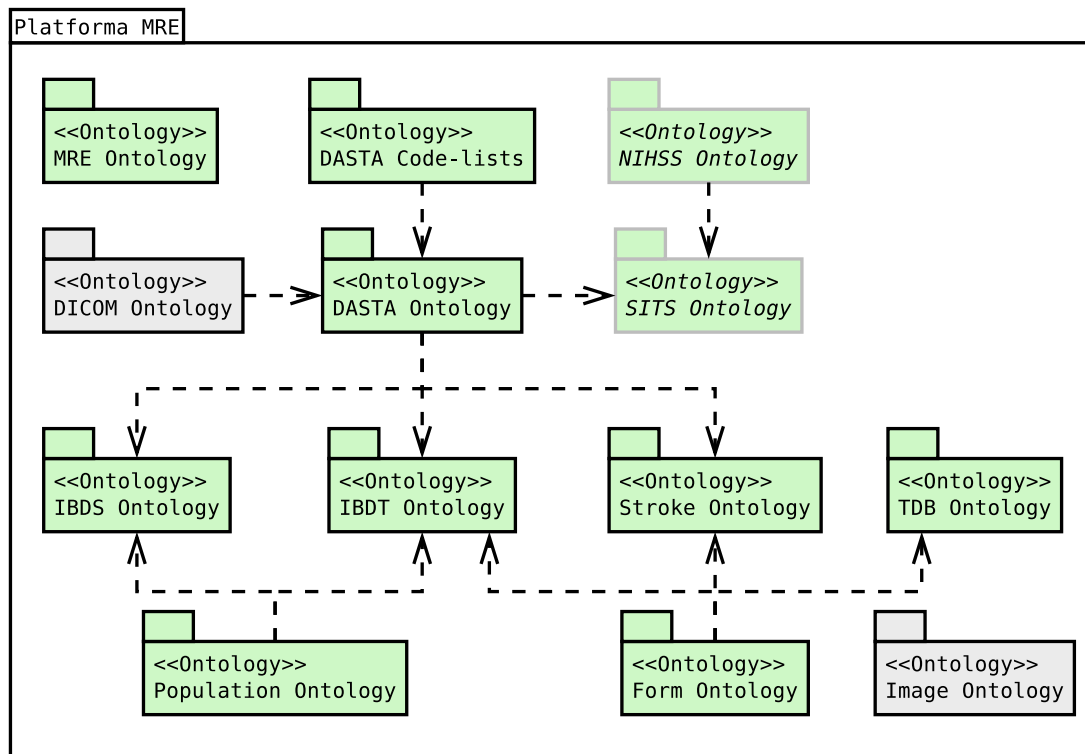
Výpis 7.1: Schéma jmenných prostorů ontologií v MRE

```
https://mre.zcu.cz/ontology/<nazev>.owl
https://mre.zcu.cz/ontology/<year>/<version>/<nazev>.owl
https://mre.zcu.cz/ontology/<identita>/<year>/<version>/<nazev>.owl
```

Výpis 7.2: Schéma URI/IRI pro RDF zdroje v MRE

```
https://mre.zcu.cz/id/<id>
```

¹Rozcestník <https://mre.zcu.cz/ontology/ontologies.html>



Obrázek 7.1: Diagram ontologií platformy MRE.

7.2 MRE Ontology

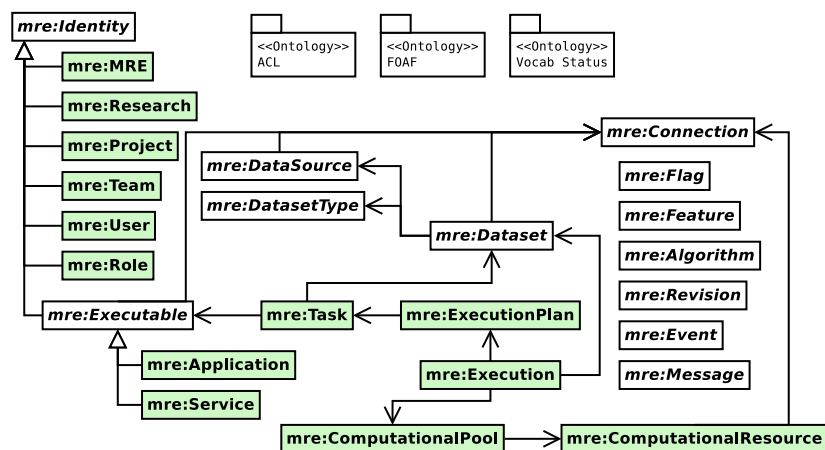
MRE Ontology má (a) základní popisnou a (b) výzkumnou část. Popisná je zaměřena na metadata obecná, o původcích, seskupování do datových sad a řízení přístupu. Výzkumná část poskytuje elementární koncepty, jejich vlastnosti, vztahy a anotace vhodné pro popis softwarových metod zpracování dat, typu algoritmů (dlouho běžící, intenzivní na CPU nebo I/O operace).

Studie proveditelnosti a analýza domény V základní popisné části ontologie je kladen důraz na větší míru abstrakce. Rozšíření jsou očekávána až v závislosti na budoucích požadavcích na výzkumnou platformu. Účelem ontologie je popsat koncepty spojené s popisem identit a datových sad. Zahrnuty jsou popisy původce dat a alespoň základní řízení oprávnění, revize, události a zprávy. Existují běžně používané slovníky *The Friend of a Friend*² (FOAF, prefix `foaf`) a *Basic Access Control Ontology*³ (ACL, prefix `acl`).

Výzkumná část popisuje životní cyklus zpracování dat prostřednictvím softwarových prostředků, který vede ke vzniku výsledků. Je to popis pro zajištění dlouhodobé udržitelnosti a použitelnosti dat. Ekonomický důsledek je v opakovaném využití předchozích výsledků nebo zopakování úloh se stejnými parametry také s novými daty. Při důsledném popisu může strukturovaný popis vést až k interoperabilní dokumentaci procesů zpracování dat. Existence popisu může pomoci minimalizovat rizika spojená s fluktuací výzkumníků. Úkolem výzkumné části je poskytnout popis dostupných metod zpracování dat, výpočetního prostředí,

²<http://xmlns.com/foaf/spec/index.rdf>

³<http://www.w3.org/wiki/WebAccessControl>



Obrázek 7.2: MRE Ontology

plánů a parametrů pro spuštění. Dekompozicí problému vzniknou opakovaně použitelné komponenty jako dílčí softwarové prostředky, které mohou být zřetězeny pro dosažení požadovaného výsledku. Vznikne popis činností (workflow) a toku dat (dataflow). Každý výzkumník může využít existující komponenty nebo přidá vlastní. Postupně budou rozšiřovány možnosti platformy k dosažení dlouhodobých záměrů a cílů určitého výzkumu. Úlohy mohou být spuštěny identitou (1) manuálně, (2) jednorázově ve stanovený čas, (3) opakovaně na základě plánu, (4) vždy jako reakce na událost nebo (5) automaticky při startu a běžet trvale (služba).

Konceptualizace a implementace *MRE Ontology*⁴ s prefixem *mre* a jmenovým prostorem viz příloha A. Konceptualizace základní části ontologie se věnuje oblastem: (1) správy identit, (2) řízení přístupu, (3) původci dat a datové zdroje, (4) popis souboru dat, (5) verze a (6) správa událostí a zpráv. Výzkumná část ontologie se zabývá oblastmi: (1) popisu software, (2) výpočetních prostředků, (3) plánu spuštění, (4) úloh a (5) propojení s výsledky zpracování. Zjednodušený model uvádí diagram na obrázku 7.2.

Správa identit Slovník FOAF poskytuje abstraktní popis založený na třídě *foaf:Agent* s potomky *foaf:Group*, *foaf:Organization* a *foaf:Person*. Právě *foaf:Agent* lze využít pro popis *identit* dle konceptu *mre:Identity*, jako jsou identita platformy *mre:MRE* nebo softwarové prostředky ve formě aplikace nebo služby (*mre:Application*, *mre:Service*). Konkrétní uživatel *mre:User* je ekvivalentem *foaf:Person*. Dále jsou identity vycházející z *foaf:Group* pro výzkumné projekty *mre:Research* a týmy *mre:Team*. Typ *foaf:Organization* lze použít k popisu institucí, z nichž uživatelé pocházejí.

Třída identity MRE (*mre:MRE*) je vyhrazena pro systémové úlohy na platformě. Identitu *mre:Research* má výzkumný projekt, který je složený z týmů *mre:Team* nebo uživatelů *mre:User*. Navíc je uvažován koncept *mre:Role* pro určení role uživatele v týmu nebo projektu. Koncepty identit jsou vhodné pro propojení s výsledky na různých úrovních – přiřazení k projektu, týmu, uživateli nebo software.

⁴<https://mre.zcu.cz/ontology/mre.owl>

Podpora řízení přístupu Slovník ACL definuje koncepty oprávnění pro čtení `acl:Read`, zápis `acl:Write`, přidávání `acl:Append` a možnost řízení oprávnění `acl:Control`. Koncept `acl:Read` je vhodné rozšířit o podtypy pro zjištění existence odpovědi `mre:Ask`, popisu `mre:Describe`, stažení `mre:Download`, získání výsledku dotazu `mre:Select` a zobrazení dat `mre:View`. Koncept `acl:Write` bude rozlišovat změnu `mre:Modify` nebo odstranění `mre>Delete`. Některé způsoby čtení nebo zápisu nemusí být podporovány nebo implementovány, protože to neumožní charakter dat. Pro SPARQL Endpoint má význam oprávnění `mre:Ask`, `mre:Describe` a `mre:Select`. Výběrové dotazy `mre:Select` se vztahují také pro relační databázová úložiště. Například oprávnění `mre:Download` může zahrnovat konkrétní soubor nebo celou DICOM studii.

Třída `acl:Authorization` je prvkem řízení přístupu ke zdrojům nebo třídám zdrojů, kde vlastnost `acl:agent` určuje identitu, jejímž rodičem je `acl:Agent`, a `acl:accessTo` pro informační zdroj, k němuž je přístup řízen prostřednictvím `acl:mode` s přiděleným oprávněním přístupu. Pro třídy zdrojů jsou to vlastnosti `acl:agentClass` a `acl:accessToClass` pro libovolnou třídu (`rdfs:Class`).

Popis datových zdrojů Způsob nebo technický prostředek, pomocí něhož lze přistupovat k datům, je koncept *datového zdroje* `mre:DataSource` s atributy nutnými pro přístup ke zdroji dat. V `mre:DatasetType` lze doplnit přístupovou metodu s podtypy pro označení souboru, adresáře, SPARQL dotazu, seznamy URI/IRI (třída, instance).

Popis původců a souboru dat Koncept datové sady `mre:Dataset` je abstraktní. Vztahuje se na heterogenní data mající společný jmenovatel – zdroj nebo typ dat, diagnózu, věkovou kategorii, období, apod. Obvykle je určen zdroj dat `mre:DataSource` nebo přístupová metoda `mre:DatasetType`. Doporučen je popis prostřednictvím `rdfs:label`, `rdfs:description`, `rdfs:comment` a `dc:issued`. Pro získané výsledky je `mre:Dataset` shodný jako pro zdrojová data, ale datovým zdrojem bude proces/úloha.

Správa verzí Koncept verzí/revizí slouží pro seskupování nebo pojmenování souvisejících datových sad označením verze. Zdrojová data jsou v archivu obvykle pouze jednou. Výsledky mohou mít variant více např. v závislosti na konfiguraci software. Více verzí mohou mít i softwarové prostředky. Pro rozlišení slouží koncept `mre:Revision` s kódovým označením, číslem verze, datem a časem (vzniku, změny) nebo identitou původce. Doporučeno je označit verzi prostřednictvím `rdfs:label` a volitelně anotace `rdfs:description` nebo `rdfs:comment`.

Správa událostí a zpráv Třída `mre:Event` je určena pro popis události s určením původce `mre:fromIdentity`, adresáta `mre:forIdentity`, kódového označení zprávy `mre:codename`, textové zprávy `mre:message`, datem a časem vzniku události `mre:datetime` a příznakem platnosti `mre:isValid`. Původce je vždy jediná identita vycházející z `mre:Identity`. Původcem systémových událostí je `mre:MRE`. Příjemce nemusí být specifikován žádný nebo jich může být i více současně. Součástí události může být vazba na zprávu `mre:Message`.

Další anotace a vlastnosti Anotační vlastnost `mre:id` je určena pro unikátní identifikaci instance. Doplněny jsou české anotace pro třídy a vlastnosti definované ve slovnících ACL a FOAF.

Softwarové prostředky Softwarové prostředky jsou odvozeny od konceptu `mre:Identity` jako spustitelné `mre:Executable` aplikace `mre:Application` a služby `mre:Service`. Popis obsahuje název `rdfs:label`, popis `dc:description` nebo doplňující komentář `rdfs:comment`, verzi, tvůrce `dc:creator`, přispěvatele `dc:contributor` a licenci `dc:rights`. U software je počítáno s klasifikací typu algoritmu `mre:Algorithm`, dat na vstupu `mre:hasDatasetInputClass` a výstupu `mre:hasDatasetOutputClass` včetně možností `mre:SoftwareFeature` nebo nároků `mre:ComputationalFeature`. Pro implementaci služeb mohou být využity atributy s kódovým označením služby `mre:provides`; pro popis služeb, které je nutné spustit `mre:requiredStart`; zastavit `mre:requiredStop`; mají běžet `mre:shouldStart`; by měly být zastaveny `mre:shouldStop`; spuštěny před startem služby `mre:runBefore`; a spouštěné až po startu služby jiné `mre:runAfter`.

Popis výpočetních prostředků Pro konkrétní stroj realizující zpracování dat je určen koncept `mre:ComputationalResource`. Seskupovat je lze prostřednictvím `mre:ComputationalPool`. Popis hardware nebo software lze doplnit přidáním `mre:ComputationalFeature` ve vlastnosti `mre:hasFeature`.

Plán spouštění Pro spuštění má existovat plán `mre:ExecutionPlan` definující konfiguraci `mre:hasSettings`, identitu úlohy `mre:hasIdentity`, prioritu `mre:hasPriority`, datovou sadu `mre:hasDataset` pro vstup a výpočetní prostředí `mre:useComputationalPool` dle `mre:ComputationalPool`. Součástí plánu mohou být adresáře potřebné pro vstupní, výstupní nebo dočasná data.

Spuštění a úlohy Spuštěním dojde k vytvoření instance `mre:Execution` odkazující na software `mre:hasApplication`, datum a čas `mre:onDatetime`, příp. prostřednictvím `mre:ExecutionPlan` na příčinu vzniku události (`mre:onCron` nebo `mre:onEvent`). Následně vznikne pro každý použitý výpočetní uzel jedna *úloha* dle konceptu `mre:Task`. Úlohy mohou mít dynamicky přidělené pracovní prostředí a priority dle `mre:TaskPriority`, pokud má být podpora prioritního zpracování. Software může běžet s identitou svojí i jakoukoliv jinou, což má vliv na zabezpečení dat, řízení priorit zpracování nebo řešení zatížení (např. na úrovni operačního systému).

Výsledek Výsledek úlohy zapsaný ve výstupním adresáři bude po skončení archivován a popsán instancí typu `mre:Result`. Úloha a výsledek budou propojeny atributem `mre:hasResult`. Podobně propojení na datové sady logů (`mre:hasLog` a `mre:hasErrorLog`). Příznak `mre:isDone` je pro úspěšné dokončení a `mre:isFailed` při selhání.

Diskuze, údržba a užití Ontologie zajišťuje základní popis na výzkumné platformě. Nejdůležitějším a nejčastěji používaným konceptem je datová sada. Ostatní

koncepty a atributy se nadále mění a vyvíjejí podle potřeb a rozšiřování implementace platformy. K většímu využití ontologie bude vhodné doplnit uživatelské rozhraní, které pomůže s generováním potřebné konfigurace pro popis algoritmů nebo plán úloh. Subsystem pro plánování a provedení úloh je nad rámec práce. Bude nutné řešit podporu distribuce dat a způsob paralelního nebo distribuovaného zpracování dat.

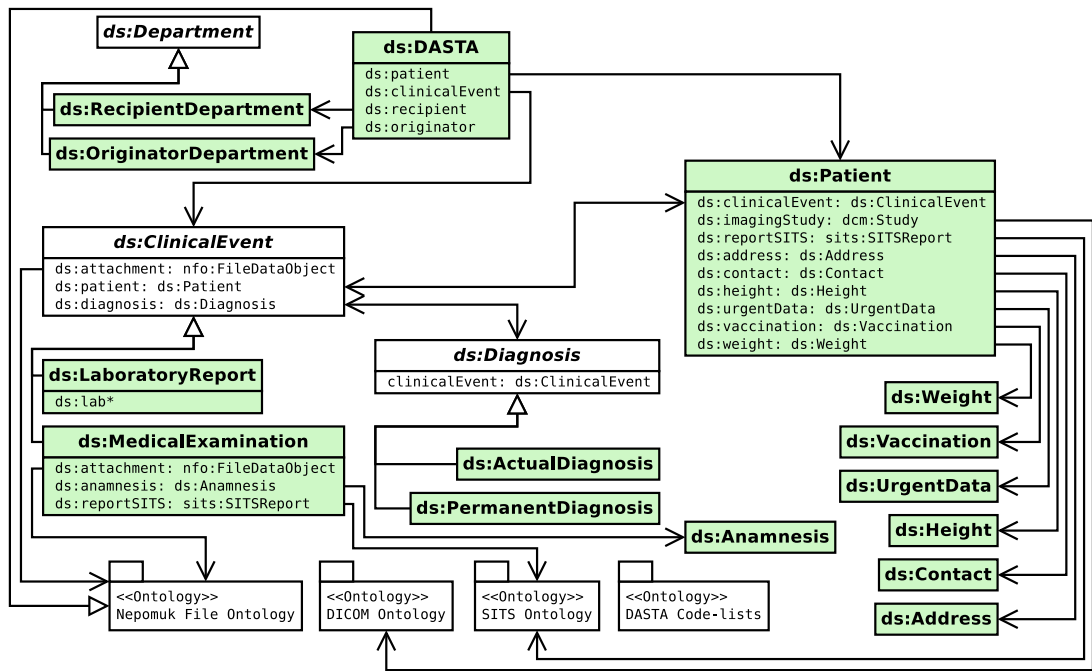
7.3 DASTA Ontology

Studie proveditelnosti a analýza domény Ontologie stejně jako formát DASTA je klíčový. Poskytuje metadata o pacientovi, konkrétní klinické události, a používají je všechny ostatní ontologie, jak ilustruje obrázek 7.1. Rizika jsou podobná s riziky formátu DASTA: (a) národní působnost, kde prakticky neexistuje jednoduchá přenositelnost do mezinárodního prostředí; (b) reálně je používáno několik verzí formátu v informačních systémech; (c) formát používá mnoho číselníků, které je nutné transformovat. Přejít k jinému formátu (5.1.1) v nejbližším období nehrozí. Možné je doplnit potřebná mapování nebo realizovat novou ontologii v souladu s budoucími standardy. Lze předpokládat, že existujících dat se změna nedotkne a zůstanou ve stávající podobě. Ontologii musí implementovat základní nástroje platformy (ETL) a respektována musí být ve všech projektech využívajících data. Potenciálním řešením je *openEHR*, kde architektura [261] podporuje medicínský výzkum i vzdělávání. Spojení *openEHR* a DASTA je vysokým rizikem, které může být příležitostí do budoucna. Přenesení DASTA do mezinárodního prostředí pro účely výzkumu je možné realizovat vlastní ontologií, která bude bez silné české lokalizace⁵ u identifikátorů i názvů konceptů a atributů. Výhoda *DASTA Ontology* je ve sjednocení verzí DASTA (DS3, DS4) do jednotné podoby, kde není cílem podpora všech informačních bloků. Nutné je počítat s konverzí používaných číselníků.

DASTA Ontology má být pouze pro výzkumné účely, kde je požadována podmnožina formátu DASTA. Důležitá jsou metadata o pacientovi, anamnéze, diagnóze, laboratorních výsledcích a terapii. Nebudou modelovány ekonomické, procesní ani jiná administrativní data, která může formát DASTA obsahovat. Používáno je celkově na 457 číselníků. Mezinárodní platnost má *Mezinárodní klasifikace nemocí* (MKN) a *Národní číselník laboratorních položek* vychází z nomenklatury IFCC. Ostatní číselníky jsou specifické pro formát a Českou republiku. Pouze zlomek číselníků má lokalizaci do angličtiny v souvislosti s eHealth.

Konceptualizace pro DASTA Pro národní formát DASTA zobrazuje obrázek 7.3 všechny definované základní koncepty, kterými jsou pacient `ds:Patient`, adresa `ds:Address`, kontakt `ds>Contact`, anamnéza `ds:Anamnesis`, diagnózy `ds:Diagnosis` se rozlišují na dva podtypy aktuální `ds:ActualDiagnosis` a trvalé `ds:PermanentDiagnosis`, klinické události `ds:ClinicalEvent` se dvěma podtypy pro lékařské vyšetření `ds:MedicalExamination` a pro laboratorní výsledky `ds:LaboratoryReport`, hmotnost `ds:Weight`, výška `ds:Height`, provedené oč-

⁵Formát DASTA je velmi silně vázán na národní prostředí. Používané elementy a atributy formátu DASTA obsahují česká slova (bez diakritiky) nebo jejich části a zkratky. Formát DASTA využívá striktně národní číselníky.



Obrázek 7.3: DASTA Ontology

kování `ds:Vaccination` a urgentní data `ds:UrgentData` s podtypy pro alergie, krevní skupinu, rizikové faktory, trvalou medikaci a očkování proti tetanu. Výše uvedená sekundární metadata spojuje koncept `ds:DASTA`, který je potomkem `nfo:FileDataObject`, pro jednu klinickou událost pacienta v DASTA souboru. Minimální záznam tvoří koncepty, kterými jsou soubor DASTA, pacient, aktuální diagnóza a jeden podtyp klinické události. Pro metadata spojená s původem dat existuje několik konceptů pro popis informačního systému zdravotnického zařízení `ds:InformationSystem`, původce dat `ds:OriginatorDepartment`, příjemce/žadatele `ds:RecipientDepartment` a garanta dat `ds:Guarantee`. Pro praktické využití ontologie při ETL procesu jsou doplněny anotace s označením názvu elementu `ds:dastaElement` nebo atributu `ds:dastaAttribute` v DASTA. Některé elementy mohou být ve více informačních blocích, a proto je možné specifikovat XPath v anotaci `ds:dastaPath`.

Konceptualizace pro slovníky a číselníky DASTA Pro účely *DASTA Ontology* je nutné adoptovat používané: (a) interní číselníky DASTA; (b) Národní číselník laboratorních položek (NČLP); a (c) Národní zdravotnický informační systém (NZIS). Prostřednictvím webových služeb lze získat číselníky⁶ v XML. Všechny číselníky jsou spojeny do *DASTA Code-lists* a doplněny o anotace pro atributy z originálních XML číselníků. Pro název konceptu je využito jméno číselníku, které se může vyskytovat v různě označeném atributu. V číselnících jsou různá označení atributů pro (a) kód slovníku, (b) jako identifikátor, (c) srozumitelný název nebo (d) popis položek slovníku. Pro jednoznačnou identifikaci hodnoty je použito spojení označení číselníku a kódu položky oddělených podtržítkem. Identifikátor pro položku určuje první existující atribut (v tomto pořadí): kod, klic, icopc, kodzaz, uzemi, histdg, zujedn, oborat, orp, cispou,

⁶<http://ciselniky.dasta.mzcr.cz/>

psc, kod1, tdgkod, slozka nebo icopcz. Srozumitelný název položky pro atribut `rdfs:label` je první existující (v tomto pořadí): `naz`, `n32`, `n32sp`, `n20`, `n90`, `znaz`, `textz`, `nazob`, `nazorp`, `nazpou`, `naz1`, `nazev`, `slozka` nebo `kodskal`.

Implementace *DASTA Ontology*⁷ používá prefix `ds`, *DASTA Code-lists*⁸ je s prefixem `dsc1` a pro oba jmenné prostory viz příloha A. Z dokumentace standardu byly převzaty české komentáře prostřednictvím `rdfs:comment`.

Diskuze, údržba a užití Modelovány jsou aspekty domény nutné pro výzkumné účely. *DASTA Ontology* a *DASTA Code-lists* jsou použity v pluginu *MetaMed Extractor DASTA* [VK12] zajišťující ETL proces výzkumné platformy. Ostatní projekty mohou metadata používat pro vyhledávání nebo je rozšiřovat a propojovat s dalšími. Vytvoření *DASTA Code-lists* je automatizováno prostřednictvím konverzního nástroje *DASTA Code-lists to RDF/OWL* (`dsc12rdf`, od 2012), který realizuje stažení⁹, extrakci, překódování do UTF8 a vygenerování *DASTA Code-lists* v jazyce OWL pro všechny číselníky¹⁰ používané v DASTA. Odděleně generuje ontologii (pouze třídy, vlastnosti a anotace) i RDF slovník s obsahem číselníků. Příloha C ve výpisu C.1 ukazuje číselník pro pohlaví POHLAV z ÚZIS, kterému odpovídá třída `dsc1:POHLAV`. V RDF je ukázka položky `dsc1:POHLAV_1` číselníku pro pohlaví s hodnotou muž v příloze C na výpisu C.2. Číselníky DASTA jsou vydávány pravidelně čtyřikrát do roka, a proto je nutná aktualizace *DASTA Code-lists* (přegenerování). *DASTA Ontology* existuje a je udržována od roku 2010. V průběhu let docházelo k rozšiřování o nové koncepty, jako byla podpora nejen číselných hodnot laboratorních výsledků. Při budoucím rozšiřování lze doplnit nepodporované informační bloky (např. ekonomická a procesní data).

7.4 DICOM Ontology

Studie proveditelnosti a analýza domény Mezi příležitosti a přínos *DICOM Ontology* lze zařadit zjednodušení a zrychlení hledání obrazových vyšetření prostřednictvím extrahovaných metadat z DICOM. Referenční informační model (RIM) DICOMu je formálním popisem domény obrazových dat, který umožňuje sdílení konzistentní podoby sémantiky metadat. Vytvoření vlastní ontologie je velkým rizikem, protože je RIM komplexní. Kahn et al. (2007) publikovali referenční informační model DICOMu [262] v rámci *DICOM Ontology Project* na Stanford Center for Biomedical Informatics Research. O tři roky později byl model publikován v [263] jako DICOM Ontology (DO). V počátcích platformy (2008–2012) existoval pouze *DICOM Ontology Project* ze Stanford Center for Biomedical Informatics Research, který splňoval požadavky pro automatizované použití. Proto je původní (první/jediná) existující implementace ontologie dle publikací [262, 263] použita pro účely výzkumné platformy. V období let 2013–2015 se objevil projekt *Semantic DICOM*¹¹ (SEDI).

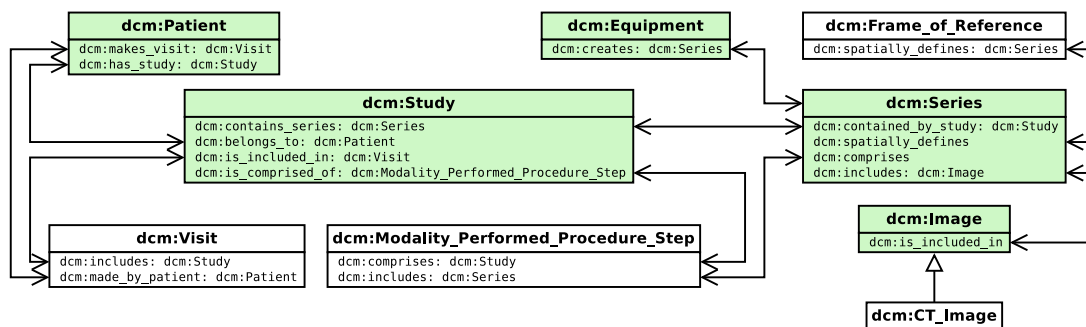
⁷<https://mre.zcu.cz/ontology/dasta.owl>

⁸<https://mre.zcu.cz/ontology/dsc1.owl>. RDF slovník s položkami číselníků lze stáhnout v komprimované podobě <https://mre.zcu.cz/ontology/dsc1.rdf.gz> (cca 51 MB).

⁹http://ciselniky.dasta.mzcr.cz/CD_DS4/CD_DS4.ZIP

¹⁰<http://ciselniky.dasta.mzcr.cz/PrehledCiselnikuXML.aspx>

¹¹<https://semantic-dicom.org/>



Obrázek 7.4: DICOM Ontology s objekty reálného světa dle [262, 263].

Konceptualizace Nejdůležitější koncepty představují objekty reálného světa `dcm:Real_World_Object`, které lze vidět v diagramu na obrázku 7.4 včetně vzájemných vazeb. Hierarchický model používá konceptů: pacienta `dcm:Patient`, studie `dcm:Study`, série `dcm:Series` a snímku `dcm:Image`. Nechybí metadata o modalitě `dcm:Equipment` a další aspekty jako jsou informační entity, moduly, reprezentace hodnot (VR), elementy nebo jejich typy.

Implementace Pro *DICOM Ontology*¹² je používán prefix `dcm` a jmenný prostor viz příloha A. Ontologie vychází z implementace dle publikací [262, 263]. Byla nutná konverze z původního zastaralého projektového formátu Protégé do OWL. U tříd a vlastností popisujících objekty reálného světa a jejich vztahy jsou doplněny české i anglické názvy prostřednictvím `rdfs:label`.

Diskuze, údržba a užití Ontologie je použita v pluginu *MetaMed Extractor DICOM* [VK12] pro zajištění ETL procesu výzkumné platformy. Plugin je plně konfigurován z ontologie prostřednictvím výběrových dotazů ve SPARQL. Do budoucna je předpokládáno nasazení *Semantic DICOM* paralelně ke stávající ontologii. Později může být plně nahrazena.

7.5 Form Ontology

Studie proveditelnosti a analýza domény Zaměření je na popis webového uživatelského rozhraní pro podporu vytváření a editace RDF dat prostřednictvím formulářů v HTML5. Předmětem zájmu jsou anotace pro obohacení datových a objektových vlastností o metadata vhodná pro realizaci uživatelského rozhraní se zaměřením na formulářové zadávání a editaci hodnot. Anotace poslouží pro zobrazení/rozvržení (rozměry nebo zástupný text vstupního pole), zjednodušení (výchozí hodnota, nebo způsob řazení položek ve výčtovém seznamu) a validaci (min/max hodnota, regulární výraz) formulářů s daty v RDF. Manuální tvorba nebo editace RDF dat není příliš častá a právě anotace mají umožnit rozvoj v tomto směru. Důsledkem bude zjednodušení možností sběru strukturovaných dat. Existující řešení nebyla nalezena. Uvažovány jsou webové technologie a HTML5.

¹²<https://mre.zcu.cz/ontology/dcm.owl>

Konceptualizace a implementace Koncepty definovány nejsou. Pouze anotace vhodné pro třídy nebo vlastnosti. *Form Ontology*¹³ používá prefix `form` a jmenný prostor viz příloha A. Určit je možné kardinalitu `form:cardinality` nebo jednotku `form:unit` pro hodnoty vlastností mimo rámec možností ontologie. Třídy číselníků lze označit anotací `form:isVocabulary`. Položky číselníku mohou uvádět kód `form:code`, skóre `form:score` nebo indikovat, že se jedná o výchozí hodnotu `form:isDefaultValue`. Při systematické tvorbě číselníků/slovníků (dle přílohy B) může být užitečné doplnit unikátní identifikaci číselníku `form:vocabularyId` a každé položky `form:vocabularyItemId`. Nebo nezávisle na slovníku použít označení řádky `form:vocabularyLineNumber` k identifikaci položek.

Pro formulářové prvky HTML5 indikují anotace pole s povinností jej vyplnit `form:htmlRequired`, pouze pro čtení `form:htmlReadOnly` a zakázané resp. neaktivní `form:htmlDisabled`. Ovlivnit lze velikost vstupního pole `form:htmlSize`, nebo počet sloupců `form:htmlCols` a řádek `form:htmlRows` pro `<textarea>`. Výchozí rozměry obrázku nastaví pro výšku `form:htmlInputImageHeight` a šířku `form:htmlInputImageWidth`. Typ vstupního pole určí `form:htmlInputType`, automatické doplňování umožní definovat `form:htmlInputAutocomplete`, nebo zaměření kurzoru `form:htmlInputAutofocus`, možnost zaškrtnout nebo vybrat více hodnot `form:htmlInputMultiple` a `form:htmlInputPlaceholder` pro zobrazení zástupného textu. V číselném vstupním poli lze nastavit výchozí hodnotu `form:htmlValue`, krok `form:htmlValueStep`, nejmenší `form:htmlValueMin` a nejvyšší `form:htmlValueMax` povolené hodnoty. Pro obecná textová vstupní pole lze určit maximální délku `form:htmlValueMaxLength` řetězce nebo regulární výraz `form:htmlValuePattern`.

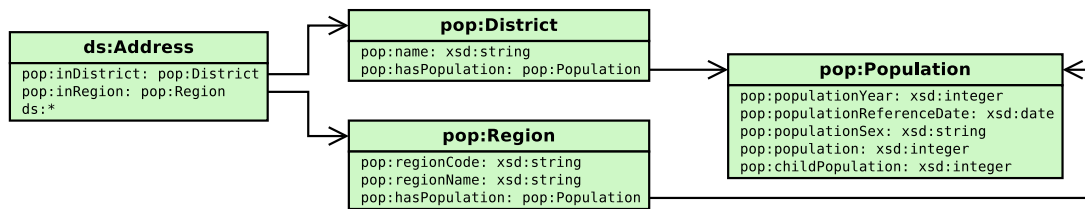
Diskuze, údržba a užití *Form Ontology* usnadňuje tvorbu uživatelského rozhraní při použití knihoven a nástrojů MRE dle kapitoly 8.6. Příklad využití je uveden v příloze B na výpisu B.1. Podporu lze implementovat ve vlastních knihovnách a aplikacích. Nové anotace mohou být vhodné pro rozšíření možností a způsobů prezentace a editace dat. Rozšiřování nemá mít vliv na existující anotace, které již mohou být v ontologiích používány.

7.6 Population Ontology

Studie proveditelnosti a analýza domény *Population Ontology* představuje formu RDF slovníku s omezeným rozsahem konceptů a jejich vlastností. Data o populaci publikuje *Český statistický úřad*¹⁴ (ČSU) v agregované podobě veřejně ve formátu CSV. Za určených podmínek lze získat (neveřejné) údaje o populaci s menší mírou agregace dat až na úroveň jednotlivých obcí, věkových kategorií nebo s odlišením pohlaví. Při návrhu je nutné počítat s různou granularitou a transformací ze zdrojového formátu dat (CSV, XLS) do RDF. Rizika mohou plynout z nevhodně zvolené granularity nebo nedostupnosti potřebných dat. Aplikace založené na *Population Ontology* ve spojení s daty o populaci mohou poskytovat významné ukazatele pro medicínský výzkum.

¹³<https://mre.zcu.cz/ontology/form.owl>

¹⁴<https://www.czso.cz/>



Obrázek 7.5: Population Ontology

Konceptualizace a implementace *Population Ontology*¹⁵ používá prefix `pop` a jmenný prostor viz příloha A. Koncepty ilustruje obrázek 7.5. Rozlišují se kraje `pop:Region` a okresy `pop:District`. Vlastnost `pop:hasPopulation` propojuje územní celky se všemi instancemi velikosti populace. Koncept `pop:Population` je pro data o populaci v oblasti s rozlišením roku, věku a pohlaví. Instance typu `pop:Population` je opatřena anotacemi s identifikací `dc:identifier` a člověku srozumitelným označením `rdfs:label`. Instance populace specifikuje referenční datum `pop:populationReferenceDate`, rok `pop:populationYear`, kraj `pop:inRegion`, okres `pop:inDistrict`, kategorii věku `pop:populationAge` a pohlaví (muž M, žena F a vše A) ve vlastnosti `pop:populationSex`. K dispozici je údaj o populaci celkové `pop:population` nebo o počtu dětí `pop:childPopulation`.

Diskuze, údržba a užití Zdrojem údajů o populaci je *Český statistický úřad*¹⁶. Použity jsou veřejné a neveřejné datové sady. Na DVD je proto uvedena pouze ukázka možné granularity dat v RDF (`data/population.ttl`). Pro využití údajů o populaci dle *Population Ontology* je nutná znalost identifikace území¹⁷ (obec, okres nebo kraj). Do budoucna lze uvažovat vyšší míru integrace nebo mapování s *Národním katalogem otevřených dat*¹⁸ (NKOD).

7.7 Stroke Ontology

Stroke Ontology vznikla v červenci 2018 jako výsledek evoluce v oblasti popisu léčby a sledování stavu pacientů po cévní mozkové příhodě. Jedná se o nástupce původně vytvořených *SITS Ontology* a *NIHSS Ontology*.

SITS Ontology a NIHSS Ontology Použití *SITS Ontology* (prefix `sits`) a *NIHSS Ontology* (prefix `nihss`) bylo v období roků 2011–2018. Poskytovaly ontologický popis včetně anotací s mapováním RDF vlastností a tříd na formuláře webového rozhraní registru SITS (5.3). Ontologie měly za celou dobu svého životního cyklu více rolí. Zpočátku sloužily pro strukturování stažených metadat z webových formulářů registru SITS do RDF. Následně pro transformaci stažených CSV dat do RDF. Poslední role byla automatické plnění a kontrola hodnot

¹⁵<https://mre.zcu.cz/ontology/pop.owl>

¹⁶<https://www.czso.cz/>

¹⁷K identifikaci území lze využít *Nomenklatury územních statistických jednotek* (NUTS), *Místních samosprávných jednotek* (LAU) nebo *Základních územních celků* (ZUJ). Například *Číselník obcí* (CISOB, <http://apl.czso.cz/iSMS/cisdet.jsp?kodcis=43>) dle ČSU.

¹⁸<https://data.gov.cz/>

v registru SITS. Relativně četné změny datového modelu nebo webového rozhraní registru SITS komplikovaly dlouhodobý hladký provoz a užití. S ontologiemi a anonymními daty pracovalo několik studentů¹⁹. Přechod k RES-Q ukončilo životní cyklus obou popsanych ontologií, které již nejsou aktivně používány.

Studie proveditelnosti a analýza domény Ontologie bude prostředkem pro analýzu strukturovaných dat spolupracujícího zdravotnického zařízení a dokumentaci transformace včetně mapování dat pro RES-Q. Mezi příležitosti patří možnost realizovat nástroj pro automatizované plnění *RES-Q*. Riziko pro *Stroke Ontology* je nižší ve srovnání se *SITS Ontology*, protože je blíže potřebám lékařů a implementace ve zdravotnickém zařízení (*StrokeXML*). Vztah k budoucímu automatickému plnění registru *RES-Q* je i rizikem, protože registr neposkytuje rozhraní pro automatizovanou výměnu dat. Nebylo identifikováno vhodné řešení. Pro strukturovaný popis je nutné modelovat doménu rozlišující několik časových období, v nichž se provádí diagnostika a potřebná vyšetření nebo se zaznamenávají zjištěné hodnoty.

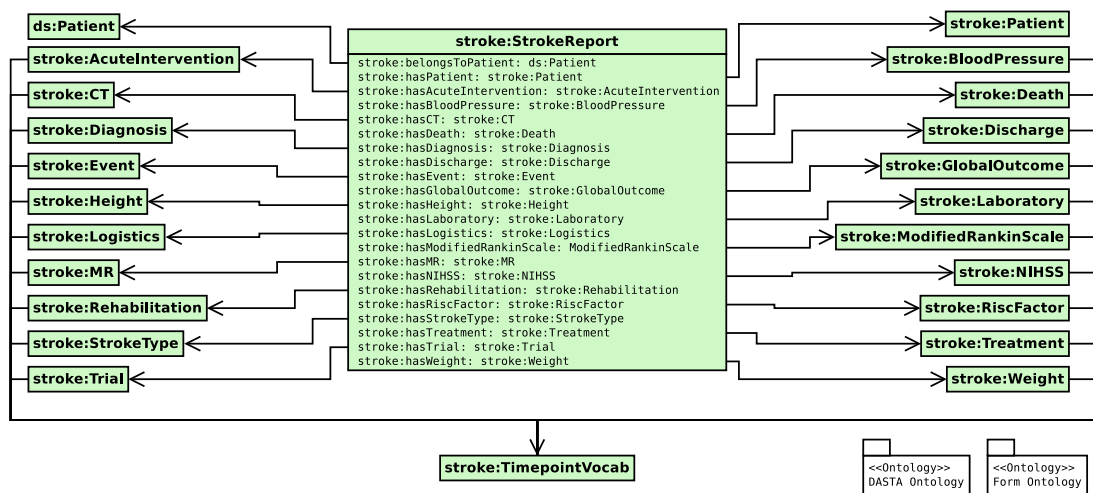
Konceptualizace a implementace Koncepty *Stroke Ontology*²⁰ dle obrázku 7.6 používají prefix `stroke` a jmenný prostor dle přílohy A. Klíčový koncept má vztah ke konkrétnímu případu cévní mozkové příhody `stroke:StrokeReport`, kterých může mít pacient více. Sledovány jsou údaje v časových bodech dle konceptu `stroke:TimepointVocab`. Údaje a souhrn údajů o stavu pacienta, realizované léčbě a dosažených výsledků poskytují koncepty: pacient `stroke:Patient`, vyšetření CT `stroke:CT` nebo MR `stroke:MR`, diagnóza `stroke:Diagnosis`, typ iktu `stroke:StrokeType`, rizikové faktory `stroke:RiscFactor`, laboratorní hodnoty `stroke:Laboratory`, krevní tlak `stroke:BloodPressure`, standardizované neurologické vyšetření dle NIHSS `stroke:NIHSS`, výška `stroke:Height` a hmotnost `stroke:Weight`, provedený akutní zákrok `stroke:AcuteIntervention`, podané medikamenty `stroke:Treatment` a rehabilitace `stroke:Rehabilitation`. Modifikovaná Rankinova škála `stroke:ModifiedRankinScale` je použita pro klasifikaci postižení. Některé údaje se vztahují na propuštění `stroke:Discharge` stejně jako výstupní hodnocení `stroke:GlobalOutcome`. Pro výzkumné účely lze pacienta zařadit do klinické studie `stroke:Trial`. Pro popis transformace dat z formátu *StrokeXML* slouží koncept `stroke:XMLElementMedicalc`. Úplná konceptualizace je na DVD²¹ a příloha B ukazuje výpis B.1 s příkladem jedné hodnoty číselníku popisující postižení zraku dle NIHSS včetně všech anotací a použití *Form Ontology* (7.5). Pro vygenerování OWL ontologie z tabulkového procesoru byla použita práce [259].

Diskuze, údržba a užití Přidávání nových atributů, které mají lékaři u pacientů nově sledovat, lze v budoucnu očekávat stejně, jako tomu bylo dosud. Definiční atributů a číselníkových hodnot spravuje zdravotnické zařízení, což se projeví okamžitou změnou formátu *StrokeXML*. Ontologie slouží jako konfigurace pluginu *MetaMed Extractor Stroke* pro zajištění ETL procesu aplikací *MetaMed* [VK12].

¹⁹Do SITS Ontology přispěli studenti Pavel Karlík (2011/2012) a Jindřich Pouba (2012/2013). Věnovali se plnění registru SITS dle XML dat, přehledům a editacím dat pro účely lékařů.

²⁰<https://mre.zcu.cz/ontology/stroke.owl>

²¹Konceptualizace *Stroke Ontology* v `./example/ontology-design-stroke.xlsx` na DVD.



Obrázek 7.6: Stroke Ontology

Praktické využití možností ontologie je ve webové aplikaci *Query Builder* (8.6.5) pro temporální data a bylo publikováno v [VKK18] (2018).

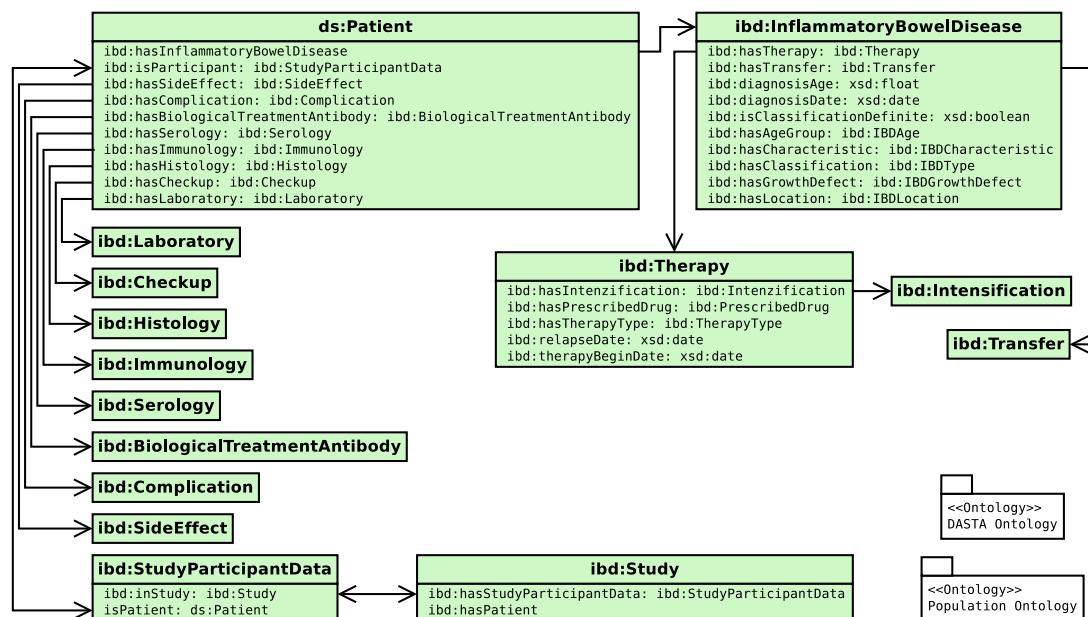
7.8 Idiopatické střevní záněty

7.8.1 IBDS Ontology

Studie proveditelnosti a analýza domény Příležitost spočívá v poskytnutí strukturovaného popisu pro realizaci studií na základě dat shromážděvaných lékařem v omezeném rozsahu. Pro prospektivní databázi nebylo existující řešení identifikováno. Existují standardizované způsoby hodnocení, klasifikace nebo škály používané při diagnostice a kontrolách pacientů, kterých je potřeba se držet při návrhu a implementaci číselníků. Některé klasifikační systémy umožňují výpočet skóre, které bude potřeba určovat.

Konceptualizace a implementace *IBDS Ontology*²² používá prefix *ibd* a jmenný prostor viz příloha A. Definováno je 39 konceptů a 132 atributů (52 datových, 70 objektových a 10 anotací). Východiskem jsou koncepty pro pacienta *ds:Patient* a laboratorní výsledky *ds:LaboratoryReport* z *DASTA Ontology*. Zjednodušený pohled na koncepty poskytuje obrázek 7.7. Případ idiopatického střevního zánětu *ibd:InflammatoryBowelDisease* je klíčovým konceptem, který poskytuje základní údaje (např. datum a věk při zařazení, typ diagnózy). Navažují koncepty pro zaznamenání vývoje a průběžnou kontrolu stavu nemoci u pacientů: kontrola *ibd:Checkup*, laboratorní výsledky *ibd:Laboratory*, histologie *ibd:Histology*, imunologie *ibd:Immunology*, sérologie *ibd:Serology*, terapie *ibd:Therapy*, protilátky biologické léčby *ibd:BiologicalTreatmentAntibody*, intenzifikace *ibd:Intensification*, komplikace *ibd:Complication* nebo vedlejší účinky *ibd:SideEffect* a předání jinam *ibd:Transfer*. Pro analýzy dat a tvorbu studií jsou určeny pomocné koncepty studie *ibd:Study* a účastníků *ibd:StudyParticipantData*. Při určování metrik (např. incidence) jsou potřeba informace o územních celcích v členění na kraje, okresy a s informací o populaci

²²<https://mre.zcu.cz/ontology/ibd.owl>



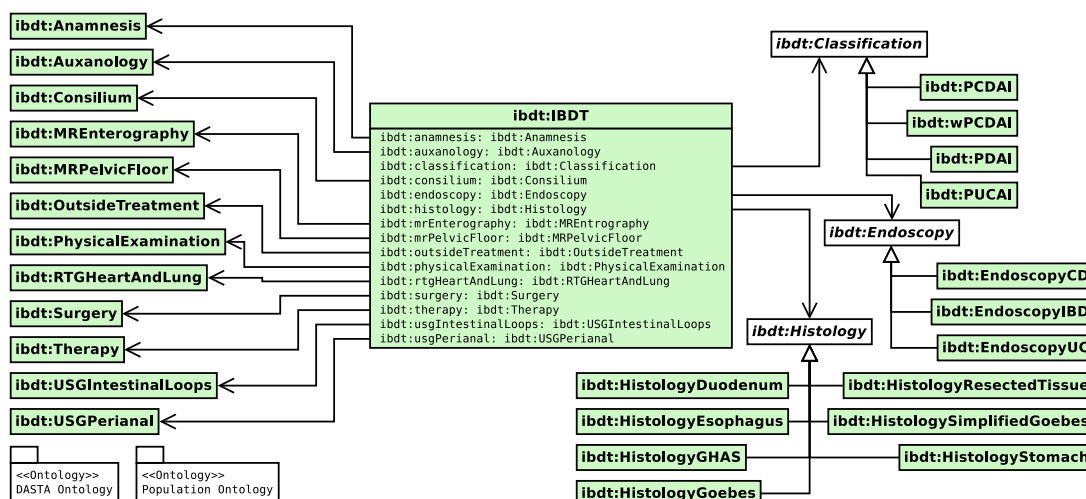
Obrázek 7.7: IBDS Ontology

v dané věkové skupině. Pro popis populace v územních celcích je určena *Population Ontology* (viz kapitola 7.6).

Diskuze, údržba a užití Ontologie je použita ve webové aplikaci prospektivní databáze pro výzkum idiopatických střevních zánětů vytvořené v rámci výzkumné skupiny. Jedná se o informační systém (spravuje Ing. Martin Kryl, KIV), prostřednictvím kterého jsou realizovány požadavky lékařů na vizualizaci, zadávání, kontrolu, ale i dotazování a analyzování dat [KVK18]. Pro výzkumné studie jsou připojena metadata o populaci pro Plzeňský kraj na velmi detailní úrovni. Výsledky prospektivní studie za období let 2000–2015 byly publikovány v [SSC⁺17]. Sběr dat a doplňování funkcionality neustále pokračuje.

7.8.2 IBDT Ontology

IBDT Ontology je řešena nezávisle na *IBDS Ontology* a vzniká ve spolupráci mezi univerzitou, lékařskou fakultou a zdravotnickým zařízením pro zajištění společného mezioborového výzkumu. Zásadním rozdílem jsou celkové požadavky, rozsah a množství detailů, které je proti IBDS několikanásobně větší. Přesto je předpokladem zajištění vzájemné kompatibility doplněním mapování pro stejné koncepty. Příležitostí *IBDT Ontology* je rozsáhlý popis domény, který by mohl pomoci k dosažení cílů spolupráce. Zapojení dalších zdravotnických zařízení umožní sběr většího objemu dat vztahujících se ke zkoumaným idiopatickým střevním zánětům. Rozsáhlý popis znamená množství atributů a číselníků, které může současně znamenat problémy a možná rizika při realizaci (manuální vyplňování velkého počtu hodnot; překlepy). Výsledky laboratorních vyšetření jsou do zdravotnického zařízení předávány ve formátu DASTA. Vhodným způsobem eliminace rizik je využití řízených slovníků a integrace s daty v DASTA (pacient, laboratorní výsledky) dle *DASTA Ontology* a *DASTA Code-lists*.



Obrázek 7.8: IBDT Ontology

Konceptualizace a implementace *IBDT Ontology*²³ používá prefix *ibdt*. Ve formě návrhu definuje 218 konceptů a 490 atributů (216 datových, 262 objektových a 12 anotací), které čekají na konzultaci a potvrzení lékařů. Vizualizace konceptů prezentuje obrázek 7.8. Celkový návrh ilustruje příloha B na obrázcích B.1 a B.2.

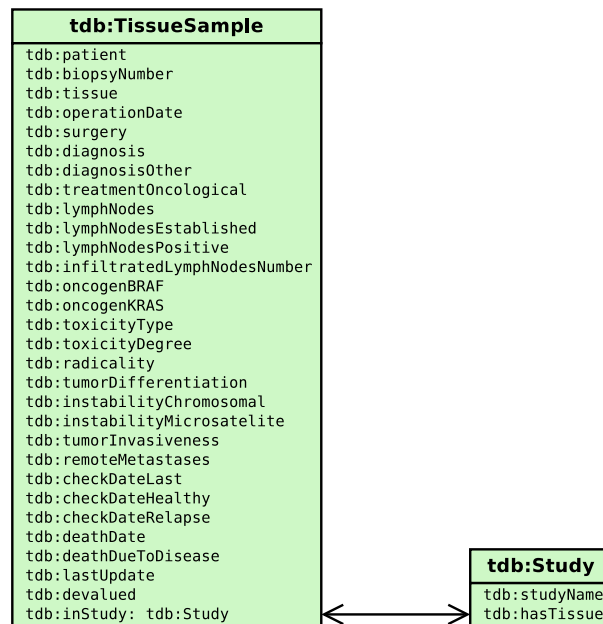
Diskuze, údržba a užití Rozsah tříd a atributů byl jedním z důvodů pro inovativní přístup [VKKK17] s cílem plně využít *potenciál metadat* dostupných ve formě ontologie pro realizaci webového rozhraní pro zadávání, tvorby a editace RDF dat. Popis frameworku *Web Form Generator* je uveden v kapitole 8.6.6 a výsledné dynamicky generované formuláře jsou na obrázcích v příloze D.3.

7.9 TDB Ontology

TDB Ontology je určena k evidenci v *Databázi tkáňových vzorků* a pro realizaci klinických studií. Jedná se o vzorky tkáně odebrané při biopsii, u kterých se sleduje množství metadat sloužících pro výběr a rozhodnutí o zařazení do klinických studií. Podobně jako u IBDS, původní požadavek vycházel z již existující vlastní databáze (Microsoft Access), kterou společně sdílelo více lékařů.

Studie proveditelnosti a analýza domény Doména je velmi úzce zaměřena na unikátní databázi, která má zejména lokální význam pro konkrétní lékaře. Vhodné řešení neexistuje. Možné komplikace přináší výskyt osobních údajů, které musí být při uložení v RDF šifrovány. Příležitostí je realizace relativně jednoduchého projektu, který bude průkopníkem pro jiný pohled na potřeby a požadavky při standardizaci popisu sbíraných metadat. S ohledem na rozsah jsou ekonomické důsledky tvorby ontologie nebo následné webové aplikace minimální. Dlouhodobý provoz nebude náročný ani s ohledem na uvažovaný objem vzorků a sledovaných atributů.

²³<https://mre.zcu.cz/ontology/ibdt.owl>



Obrázek 7.9: TDB Ontology

Konceptualizace a implementace *TDB Ontology*²⁴ má prefix `tdb` a jmenný prostor viz příloha A. Pro databázi tkáňových vzorků existují koncepty vzorek tkáně `tdb:TissueSample` a studie `tdb:Study` dle obrázku 7.9. Vlastnosti jsou vidět na obrázcích 7.9 a D.9. Datové vlastnosti jsou identifikace pacienta `tdb:patient`, číslo biopsie `tdb:biopsyNumber`, typ tkáně `tdb:tissue`, datum operace `tdb:operationDate`, popis provedeného výkonu `tdb:surgery`, diagnóza `tdb:diagnosis`, anebo slovní popis jiné diagnózy `tdb:diagnosisOther`, onkologická léčba `tdb:treatmentOncological`, lymfatické uzliny `tdb:lymphNodes` stanovené `tdb:lymphNodesEstablished` a pozitivní `tdb:lymphNodesPositive`, počet infiltrovaných mízních uzlin `tdb:infiltratedLymphNodesNumber`, onkogeny B-RAF `tdb:oncogenBRAF` a K-RAS `tdb:oncogenKRAS`, typ `tdb:toxicityType` a stupeň toxicity `tdb:toxicityDegree`, radikalita `tdb:radicality`, diferenciací `tdb:tumorDifferentiation`, instabilita chromosomální a současně mikrosatelitní (`tdb:instabilityChromosomal` a `tdb:instabilityMicrosatelite`), invazivita `tdb:tumorInvasiveness`, metastázy `tdb:remoteMetastases`, datum kontroly `tdb:checkDateLast`, do kdy byl zdrav `tdb:checkDateHealthy`, recidiva `tdb:checkDateRelapse`, úmrtí `tdb:deathDate`, příznak, zda byla diagnóza příčinou úmrtí `tdb:deathDueToDisease`, datum aktualizace `tdb:lastUpdate` a znehodnocení vzorku `tdb:devalued`. Objektová vlastnost `tdb:inStudy` ukazuje na studii `tdb:Study`, v níž byl vzorek použit. Studie má název `tdb:studyName` a odkazy na vzorky `tdb:hasTissue`, které jsou ve studii použity. Používané anotace jsou identifikátor `tdb:id`, kód `tdb:code` a poznámka `tdb:note`.

Diskuze, údržba a užití Model vychází z existujících dat a má dlouhodobě vyhovující strukturu. Uvažovat lze přidání atributů pro určení období studie, jména zodpovědného lékaře nebo odkazu na související výsledky a publikace.

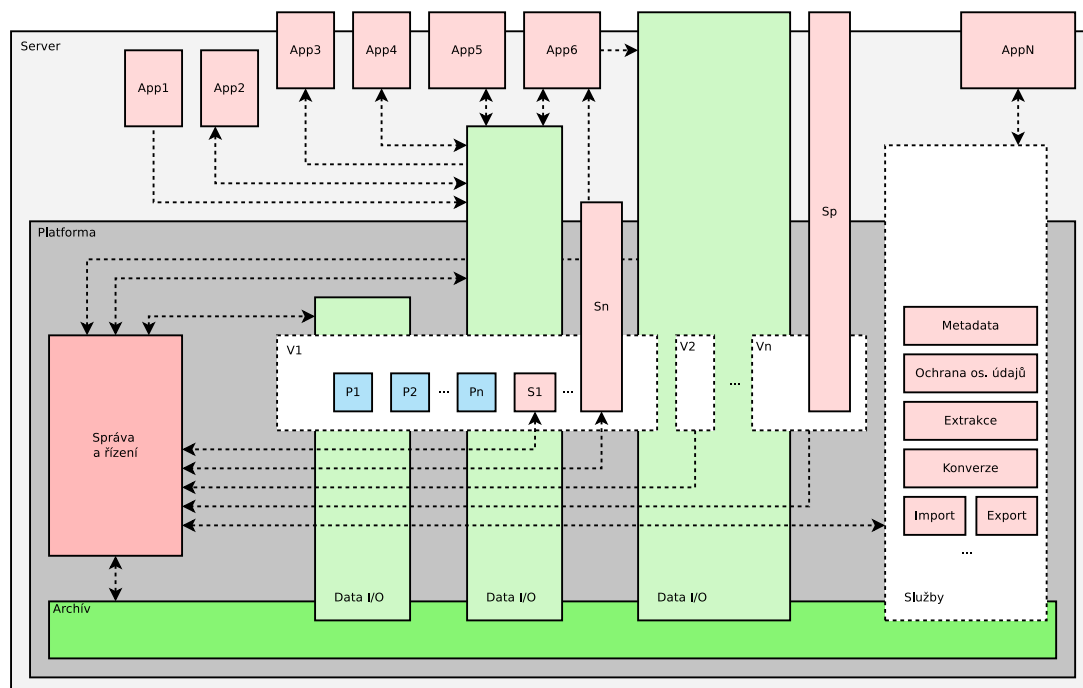
²⁴<https://mre.zcu.cz/ontology/tdb.owl>

8 Standardizace procesů medicínského výzkumu

Platforma *Medical Research & Education* (MRE) řeší standardizaci procesů medicínského výzkumu. Poskytuje potřebný rámec pro dlouhodobou archivaci a budoucí zpracování dat určených pro výzkumné účely. Jedná se o výzkum zahrnující heterogenní data s množstvím často opakujících se úloh, které jsou společné pro různé typy a zdroje dat. Klíčové požadavky a metriky uvádí příloha G. Použitou metodou je *systémový přístup*. Strukturu a vlastnosti platformy popisuje kapitola 8.1. Hlavní subsystémy tvoří často opakující se úlohy – zajišťující zpracování nových dat (8.2), ochranu osobních údajů (8.3) a interakci s okolím (8.5). V kapitole 8.6 jsou popsány další softwarové prostředky a knihovny využívající možnosti RDF dat a ontologií nejen z předchozí kapitoly 7.

8.1 Výzkumná platforma

Schéma výzkumné platformy je ve formě diagramu uvedeno na obrázku 8.1. Klíčovým prvkem je *archiv* pro perzistentní uchovávání dat, metadat, softwarových prostředků, konfigurace a výsledků i ve více verzích. Platforma poskytuje přístup a vstupně/výstupní (I/O) operace na třech úrovních: (a) interně v rámci platformy, (b) lokálně na konkrétním serveru a (c) veřejně přístupné vnější rozhraní. Blok *Správa a řízení* zajišťuje realizaci procesů a spouštění softwarových prostředků. Pro oddělení dat a úloh jsou rozlišovány **identity**, jimiž mohou být platforma, její uživatelé, projekty, týmy, softwarové prostředky.



Obrázek 8.1: Schéma výzkumné platformy

Softwarové prostředky jsou aplikace `mre:Application` a služby `mre:Service` dle *MRE Ontology* (7.2). *Aplikace* představují software spouštěný jednorázově pro zpracování dat a na obrázku 8.1 jsou zobrazeny bloky `App1..AppN`. Pouze lokálně na daném serveru jsou dostupné aplikace `App1` a `App2`. Ostatní poskytují veřejné rozhraní, ale nemusí být volně dostupné (zabezpečeny). *Služby* jsou trvale běžící úlohy, které lze využívat opakovaně a na obrázku 8.1 jsou ilustrovány bloky `S1..Sn` a `Sp`. Hranice mezi platformou a vnějším okolím zprostředkovává rozhraní datové `Data I/O` a aplikační (`App1..AppN`, `Sn` a `Sp`) v rámci serveru i mimo něj.

Diagram na obrázku 8.1 ukazuje současnou existenci více výzkumných projektů `V1, V2..Vn`, které mohou být vytvořeny v rámci různých týmů nebo pracovních skupin (`P1..Pn`). Mezi jednotlivými týmy a skupinami jsou očekávány vazby, které na diagramu zachyceny nejsou (pro přehlednost). Platforma poskytuje přístup k datům a metadatům uživatelům přímo nebo prostřednictvím softwarových prostředků.

Schéma na obrázku 8.1 zobrazuje několik vrstev, které se mohou různě protínat v závislosti na potřebách výzkumných projektů. Nejnižší vrstvu `L1` tvoří bloky *Správa a řízení* a *Archiv*. Navazuje vrstva `L2` v podobě přístupových metod k archivu – *Data I/O*, obecné služby (např. manipulace s metadaty, ochrana osobních údajů) a aplikace (`App1..AppN`). Vrstva `L3` představuje výzkumné projekty (`V1..Vn`) včetně jejich softwarových prostředků (`S1..Sn` a `Sp`). Vrstva `L4` je speciálním případem, kam by bylo možné zařadit softwarové prostředky, které vznikly v rámci výzkumných projektů, ale mají širší dosah jako jsou okolím použitelné `Sn` a `Sp`.

8.1.1 Archiv

Archiv slouží pro dlouhodobou archivaci dat od roku 2010. Na základě OAIS [116, 117] a PAIMAS [118, 119] je striktně oddělen zápis a čtení. Přímý zápis do archivu povolen není. K zápisu dat je vyhrazen *vstupní bod*. Analogicky je definován *výstupní bod* určený pro exporty a výměnu dat, který může být vyhrazen konkrétní přístupové metodě. Právě `Data I/O`, dle schématu na obrázku 8.1, představuje vstupní a výstupní body, jejichž obsluhu zajišťuje blok *Správa a řízení*.

Vstupní a výstupní body jsou rozlišovány na vnitřní a vnější, dle možné viditelnosti z okolního prostředí. Vnitřními body lze předávat data pouze mezi úlohami a identitami. Vnějšími body lze provádět výměnu dat s okolím. Vstupní vnější body jsou určeny pro příjem a import dat. Pro sdílení nebo export dat slouží výstupní vnější body. Detail vnějšího vstupního bodu archivu je blok označený *Příchozí* na obrázku 8.2. Vnější *zdroj dat* odešle data vhodnou přístupovou metodou na pojmenovaný vstupní bod `S1..Sn`, který je vyhrazený pro konkrétní zdroj dat nebo projekt.

Konfigurační soubory Konfigurační soubory platformy jsou umístěny v systému (`/etc/mre/`). Správa *identit* je možná vytvořením souboru v podadresáři `ident.name.d/`. Schéma archivu rozlišuje na nejvyšší úrovni dle *typu přístupu* (`access.type.d/`) na lokální data archivu `local`, vnější vstupní bod `received`, vnější výstupní bod `send` a pracovní adresář běžících úloh `run`, ale možné jsou další přístupové metody. Pro dlouhodobou udržitelnost archivu je nutné udržo-

vat data použitelná. Možné to je prostřednictvím speciálních interních vstupních bodů, které mohou zajišťovat obnovu dat nebo transformace formátů. Aktualizaci existujících dat v archivu umožňuje přístupová metoda `reupload`, protože výchozí `received` je primárně určen k archivaci nových příchozích dat. Rozlišovány jsou *typy dat* (`data.type.d/`) pro software `app`, zálohy `backup`, konfigurace `conf`, dokumentaci `doc`, obecný domovský adresář identity `home`, uživatelem nebo systémem generované seznamy `list`, logy `log`, ontologie `ontology`, zdrojové soubory `raw`, RDF data `rdf`, trvale uložené výsledky zpracování dat `result` a data ke sdílení prostřednictvím webu `www`.

Pro činnost platformy jsou k dispozici *funkce* (`functions.d/`) pro definování proměnných, logování, import RDF dat, nebo řízení přístupu ke sdíleným datům (ochrana kritické sekce, možnost vyloučení násobného spuštění).

Zpracování dat je řešeno prostřednictvím *událostí*. Definice procesů a úloh je v adresáři `event.d/` např. pro zpracování dat z vnějších vstupních bodů. S událostmi souvisí *spouštěče* pro provedení před `trigger.before.d/` nebo po `trigger.after.d/` úloze. Identita může definovat funkce `conf/functions.d/`, události `conf/event.d/`, spouštěč před `conf/trigger.before.d/` a po události `conf/trigger.after.d/`.

Datové úložiště Základem archivu je datové úložiště, které může být tvořeno disky připojenými lokálně nebo po síti (např. cloud, DRBD, iSCSI). Postupně jsou do systému v `/storage/data-<code>` připojovány disky, kde `<code>` je unikátní označení (rok připojení). Existuje symbolický odkaz `/storage/current` na poslední připojený disk, který byl nejprve inicializován vytvořením adresářové struktury dle konfigurace. Starší připojené disky archivu jsou pouze pro čtení. Archiv má adresářovou strukturu na připojeném disku obecně podle prvního vzoru ve výpisu 8.1. Použity jsou zástupné znaky pro typ přístupu `<accessType>`, identitu `<ident>`, typ dat `<dataType>`, kód nebo pojmenování `<name>`, rok `<year>`, měsíc `<month>`, den `<day>`, celý datum `<date>` nebo čas `<time>`. Další řádky výpisu 8.1 zobrazují příklady vybraných typů přístupu a dat. Výchozí pro zařazení a seskupení dat je datum poslední změny souboru. Datum modifikace zajistí uložení více verzí stejných dat. Výchozí chování lze modifikovat podle požadavků projektu, tj. může být specifické pro jednotlivé vstupní body.

Výpis 8.1: Schéma a příklady adresářové struktury archivu

```
./<accessType>/<ident>/<dataType>/[<name>/][<year>/<month>/][<day>/][<hour>/]

./<accessType>/<ident>/raw/<year>/<month>/<day>/<hour>
./<accessType>/<ident>/rdf/<date>
./<accessType>/<ident>/result/<name>/<year>-<month>-<day>
./<accessType>/<ident>/list/date-received/<year>/<month>/<date>-<time>.lst
./run/<ident>/
./send/<ident>/[<year>/]
```

8.1.2 Softwarové prostředky

Rozšiřitelnost platformy je založena na softwarových prostředcích. Přidáním nového software dojde k rozšíření o nové možnosti zpracování dat a metadat. Pro

konfiguraci software lze využít konfigurační soubor nebo parametry při spuštění. Vhodný popis software pomůže s integrací. Dokumentace uživatelům umožní správné použití (očekávané vstupy, předpokládané výstupy). Doporučeno je popsat softwarové prostředky prostřednictvím metadat: (1) základního popisu – rozlišení mezi aplikací a službou (`mre:Application`, `mre:Service`), uvádí identifikační a popisné informace o aplikaci jako název (`rdfs:label`, `dc:title` nebo `skos:prefLabel`), autora (`dc:creator`), přispěvatele (`dc:contributor`), licence (`dc:rights`), verze nebo data vydání (`dc:modified`); (2) umístění včetně potřebných knihoven a parametrů; (3) definice události a spouštěčů; (4) aplikačních dat; a (5) dokumentace.

Doporučená adresářová struktura software předpokládá umístění v adresáři `app`, kde může být libovolná struktura adresářů včetně knihoven nebo aplikačních dat. Další adresáře jsou určeny pro sdílenou dokumentaci (`doc/`), konfiguraci (`conf/`), aplikační data (`data/`), anebo ontologii (`ontology/`). Pro spuštění by měl existovat skript `run.sh` a nejlepší je definovat událost v adresáři `conf/event.d/` pro spuštění. Volitelně lze doplnit metadata jako jsou klasifikace algoritmu, vstupů, výstupů nebo používané metody a nároky v souboru `description.xml`.

8.1.3 Úlohy a procesy zpracování dat

Spuštěné softwarové prostředky jsou *úlohy*. Pro komplexní zpracování dat jsou určeny *procesy* představující předpis s posloupností úloh. Zahájení úlohy nebo celého procesu je prostřednictvím mechanismu událostí (`event`) a spouštěčů (`triggerů`). Spouštěče se aktivují *před* vznikem, anebo *po* skončení události, pokud budou pro danou událost definovány. Implementace využívá tři nástroje pro (1) kontrolu dat, (2) obsluhu událostí a (3) spouštěče. Blok *Správa a řízení*, dle kapitoly 8.1, je přímo založen na implementaci operačního systému, který zajišťuje oddělení uživatelů a provádí kontrolu vstupních bodů včetně vyvolání události. Další metody spuštění, např. na základě konkrétního času, nejsou běžné, ale jejich použití je možné. Událost může být ignorována (duplicitní) nebo čekat (souběh nebo případy prioritního zpracování dat) na obslužení. Každá úloha má definovanou identitu, vstupní a výstupní bod, které mohou být přidělovány dynamicky (nebo distribuovaně) stejně jako adresář pro dočasná data nebo logy. Je-li detekována existence dat, je vyvolána odpovídající událost pro jejich zpracování. Formálně události¹ procesů vždy začínají klíčovým slovem `processing` a následuje identifikace vstupního bodu a identity. Události jsou uloženy v systému nebo domovském adresáři identity (`/etc/mre/event.d/`, `conf/event.d/`). Před a po události jsou volány spouštěče, kterých může být pro jednu událost více. Definice spouštěče před událostí je v `/etc/mre/trigger.before.d/` nebo u identity (`./conf/trigger.before.d/`). Po dokončení události se jedná o spouštěče v `/etc/mre/trigger.after.d/` nebo u identity (`./conf/trigger.after.d/`). Spouštěč² má vždy prefix shodně s názvem události. Za tečkou následuje dvoumístné pořadové číslo doplněné o název identifikující spouštěč. Implementace spouštěčů umožňuje např. filtrovat a přesouvat data mezi vstupními body i přes různé identity (pokud je povoleno). Data z výstupního bodu jsou po dokončení

¹Název události `processing.received.ident1.raw`.

²Název spouštěče `processing.received.ident1.raw.10-forward-ident2`.

úlohy archivována, ale mohou být i přímo použita jako vstupní bod pro další úlohu.

8.1.4 Metadata

Pro metadata jsou použity *OpenLink Virtuoso*³ a *Apache Jena Fuseki*⁴. Preferovaný způsob přístupu je prostřednictvím SPARQL Endpointu⁵. Ve specifických případech lze využít i JDBC. Pro správu metadat a ontologického modelu v *OpenLink Virtuoso* jsou určeny připravené skripty pro prvotní inicializaci prefixů a jmenných prostorů (`mreadm-database-init-namespaces`), definice pravidel pro odvozování (`mreadm-database-init-ruleset`), nebo manipulace s grafy – zjištění velikosti (`mredata-graph-size`, `mredata-graph-size-server`), odstranění grafu (`mreadm-graph-drop`). Ontologie se mohou měnit v čase, a proto musí být možnost graf s ontologiemi aktualizovat (`mreadm-update-ontology_dataset`).

8.2 Proces zpracování nových dat

Na obrázku 8.2 je *proces zpracování nových dat* jako předpis posloupnosti úloh. Existence dat ve vnějším vstupním bodě S1 znamená vznik události a spuštění procesu platformou MRE. Pro některé body může být proces blokován až do okamžiku vytvoření speciálního souboru, který indikuje připravenost nebo úplnost dat. První krok procesu je *Extrakce*, kde jsou získána metadata nutná pro prvotní kontrolu dat. Blok *Kontrola* je významným vstupním filtrem provádějícím akceptaci dat na vstupu. Neakceptovaná data (nepodporovaný formát/verze, poškozená nebo nekonzistentní) mohou být přímo odstraněna, anebo přejít do karantény, kde rozhoduje odpovědný uživatel. Zdroj může produkovat současně data, která patří do různých projektů. Blok *Kontrola* zajišťuje požadované správné zařazení přesunutím dat mezi pojmenovanými vstupními body, protože nelze předpokládat, že by zdroj musel znát členění projektů, které se může v čase měnit. Následuje blok *Ochrana osobních údajů* (kapitola 8.3), který produkuje **sekundární data** uchovávaná v archivu. Následuje ETL proces (kapitola 8.4), který z archivovaných dat získá metadata, umístí je do archivu a zpřístupní prostřednictvím SPARQL Endpointu.

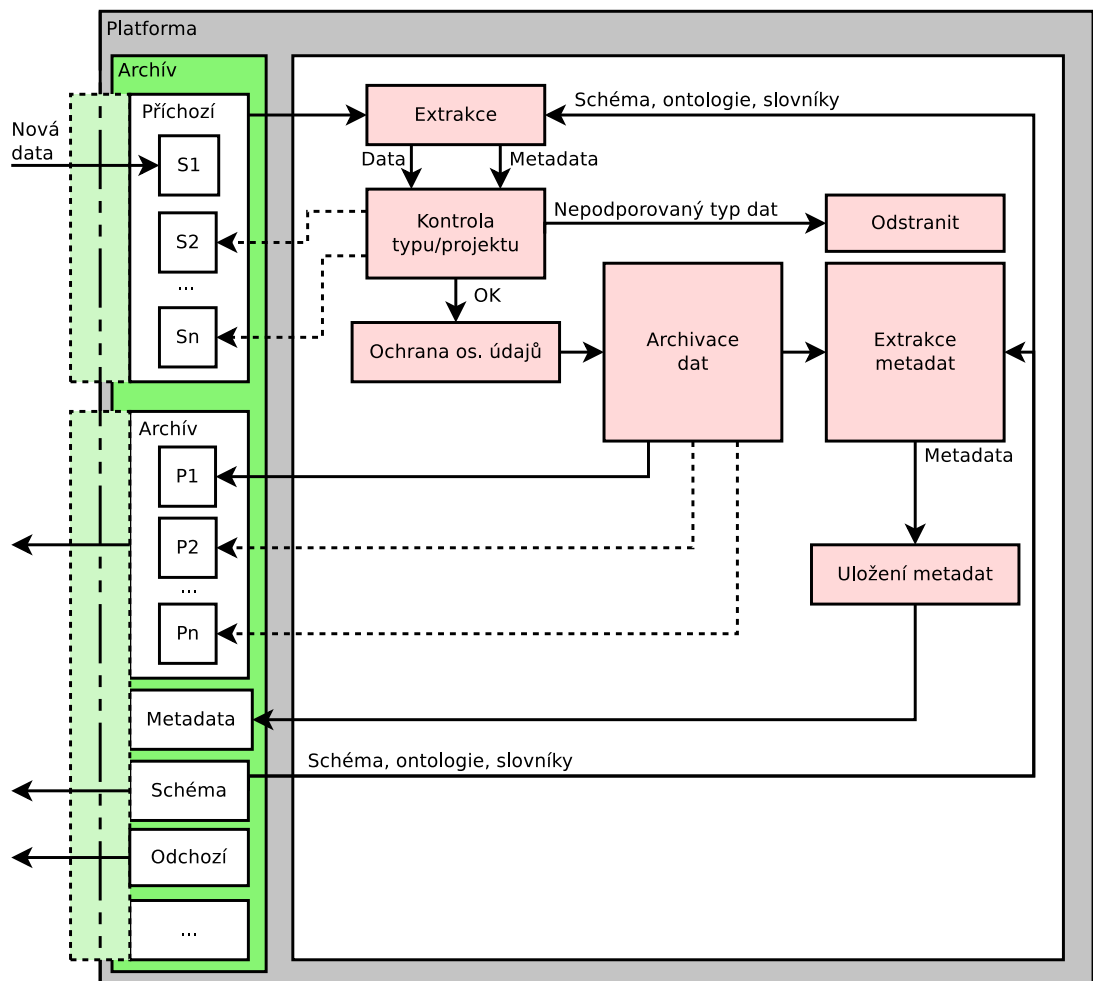
8.3 Ochrana osobních údajů

Ochrana osobních údajů je nezbytnou součástí procesu zpracování nových dat (viz kapitola 8.2), ale nástroje mohou být použity i samostatně pro dodatečnou modifikaci dat a metadat. Používají se metody anonymizace, de-identifikace nebo pseudonymizace dat v závislosti na požadavcích. Důležité je nepoškodit data včetně anatomických struktur a zachovat vazby, které byly v původních datech. Informační systém zdravotnického zařízení exportuje primární data na vnější vstupní bod platformy S1. Data akceptovaná v bloku *Kontrola*, dle diagramu na obrázku 8.2, se dostávají k procesu zajišťujícím ochranu osobních údajů, který je vysvětlen

³<https://github.com/openlink/virtuoso-opensource>

⁴<https://jena.apache.org/documentation/fuseki2/>

⁵SPARQL Endpoint např. <https://mre.zcu.cz/sparql>

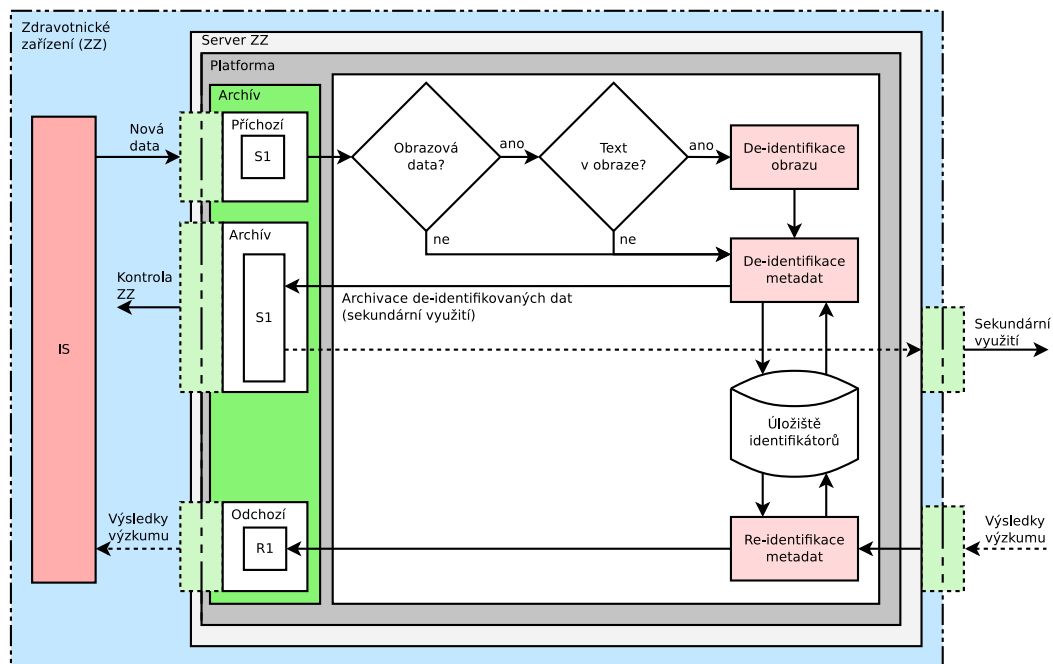


Obrázek 8.2: Proces zpracování dat z vnějšího vstupního bodu

v diagramu na obrázku 8.3. Podle typu dat je posouzeno, zda se jedná o obrazová data (DICOM, kapitola 5.1.2). V obrazových datech je hledán výskyt textu prostřednictvím software *PureImage* (viz kapitola 8.3.2). Následně je provedena ochrana osobních údajů v (semi-)strukturovaných metadatech prostřednictvím software *AnonMed*, který je popsán v kapitole 8.3.1. Snímky, v nichž byly detekovány osobní údaje, přechází do karantény (prevence poškození). Použito je *Úložiště identifikátorů* s generovanými identifikátory. Přiřazen je unikátní identifikátor, který je obvykle používán opakovaně i pro budoucí data stejných osob. Výstupem procesu jsou sekundární data dle specifikovaných pravidel ochrany osobních údajů, které se mohou lišit v závislosti na projektu. Dalším případem užití úložiště pro identifikátory je plánovaná možnost doplnění identity pacienta do dat a výsledků vrácených zpět do zdravotnického zařízení v budoucnu. Blok *Re-identifikace metadat* zajišťuje možnost obnovit původní identifikátory osob.

8.3.1 AnonMed – de-identifikace dat a metadat

Software *AnonMed* (*anonmed*) je od roku 2010 vytvořen pro zajištění ochrany osobních údajů jejich odstraněním nebo nahrazením (de-identifikace, pseudonymizace, šifrování) ve strukturovaných a semi-strukturovaných metadatech pod-



Obrázek 8.3: Proces ochrany osobních údajů

porovaných formátů souborů. Cílem je odstranění přímých a nepřímých identifikátorů osob.

*AnonMed*⁶ je neinteraktivní konzolová a pravidly řízená aplikace. V závislosti na konfiguraci může sloužit jako: (a) nástroj pro akceptaci dat; (b) de-identifikaci; (c) anonymizaci dat; a (d) vytváření tzv. slepých kopií dat např. pro klinické studie. *AnonMed* je konfigurován sadou pravidel tvořících tzv. profil. Podporováno je množství operací s metadaty. K nahrazení za anonymní identifikátor slouží operace IDENTIFICATION. Celé pole lze změnit CHANGE, ponechat prázdné EMPTY, pokud to datový formát umožňuje, nebo zcela odstranit REMOVE. Manipulace s obsahem polí nabízí operace přidání textového řetězce před APPEND_BEFORE nebo za APPEND_AFTER stávající obsah, použít lze podřetězec původního SUBSTRING, provést nahrazení části řetězce jiným SUBSTITUTE nebo převést na velké/malé UPPERCASE/LOWERCASE znaky. Podporováno je šifrování ENCRYPT a DECRYPT. Použita může být externí aplikace EXTERNAL nebo provést operaci SPECIFIC, vztahující se pouze pro daný datový formát. Operace KEEP a NONE slouží k deklaraci, že takový element nebo atribut existuje, ale nedělají nic. Pravidlo KEEP se projeví pro ponechání pole, v případě spuštění *AnonMed* ve striktním režimu, kde jsou odstraněna nejen pole REMOVE, ale i všechna ostatní, která nejsou v profilu uvedena. Příklady použití a profilů jsou uvedeny v dokumentaci⁷. Při tvorbě vlastního profilu je nutné znát datový formát a použít vhodné operace. Po všech provedených změnách musí být vytvořený modifikovaný soubor stále validní. Nevhodná pravidla mohou způsobit poškození vazeb, pokud je dat více a mají vzájemnou souvislost.

Software *AnonMed* [VK10] (2009–2010) poskytuje podporu používaných datových formátů a umožňuje automatizované neinteraktivní provádění de-identifikace

⁶<https://mre.zcu.cz/anonmed/>

⁷<https://mre.zcu.cz/anonmed/profile.html>

se zachováním vazeb mezi daty. Existující nástroje, kterými jsou *DicomCleaner*⁸, *De-Identification GUI*⁹ a *DicomBrowser*¹⁰, neposkytovaly požadovanou funkcionalitu¹¹.

Autorizovaný software *AnonMed* [VK10] je používán v produkčním prostředí od ledna 2010 a o rok později byl publikován v [VK11c]. *AnonMed* pomáhá řešit ochranu osobních údajů dlouhodobě a systematicky. Umožňuje tvorbu slepé kopie dat [SJV⁺15], má praktický přínos v projektu zjištění koncentrace hladiny glukózy v krvi [KKK⁺16a, KKK⁺16b], ale může sloužit i pro ochranu exportovaných dat (doplnění cíle nebo účelu). Aktuálně podporované datové formáty jsou CSV/TSV, DICOM, XML (DASTA), sešity tabulkového procesoru a z nich odvozené specializované formáty například pro data měření hladiny glukózy v krvi [KKK⁺16a] (*Diasend*¹², *Medtronic*¹³). Příklady de-identifikovaných souborů jsou přiloženy na DVD (I, data/data-*/).

8.3.2 PureImage – osobních údaje v obrazových datech

Aplikace *PureImage* (`pureimage`) se zabývá klasifikací DICOM snímků na základě detekce vypálených osobních údajů v obrazové informaci od roku 2016. Použita je adaptivní-iterativní metoda založená na optickém rozpoznávání znaků. Jakýkoliv rozpoznatý textový řetězec vede k přesnějšimu (a náročnějšimu) zpracování v další iteraci. Detaily včetně podrobných výsledků byly publikovány v [VKKK19]. Klasifikace snímků rozlišuje třídy (a) čisté bez textu, (b) s textem a (c) s osobními údaji. Výsledky klasifikace DICOM souborů bez použitelných metadat v hlavičce (metadata byla de-identifikována), vykazují falešně pozitivní míru pod 4,00 % u všech tříd. Pro kritickou třídu osobních údajů je falešně pozitivní míra pod 1,81 %. Dosahovat lze ještě lepších výsledků, protože *PureImage* umí využít existující metadata v hlavičce DICOM souborů. Vhodné je *PureImage* zařadit před *AnonMed*, jak ukazuje diagram na obrázku 8.3. V příloze E jsou zobrazeny dva ilustrační DICOM snímky (E.1 a E.2), které byly klasifikovány do třídy s osobními údaji. *PureImage* má implementovanou podporu odstranění osobních údajů začerněním (přerytí černým obdélníkem). Na obrázcích byla pro zvýraznění použita oranžová/žlutá barva. Klasifikace snímků je v produkční verzi, ale funkce jejich de-identifikace zatím nikoliv, protože je nutné důkladně vyhodnotit kvalitu a přesnost ve vztahu k jednotlivým textovým řetězcům ve snímcích. Posoudit je nutné případnou nežádoucí míru poškození zobrazovaných anatomických struktur. DICOM snímek je v karanténě až do doby posouzení pověřenou osobou, pokud byl klasifikován do jiné třídy než první *čisté bez textu*. Příklady de-identifikovaných snímků jsou v příloze I a na DVD (`screenshots/pureimage/`).

⁸<http://www.dclunie.com/pixelmed/software/webstart/DicomCleanerUsage.html>

⁹<http://www.softpedia.com/get/Science-CAD/De-Identification-GUI.shtml>

¹⁰<https://wiki.xnat.org/xnat-tools/dicombrowser>

¹¹Aplikace s grafickým uživatelským rozhraním nejsou pro automatizaci vhodné. *DicomBrowser* umožňuje v linuxové verzi i dávkovou anonymizaci dat a tvorbu vlastních skriptů ve vlastním jazyce *DicomEdit*, ale pracuje pouze s metadaty v souborech DICOM a odstranění v obrazu vypálené identifikace možné není. V roce 2010 byl v [264] publikován toolkit pro de-identifikaci medicínských obrazových dat.

¹²<https://www-int.glooko.com/>

¹³<https://www.medtronic.com/>

8.4 MetaMed – ETL proces

Od roku 2011 je software *MetaMed*¹⁴ (`metamed`) [VKKR12] určen pro zajištění ETL procesu s využitím ontologií. Umí provádět konverze RDF dat nebo realizovat SPARQL dotazy. Podporovány jsou úložiště v operační paměti a databáze OpenLink Virtuoso. Podpora datových formátů je prostřednictvím pluginů. Hlavním z nich je plugin *MetaMed Extractor* (`metamed-extractor`), který poskytuje rozhraní `Extractor` a abstraktní třídu `AbstractExtractor`. Jejich implementací lze doplnit podporu o požadované verze formátů. Nový formát je potřeba doplnit do `FileTypeDeterminer` z knihovny `mrelib` (viz 8.6.1). U všech rozšíření je kladen maximální důraz na používání ontologií. Anotace s mapováním v ontologii velmi zjednodušují a zrychlují konfiguraci daného rozšíření, protože často stačí modifikovat ontologii. Příklady metadat jsou v příloze I a na DVD (`data/`).

MetaMed Extractor Generic Výchozí plugin pro získání obecných metadat o souboru ze souborového systému. Jedná se o název `nfo:fileName` a cestu `mre:filePath` k souboru, formát `dcterms:format`, velikost `nfo:fileSize` a datum poslední změny `nfo:fileLastModified`. Přidán je hash prostřednictvím `nfo:hashCode` na instanci třídy `nfo:FileHash` s hodnotami `nfo:hashAlgorithm` a `nfo:hashValue`.

MetaMed Extractor DASTA Plugin podporuje formát DASTA ve shodě s *DASTA Ontology* (7.3) včetně číselníků *DASTA Code-lists*. Používány jsou anotace s názvy elementů, atributů nebo XPath z ontologie.

MetaMed Extractor DICOM Extraktor pro DICOM soubory je plně konfigurovatelný prostřednictvím SPARQL dotazu nad *DICOM Ontology* (7.4). Datové vlastnosti popisující elementy získá výběrový dotaz z výpisu D.1 v příloze D.1. RIM získá z ontologie druhý dotaz, který je na výpisu D.2. Výsledná RDF data jsou ve shodě s diagramem na obrázku 7.4 a bez duplicitních metadat.

MetaMed Extractor DirectoryStructure Plugin automaticky rekurzivně generuje hierarchii odpovídající adresářové struktuře od zadaného umístění. Popis adresáře má koncept `nfo:Folder` obsahující identifikátor `dc:identifier` s absolutní cestou a `nfo:belongsToContainer` pro propojení adresářové struktury.

MetaMed Extractor Image Plugin je určen pro získání metadat z obrázků ve formátech JPEG, PNG a TIFF. Autorem je Martin Kryl [260] a využívá *Nepomuk EXIF Ontology*¹⁵ s prefixem `nexif`¹⁶. Definována je *Image Ontology*¹⁷ s prefixem `image` a *Image Ontology Mapping*¹⁸ s prefixem `imm`.

¹⁴<https://mre.zcu.cz/metamed/>

¹⁵<http://oscaf.sourceforge.net/nexif.html>

¹⁶Jmenný prostor <http://www.semanticdesktop.org/ontologies/2007/05/10/nexif#>.

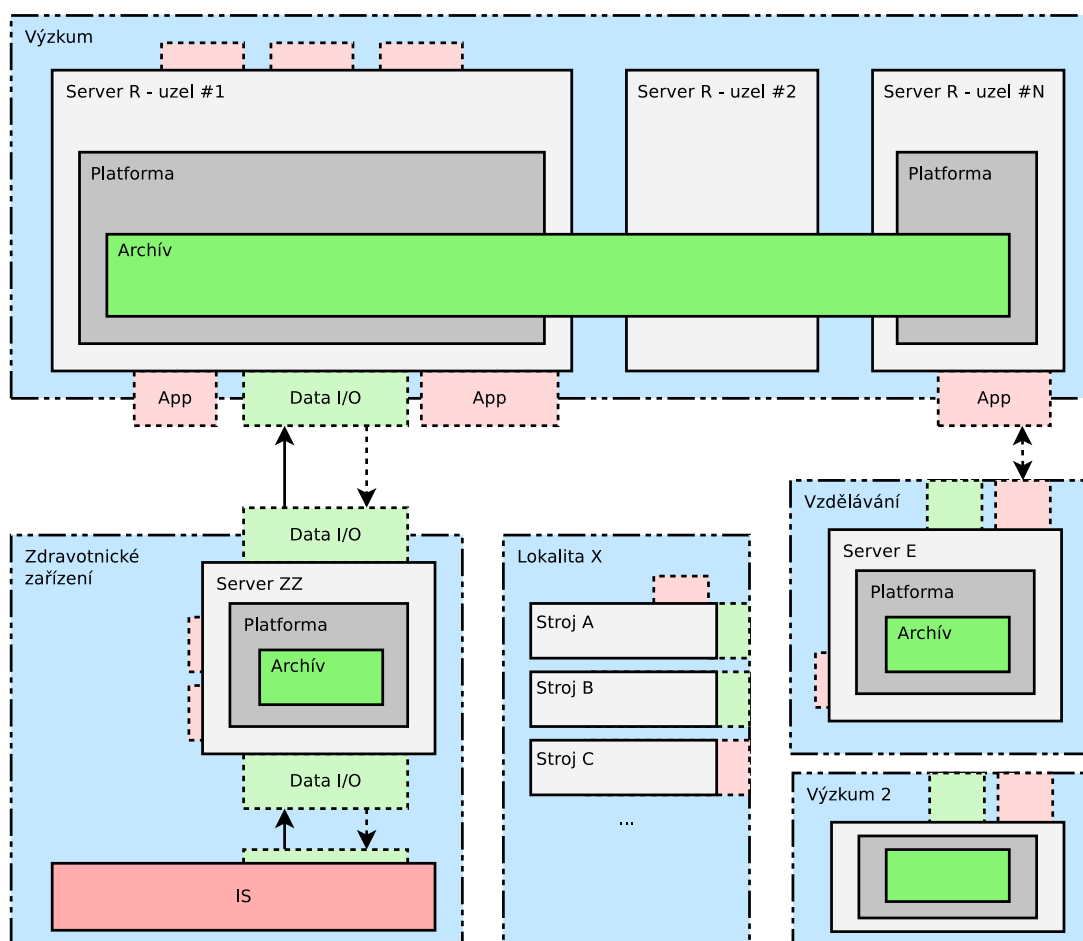
¹⁷<https://mre.zcu.cz/ontology/image.owl>

¹⁸<https://mre.zcu.cz/ontology/image-mapping.owl>

MetaMed Extractor Miscellaneous Plugin podporuje získání metadat na úrovni *MetaMed Extractor Generic* pro soubory jako jsou např. FLAC, MPEG, EXE, Java, PDF, XML nebo ZIP.

MetaMed Extractor Stroke Plugin podporuje formát **StrokeXML** a produkuje metadata v souladu se *Stroke Ontology*. Neaktivní, ale stále dostupný je extraktor používající *SITS Ontology* a *NIHSS Ontology*.

8.5 Interakce s okolím



Obrázek 8.4: Interakce platformy s okolím

Na obrázku 8.4 je diagram a uvažované interakce s okolním světem, kde je důraz na aplikační **App** a datové rozhraní platformy zahrnující vstupní a výstupní body v **Data I/O**. S vnějším aplikačním i datovým rozhraním mohou komunikovat samozřejmě jakékoliv jiné informační systémy a zdroje dat. Bloky *Výzkum*, *Zdravotnické zařízení* a *Vzdělávání* ilustrují rozsáhlý informační systém rozptýlený přes více lokalit. Blok *Výzkum* je zjednodušený pohled na výzkumnou organizaci, kde je prováděn výzkum. Archiv je vyobrazen s přesahem přes více serverů/výpočetních uzlů, které mohou být vhodné pro rozložení zátěže. Blok *Zdravotnické zařízení* obsahuje instanci platformy řešící příjem dat z jiného informačního systému a ochranu osobních údajů s následným přenosem do výzkumné

organizace. Sekundární data jsou ze zdravotnického zařízení předávána zabezpečeně na vnější vstupní bod instance *Výzkum*. Informační systém by mohl data exportovat přímo, ale je kladen důraz na fyzické oddělení, kam se mohou dostat primární data. Čárkovane je naznačena možná cesta výsledků zpět lékařům do zdravotnického zařízení. Propojení bloků *Výzkum* a *Vzdělávání* je použito na základě aplikačního rozhraní *App*, které může poskytovat například kurzy a příklady odvozené ze sekundárních dat. Bloky *Lokalita X* a *Výzkum 2* zastupují možnosti dalšího budoucího rozšiřování a využití. Další instance mohou fungovat nezávisle, ale nic nebrání ani jejich budoucímu libovolnému propojení.

8.6 Software vs. standardizace metadat

Metadata a ontologie jsou dostupná prostřednictvím SPARQL Endpointu. Graf s ontologiemi je <https://mre.zcu.cz/dataset/ontology>. Pro ostatní grafy se používá schéma <https://mre.zcu.cz/dataset/<identity>/<type>>, kde jsou identita `<identity>` a typ dat `<type>`. V kapitole jsou popsány knihovny, software a webové aplikace, které poskytují možnosti standardizace na aplikační úrovni s využitím ontologického popisu.

8.6.1 MRE Library

Knihovna *MRE Library* (`mrelib`) v jazyce Java implementuje funkcionalitu pro základní nástroje platformy. Poskytuje třídy a metody pro (a) detekci podporovaných formátů souborů `FileTypeDeterminer`, (b) mapu používaných jmenných prostorů včetně prefixů `PrefixMap`, (c) abstraktní třídy `Farmer` a `Worker` pro podporu paralelního zpracování dat (d) připojení k databázi OpenLink Virtuoso (e) předdefinovaný `XMLParser`, (f) šifrování `CipherService` a (g) další nástroje jako jsou manipulace s datem `DateUtil`, soubory `FileUtil`, tvorbou kontrolních součtů `HashGenerator` nebo RDF. Knihovna poskytuje metody přístupu k RDF datům `RDFDataAccess` včetně dotazování `RDFQuery`.

8.6.2 MRE Ontology

Knihovna *MRE Ontology* (`ontology`) v jazyce Java poskytuje jmenné prostory a URI/IRI součásti ontologií používaných v rámci platformy dle kapitoly 7. Prostřednictvím knihovny je zajištěno (a) vytvoření HTML dokumentace ontologií (`owl2doc`), (b) automatické vygenerování Java tříd s identifikátory představující definovaná URI tříd i vlastností a (c) publikace ontologií včetně dokumentace¹⁹.

8.6.3 MRE Core

Knihovna *MRE Core* (`mrecore`) je rozšířením *MRE Library* o třídy a metody pro webové aplikace a služby. Využívá knihovny *MRE Ontology*, kterou rozšiřuje o možnosti pro přímé využití v aplikacích s uživatelským rozhraním. Poskytovány jsou data nebo texty popisků vhodné pro uživatelské rozhraní: názvy vlastností a tříd `LabelService`. Slovníky a číselníky zpřístupňuje `VocabularyService`, ale

¹⁹<https://mre.zcu.cz/ontology/ontologies.html>

nechybí `OntologyService` zprostředkovávající přístup k ontologiím, obecná datová služba `DataService` nebo `SPARQLBuilder` pro manipulaci s RDF daty. Pro uživatelské rozhraní a přehledové tabulky nebo výpisy slouží implementace backendu `DataTables`²⁰. Pro integraci jazyka Java a RDF jsou anotace pro datové třídy (Plain Old Java Object, POJO) nebo jejich mapování na ontologie. Nechybí podpůrné třídy DAO (`OntologyRepository` a `SPARQLRepository`), kontroléry (např. `AccountController`, `RegistrationController`) nebo konvertory (např. `ResourceToStringConverter`, `StringToResourceConverter`).

8.6.4 Sparkle

*Sparkle*²¹ (`sparkle`) je aplikace grafického SPARQL editoru pro pomoc s tvorbou dotazů. Software vznikl jako studentský projekt, na kterém spolupracovala řada studentů ve formě diplomových prací nebo oborových projektů [265–269]. Nechybí podpora konstrukcí SPARQL 1.0 a 1.1. Uživatelské rozhraní používá JavaFX a poskytuje formulářové prostředí, textový editor se zvýrazněním syntaxe a chyb, které lze vidět v příloze D.2 na obrázcích D.1 a D.2. U obou přístupů lze využít našeptávačů nebo provést tvorbu dotazu graficky a vizualizovat jej jako graf. Aplikace poskytuje správce jmenných prostorů s prefixy (nebo alternativním URL pro stažení). Třídy a vlastnosti jsou zobrazeny formou identifikátorů (URI/IRI) nebo použitím textových názvů/titulků z `rdfs:label`, `dc:title` nebo `skos:prefLabel`. Hotové stejně jako rozpracované dotazy lze uložit/načíst prostřednictvím vlastního formátu souboru nebo je přímo vyzkoušet a výsledky si v aplikaci prohlédnout. Podporovány jsou čtyři úložiště: (a) v operační paměti, (b) v souborovém systému, (c) databáze OpenLink Virtuoso a (d) SPARQL Endpoint.

Část *SPARQL Visual Query Builder* je dostupný i samostatně²². Snímky aplikace jsou v příloze D.2, kde je na obrázku D.3 příklad vizualizace dotazu, a zobrazené výsledky ukazuje obrázek D.4, kde sloupec `CT_inTimepoint` obsahuje srozumitelný text a nikoliv jen URI/IRI zdroje. Další snímky jsou na DVD v adresáři `screenshots/sparkle/`.

8.6.5 Query Builder

*Query Builder*²³ publikovaný v [VKK18] byl předchůdcem části *SPARQL Visual Query Builder* ze Sparkle. Cílem bylo poskytnout jednoduché uživatelské rozhraní, které by mohli využívat přímo lékaři bez větších znalostí technologií sémantického webu a zajistilo by automatickou tvorbu dotazu podle ontologie umožňující získání dat ze SPARQL Endpointu. Lékaři potřebují získaná data vyhodnocovat, statisticky zpracovávat nebo testovat hypotézy. Příklad užití lze popsat tak, že lékař vybere třídy/koncepty a jejich atributy, z nichž bude na základě ontologie automaticky vygenerován výběrový dotaz ve SPARQL. Výsledná data si lékař stáhne v CSV nebo XLS pro další analýzu. Původně bylo počítáno

²⁰<https://datatables.net/>

²¹<https://mre.zcu.cz/sparkle/>

²²<https://mre.zcu.cz/sparkle-webview/>

²³Query Builder je dostupný na adrese <https://medical.zcu.cz/qbuilder>, ale SPARQL Endpoint je přístupný pouze vybraným uživatelům.

s filtrováním, seskupováním nebo agregacemi dat, ale po uživateli by to vyžadovalo znalost SPARQL nebo alespoň povědomí o SQL. Cílová skupina uživatelů využívá zpracování ve statistickém software nebo v tabulkovém procesoru, kde lze v případě potřeby filtrování i agregace realizovat.

Ve webové aplikaci je plně využito *Stroke Ontology* (kapitola 7.7) pro specifickou doménu s temporálními daty, z níž se uživateli nabízí třídy a vlastnosti odpovídající vybranému časovému období. Pro provedení dotazu lze nastavit SPARQL Endpoint a výsledky zobrazit v HTML, stáhnout jako CSV nebo soubor tabulkového procesoru. V příloze D.4 je na obrázku D.8 ukázka uživatelského rozhraní. Na DVD viz příloha I jsou datový model, příklad dat i výsledný dotaz (`data/query-builder/`) a snímky obrazovky (`screenshots/qbuilder/`).

8.6.6 Web Form Generator

Tvorba, zobrazení, zpracování a použití RDF dat je rozšířené. Změnám RDF dat, jako jsou třeba aktualizace v průběhu času nebo opravy, není věnována dostatečná pozornost. Tvorba strukturovaných sémantických dat, které mohou být kontrolovány a validovány, je významným prvkem pro jejich následné použití a sdílení. *Web Form Generator* [VKKK17] je prototypem webové aplikace s cílem umožnit sběr a editaci RDF dat prostřednictvím webových formulářů generovaných automaticky na základě ontologií. Jedná se o praktické využití ontologie pro tvorbu uživatelského rozhraní. Všechny potřebné informace nelze vždy jednoduše odvodit z ontologického jazyka a použitých konstrukcí. Potřebné anotace lze doplnit využitím *Form Ontology* (kapitola 7.5) v ontologiích. Princip fungování je uveden v diagramu na obrázku D.5. Vstupem jsou ontologie, anotace dle *Form Ontology*, a potřeba je i přístup k datům, pokud chceme mít možnost je editovat. Pro vytvoření nových dat stačí zadat IRI třídy a následně správně vyplnit automaticky vygenerovaný formulář. Příklady formulářů jsou vidět v příloze D.3 na obrázcích D.6 a D.7. Příklad příložený na DVD je v adresáři `data/web-form-generator/`.

8.6.7 Databáze tkáňových vzorků

Databáze tkáňových vzorků je webová aplikace s uživatelským rozhraním založená na *TDB Ontology* dle kapitoly 7.9. Aplikaci mají používat lékaři také z dalších oddělení. Konkrétní identifikaci pacientů mohou vidět a editovat pouze ošetřující lékaři. Interně je identifikace pacientů zašifrována. Hlavním prvkem uživatelského rozhraní je přehledová tabulka vzorků, v níž je potřeba vyhledávat. Pro realizaci byla použita implementace *DataTables* pluginu pro *jQuery*. Propojení ontologie a RDF dat bylo realizováno prostřednictvím Java tříd s anotacemi a *MRE Core* (8.6.3). Třídy a proměnné byly doplněny o anotace `@MREDataOntology`, `@MREDataURIBaseName`, `@MREDataPropertyName` včetně integritních omezení (balík `javax.validation.constraints` s anotacemi `@NotNull` nebo `@Size`). Obrázek D.9 ukazuje výslednou tabulku, která je realizována prostřednictvím *DataTables* a backendem využívajícím RDF data. Názvy sloupců a způsob zobrazení odpovídají metadatům a anotacím v *TDB Ontology*. Používán je SPARQL Endpointem z *Apache Jena Fuseki*²⁴.

²⁴<https://jena.apache.org/documentation/fuseki2/>

9 Ověření výsledků a zhodnocení

Hodnocení metrik pro klíčové požadavky je uvedeno v příloze H, kde tabulka uvádí hodnocení na škále *ano*, *ne* a *částečně* s uvedením detailu v závorce.

Ontologie Vyhovují požadavkům pro modelování řešené domény s popisem konceptů, vlastností, vztahů a používaných číselníků. Ontologie byly validovány v Protégé s použitím Fact++ (v1.6.5), HermiT (v1.4.3.456) a Pellet (v2.2.0). Navržené a implementované ontologie jsou dlouhodobě použity pro metadata v ostrém provozu, a proto již byly nedostatky eliminovány. Prováděné změny jsou v souladu s předpokládaným vývojem a změnami požadavků v popisované doméně.

Způsobem ověření návrhu/implementace je použití a provedená integrace existujících i implementovaných ontologií (RDF slovníků) s výzkumnou platformou MRE, dalším vytvořeným softwarem nebo zpracováním (dotazováním) získaných metadat. V praxi jsou ontologie používány pro konfiguraci v nástrojích jako jsou extraktory pro software *MetaMed (DASTA Ontology, DICOM Ontology a Stroke Ontology)* i ve vytvořených webových aplikacích *Query Builder* (8.6.5) a *Web Form Generator* (8.6.6, využívá *Form Ontology* viz 7.5).

Pro výzkum existují specifické a na ontologiích založené informační systémy. Příkladem je *Databáze tkáňových vzorků* (8.6.7) nebo webová aplikace prospektivní databáze pro výzkum idiopatických střevních zánětů, kde jsou realizovány požadavky lékařů na vizualizaci, zadávání, kontrolu, ale i dotazování a analyzování dat. Výsledky prospektivní studie idiopatických střevních zánětů za období 2000–2015 byly publikovány v [SSC⁺17] s využitím *IBDS Ontology* a metadat o populaci prostřednictvím *Population Ontology* (7.6). *Stroke Ontology* modeluje doménu pacienta po cévní mozkové příhodě zahrnující diagnostiku a sledování jeho stavu. Jedná se o temporální data, nad kterými provádí lékaři další analýzy. Využít mohou webové aplikace *Query Builder* pro zpřístupnění dat i bez znalosti OWL a RDF.

V příloze F jsou uvedeny metriky ontologií, které zahrnují počty tvrzení a úroveň expresivity ontologie (příp. RDF slovníku). Mezi počty patří celkový počet tvrzení a počty tříd, vlastností (anotace, datové, objektové) a individuí. Nechybí počty tvrzení o disjunkci, ekvivalenci, inverzi nebo vztazích rodič/potomek pro třídy a vlastnosti.

Server a datové úložiště Výzkumná platforma MRE včetně archivu je na serveru `mre.zcu.cz`, který je provozován ve virtualizovaném prostředí prostřednictvím *Kernel-based Virtual Machine* na Katedře informatiky a výpočetní techniky. Hardwarové parametry serveru jsou 4x vyhrazený Intel(R) Xeon(R) CPU E5-4620 v2 @ 2,60 GHz (5 199,99 bogomips), 16 GB RAM a 1 GB swap. Operační systém *Debian GNU/Linux* ze stabilní větve s jádrem 4.9.0-7-amd64. Archiv dat je umístěn na síťových discích diskového pole, které jsou k serveru připojeny prostřednictvím *Distributed Replicated Block Device* (DRBD) a *Internet Small Computer System Interface* (iSCSI). Server je s disky spojen rychlostí přes 20 Gbps. Pro data archivu jsou aktuálně používány čtyři disky s hardwarovým RAID 10 o kapacitě 5,9 TB (2011), 1,8 TB (2018), 1,5 TB (2019) a 1 TB (2020). Disková

kapacita celkem 10,2 TB, z níž je obsazeno 8,1 TB. Sekundární data zabírají 71,46 % (7,3 TB). Navyšování je prováděno operativně v případě potřeby. Data jsou zálohována denně na úrovni souborového systému a kompletní obraz disků je zálohován třikrát týdně. Rozsah dat a metadat je uveden v tabulce 9.1 ve druhém a čtvrtém sloupci. Mezi XML data patří právě DASTA a obecný XML soubor označovaný **StrokeXML** pro další strukturovaná klinická data nepodporovaná v DASTA.

Metadata jsou spravována v úložišti *OpenLink Virtuoso Open-Source Edition*¹ s databází umístěnou na síťovém SSD disku s RAID 10 o kapacitě 100 GB. Rozsah metadat je pro jednotlivé projekty a typy dat uveden v tabulce 9.1 ve třetím a pátém sloupci.

Projekt	XML	XML metadata	DICOM	DICOM metadata
CMP-img	6 966	181 803	12 783 678	754 125 417
CMP-stroke	376 432	3 221 306	0	0
IBD-S	12 478	3 096 863	0	0
Liver	191	10 415	576 302	33 852 901
Celkem	396 067	6 510 387	13 359 980	787 978 318

Tabulka 9.1: Rozsah dat a metadat archivu dle projektu a typu. Sloupce *DICOM* a *XML* představují počet souborů. Ve sloupcích s metadaty se jedná o počet RDF trojic.

Ochrana osobních údajů Od počátku řešení (2010) je nedílnou součástí platformy zajištění ochrany osobních údajů a soulad s platnou legislativou včetně zákona č. 110/2019 Sb. upravující zpracování takových údajů dle nařízení Evropského parlamentu a Rady (EU) 2016/679. První příspěvek k problematice byl publikován v [VKR11]. Definovaný způsob a nástroje se stanovenými pravidly pro ochranu osobních údajů se ukázaly být s odstupem času zcela legitimní a správné. Zajištěna je nejen ochrana pacientů, ale současně lékařského i nelékařského personálu a zdravotnického zařízení. Architektura platformy umožňuje provést zajištění ochrany osobních údajů ještě na straně zdravotnického zařízení.

Navržený model dle obrázku 8.2 v kapitole 8.2 je univerzální. Umožňuje realizovat akceptaci pouze podporovaných datových formátů bezprostředně na vstupu po přijetí dat. Nepodporované soubory jsou přesunuty do karantény a nedojde k jejich automatickému postoupení ke zpracování. Mohou být akceptovány a dále zpracovány až po doplnění podpory daného formátu. Proces zajišťující ochranu osobních údajů dle kapitoly 8.3 je flexibilní a jednotlivé části lze modifikovat dle datových formátů a požadavků na výzkum v jednotlivých projektech, aby byly zachovány potřebné informace byť například v pseudonymizované podobě. Dedikovaný server platformy se zaměřením na ochranu osobních údajů ve zdravotnickém zařízení je 8x Intel(R) Core(TM) i7 CPU 950 @ 3,07 GHz (6 147,21 bogomips), 12 GB RAM, 24 GB swap, a 2x 4 TB HDD pro data lokálního archivu. Pseudonymizace a šifrování umožňují ponechat v datech důležité informace, které mohou pomoci s určením vazeb mezi daty nebo následně sloužit k návratu výsledků

¹<https://github.com/openlink/virtuoso-opensource>

do rukou lékařů a zdravotnického zařízení. K případnému obnovení identifikace slouží také aplikace *AnonMed*. Část ochrany osobních údajů je možné aplikovat pro tvorbu tzv. slepé kopie dat sloužící pro různé studie, jak bylo realizováno pro [SJV⁺15]. S velmi přísnými pravidly může *AnonMed* produkovat anonymní data. Do budoucna je předpoklad doplnit další vhodné způsoby modifikace dat právě pro použití při tvorbě studií.

Pro identifikaci snímků obrazových vyšetření, kde může být zapsána textová informace nebo citlivé údaje přímo v obraze je určena aplikace *PureImage* (8.3.2) [VKKK19]. Časová náročnost identifikace textu je uvedena v tabulce 9.2 jako řádek *Klasifikace snímků*. Potřebný čas reálně závisí na rozlišení snímků, typu vyšetření a použité konfiguraci aplikace. Publikovaný článek [VKKK19] uvádí, že v datové sadě s 7 138 750 DICOM soubory se snímek s textem v obraze vyskytoval v 0,33 % (23 872) případech. Jakýkoliv osobní údaj byl v této datové sadě detekován u 0,23 % (16 279) souborů. DICOM snímky po provedení de-identifikace vypálených osobních údajů v obrazové části jsou uvedeny v příloze E. Právě část zajišťující de-identifikaci nestrukturovaných osobních údajů v obraze ještě bude vyžadovat další otestování a vyhodnocení.

V závislosti na účelu dat a metadat je nutné některé z nich stále považovat za citlivé a to i v datech sekundárních, které jsou na první pohled anonymní. Zajištění zabezpečení přístupu k datům se týká problematika k-anonymity dle [225, 226]. Zejména budou-li sdíleny nebo publikovány, anebo mohou být dále analyzovány, filtrovány nebo doplněny o data z dalšího zdroje. Příkladem je datum narození nebo poštovní směrovací číslo, které mohou, v malém vzorku dat, identifikovat konkrétní fyzickou osobu. Řešením může být pseudonymizace, odvození jiné hodnoty (věk) nebo různé formy agregace a seskupování do obsáhlejších kategorií².

Proces zpracování dat a metadat Jednotlivé měřené části zpracování dat jsou uvedeny v tabulce 9.2 a byly získány na základě provozních logů v období od září do listopadu 2019.

Příchozí data Nová data určená pro výzkumné účely jsou nahrávána do vnějšího vstupního bodu platformy ve zdravotnickém zařízení. Informační systém zdravotnického zařízení exportuje data rychlostí 1 020–2 242 kB/s s průměrem **1 618 kB/s** a mediánem 1 725 kB/s, jak uvádí tabulka 9.2. Maximum odpovídá rychlosti 17,5 Mbps (7,7 GB/hod.).

Zpracování příchozích dat Ve druhém a třetím řádku tabulky 9.2 jsou uvedeny počty souborů přenesených celkově za hodinu a pouze v jedné iteraci, která zajišťuje ochranu osobních údajů. Iterace je prováděna pravidelně každé dvě minuty a řeší akceptaci dat současně s jejich de-identifikací. Klasifikace snímků je v testovacím provozu a detekuje snímky s výskytem textu v obrazové informaci. Provedení může být spuštěno na základě metadat i jen pro specifické modality nebo typy snímků. Následuje de-identifikace dat použitím software *AnonMed*, která trvá v průměru **5 ms** pro DICOM a **50 ms** u XML souboru. Po de-

²Místo původního data narození tak bude k dispozici např. pouze věk nebo věková kategorie. Náhradou za směrovací čísla může být nahrazení označením kraje.

Úloha	Typ dat	Min	Max	Průměr	Medián
Export ze ZZ [kB/s]	celkem	1 020	2 242	1 618	1 725
Data – celkem [počet/hod.]	DICOM	131	13 927	3 777	1 771
	XML	4	128	46	40
Data – iterace [počet]	DICOM	11	690	400	405
	XML	1	78	27	26
Klasifikace snímků [ms] [VKKK19]	DICOM	51	99 587	1 146	969
De-identifikace [ms]	DICOM	4	42	5	4
	XML	15	2 125	50	31
Rychlost přenosu [kB/s]	celkem	1 978	9 599	7 320	7 346
Extrakce, transformace a import [ms]	DASTA	1	9 865	142	9
	XML	1	792	16	16
Extrakce a transformace [ms]	DICOM	5	72	12	9
Persistence metadat [ms/iter.]	DICOM	259	4 960	893	833
Kompresa metadat [s/iter.]		0	14	1	0
Import metadat [s/iter.]		0	91	8	7

Tabulka 9.2: Rychlosti zpracování metadat.

identifikaci jsou vytvořena sekundární data umístěna do archivu pro výzkumné účely.

Přenos de-identifikovaných dat pro výzkumné účely Pouze předávání úspěšně de-identifikovaných dat probíhá v pravidelném pětiminutovém intervalu na server MRE prostřednictvím komunikace zabezpečené šifrováním. Průměr přenosové rychlosti je **7 320 kB/s** (57 Mbps).

Kontrola nových dat ve vstupních bodech Platforma na serveru MRE v pravidelném minutovém intervalu kontroluje existenci nových dat ve všech vnějších vstupních bodech. Pro každý vstupní bod (např. **raw**, **rdf**) je definován proces zpracování dat specifický pro identitu (uživatel, projekt) nebo typ dat.

Metadata – ETL proces XML souborů Pro formáty XML je import metadat převeden přímo do odpovídajícího grafu v *OpenLink Virtuoso* a zpracování probíhá sekvenčně v jediném vlákně. Pro formát DASTA je ETL prostřednictvím aplikace *MetaMed* rychlostí od 1–9 865 ms na soubor. Maximální hodnoty bylo dosaženo pravděpodobně v důsledku vyšší zátěže na serveru, protože průměrná hodnota je **142 ms** a medián 9 ms na soubor ve formátu DASTA. Právě u **StrokeXML** je možnost modifikovat výchozí proces zpracování příchozích dat důležitý. Soubory **StrokeXML** společně s DASTA souborem je nutné zpracovávat v pořadí jejich přijetí, protože obsahují průběh stavu a léčby pacienta. Čas zpracování obou souborů současně se pohybuje celkově v rozmezí 1–792 ms se stejnou hodnotou **16 ms** pro průměr i medián.

Uvedené časy jsou oproti modelu v operační paměti vyšší kvůli komunikaci s *OpenLink Virtuoso*.

Metadata – ETL proces DICOM souborů DICOM souborů je obrovské množství a každý obsahuje všechna potřebná metadata. Výsledkem optimalizace ETL procesu pro tento formát souboru je: (1) paralelní zpracování ve více vláknech pro extrakci a transformaci do RDF modelu v operační paměti, (2) perzistence RDF modelu do souboru ve formátu N-TRIPLE, (3) komprese RDF dat, (4) import do pojmenovaného grafu *OpenLink Virtuoso* a (5) archivace RDF souboru v archivu. Rychlost extrakce metadat se pohybuje v intervalu 5–72 ms s průměrem **12 ms** a mediánem 9 ms na DICOM soubor. Další hodnoty se vztahují ke všem současně zpracovávaným souborům, jejichž počet je omezen maximálně na 10 000 souborů pro jednu iteraci. Export modelu ve formátu N-TRIPLE se pohybuje mezi 259–4 960 ms s průměrem **893 ms** a mediánem 833 ms. Čas komprese je mezi 0–14 s a průměrnou hodnotou **1 s**. Import komprimovaného RDF souboru do *OpenLink Virtuoso* je v intervalu 0–91 s a průměr **8 s**, medián 7 s.

Import metadat v RDF do OpenLink Virtuoso Pro ověření úložiště *OpenLink Virtuoso* na serveru MRE byl měřen čas nutný pro naplnění grafů dle jednotlivých identit. Tabulka 9.3 ukazuje výsledné časy nutné pro import uvedeného množství RDF trojic do grafu. Rozlišován je čas (a) pro vlastní import RDF dat a (b) pro indexování. V posledním sloupci je uvedena dosažená rychlost. Druhý řádek tabulky ukazuje výsledek pro graf o velikosti 754M trojic, kde byla rychlost importu 132 721 RDF trojic/s. Poslední řádek ukazuje součet časů představující plnění všech grafů současně a průměrnou rychlost importu **134 886 RDF trojic/s**.

Identita	Graf	Velikost [tpl]	Import [s]	Index [s]	Celkem [s]	Rychlost [tpl/s]
mre	ont.	9 864 823	00:01:00	00:00:22	00:01:22	323 158
CMP-img	dcm	754 251 713	01:24:59	00:09:44	01:34:43	132 721
CMP-stroke	med.	3 221 306	00:00:24	00:00:09	00:00:33	97 615
Liver	dcm	33 852 899	00:02:14	00:00:10	00:02:24	235 090
Liver	med.	10 415	00:00:00	00:00:02	00:00:02	5 208
Celkem		794 614 999	01:27:58	00:10:13	01:38:11	134 886

Tabulka 9.3: Rychlost importu metadat v RDF

Celkové hodnocení a diskuze Počet souborů DICOM studie se významně liší podle modality, vyšetřované oblasti těla, tloušťce řezu, ale i počtu sérií. V tabulce 9.1 pro projekt Liver vychází v průměru cca 3 000 souborů na DICOM studii (191x DASTA odpovídá stejnému počtu DICOM studií s celkem 576 302 soubory). Pro obrazová vyšetření hlavy je průměrná velikost studie 1 835 DICOM souborů.

Pro jednu iteraci je nastavený limit 10 000 souborů, který zajišťuje plynulé zpracování i v mezních případech, kdy je nutné mimořádně zpracovávat větší objemy dat najednou (obnovení po výpadku). Příklad časové náročnosti celého

procesu zpracování dat od de-identifikace až po možnost použití metadat prostřednictvím SPARQL Endpointu je následující. Při počtu 1x DASTA a souvisejících 3 000 DICOM souborů je velikost 1 617 MB a použitím průměrných hodnot rychlosti zjistíme, že je de-identifikace 50 ms pro DASTA a 15 000 ms pro DICOM (3 000 * 5); přenos 226,20 s; extrakce 142 ms DASTA a 36 000 ms DICOM (3 000 * 12); export modelu 893 ms; komprese 1 s; a import do RDF úložiště 8 s. Kompletní data a metadata klinické události jsou k dispozici za 287,285 s (4m47s).

Koncentrace hladiny glukózy v krvi Software *AnonMed* [VK10] má praktický přínos v projektu zjištění koncentrace hladiny glukózy v krvi [KKK⁺16a, KKK⁺16b], který je dostupný online³. Pro zajištění ochrany osobních údajů uživatelů s diabetem je *AnonMed* používán jako knihovna. Podporuje formáty dat z glukometrů (CSV a XLS).

Rozšiřitelnost platformy Platforma je použita v praxi pro data projektu *Aplikace moderních technologií v medicíně a průmyslu*⁴ (AMTMI). Interdisciplinární tým se dlouhodobě zabývá regenerací jaterní tkáně a pracuje na vytvoření modelů se zaměřením na makroskopický i mikroskopický popis tkáně a mechanických vlastností jako jsou deformace nebo pružnost, které se projevují běžně při dýchání. Používají se obrazová vyšetření oblasti břicha získaná prostřednictvím počítačové tomografie a z experimentálního CT přístroje pro práci s velkými zvířaty nebo data získaná ze vzorků zvířecího jaterního parenchymu i na mikrostrukturální úrovni (μ CT). Modely/výsledky jsou odvozeny ze zdrojových dat v množství datových formátů (DICOM, OBJ, PVSM, PKLZ, STL, VTK, AVI, ale i další). Modelů může být více s ohledem na další faktory, zkoumané parametry a trénování, ladění nebo validaci. Řádky označené *Liver* v tabulce 9.1 uvádí rozsah dat a metadat projektu. Aplikace AnonMed [VK10] umožnila tvorbu slepé kopie dat použitých pro publikaci [SJV⁺15]. Ontologie pro popis modelů vzniká a ve spojení s *Web Form Generator* je možné jednoduše vytvářet strukturovaná metadata uživateli i bez znalosti RDF a OWL. Interdisciplinární tým dlouhodobě využívá možnosti výzkumné platformy MRE pro zpřístupnění dat.

Princip založený na technologiích sémantického webu, propojených dat a platformě MRE byl zahrnut jako součást žádosti o grant na *LiverCloud* podaný v roce 2018. Projekt byl podán ve výzvě Horizon 2020 v konsorciu s celkem 18 účastníky z Belgie, České republiky, Francie, Itálie, Německa, Norska a Španělska. Podmínky výzvy byly splněny, ale projekt nebyl vybrán k financování.

³Uživatelé mohou nahrát svá data z glukometrů anonymně nebo v rámci provedené registrace na adrese <https://diabetes.zcu.cz/>. Bezprostředně po uploadu jsou data automaticky de-identifikována a teprve následně je proveden výpočet koncentrace hladiny glukózy v krvi.

⁴Projekt (reg. č. CZ.02.1.01/0.0/0.0/17_048/0007280, 2018–2022) ve spolupráci Západočeské univerzity v Plzni (ZČU) a Lékařské fakulty v Plzni (LFP) Univerzity Karlovy.

10 Diskuze

Pro standardizaci popisu sekundárních zdravotnických dat a metadat určených pro výzkumné účely byl použit model RDF, ontologický jazyk OWL a princip spojených dat. Standardizovány jsou procesy zahrnující interakci s okolím, ochranu osobních údajů (AnonMed, PureImage), podporu ETL procesu (MetaMed) a archivaci heterogenních dat. Maximální využití ontologií v ETL procesu poskytují pluginy *Metamed Extractor DASTA*, *Metamed Extractor DICOM* a *Metamed Extractor Stroke* v software *MetaMed* [VKKR12]. Ontologie nepomáhají pouze s automatickou extrakcí metadat. Z ontologického jazyka jsou odvozeny další praktické případy užití, které přímo využívají možností a vlastností ontologického popisu pro (a) generátor SPARQL dotazů s temporálními daty a (b) webové formuláře dynamicky generované na základě ontologií. V prvním případě prostřednictvím *Query Builderu* uživatel snadno vybere kombinace třídy, vlastnosti a období, které jsou mu nabízeny na základě jedné nebo více ontologií. Generátor automaticky sestaví dotaz a umožní jej vykonat přímo nad SPARQL Endpointem. Praktické ověření proběhlo v souvislosti s projektem výzkumu cévních mozkových příhod využívající *Stroke Ontology* a bylo publikováno v [VKK18]. Ve druhém případě je použito ontologií a anotací z *Form Ontology* pro sestavení HTML5 formuláře, který umožňuje zobrazení a editaci RDF dat, jak bylo publikováno v [VKKK17]. Přístup může být aplikován při realizaci informačního systému využívajícím sémantické technologie. Na práci s ontologiemi se zaměřuje množství popsaných knihoven a nástrojů. Společně s dalšími studenty byla v rámci kvalifikačních prací realizována aplikace *Sparkle*, která poskytuje formulářový, textový i grafický způsob tvorby dotazu v jazyce SPARQL. Vedle snazšího vytváření dotazů nechybí ani přímá možnost dotaz vyzkoušet proti SPARQL Endpointu nebo úložišti RDF.

Zmíněny jsou existující rozsáhlé informační systémy. Na mezinárodní úrovni Evropské unie se jedná například o CIS, EURODAC, EUROPOL, SIS nebo VIS. V oblasti zdravotnictví existuje doporučení Evropské komise o přeshraniční interoperabilitě systémů elektronické zdravotní dokumentace (EHR) [193] od roku 2008. Zavedení přeshraničních služeb eHealth v České republice vzniká až nyní, jak uvádí dokumenty [194, 195]. Požadavek na přeshraniční interoperabilitu systémů EHR ukazuje souvislosti spojené se standardizací metadat. Národní formát DASTA [198] pro klinická data i laboratorní výsledky je v České republice hluboce zakořeněn (od roku 1997). Implementují jej informační systémy zdravotnických zařízení a laboratoří nebo slouží pro ekonomické účely jako je např. vykazování ve zdravotních pojišťovnách. Za dobu existence sdružení HL7 Česká republika není znám výstup [207]. Nejnovější příspěvek za HL7 ČR z roku 2018 [206] obsahuje jen obecné informace z jiných států. Rozšíření formátu DASTA v letech 2018–2019 o podporu patientského souhrnu a související infrastrukturu pro přeshraniční sdílení EHR [200, 210] uzavírá diskuzi DASTA vs. HL7. Teprve při mezinárodním předávání EHR je realizována transformace do HL7 a to pouze vybrané části povinného patientského souhrnu. Projekt DASTA lze bez pochyb označit za fungující a unikátní. Svým způsobem DASTA chrání konkurenční prostředí v ČR před vstupem dalších významných hráčů na trh nabízející informační systémy nad HL7. Přejít k HL7 je možný, ale rozhodně není jednoduchý ani

levný. Případ DASTA vs. HL7 poukazuje na obecnější komplikace a problémy v souvislosti se standardizací metadat.

Mezi rozsáhlé informační systémy patří množství mezinárodních nebo geograficky rozlehlých informačních systémů. Výzkumnou platformu lze mezi takové systémy zařadit, protože je navržena pro distribuované prostředí s množstvím instancí platformy vyskytujících se ve výzkumných organizacích a zařízeních produkujících data. Pojem velká data (big data) odkazuje dle [224] na „*rostoucí technologickou schopnost shromažďovat a získávat nové znalosti z velkého objemu, množství, rychlosti a rozmanitosti dat*“. V medicínském výzkumu mohou velká data sloužit pro tvorbu modelu nebo zpřesnění stávajících znalostí a algoritmů. Dlouhodobý sběr nebo sledování stavu pacienta a průběhu nemoci, jako byly publikace [242, 243] nebo [SSC⁺17], může přinést důležité poznatky nebo podklady pro změnu způsobu léčby v závislosti na dalších faktorech. Výsledek se nemusí projevit okamžitě. Provádějí se klinické studie nebo výroční zprávy [228]. S odstupem času může být identifikován nový rizikový faktor nebo jiná souvislost s danou diagnózou [227].

Výzkum v oblasti medicíny a zdravotní péče vede na potřebu sekundárního využití dat realizovaného také samostatnými registry, datovými archivy nebo platformami. Do kategorie registrů patří *REgistry of Stroke Care Quality* (RES-Q) [234, 235], *Rick Hansen Spinal Cord Injury Registry* [241–243], nebo *Safe Implementation of Treatments in Stroke* (SITS) [238–240]. Registr je vždy úzce zaměřen na konkrétní doménu. Mezi registry lze zařadit *Databázi tkáňových vzorků* (TDB; 8.6.7). V disertační práci uvedené informační systémy pro idiopatické střevní záněty nebo cévní mozkové příhody mohou sloužit jako prospektivní databáze i pro retrospektivní analýzu dat. Lze zmínit existenci *re3data.org*¹ jako rozcestník na repositáře výzkumných dat. Datové archivy mohou být obecné nebo se specializují na konkrétní obor. Archiv *The Cancer Imaging Archive* [188] nabízí kolekce de-identifikovaných snímků ve formátu DICOM z množství modalit. Archiv *Segmentation of the Liver Competition 2007* [189, 190] poskytuje medicínské snímky pro analýzu za účelem porovnávání různých algoritmů nad stejnými daty. Platformy obvykle poskytují rozsáhlejší možnosti práce s daty. Projekt *Global Research Platform* [241, 255] se orientuje na webově orientovaný sběr strukturovaných dat prostřednictvím formulářů. Rozdíl oproti podobnému projektu *Google Forms* je v umístění/zabezpečení databáze. Formuláře mohou sloužit pro realizaci registru, ale i na klinické pokusy, klinický výzkum, apod. Projekt *CARMEN Virtual Laboratory* [233] fungoval (do roku 2014) skutečně jako virtuální laboratoř sdílející výzkumná data a služby mezi uživateli pro neurofyziologii. Uživatel mohl vytvářet vlastní data a služby, mohl je sdílet, ale existovalo množství omezení [233] včetně limitu na poskytovaný hardware. Prezentovaná výzkumná platforma má podobnost v obou výše popsáných projektech. *CARMEN Virtual Laboratory* poskytovala centralizovanou správu a zpracování dat. Při datově orientovaném způsobu výzkumu je nutné uchovávat data a výsledky, ale i popis jejich vzniku. Platforma MRE je archivem pro heterogenní data, která lze členit nebo oddělovat například podle projektů, a je rozšiřitelná do podoby virtuální laboratoře. Od počátku je prioritou medicínský výzkum, ale nic nebrání využít stejné principy platformy MRE ani v jiných oblastech. V případě knihovny *MRE Core* (8.6.3) a aplikace *Web Form Generator* publikované v [VKKK17] existuje podobnost

¹<https://www.re3data.org/>

s *Global Research Platform*. Budou-li atributy a číselníky realizovány v ontologii, aplikace může sloužit pro sběr strukturovaných dat v souladu s ontologií a uchovávat v RDF. Projekt IASIS² ukazuje podobný směr v oblasti velkých dat a medicíny, konkrétně pro rakovinu plic a Alzheimerovu chorobu.

Standardizace metadat je komplexní problematika silně závislá na konkrétní řešené doméně a míře požadované granularity dat. Kvalitu metadat lze zvýšit, existují-li standardy, tezaury a strukturování termínů, nebo je-li možné vytvořit a sdílet s ostatními vlastní řízený slovník. V delším časovém období je nutné počítat s existencí změn schématu. Příkladem je výzkum cévních mozkových příhod v období 2010–2020. Na prvopočátku byla data stahována z HTML dokumentu v SITS registru, následně bylo možné získat CSV soubor s daty. Později byla výchozím bodem právě výzkumná platforma a do registru byla data importována. Na konci roku 2018 byl realizován přechod k RES-Q, který je aktuálně v řešení členy výzkumné skupiny. Při existenci anotací v *SITS Ontology* mohlo být stahování nebo následně plnění registru automatické. Po realizaci *Stroke Ontology* bylo možné zdrojová data získávaná ve *StrokeXML* zpracovat aplikací *MetaMed* [VKKR12] s pluginem *MetaMed Extraktor Stroke* pro získání odpovídajících metadat do RDF. Pro zpětnou kompatibilitu mohou existovat obě schémata vedle sebe. Důraz na ontologický popis a technologie sémantického webu umožní zjednodušení, kdy po změně publikované ontologie stačí webovou aplikaci restartovat. Takto fungují *Query Builder*, *Web Form Generator* a pluginy *MetaMed* [VKKR12] jako jsou *MetaMed Extraktor DICOM* a *MetaMed Extraktor Stroke*. Na týmech uživatelů přesto zůstává důležitý úkol – doplnit metadata, která nelze odvodit automaticky. Přesto snahou platformy je, aby bylo do budoucna možné takových metadat odvodit co nejvíce (popis aplikací, služeb a plánu jejich spouštění).

Inspirací výzkumné platformy byly referenční modely Open Archival Information System (OAIS, ISO 14721:2012) [116, 117] a Producer-Archive Interface Methodology Abstract Standard (PAIMAS) [118, 119]. Zejména druhý jmenovaný je orientován na role producenta a archivu. Producentem je zdravotnické zařízení nebo jakýkoliv proces/úloha produkující výsledky (data). Uváděný archiv využívá metadat dle ontologického modelu pro dlouhodobé uchování, ochranu a zpřístupnění dat. Obory archivnictví a knihovnictví [102, 125] jsou velmi dobře unifikovanou oblastí z reálného světa, kde je standardizovaný systematický popis jasně ohraničených objektů, jako jsou knihy, periodika nebo další objekty. Způsoby zpřístupnění se omezují na práci s metadaty – funkce jako je vyhledávání autorit, exemplářů ve fondu nebo rezervace a výpůjčky dokumentů. Při řešení výzkumných projektů vyplynou některé požadavky až v průběhu realizace nebo se mohou měnit a zpřesňovat. Každý projekt může používat zcela jiné objekty nebo pouze jejich části. Na komplexnost domény zdravotnictví poukazují analýzy publikací realizované Andargoli et al. (2017) [270] a Gesulga et al. (2017) [271]. Potenciál zlepšení zdravotní péče i existující problémy zmiňují Gesulga et al. [271]. Dle Abril-Gonzalez et al. (2017) [272] dosud některé oblasti zdravotní péče (např. elektronický zubní záznam) nemají podporu v HL7v3 CDA.

Se zdravotními záznamy souvisí zabezpečení dat. Problematice se věnovala Thurston (2014) [273], a přezkoumáním prošly právní a etické otázky, jak publikoval Ben-Assuli (2015) [274], který zmiňuje souvislost mezi aspekty ochrany osobních údajů a nutnými činnostmi zaneprázdněných lékařů. Pro účely medicín-

²<http://project-iasis.eu/>

ského výzkumu se jedná o sekundární využití dat, které byly pořízeny primárně pro stanovení diagnózy, sledování léčby a vývoje stavu pacienta. Vuokko et al. (2017) [275] se věnuje publikacím ve vztahu k možnostem sekundárního používání dat. Lékaři využívají nebo jsou zavázáni vyplňovat množství lokálních nebo mezinárodních registrů s evidencí případů, které jsou často zaměřeny na specifickou zájmovou oblast. Pro tvorbu klinických studií nebo dalších výzkumných projektů nebývá nutné znát skutečnou identitu pacientů. Prvním standardem pro ochranu osobních údajů ve zdravotních záznamech byl *Health Insurance Portability and Accountability Act* [218, 219] známý pod označením HIPAA (1996). Od té doby se věnovalo zabezpečení ochrany soukromí pacientů mnoho úsilí. Fernández-Alemán publikoval rozsáhlou recenzi [221] (2013) existujících řešení problematiky sekundárního využití dat a ochrany osobních údajů. Svým přístupem přispěli Danciu et al. (2014) [222]. Nařízení Evropského parlamentu a Rady Evropské unie č. 2016/679 z dubna 2016 se věnuje ochraně fyzických osob v souvislosti se zpracováním osobních údajů a o volném pohybu těchto údajů známé jako GDPR. Význam GDPR je srovnatelný s HIPAA. Informace o zdraví pacienta patří dle GDPR do zvláštní kategorie osobních údajů [224]. Důsledky GDPR na vědecký výzkum zmiňuje např. [223]. Česká legislativa zareagovala na GDPR prostřednictvím zákona č. 110/2019 Sb.

Na ochranu osobních údajů byl kladen důraz od samého počátku projektu MRE. Právě pro účel de-identifikace dat i možné úplné anonymizace vznikl software *AnonMed* (2010) [VK10]. Pro detekci snímků, kde se vyskytují osobní údaje vypálené v obrazové informaci, slouží *PureImage* (2019) [VKKK19]. V případě tvorby anonymní sady dat je stále nejvhodnější držet se pravidel dle HIPAA, protože GDPR je poměrně všeobecné.

V oblasti medicíny, zdravotní péče nebo zdravotnických aplikací není důraz kladen na sémantický web, ale obecněji na princip propojených dat a související technologie. Nespornou výhodou je sdílení popisu, znalostí a možnost další integrace nejen informací zdravotní péče. Ontologie elegantně řeší rozšiřitelnost datového modelu potřebného pro výzkumné účely na projektech, které nemusí být v době návrhu známy (datové formáty nebo jejich verze). Martin-Rodilla a Gonzalez-Perez (2019) v [276] mapují metadatové přístupy pro konceptuální modelování včetně RDF/OWL. Možnými problémy se zabývali autoři Halpin (2017) [277] a Kirrane et al. (2018) ve článku [278]. Halpin (2017) [277] poukazuje na použití protokolu HTTP vs. HTTPS jako součásti URI/IRI [47, 53, 144, 145]. Lokalizační funkce URI/IRI může spoléhat na infrastrukturu, která zajistí přesměrování, jak doporučuje organizace W3C na blogu³. Tvrzení W3C a Halpin (2017) [277] považují za nejlepší způsob je zachování ekvivalence mezi HTTP a HTTPS v URI/IRI. Přesto tato úprava může způsobit problémy a vést k nekonzistentním výsledkům při odvozování v závislosti na implementaci ekvivalence v URI/IRI. Platforma MRE používá ve všech jmenných prostorech protokol HTTPS, protože se důraz a jeho podpora v posledních letech významně rozšířila.

Zapojení technologií sémantického webu, principu propojených dat nebo RDF a ontologií se věnovalo množství autorů v [279–286]. Využití principu propojených dat (linked data) pro informační systém uvažovali autoři O’Riain et al. [279] pro integraci finančních dat. Nardon a Moura (2004) ve svém příspěvku [280] používají ontologie a deduktivní databázi pro integraci a sdílení informací zdravotní

³<https://www.w3.org/blog/2016/05/https-and-the-semantic-weblinked-data/>

péče. Podobně se vyjadřují Lindemann et al. (2007) v [281] označující RDF jako vhodnou moderní strukturu pro medicínská data. Klinickými studii se sémantickým modelem se věnovali Muecke et al. (2009) v příspěvku [282]. Ve stejném roce (2009) se Bhatt et al. [283] zabývají na ontologiích založenou sémanticky interoperabilní analýzou a získáním informací. V roce 2010 Ciuciu et al. uvádějí případovou studii [284] zaměřenou na systém pro správu obsahu (CMS) doplněný o anotace anatomických dat v obrazových vyšetřeních, kde sémantická informace vychází z ontologie. Autoři Pierce et al. (2012) prezentovali článek [285] na téma sémantické databáze založené na infrastruktuře sémantického webu se zaměřením na klinický výzkum a přehledy pro hodnocení kvality poskytované zdravotní péče. Skupina Semantic Web Health Care and Life Sciences Interest Group⁴ (HCLS) pod W3C vydala koncept příručky *Health Care and Life Science (HCLS) Linked Data Guide* [286] na základě zkušeností při tvorbě *Linked Open Drug Data* (LODD). Důraz byl na možnosti využití sémantických technologií pro tvorbu a sdílení dat zdravotní péče a zdravotních věd. V roce 2018 skupina skončila, ale její činnost pokračuje v komunitní skupině.

Landwehrmeyer et al. (2017) v publikaci [287] o výzkumné platformě pro Huntingtonovu nemoc potvrzují odklon od prostého testování hypotéz k rozsáhlým multimodálním přístupům vyžadující standardizované úsilí a širokou spolupráci, které plně odpovídá výzkumné platformě MRE.

Interoperabilitě se ve světě věnuje více autorů. Z posledního období Mandel et al. (2016) [288] se věnují projektu Substitutable Medical Applications and Reusable Technologies (SMART) spojující RDF a Fast Health Interoperability Resources (FHIR) známé z [289, 290]. FHIR⁵ vychází z HL7, který není kompatibilní s DASTA. Dvojice autorů Jessica Wan-Yi Kuo a Alex Mu-Hsing Kuo (2017) [291] řeší integraci informačních systémů prostřednictvím HL7 a dávají za pravdu FHIR z důvodu snazší implementace a robustnější infrastruktury. Nutné je podotknout, že uváděná integrace se zaměřuje na informační systémy ve zdravotnictví, ale nikoliv přímo pro výzkumné účely.

⁴<https://www.w3.org/2001/sw/hcls/>

⁵<http://hl7.org/fhir/documentation.html>

11 Přínosy práce

Přínosy zasahují do oblastí: (1) integrace, sdílení a znovupoužitelnost dat; (2) rozvoje informačních systémů v oblasti technologií sémantického webu a popisu prostřednictvím ontologií; (3) ekonomické nebo technické praxe; a (4) diagnostiky a léčby pacientů nebo ochrany zdraví obyvatel.

- Doménové ontologie významně přesahují běžný nestrukturovaný popis klinických událostí pro cévní mozkové příhody a idiopatické střevní záněty.
- Platforma MRE standardizuje procesy s cílem eliminovat lidský faktor a ad-hoc řešení. Systematický a automatický přístup umožnil minimalizovat poškození a vznik nekonzistentních dat nebo vazeb mezi nimi, který lze uplatnit i mimo doménu medicínského výzkumu.
- Definován je rámec pro manipulaci s primárními daty na principech: (1) nezasahovat do IS zdroje, (2) nutná akceptace dat, (3) osobní údaje nesmí opustit zdroj a (4) dlouhodobá archivace.
- Pozornost byla věnována nevýhodám metadat, které mohou působit komplikace při ochraně osobních údajů. Software *AnonMed* je pravidly řízený a umožňuje modifikaci metadat v množství datových formátů. Identifikaci nebo začernění míst s osobními údaji ve snímcích diagnostických zobrazovacích metod řeší software *PureImage*.
- Systematické a automatizované provedení ETL procesu pro metadata zabezpečuje software *MetaMed*, který je rozšiřitelný prostřednictvím pluginů nejlépe ve spojení s ontologiemi pro jejich konfiguraci.
- Automatické konstruování výběrových dotazů představuje *Query Builder*, který uživateli zpřístupní komplexní temporální data bez znalosti RDF, OWL a SPARQL.
- Ontologický popis, příp. doplněný o anotace dle *Form Ontology*, poskytuje možnost plně automatického dynamického online generování HTML5 formulářů pro zobrazení, tvorbu a editaci RDF dat (*Web Form Generator*).
- Častý zájem lékařů směřuje k tvorbě registru nebo databáze určené pro sběr a analýzu dat (retrospektivní, prospektivní). Možnou cestou realizace technologického řešení univerzálního informačního systému nebo registru jako *Databáze tkáňových vzorků*, je spojení knihovny *MRE Core*, aplikace *Web Form Generator* a ontologický popis domény s *Form Ontology*.
- Aplikace *Sparkle* umožní tvorbu dotazů v souladu se SPARQL 1.1. K dispozici je formulářový přístup nebo klasický textový editor se zvýrazňováním syntaxe včetně množství našeptávačů.
- *DASTA Ontology* se zaměřuje na datové bloky DASTA, které jsou významné pro medicínský výzkum, a odstraňuje rozlišení verze DASTA. V tomto ohledu může být ontologie interoperabilní alternativou konverze DASTA do HL7.

12 Doporučení

Většina výzkumných projektů v medicíně začíná od jednotlivců a velmi často pro konkrétní úzce zaměřenou doménu. Potřeba standardizace sledovaných dat a metadat včetně systematického přístupu se projeví krátce po překročení hranice, kdy se další rozšiřování nebo spolupráce stává složitou.

- Platforma MRE podporuje dekompozici řešených projektů na jednodušší a jasně ohraničené části, které mohou být zřetězeny. Používat data, metadata a dosažené výsledky opakovaně pro zpracování je vhodné a možné. Pro dlouhodobou udržitelnost výzkumu, výsledků a jednoznačný kontext jejich vzniku je přínosem zachování popisu a vztahu mezi vytvořeným výsledkem, zdrojovými daty, softwarovými prostředky a jejich konfigurací. *MRE Ontology* představuje koncept možného řešení. Detailnímu popisu a implementaci v MRE by měla být věnována pozornost v budoucnu. Popis pomůže dokumentovat požadavky software. Vazba na použitou konfiguraci a běhové prostředí poslouží jako dokumentace, která umožní jednoznačné odlišení a porovnání výsledků mezi sebou.
- Softwarové prostředky s vhodně popsányými typy vstupních a výstupních dat může být cestou pro odvození toku dat a následnou realizaci plánování, spouštění a provádění úloh.
- Formát DASTA pro klinická data a laboratorní výsledky je specifický pro národní prostředí v České republice. Odklon od DASTA není pravděpodobný (eHealth), proto je *DASTA Ontology* variantou, jak zajistit integraci sekundárních dat pro výzkumné účely. V budoucnu je vhodné opakovaně vyhodnotit situaci s transformací do *HL7* nebo *openEHR*, kde bude nezbytná rozsáhlá odborná diskuze jako prevence zanesení nepřesností a chyb.
- Doporučit lze ověření souladu *DICOM Ontology* (2010) s alternativní *Semantic DICOM* (2015) pro možnost nahrazení s důrazem na použitelnost při ETL procesu.
- Z platformy MRE, tvořené *MRE Core*, *Web Form Generator* a *Form Ontology*, lze vycházet pro významnější využití technologií sémantického webu v informačních systémech. Inspirací mohou být aplikace *SPARQL Visual Query Builder*, *Query Builder* nebo *Web Form Generator* pro tvorbu formulářů u dotazníkových šetření se sběrem dat přímo do RDF s možností následného zpracování.

13 Závěr

Disertační práce se zabývala standardizací metadat ve spojení informačních systémů a multidisciplinárního datově orientovaného medicínského výzkumu. Představen byl přehled stavu a současných přístupů ve vztahu k metadatům, jejich výhodám a nevýhodám, standardizaci, technologiím sémantického webu, oblasti zdravotnictví a výzkumu v medicíně. Pojmy interoperabilita a standardizace jsou velmi často skloňovaným tématem na úrovni národní i mezinárodní, a jsou nutným předpokladem pro dosažení schopnosti systémů spolupracovat, sdílet a správně pochopit data nebo informace. Ontologický jazyk OWL (Web Ontology Language) se používá na webu, pro účely umělé inteligence a je hojně rozšířen v biomedicínské informatice.

Standardizace metadat nespočívá pouze v samotném vhodném, jednoznačném a srozumitelném popisu. Neméně důležitá jsou doporučení, systematické postupy a procedury nebo celé procesy, které definují pravidla tvorby, získávání a další manipulace s metadaty.

Formulovány a splněny byly tři cíle, což je prokázáno dlouhodobým provozem s rozsáhlými heterogenními daty na výzkumných projektech, jak uvádí kapitola 9. Standardizace procesů pro medicínský výzkum je realizována prostřednictvím vytvořené platformy *MRE* definující základní pravidla, nástroje a postupy, které lze měnit a rozšiřovat dle potřeb projektů a uživatelů. Pro výzkumné projekty platforma poskytuje sjednocení společných nebo často opakujících se úloh jako jsou: (1) akceptace dat, (2) ochrana osobních údajů, (3) tvorba metadat, (4) ETL, (5) ontologický popis, (6) dlouhodobá archivace a manipulace s daty, výsledky a softwarem, (7) ostatní praktické použití dat, metadat a jejich popisu. Princip procesů s konkrétními úlohami pro zpracování dat lze uplatnit zcela obecně i pro data z jiného oboru. V metadatech zůstávají zachovány nejen původní vazby z dat, ale lze je rozšířit o další vztahy a propojení na data z jiných zdrojů. Druhý cíl představuje doménová ontologie *DAŠTA Ontology* určená k sekundárnímu využití zdravotnických dat používaných v České republice. Třetí cíl souvisel s návrhem a implementací doménových ontologií pro metadata o anamnéze, diagnostice, terapii nebo dispenzarizaci u pacientů po cévní mozkové příhodě (*Stroke Ontology*) nebo s idiopatickými střevními záněty (*IBDS Ontology*, *IBDT Ontology*). Nad rámec stanovených cílů byla doplněna *Population Ontology* pro možnost určení incidence. Realizována byla *TDB Ontology* pro databázi vzorků tkáně. Anotace v ontologiích jsou užitečné pro ETL proces, realizaci uživatelského rozhraní nebo validaci dat, což bylo prokázáno prostřednictvím realizovaného software a webových aplikací.

Literatura

- [1] **PICCOLI, Gabriele a PIGNI, Federico:** *Information Systems for Managers: With Cases, Edition 4.0.* Prospect Press, 2018, ISBN 978-1-94-315350-3.
- [2] **LAUDON, Kenneth C. a LAUDON, Jane P.:** *Management Information Systems: Managing the Digital Firm, 15th Edition.* Pearson, 2018, ISBN 978-0-13-463971-0.
- [3] **VALACICH, Joseph a SCHNEIDER, Christoph:** *Information Systems Today: Managing in the Digital World.* Pearson Education, 2017, ISBN 978-0-13-460636-1.
- [4] **BULGACS, Simon:** *The first phase of creating a standardised international innovative technological implementation framework/software application.* International Journal of Business and Systems Research, 7(3):250–265, 7/2013. DOI 10.1504/IJBSR.2013.055312.
- [5] **ALTER, Steven:** *Work system theory: overview of core concepts, extensions, and challenges for the future.* Journal of the Association for Information Systems, 14:72–121, Feb 2013, ISSN 1536–9323.
- [6] **ALTER, Steven:** *The Work System Method: Connecting People, Processes, and IT for Business Results.* Works System Press, CA, United States of America, October 2006, ISBN 0-9778497-0-8.
- [7] **ALTER, Steven:** *Defining Information Systems as Work Systems: Implications for the IS Field.* European Journal of Information Systems, 17(5):448–469, 10/2008. DOI 10.1057/ejis.2008.37. Dostupné z <https://works.bepress.com/stevenalter/48>.
- [8] **O’HARA, Margaret, WATSON, Richard a KAVAN, C.:** *Managing the three Levels of Change.* IS Management, 16:63–70, 06/1999. DOI 10.1201/1078/43197.16.3.19990601/31317.9.
- [9] **LANGFORS, Börje:** *Theoretical Analysis of Information Systems.* Auerbach, 1973, ISBN 978-0-87-769151-8.
- [10] **BOURGEOIS, David:** *Information systems for business and beyond.* The Saylor Foundation, 2014. Dostupné z <https://resources.saylor.org/wwwresources/archived/site/textbooks/Information%20Systems%20for%20Business%20and%20Beyond.pdf>.
- [11] **IVES, Blake a LEARMONTH, Gerard P.:** *The information system as a competitive weapon.* Communications of the ACM, 27(12):1193–1201, 1984. Dostupné z <https://aisel.aisnet.org/icis1984/3>.
- [12] **DELONE, William H. a MCLEAN, Ephraim R.:** *Information systems success: The quest for the dependent variable.* Information systems research, 3(1):60–95, 1992, ISSN 1047–7047.

- [13] **DELONE, William H.** a **MCLEAN, Ephraim R.:** *Information systems success revisited.* V *Proceedings of the 35th Annual Hawaii International Conference on System Sciences*, strany 2966–2976. IEEE, 2002, ISBN 0-7695-1435-9. DOI 10.1109/HICSS.2002.994345.
- [14] **DELONE, William H.** a **MCLEAN, Ephraim R.:** *The DeLone and McLean model of information systems success: a ten-year update.* *Journal of management information systems*, 19(4):9–30, 2003, ISSN 0742-1222. Taylor & Francis. DOI 10.1080/07421222.2003.11045748.
- [15] **PETTER, Stacie, DELONE, William** a **MCLEAN, Ephraim R.:** *Information systems success: The quest for the independent variables.* *Journal of management information systems*, 29(4):7–62, 2013. DOI 10.2753/MIS0742-1222290401.
- [16] **O'BRIEN, James** a **MARAKAS, George:** *Introduction to Information Systems, 15th Edition [Loose Leaf].* McGraw-Hill/Irwin, Kansas, November 2009, ISBN 978-0073376776.
- [17] **MARAKAS, George** a **O'BRIEN, James:** *Introduction to Information Systems – Loose Leaf [Loose Leaf].* McGraw-Hill/Irwin, Kansas, 16th edice, January 2012, ISBN 978-0073376882.
- [18] **FROST, Raymond, PIKE, Jacqueline, KENYO, Lauren** a **PELS, Sarah:** *Business Information Systems: design an app for that.* Saylor Foundation, 2011, ISBN 978-1-45-331157-8. [online]. 4. duben 2013 [cit. 2019-06-15]. Dostupné z <http://www.saylor.org/site/textbooks/Business%20Information%20Systems.pdf>.
- [19] **AVISON, David** a **FITZGERALD, Guy:** *Information systems development: methodologies, techniques and tools (4th edition).* McGraw Hill, 2006, ISBN 978-0077114176.
- [20] **HASSELBRING, Wilhelm:** *Information system integration.* *Communications of the ACM*, 43(6):32–36, 2000. DOI 10.1145/336460.336472.
- [21] **EBERT, Nico, WEBER, Kristin** a **KORUNA, Stefan:** *Integration platform as a service.* *Business & Information Systems Engineering*, 59(5):375–379, 6/2017. DOI 10.1007/s12599-017-0486-0.
- [22] **GUTTRIDGE, Keith, PEZZINI, Massimo, GOLLUSCIO, Elizabeth, THOO, Eric, IJIMA, Kimihiko** a **WILCOX, Mary:** *Magic quadrant for enterprise integration platform as a service.* [online]. 30. březen 2017 [cit. 2019-06-15]. Dostupné z <https://www.gartner.com/en/documents/3645397/magic-quadrant-for-enterprise-integration-platform-as-a->, 2017.
- [23] **ČERVENÁ, V., FILIPEC, J., HAVLOVÁ, F., CHURAVÝ, M., JANSKÝ, L., KOZLOVÁ, K., KROUPOVÁ, L., MAREŠOVÁ, J. Machačand H., MEJSTRÍK, V., MICHÁLEK, E., PAPÍRNÍKOVÁ, B., POKORNÁ, E., POŠTOLKOVÁ, B., ROUDNÝ, M., SOCHOVÁ, Z., SVOZILOVÁ, N., VODRÁŽKOVÁ, E.** a **ZIMA, J.:** *Slovník spisovného jazyka českého.*

- [online]. 2011 [cit. 2019-06-18]. Dostupné z <https://ssjc.ujc.cas.cz/search.php?db=ssjc>, 2011.
- [24] **BAGHERI, Hamid, TORKAMANI, Mohammad Ali a GHAFARI, Zhaleh:** *Multi-Agent Approach for facing challenges in Ultra-Large Scale systems*. International Journal of Electrical and Computer Engineering, 4(2):151–154, 4/2014, ISSN 2088-8708.
- [25] **GABRIEL, Richard P., NORTHROP, Linda, SCHMIDT, Douglas C. a SULLIVAN, Kevin:** *Ultra-large-scale systems*. V *Companion to the 21st ACM SIGPLAN symposium on Object-oriented programming systems, languages, and applications*, strany 632–634. ACM, 10/2006, ISBN 1-59593-491-X. DOI 10.1145/1176617.1176645.
- [26] **CHEN, Hong Mei a KAZMAN, Rick:** *Architecting ultra-large-scale green information systems*. V *2012 First International Workshop on Green and Sustainable Software (GREENS)*, strany 69–75. IEEE, 2012, ISBN 978-1-4673-1832-7.
- [27] **JORMANAINEN, Vesa:** *Large-scale implementation and adoption of the Finnish national Kanta services in 2010–2017: a prospective, longitudinal, indicator-based study*. Finnish Journal of eHealth and eWelfare, 10(4):381–395, 12/2018. DOI 10.23996/fjhw.74511.
- [28] **REZAEI, Reza, CHIEW, Thiam Kian a LEE, Sai Peck:** *An interoperability model for ultra large scale systems*. Advances in Engineering Software, 67:22–46, 2014. ISSN 0965-9978. DOI 10.1016/j.advengsoft.2013.07.003. Dostupné z <http://www.sciencedirect.com/science/article/pii/S096599781300121X>.
- [29] **EUROPEAN COMMISSION:** *Towards interoperability for European public services*. [online]. 16. prosinec 2010 [cit. 2019-06-22]. Dostupné z https://ec.europa.eu/isa2/sites/isa/files/docs/publications/communication_on_interoperability.pdf, 2010.
- [30] **BUYLE, Raf:** *Towards interoperability in the public sector*. V *ISWC2017, the 16e International Semantic Web Conference*, svazek 1931, strany 1–8, 2017. ISSN 1613-0073.
- [31] **BUTTARELLI, Giovanni:** *Reflection paper on the interoperability of information systems in the area of Freedom, Security and Justice*. [online]. 17. listopad 2017 [cit. 2019-06-22]. Dostupné z https://edps.europa.eu/sites/edp/files/publication/17-11-16_opinion_interoperability_en.pdf, 2017.
- [32] **BUTTARELLI, Giovanni:** *Shrnutí stanoviska evropského inspektora ochrany údajů k návrhům dvou nařízení, kterými se zřizuje rámec pro interoperabilitu mezi rozsáhlými informačními systémy EU*. [online]. 19. březen 2018 [cit. 2019-06-22]. Dostupné z [https://eur-lex.europa.eu/legal-content/CS/TXT/HTML/?uri=CELEX:52018XX0704\(02\)&from=EN](https://eur-lex.europa.eu/legal-content/CS/TXT/HTML/?uri=CELEX:52018XX0704(02)&from=EN), 2018.

- [33] **EUROPEAN UNION AGENCY FOR FUNDAMENTAL RIGHTS:** *Fundamental rights and the interoperability of EU information systems: borders and security*. European Union Agency for Fundamental Rights, 2017, ISBN 978-92-9491-721-8. [online]. 7. červenec 2017 [cit. 2019-06-22]. Dostupné z https://fra.europa.eu/sites/default/files/fra_uploads/fra-2017-interoperability-eu-information-systems_en-1.pdf.
- [34] **EUROPEAN UNION AGENCY FOR FUNDAMENTAL RIGHTS:** *Under watchful eyes: biometrics, EU IT systems and fundamental rights*. European Union Agency for Fundamental Rights, 2018, ISBN 978-92-9491-925-0. [online]. 28. březen 2018 [cit. 2019-06-22]. Dostupné z https://fra.europa.eu/sites/default/files/fra_uploads/fra-2018-biometrics-fundamental-rights-eu_en.pdf.
- [35] **JÄRVSOO, Maris, NORTA, Alexander, TSAP, Valentyna, PAPPEL, Ingrid a DRAHEIM, Dirk:** *Implementation of Information Security in the EU Information Systems*. V **AL-SHARHAN, Salah A., SIMINTIRAS, Antonis C., DWIVEDI, Yogesh K., JANSSEN, Marijn, MÄNTYMÄKI, Matti, TAHAT, Luay, MOUGHRABI, Issam, ALI, Taher M. a RANA, Nripendra P.** (editoři): *Challenges and Opportunities in the Digital Era*, strany 150–163, Cham, 2018. Springer International Publishing, ISBN 978-3-030-02131-3.
- [36] **CRESSWELL, Kathrin M., BATES, David W. a SHEIKH, Aziz:** *Ten key considerations for the successful implementation and adoption of large-scale health information technology*. Journal of the American Medical Informatics Association, 20(e1):e9–e13, 2013. DOI 10.1136/amiajnl-2013-001684.
- [37] **SLIGO, Judith, GAULD, Robin, ROBERTS, Vaughan a VILLA, Luis:** *A literature review for large-scale health information system project planning, implementation and evaluation*. International journal of medical informatics, 97:86–97, 2017, ISSN 1386-5056. DOI 10.1016/j.ijmedinf.2016.09.007.
- [38] **JONÁK, Zdeněk:** *Data. Česká terminologická databáze knihovnictví a informační vědy (TDKIV)*. [online]. Praha: Národní knihovna ČR, 10. června 2019 [cit. 2019-06-10]. Dostupné z https://aleph.nkp.cz/F/?func=direct&doc_number=000000442&local_base=KTD, 2019.
- [39] **ČESKÝ NORMALIZAČNÍ INSTITUT:** *ČSN ISO 5127 (010162). Informace a dokumentace – slovník*, 2003.
- [40] **KLIMEŠ, Lumír:** *Slovník cizích slov*, svazek 1. SPN, 2010, ISBN 978-80-7235-446-7.
- [41] **JONÁK, Zdeněk:** *Informace. Česká terminologická databáze knihovnictví a informační vědy (TDKIV)*. [online]. Praha: Národní knihovna ČR, 10. června 2019 [cit. 2019-06-10]. Dostupné z https://aleph.nkp.cz/F/?func=direct&doc_number=000000456&local_base=KTD, 2019.

- [42] **ČESKÝ NORMALIZAČNÍ INSTITUT: ČSN ISO 1087-2 (010501).** *Terminologická práce - Slovník - Část 2: Počítačové aplikace.* Praha, 2003.
- [43] **JONÁK, Zdeněk:** *Znalost. Česká terminologická databáze knihovnictví a informační vědy (TDKIV).* [online]. Praha: Národní knihovna ČR, 10. června 2019 [cit. 2019-06-10]. Dostupné z https://aleph.nkp.cz/F/?func=direct&doc_number=000000498&local_base=KTD, 2019.
- [44] **CELBOVÁ, Ludmila:** *Data. Česká terminologická databáze knihovnictví a informační vědy (TDKIV).* [online]. Praha: Národní knihovna ČR, 10. června 2019 [cit. 2019-06-10]. Dostupné z https://aleph.nkp.cz/F/?func=direct&doc_number=000000543&local_base=KTD, 2019.
- [45] **ČESKÝ NORMALIZAČNÍ INSTITUT: ČSN ISO 8459-5. Informace a dokumentace – Sborník bibliografických datových prvků – Část 5: Datové prvky pro výměnu katalogizačních dat a metadata,** 2003.
- [46] **JONÁK, Zdeněk:** *Metainformace. Česká terminologická databáze knihovnictví a informační vědy (TDKIV).* [online]. Praha: Národní knihovna ČR, 10. června 2019 [cit. 2019-06-10]. Dostupné z https://aleph.nkp.cz/F/?func=direct&doc_number=000000479&local_base=KTD, 2019.
- [47] **BERNERS-LEE, Tim, MASINTER, Larry a FIELDING, Roy T.:** *RFC3986: Uniform Resource Identifier (URI): Generic syntax,* 2005.
- [48] **MISSLER, Daniel:** *The Difference Between URLs, URIs, and URNs.* [online]. 4. květen 2019 [cit. 2019-06-13]. Dostupné z <https://danielmiessler.com/study/url-uri/>, 2019.
- [49] **CELKOVÁ, Ludmila:** *PURL. Česká terminologická databáze knihovnictví a informační vědy (TDKIV).* [online]. Praha: Národní knihovna ČR, 10. června 2019 [cit. 2019-06-14]. Dostupné z https://aleph.nkp.cz/F/?func=direct&doc_number=000000551&local_base=KTD, 2019.
- [50] *ISO/IEC 9834-8:2014. Information technology – Procedures for the operation of object identifier registration authorities – Part 8: Generation of universally unique identifiers (UUIDs) and their use in object identifiers,* 2014.
- [51] **BEST, K. a WALSH, N.:** *RFC3121: A URN Namespace for OASIS,* 2001.
- [52] **INTERNATIONAL DOI FOUNDATION: DOI Handbook - Resolution.** [online]. International DOI Foundation, 21. březen 2017 [cit. 2019-06-14]. Dostupné z https://www.doi.org/doi_handbook/3_Resolution.html, 2019.
- [53] **DUERST, M. a SUIGNARD, M.:** *RFC3987: Internationalized Resource Identifiers (IRIs),* 2005.
- [54] **RILEY, Jenn:** *Understanding Metadata: What is Metadata, and What is it For?: A Primer.* National Information Standards Organization, Baltimore, 2017, ISBN 978-1-937522-72-8. National Information Standards Organization. [online]. 18. ledna 2017 [cit. 2019-06-11].

Dostupné z https://groups.niso.org/apps/group_public/download.php/17443/understanding-metadata.

- [55] **LAMBE, Patric**: *Organising Knowledge: Taxonomies, Knowledge and Organisational Effectiveness*. Chandos Publishing, January 2014, ISBN 978-1-84-334228-1.
- [56] **BARGMEYER, Bruce**: *eXtended Metadata Registry (XMDR): Input for Open Ontology Repository*. [online]. Library of Congress, 28. únor 2008 [cit. 2019-07-06]. Dostupné z http://ontolog.cim3.net/file/work/OpenOntologyRepository/2008-02-28_Ontology-Repository-Landscape/XMDR-input-to-Open-Ontology-Registry_v2--BruceBargmeyer_20080228.pdf, 2008.
- [57] **HÜNER, Kai M., OTTO, Boris a ÖSTERLE, Hubert**: *Collaborative management of business metadata*. International Journal of Information Management, 31(4):366–373, 8/2011, ISSN 0268-4012. DOI 10.1016/j.ijinfomgt.2010.12.002.
- [58] **BARGMEYER, Bruce E. a GILLMAN, Daniel W.**: *Metadata Standards and Metadata Registries: An Overview*. Bureau Of Labor Statistics, 2000. Dostupné z <https://www.bls.gov/osmr/research-papers/2000/st000010.htm>.
- [59] **NATIONAL INFORMATION STANDARDS ORGANIZATION**: *Understanding metadata*. National Information Standards, strany 1–20, 2004. ISBN 1-880124-62-9. [online]. NISO Press, Bethesda, USA, 23. června 2004 [cit. 2019-06-10]. Dostupné z https://www.lter.uaf.edu/metadata_files/UnderstandingMetadata.pdf.
- [60] **BRETHERTON, Francis P. a SINGLEY, Paul T.**: *Metadata: A User's View*. V *Seventh International Working Conference on Scientific and Statistical Database Management*, strany 166–174, Washington, DC, USA, sep 1994. IEEE Computer Society, ISBN 0-8186-6610-2. DOI 10.1109/SSDM.1994.336950.
- [61] **BERNERS-LEE, Tim**: *The world-wide web*. Computer networks and ISDN systems, 25(4-5):454–459, 1992. Dostupné z <https://web.stanford.edu/class/cs344g/www-1992.pdf>.
- [62] **CHOUDHURY, Nupur**: *World Wide Web and Its Journey from Web 1.0 to Web 4.0*. International Journal of Computer Science and Information Technologies, 5(6):8096–8100, 2014. ISSN 0975-9646.
- [63] **WORLD WIDE WEB CONSORTIUM**: *Facts About W3C*. [online]. 27. prosince 2019 [cit. 2019-06-12]. Dostupné z <https://www.w3.org/Consortium/facts>, 2018.
- [64] **KIMBALL, Ralph, ROSS, Margy, THORNTHWAITE, Warren, MUNDY, Joy a BECKER, Bob**: *The Data Warehouse Lifecycle Toolkit*. Willey Publishing, Inc., Indianapolis, second edice, 2008, ISBN 978-0470149775.

- [65] **PREMIS EDITORIAL COMMITTEE:** *PREMIS Data Dictionary for Preservation Metadata, version 3.0*. PREservation Metadata: Implementation Strategies (PREMIS), Washington, 2015. [online]. The Library of Congress. Preservation Metadata: Implementation Strategies (PREMIS). 1. květen 2019 [cit. 2019-06-15]. Dostupné z <https://www.loc.gov/standards/premis/v3/index.html>.
- [66] **NILSSON, Mikael, BAKER, Thomas a JOHNSTON, Pete:** *Interoperability Levels for Dublin Core Metadata*. [online]. Dublin Core Metadata Initiative. 1. květen 2009 [cit. 2019-06-24]. Dostupné z <http://www.dublincore.org/specifications/dublin-core/interoperability-levels/>, 2009.
- [67] **TOLK, Andreas a MUGUIRA, James A.:** *The levels of conceptual interoperability model*. V *Proceedings of the 2003 fall simulation interoperability workshop*, svazek 7, strany 1–11, 9/2003.
- [68] **TOLK, Andreas, DIALLO, Saikou Y. a TURNITSA, Charles D.:** *Applying the levels of conceptual interoperability model in support of integrability, interoperability, and composability for system-of-systems engineering*. *Journal of Systemics, Cybernetics and Informatics*, 2007, ISSN 1690-4524.
- [69] **NATIONAL INFORMATION STANDARDS ORGANIZATION:** *A Framework of Guidance for Building Good Digital Collections*. National Information Standards, 2007. [online]. 10. prosinec 2007 [cit. 2019-06-15]. Dostupné z <https://www.niso.org/sites/default/files/2017-08/framework3.pdf>.
- [70] **IMS GLOBAL LEARNING CONSORTIUM:** *Learning Resource Meta-data Specification*. [online]. Visual Resources Association Foundation (VRAF), 30. duben 2012 [cit. 2019-07-04]. Dostupné z <http://www.imslobal.org/metadata/index.html>, 2012.
- [71] **GREENBERG, Jane:** *Understanding metadata and metadata schemes*. *Cataloging & classification quarterly*, 40(3-4):17–36, 2005, ISSN 0163-9374. DOI 10.1300/J104v40n03_02.
- [72] **FEDERAL GEOGRAPHIC DATA COMNMITEE (FGDC):** *FGDC Geospatial Standards*. [online]. RDA Steering Committee, 4. červenec 2019 [cit. 2019-07-04]. Dostupné z <https://www.fgdc.gov/standards/list>, 2019.
- [73] *ISO 19115-1:2014. Geographic information – Metadata – Part 1: Fundamentals*, 2014.
- [74] **CATHRO, Warwick:** *Metadata: An Overview*. [online]. National Library of Australia Staff Papers, [cit. 2011-07-20]. Dostupné z <http://www.nla.gov.au/nla/staffpaper/cathro3.html>, 1997.
- [75] **INTERNATIONAL ORGANIZATION FOR STANDARDIZATION:** *ISO 15836-1:2017. Information and documentation – The Dublin Core*

- metadata element set – Part 1: Core elements*, 2017. ICS: 35.240.30 IT applications in information, documentation and publishing.
- [76] *ISO 15489-1:2016. Information and documentation – Records management – Part 1: Concepts and principles*, 2016.
- [77] *ISO 23081-1:2017. Information and documentation – Records management processes – Metadata for records – Part 1: Principles*, 2017.
- [78] *ISO 23081-2:2009. Information and documentation – Managing metadata for records – Part 2: Conceptual and implementation issues*, 2009.
- [79] *ISO 23081-3:2011. Information and documentation – Managing metadata for records – Part 3: Self-assessment method*, 2011.
- [80] **AUSTRALIAN NATIONAL DATA SERVICE: Metadata**. [online]. 23. leden 2018 [cit. 2019-06-28]. Dostupné z https://www.ands.org.au/_data/assets/pdf_file/0004/728041/Metadata-Workinglevel.pdf, 2018.
- [81] **UNIVERSITY OF NORTH CAROLINA: Basic Elements – Metadata for Data Management: A Tutorial**. [online]. 10. červen 2019 [cit. 2019-06-28]. Dostupné z <https://guides.lib.unc.edu/metadata/basic-elements>, 2019.
- [82] **GENESERETH, Michael R. a NILSSON, Nils J.:** *Logical foundations of artificial intelligence*, svazek 9. Morgan Kaufmann Los Altos, Los Altos, California, 1987, ISBN 978-0-934613-31-6. DOI 10.1016/C2009-0-27551-9.
- [83] **GRUBER, Thomas R. a kol.:** *A translation approach to portable ontology specifications*. Knowledge acquisition, 5(2):199–220, 1993, ISSN 1042-8143. DOI 10.1006/knac.1993.1008.
- [84] **NORDMANN, Kore:** *Standardization of Ontologies*. Online entry, 13, 5/2009. Dostupné z https://kore-nordmann.de/talks/09_04_standardization_of_ontologies_paper.pdf.
- [85] **NEWELL, Allen:** *The knowledge level*. Artificial intelligence, 18(1):87–127, 1982, ISSN 0004-3702.
- [86] **RUSSELL, Stuart a NORVIG, Peter:** *Artificial Intelligence: A Modern Approach*. Pearson Education, Upper Saddle River: Prentice Hall, 3rd edition, international edition edice, 2013, ISBN 978-1-292-02420-2.
- [87] **MATUSZEK, C., CABRAL, J., WITBROCK, M. a DEOLIVEIRA, J.:** *An introduction to the syntax and content of Cyc*. V *Proceedings of the 2006 AAAI spring symposium on formalizing and compiling background knowledge and its applications to knowledge representation and question answering*, svazek 3864, 2006. DOI 10.13016/M2J09W76T.
- [88] **SMITH, Barry, ASHBURNER, Michael, ROSSE, Cornelius, BARD, Jonathan, BUG, William, CEUSTERS, Werner, GOLDBERG, Louis J., EILBECK, Karen, IRELAND, Amelia a MUNGALL, Christopher J.:** *The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration*. Nature Biotechnology, 25(11):1251–1255, 2007, ISSN 1546-1696. DOI 10.1038/nbt1346.

- [89] **TIRMIZI, Syed Hamid, AITKEN, Stuart, MOREIRA, Dilvan A., MUNGALL, Chris, SEQUEDA, Juan, SHAH, Nigam H. a MIRANKER, Daniel P.:** *Mapping between the OBO and OWL ontology languages*. Journal of biomedical semantics, 2(Suppl 1):S3, 2011, ISSN 2041-1480. DOI 10.1186/2041-1480-2-S1-S3.
- [90] **DIGITAL CURATION CENTER:** *Disciplinary Metadata*. [online]. 2019 [cit. 2019-06-28]. Dostupné z <http://www.dcc.ac.uk/drupal/resources/metadata-standards>, 2019.
- [91] **DIGITAL CURATION CENTER:** *All Standards for Any Lifecycle Action*. [online]. 2009 [cit. 2019-06-28]. Dostupné z <http://www.dcc.ac.uk/resources/standards/diffuse/standards>, 2009.
- [92] **HAMILL, Lois:** *Archival Arrangement and Description: Analog to Digital*. Rowman & Littlefield Publishers, 2017, ISBN 978-1-44-227917-9.
- [93] **SOCIETY OF AMERICAN ARCHIVISTS a STAATSBIBLIOTHEK ZU BERLIN:** *EAC-CPF: Schemata and Tag Library*. [online]. Staatsbibliothek zu Berlin, 1. prosinec 2018 [cit. 2019-07-04]. Dostupné z <https://eac.staatsbibliothek-berlin.de/schemata-and-tag-library/>, 2018.
- [94] **LIBRARY OF CONGRESS:** *EAD: Encoded Archival Description*. [online]. 10. duben 2019 [cit. 2019-07-04]. Dostupné z <https://www.loc.gov/ead/>, 2019.
- [95] **INTERNATIONAL COUNCIL ON ARCHIVES:** *ISAD(G): General International Standard Archival Description – Second edition*. [online]. 1. září 2011 [cit. 2019-07-04]. ISBN 0-9696035-5-X. Dostupné z <https://www.ica.org/en/isadg-general-international-standard-archival-description-second-edition>, 2011.
- [96] **LIBRARY OF CONGRESS:** *PREMIS: Preservation Metadata Maintenance Activity*. [online]. 1. květen 2019 [cit. 2019-07-04]. Dostupné z <https://www.loc.gov/standards/premis/>, 2019.
- [97] **INTERNATIONAL FEDERATION OF LIBRARY ASSOCIATIONS AND INSTITUTIONS:** *Functional Requirements for Bibliographic Records: Final Report*. [online]. International Federation of Library Associations and Institutions, 4. duben 2019 [cit. 2019-07-04]. Dostupné z https://www.ifla.org/files/assets/cataloguing/frbr/frbr_2008.pdf, 2009.
- [98] **LIBRARY OF CONGRESS:** *MARC Standards*. [online]. 22. květen 2019 [cit. 2019-07-04]. Dostupné z <https://www.loc.gov/marc/>, 2019.
- [99] **LIBRARY OF CONGRESS:** *METS: Metadata Encoding and Transmission Standard*. [online]. 17. leden 2019 [cit. 2019-07-04]. Dostupné z <http://www.loc.gov/standards/mets/>, 2019.
- [100] **LIBRARY OF CONGRESS:** *MODS: Metadata Object Description Schema*. [online]. 14. září 2018 [cit. 2019-07-04]. Dostupné z <http://www.loc.gov/standards/mods/>, 2018.

- [101] **EDITEUR**: *ONIX: ONline Information eXchange*. [online]. 2019 [cit. 2019-07-04]. Dostupné z <https://www.editeur.org/8/ONIX/>, 2019.
- [102] **RESOURCE DESCRIPTION AND ACCESS STEERING COMMITTEE**: *Welcome to the RSC website*. [online]. RDA Steering Committee, 21. červen 2019 [cit. 2019-07-04]. Dostupné z <http://www.rda-rsc.org/>, 2019.
- [103] **DARWIN CORE MAINTENACE GROUP**: *Darwin Core*. [online]. RDA Steering Committee, 23. duben 2019 [cit. 2019-07-04]. Dostupné z <https://dwc.tdwg.org/>, 2019.
- [104] **DARWIN CORE MAINTENACE GROUP**: *Ecological Metadata Language (EML)*. [online]. RDA Steering Committee, 26. červen 2019 [cit. 2019-07-04]. Dostupné z <https://knb.ecoinformatics.org/external//emlparser/docs/eml-2.1.1/index.html>, 2019.
- [105] **VISUAL RESOURCES ASSOCIATION FOUNDATION**: *VRA Core Schemas and Documentation*. [online]. 9. duben 2007 [cit. 2019-07-04]. Dostupné z <https://www.loc.gov/standards/vracore/schemas.html>, 2007.
- [106] **GONZALEZ-PEREZ, Cesar** a **MARTIN-RODILLA, Patricia**: *An Alternative Approach to Metainformation Conceptualisation and Use. V Conceptual Modeling*, strany 92–105. Springer, 2017, ISBN 978-3-319-69903-5. ISSN 0302-9743. DOI 10.1007/978-3-319-69904-2.
- [107] **WORLD WIDE WEB CONSORTIUM**: *Data Catalog Vocabulary (DCAT)*. [online]. 16. leden 2014 [cit. 2019-07-04]. Dostupné z <https://www.w3.org/TR/vocab-dcat/>, 2014.
- [108] *ISO/IEC 11179-1:2015. Information technology – Metadata registries (MDR) – Part 1: Framework*, 2015. [online]. International Organization for Standardization (ISO), 15. prosinec 2015 [cit. 2019-06-12]. Dostupné z https://standards.iso.org/ittf/PubliclyAvailableStandards/c061932_ISO_IEC_11179-1_2015.zip.
- [109] *ISO/IEC 11179-2:2005. Information technology – Metadata registries (MDR) – Part 2: Classification*, 2005. [online]. International Organization for Standardization (ISO), 15. listopad 2005 [cit. 2019-06-12]. Dostupné z [https://standards.iso.org/ittf/PubliclyAvailableStandards/c035345_ISO_IEC_11179-2_2005\(E\).zip](https://standards.iso.org/ittf/PubliclyAvailableStandards/c035345_ISO_IEC_11179-2_2005(E).zip).
- [110] *ISO/IEC 11179-3:2013. Information technology – Metadata registries (MDR) – Part 3: Registry metamodel and basic attributes*, 2013.
- [111] *ISO/IEC 11179-4:2004. Information technology – Metadata registries (MDR) – Part 4: Formulation of data definitions*, 2004.
- [112] *ISO/IEC 11179-5:2015. Information technology – Metadata registries (MDR) – Part 5: Naming principles*, 2015.

- [113] *ISO/IEC 11179-6:2015. Information technology – Metadata registries (MDR) – Part 5: Registration*, 2015.
- [114] **WORLD WIDE WEB CONSORTIUM: SKOS Simple Knowledge Organization System - Home Page**. [online]. 13. prosinec 2012 [cit. 2019-07-04]. Dostupné z <https://www.w3.org/2004/02/skos/>, 2012.
- [115] **OPEN ARCHIVES INITIATIVE: Protocol for Metadata Harvesting**. [online]. 27. červenec 2018 [cit. 2019-07-04]. Dostupné z <https://www.openarchives.org/pmh/>, 2018.
- [116] *ISO 14721:2012. Space data and information transfer systems – Open archival information system (OAIS) – Reference model*, 2012.
- [117] **CONSULTATIVE COMMITTEE FOR SPACE DATA SYSTEMS: Reference Model for an Open Archival Information System (OAIS)**. [online]. 14. červen 2012 [cit. 2019-07-04]. Dostupné z <https://public.ccsds.org/pubs/650x0m2.pdf>, 2012.
- [118] **HUC, C., BOUCON, D., SAWYER, D.M. a GARRETT, J.G.:** *The Producer-Archive Interface Methodology Abstract Standard (PAIMAS)*. Space OPS 2004 Conference, 2014. DOI 10.2514/6.2004-649-446. Dostupné z <https://arc.aiaa.org/doi/pdf/10.2514/6.2004-649-446>.
- [119] **CONSULTATIVE COMMITTEE FOR SPACE DATA SYSTEMS: Producer-Archive Interface Methodology Abstract Standard (PAIMAS)**. [online]. 18. červen 2010 [cit. 2019-07-04]. Dostupné z <https://public.ccsds.org/Pubs/651x0m1.pdf>, 2004.
- [120] *ISO/TR 18492:2005. Long-term preservation of electronic document-based information*, 2005.
- [121] **WORLD WIDE WEB CONSORTIUM: XML Schema**. [online]. 9. březen 2018 [cit. 2019-07-04]. Dostupné z <https://www.w3.org/XML/Schema>, 2018.
- [122] *ISO 3166:2013. Country Codes*, 2013.
- [123] *ISO 639. Language Codes*, 2010.
- [124] *ISO 8601:2019. Date and time format*, 2019.
- [125] **LIBRARY OF CONGRESS: MADS: Metadata Authority Description Schema**. [online]. Library of Congress, 23. květen 2018 [cit. 2019-07-04]. Dostupné z <http://www.loc.gov/standards/mads/>, 2018.
- [126] **EUROPEAN COMMISSION: European Interoperability Timeline**. Publication Office of the European Union, Luxembourg, 2017, ISBN 978-9-27-963756-8. DOI 10.2799/78681. [online]. 2. červen 2017 [cit. 2019-07-04]. Dostupné z https://ec.europa.eu/isa2/publications/europe-an-interoperability-timeline-poster_en.

- [127] **EUROPEAN COMMISSION:** *New European Interoperability Framework: Promoting seamless services and data flows for European public administrations.* Publication Office of the European Union, Luxembourg, 2017, ISBN 978-9-27-963756-8. DOI 10.2799/78681. [online]. 23. březen 2017 [cit. 2019-07-04]. Dostupné z https://ec.europa.eu/isa2/sites/isa/files/eif_brochure_final.pdf.
- [128] **EUROPEANA FOUNDATION:** *Definition of the Europeana Data Model v5.2.8.* [online]. Europeana, 6. říjen 2017 [cit. 2019-07-04]. Dostupné z https://pro.europeana.eu/files/Europeana_Professional/Share_your_data/Technical_requirements/EDM_Documentation/EDM_Definition_v5.2.8_102017.pdf, 2017.
- [129] **EUROPEANA FOUNDATION:** *Europeana Data Model – Mapping Guidelines v2.4.* [online]. Europeana, 6. říjen 2017 [cit. 2019-07-04]. Dostupné z https://pro.europeana.eu/files/Europeana_Professional/Share_your_data/Technical_requirements/EDM_Documentation/EDM_Mapping_Guidelines_v2.4_102017.pdf, 2017.
- [130] **KIM, Jeong Dong, SON, Jiseong a BAIK, Doo Kwon:** *XMDR+: An extended XMDR model for supporting diverse ontological relations.* IEICE transactions on information and systems, 94(3):515–524, 3/2011, ISSN 0916-8532. DOI 10.1587/transinf.E94.D.515.
- [131] **INTERNATIONAL ORGANIZATION FOR STANDARDIZATION:** *ISO 15836-2. Information and documentation – The Dublin Core metadata element set – Part 2: DCMI Properties and classes,* 2019. Dostupné z <https://www.iso.org/standard/71341.html>.
- [132] **FREED, Ned a KUCHERAWY, Murray:** *Media Types.* [online]. Internet Assigned Numbers Authority (IANA), 25. červen 2019 [cit. 2019-07-06]. Dostupné z <https://www.iana.org/assignments/media-types/media-types.xhtml>, 2019.
- [133] **LIBRARY OF CONGRESS:** *METS: Metadata Encoding and Transmission Standard: Primer and Reference Manual.* [online]. 27. únor 2018 [cit. 2019-07-06]. Dostupné z <http://www.loc.gov/standards/mets/METSPrimer.pdf>, 2018.
- [134] **PREMIS EDITORIAL COMMITTEE:** *Data Dictionary for Preservation Metadata: PREMIS version 3.0.* [online]. Library of Congress, 7. leden 2016 [cit. 2019-07-06]. Dostupné z <https://www.loc.gov/standards/premis/v3/premis-3-0-final.pdf>, 2015.
- [135] **OPEN ARCHIVES INITIATIVE:** *The Open Archives Initiative Protocol for Metadata Harvesting.* [online]. 8. leden 2015 [cit. 2019-07-05]. Dostupné z <http://www.openarchives.org/OAI/openarchivesprotocol.html>, 2015.
- [136] **DEMPSEY, Lorcan:** *Google and OAI-PMH.* [online]. 24. duben 2008 [cit. 2020-02-08]. Dostupné z <http://orweblog.oclc.org/google-and-oai-pmh/>, 2008.

- [137] **LAVOIE, Brian F.:** *The open archival information system (OAIS) reference model: introductory guide*. Digital Preservation Coalition, 2014. ISSN 2048-7916. DOI 10.7207/twr14-02.
- [138] **OTTE, Maxwell:** *Building Trustworthy Digital Repositories: Theory and Implementation*. Archivaria, 86(86):187–190, 2018. ISSN 0318-6954.
- [139] **INTERNATIONAL COUNCIL FOR SCIENCE WORLD DATA SYSTEM:** *DSA–WDS Partnership Working Group – Catalogue of Common Requirements*. [online]. International Council for Science, 16. únor 2016 [cit. 2019-07-05]. Dostupné z https://assessment.datasealofapproval.org/sitemedia/files/DSA_booklets/DSA-booklet_2017-2019.pdf, 2016.
- [140] **FOJTU, Andrea, HUTAŘ, Jan, MELICHAR, Marek a PAVLÁSKOVÁ, Eliška:** *překl. Data Seal of Approval Board: Pečeť kvality digitálního repozitáře. Zásady verze 2*. [online]. 19. červenec 2013 [cit. 2019-07-05]. Dostupné z <https://dsa.cuni.cz/DSA-3-version1-datasealofapproval2.pdf>, 2013.
- [141] **LAUGHTON, Paul Arthur:** *Open Archival Information System (OAIS) as a data curation standard in the World Data Centre*. diplomová práce, Centre for Information and Knowledge Management, University of Johannesburg, 2011.
- [142] **KUNZE, J., LITTMAN, J., MADDEN, E., SCANCELLA, J. a ADAMS, C.:** *RFC8493: The BagIt File Packaging Format (V1.0)*, 2018.
- [143] *ISO 20652:2006. Space data and information transfer systems – Producer-archive interface – Methodology abstract standard*, 2006.
- [144] **WORLD WIDE WEB CONSORTIUM:** *Resource Description Framework (RDF): Concepts and Abstract Syntax*. [online]. 10. únor 2004 [cit. 2019-07-07]. Dostupné z <http://www.w3.org/TR/rdf-concepts/>, 2004.
- [145] **WORLD WIDE WEB CONSORTIUM:** *RDF 1.1 Concepts and Abstract Syntax*. [online]. 25. únor 2014 [cit. 2019-07-07]. Dostupné z <https://www.w3.org/TR/rdf11-concepts/>, 2014.
- [146] **WORLD WIDE WEB CONSORTIUM:** *RDF 1.1 Whats New*. [online]. 25. únor 2014 [cit. 2019-07-08]. Dostupné z <https://www.w3.org/TR/rdf11-new/>, 2014.
- [147] **WORLD WIDE WEB CONSORTIUM:** *RDF Primer 1.1*. [online]. 24. červen 2014 [cit. 2019-07-08]. Dostupné z <https://www.w3.org/TR/rdf11-primer/>, 2014.
- [148] **WORLD WIDE WEB CONSORTIUM:** *RDF 1.1 Semantics*. [online]. 25. únor 2014 [cit. 2019-07-08]. Dostupné z <https://www.w3.org/TR/rdf11-mt/>, 2014.
- [149] **ALESSO, H. Peter a SMITH, Craig F.:** *Thinking on the Web*. John Wiley& Sons, Ltd, 2006. ISBN 978-0-47-176814-2.

- [150] **ANDERSON, Paul**: *Web 2.0 and beyond: Principles and technologies*, svazek 7. CRC Press, 2012, ISBN 978-1-43-982867-0.
- [151] **WORLD WIDE WEB CONSORTIUM**: *Cool URIs for the Semantic Web*. [online]. 3. prosinec 2008 [cit. 2019-07-07]. Dostupné z <https://www.w3.org/TR/cooluris/>, 2008.
- [152] **WORLD WIDE WEB CONSORTIUM**: *RDF Schema 1.1*. [online]. 25. únor 2014 [cit. 2019-07-08]. Dostupné z <https://www.w3.org/TR/rdf-schema/>, 2014.
- [153] **WORLD WIDE WEB CONSORTIUM**: *RDF 1.1 XML Syntax*. [online]. 25. únor 2014 [cit. 2019-07-08]. Dostupné z <https://www.w3.org/TR/rdf-syntax-grammar/>, 2014.
- [154] **WORLD WIDE WEB CONSORTIUM**: *RDF 1.1 N-Quads*. [online]. 25. únor 2014 [cit. 2019-07-08]. Dostupné z <https://www.w3.org/TR/n-quads/>, 2014.
- [155] **WORLD WIDE WEB CONSORTIUM**: *RDF 1.1 N-Triples*. [online]. 25. únor 2014 [cit. 2019-07-08]. Dostupné z <https://www.w3.org/TR/n-triples/>, 2014.
- [156] **WORLD WIDE WEB CONSORTIUM**: *RDF 1.1 TriG*. [online]. 25. únor 2014 [cit. 2019-07-08]. Dostupné z <https://www.w3.org/TR/trig/>, 2014.
- [157] **BECKETT, David, BERNERS-LEE, Tim, PRUD'HOMMEAUX, Eric a CAROTHERS, Gavin**: *RDF 1.1 Turtle: Terse RDF Triple Language*. [online]. World Wide Web Consortium, 25. únor 2014 [cit. 2019-07-08]. Dostupné z <https://www.w3.org/TR/turtle/>, 2014.
- [158] **WORLD WIDE WEB CONSORTIUM**: *RDF 1.1 JSON Alternate Serialization (RDF/JSON)*. [online]. 7. listopad 2013 [cit. 2019-07-08]. Dostupné z <https://www.w3.org/TR/rdf-json/>, 2013.
- [159] **WORLD WIDE WEB CONSORTIUM**: *RDFa 1.1 Primer – Third Edition*. [online]. 17. březen 2015 [cit. 2019-07-08]. Dostupné z <https://www.w3.org/TR/rdfa-primer/>, 2015.
- [160] **WORLD WIDE WEB CONSORTIUM**: *SPARQL 1.1 Overview*. [online]. 21. březen 2013 [cit. 2019-07-08]. Dostupné z <https://www.w3.org/TR/sparql11-overview/>, 2013.
- [161] **WORLD WIDE WEB CONSORTIUM**: *SPARQL Query Language for RDF*. [online]. 15. leden 2008 [cit. 2019-07-08]. Dostupné z <https://www.w3.org/TR/rdf-sparql-query/>, 2008.
- [162] **WORLD WIDE WEB CONSORTIUM**: *SPARQL 1.1 Query Language*. [online]. 21. březen 2013 [cit. 2019-07-08]. Dostupné z <https://www.w3.org/TR/sparql11-query/>, 2013.

- [163] **WORLD WIDE WEB CONSORTIUM: SPARQL 1.1 Update.** [online]. 21. březen 2013 [cit. 2019-07-08]. Dostupné z <https://www.w3.org/TR/sparql11-update/>, 2013.
- [164] **WORLD WIDE WEB CONSORTIUM: SPARQL 1.1 Service Description.** [online]. 21. březen 2013 [cit. 2019-07-08]. Dostupné z <https://www.w3.org/TR/sparql11-service-description/>, 2013.
- [165] **WORLD WIDE WEB CONSORTIUM: SPARQL 1.1 Federated Query.** [online]. 21. březen 2013 [cit. 2019-07-08]. Dostupné z <https://www.w3.org/TR/sparql11-federated-query/>, 2013.
- [166] **WORLD WIDE WEB CONSORTIUM: SPARQL 1.1 Entailment Regimes.** [online]. 21. březen 2013 [cit. 2019-07-08]. Dostupné z <https://www.w3.org/TR/sparql11-entailment/>, 2013.
- [167] **WORLD WIDE WEB CONSORTIUM: SPARQL 1.1 Protocol.** [online]. 21. březen 2013 [cit. 2019-07-08]. Dostupné z <https://www.w3.org/TR/sparql11-protocol/>, 2013.
- [168] **WORLD WIDE WEB CONSORTIUM: SPARQL 1.1 Graph Store HTTP Protocol.** [online]. 21. březen 2013 [cit. 2019-07-08]. Dostupné z <https://www.w3.org/TR/sparql11-http-rdf-update/>, 2013.
- [169] **WORLD WIDE WEB CONSORTIUM: OWL 2 Web Ontology Language Document Overview (Second Edition).** [online]. 11. prosinec 2012 [cit. 2019-07-08], Dostupné z <https://www.w3.org/TR/owl2-overview/>, 2012.
- [170] **FENSEL, Dieter, HARMELEN, Frank Van, HORROCKS, Ian, MCGUINNESS, Deborah L. a PATEL-SCHNEIDER, Peter F.: OIL: An ontology infrastructure for the semantic web.** Intelligent Systems, IEEE, 16(2):38–45, 2001, ISSN 1094-7167. DOI 10.1109/5254.920598.
- [171] **WORLD WIDE WEB CONSORTIUM: OWL Web Ontology Language Guide.** [online]. 10. únor 2004 [cit. 2019-07-08], Dostupné z <https://www.w3.org/TR/owl-guide/>, 2004.
- [172] **WORLD WIDE WEB CONSORTIUM: OWL 2 Web Ontology Language New Features and Rationale (Second Edition).** [online]. 11. prosinec 2012 [cit. 2019-07-08], Dostupné z <https://www.w3.org/TR/owl2-new-features/>, 2012.
- [173] **HORROCKS, Ian, KUTZ, Oliver a SATTLER, Ulrike: The Even More Irresistible SROIQ.** Kr, 6:57–67, 2006. ISBN 978-1-57735-271-6.
- [174] **WORLD WIDE WEB CONSORTIUM: OWL 2 Web Ontology Language Profiles (Second Edition).** [online]. 11. prosinec 2012 [cit. 2019-07-08], Dostupné z <https://www.w3.org/TR/owl2-profiles/>, 2012.
- [175] **BAADER, Franz, BRANDT, Sebastian a LUTZ, Carsten: Pushing the EL envelope.** V *Proceedings of the 19th International Joint Conference on Artificial Intelligence (IJCAI)*, svazek 5, strany 364–369, San Francisco, CA, USA, 1/2005. Morgan Kaufmann Publishers Inc.

- [176] **BAADER, Franz, BRANDT, Sebastian a LUTZ, Carsten:** *Pushing the EL envelope further*. OWL: Experiences and Direction, 2008.
- [177] **CALVANESE, Diego, GIACOMO, Giuseppe De, LEMBO, Domenico, LENZERINI, Maurizio a ROSATI, Riccardo:** *Tractable reasoning and efficient query answering in description logics: The DL-Lite family*. Journal of Automated reasoning, 39(3):385–429, 7/2007, ISSN 0168-7433. DOI 10.1007/s10817-007-9078-x.
- [178] **GROSOFF, Benjamin N., HORROCKS, Ian, VOLZ, Raphael a DECKER, Stefan:** *Description logic programs: combining logic programs with description logic*. V *Proceedings of the 12th international conference on World Wide Web*, strany 48–57. ACM, 2003, ISBN 1-58113-680-3. DOI 10.1145/775152.775160.
- [179] **HORST, Herman J. Ter:** *Completeness, decidability and complexity of entailment for RDF Schema and a semantic extension involving the OWL vocabulary*. Web Semantics: Science, Services and Agents on the World Wide Web, 3(2-3):79–115, 2005, ISSN 1570-8268. DOI 10.1016/j.websem.2005.06.001.
- [180] **BERNERS-LEE, Tim, HENDLER, James a LASSILA, Ora:** *The Semantic Web*. Scientific American, 284(5):28–37, 5/2001.
- [181] **BERNERS-LEE, Tim:** *Linked Data*. [online]. Cambridge, 6. listopad 2017 [cit. 2019-07-07], Dostupné z <http://www.w3.org/DesignIssues/LinkedData.html>, 2017.
- [182] **WORLD WIDE WEB CONSORTIUM:** *W3C Semantic Web Frequently Asked Questions*. [online]. 12. listopad 2009 [cit. 2019-07-08]. Dostupné z <https://www.w3.org/RDF/FAQ>, 2009.
- [183] **WORLD WIDE WEB CONSORTIUM:** *W3C Semantic Web Activity*. [online]. 11. prosinec 2013 [cit. 2019-07-09]. Dostupné z <https://www.w3.org/2001/sw/>, 2013.
- [184] **WORLD WIDE WEB CONSORTIUM:** *W3C Data Activity*. [online]. 1. červenec 2019 [cit. 2019-07-09]. Dostupné z <https://www.w3.org/2013/data/>, 2019.
- [185] **SAUERMAN, Leo, BERNARDI, Ansgar a DENGEL, Andreas:** *Overview and outlook on the semantic desktop*. V *Proceedings of the 2005 International Conference on Semantic Desktop Workshop: Next Generation Information Management D Collaboration Infrastructure*, svazek 175, strany 74–91. CEUR-WS.org, 2005.
- [186] **LINKED DATA FOR PROFESSIONAL EDUCATION:** *Linked Data Competency Index*. [online]. 28. červen 2017 [cit. 2019-07-07], Dostupné z <https://dcmi.github.io/ldci/D2695955/>, 2017.

- [187] **FERNANDEZ-GARCIA, Javier D., EKAPUTRA, Fajar J., ARYAN, Peb Ruswono, AZZAM, Amr a KIESLING, Elmar:** *Privacy-aware Linked Widgets*. Companion Proceedings of The 2019 World Wide Web Conference, strany 609–514, March 2019. DOI 10.1145/3308560.3317591.
- [188] **CLARK, Kenneth, VENDT, Bruce, SMITH, Kirk, FREYMAN, John, KIRBY, Justin, KOPPEL, Paul, MOORE, Stephen, PHILLIPS, Stanley, MAFFITT, David, PRINGLE, Michael a kol.:** *The Cancer Imaging Archive (TCIA): maintaining and operating a public information repository*. Journal of digital imaging, 26(6):1045–1057, 2013, ISSN 0897-1889. DOI 10.1007/s10278-013-9622-7.
- [189] **HEIMANN, Tobias, VAN GINNEKEN, Brain, STYNER, Martin A., ARZHAIEVA, Yulia, AURICH, Volker, BAUER, Christian, BECK, Andreas, BECKER, Christoph, BEICHEL, Reinhard, BEKES, György, BELLO, Fernando, BINNIG, Gerd, BISCHOF, Horst, BORNIK, Alexander, CASHMAN, Peter M.M., CHI, Ying, CÓRDOVA, Andrés, DAWANT, Benoit M., FIDRICH, Márta, FURST, Jacob D., FURUKAWA, Daisuke, GRENACHER, Lars, HORNEGGER, Joachim, KAINMÜLLER, Dagmar, KITNEY, Richard I., KOBATAKE, Hidefumi, LAMECKER, Hans, LANGE, Thomas, LEE, Jeongjin, LENNON, Brian, LI, Rui, LI, Senhu, MEINZER, Hans Peter, NÉMETH, Gábor, RAICU, Daniela S., RAU, Anne Mareike, VAN RIKXOORT, Eva M., ROUSSON, Mikaël, RUSKÓ, László, SADDI, Kinda A., SCHMIDT, Günter, SEGHERS, Dieter, SHIMIZU, Akinobu, SLAGMOLEN, Pieter, SORANTIN, Erich, SOZA, Grzegorz, SUSOMBOON, Ruchaneewan, WAITE, Jonathan M., WIMMER, Andreas a WOLF, Ivo:** *Comparison and evaluation of methods for liver segmentation from CT datasets*. IEEE Transactions on Medical Imaging, 28(8):1251–1265, 8/2009. ISSN 0278-0062. DOI 10.1109/TMI.2009.2013851.
- [190] **GINNEKEN, Bram Van, HEIMANN, Tobias a STYNER, Martin:** *3D segmentation in the clinic: A grand challenge*. 3D segmentation in the clinic: a grand challenge, strany 7–15, 2007. [online]. 3. říjen 2007 [cit. 2019-07-08]. Dostupné z <http://sliver07.org/p7.pdf>.
- [191] **GORDON, Evan M., LAUMANN, Timothy O., GILMORE, Adrian W., NEWBOLD, Dillan J., GREENE, Deanna J., BERG, Jeffrey J., ORTEGA, Mario, HOYT-DRAZEN, Catherine, GRATTON, Caterina, SUN, Haoxin, HAMPTON, Jacqueline M., COALSON, Rebecca S., NGUYEN, Annie L., MCDERMOTT, Kathleen B., SHIMONY, Joshua S., SNYDER, Abraham Z., SCHLAGGAR, Bradley L., PETERSEN, Steven E., NELSON, Steven M. a DOSENBACH, Nico U.F.:** *Precision Functional Mapping of Individual Human Brains*. Neuron, 95(4):791–807.e7, 2017, ISSN 0896-6273. <http://www.sciencedirect.com/science/article/pii/S089662731730613X>.

- [192] **NEJEDLÁ, Marie:** *Klinická propedeutika: pro studenty zdravotnických oborů.* Grada Publishing, Praha, seštra edice, 2015, ISBN 978-80-247-4402-5.
- [193] **COMMISSION OF THE EUROPEAN COMMUNITIES:** *Commission Recommendation of 2 July 2008 on cross-border interoperability of electronic health record systems (notified under document number C(2008) 3282).* OJ, L 190/37:1–7, 2008-07-18.
- [194] **MINISTERSTVO ZDRAVOTNICTVÍ ČESKÉ REPUBLIKY:** *Souhrnná informace k celému dokumentu NSEZ.* [online]. 2016 [cit. 2019-07-13]. Dostupné z http://www.nsez.cz/obsah/uvod_3556_31.html, 2016.
- [195] **EUROPEAN PARLIAMENT:** *Directive 2011/24/EU of the European Parliament and of the Council of 9 March 2011 on the application of patients' rights in cross-border healthcare.* OJ, L 88/45:1–21, 2011-04-04.
- [196] **JIRÁKOVÁ, Klára:** *Aktuální stav projektu přeshraničních služeb eHealth.* [online]. 3. duben 2019 [cit. 2019-07-13]. Dostupné z https://www.nixzd.cz/prilohy/90_narodni_kontaktني_misto_pro_eHealth_%20stav_duben_2019.ppt, 2019.
- [197] **MINISTERSTVO ZDRAVOTNICTVÍ ČESKÉ REPUBLIKY:** *Fakta a DASTA.* MZ ČR. [online]. 5. dubna 2019 [cit. 2019-07-14]. Dostupné z <https://www.dastacr.cz/fakta.html>, 2019.
- [198] **ZÁMEČNÍK, Miroslav:** *Přehled stěžejních verzí DASTA.* [online]. 5. duben 2019 [cit. 2019-07-14]. Dostupné z <https://www.dastacr.cz/info-2.html>, 2019.
- [199] **MINISTERSTVO ZDRAVOTNICTVÍ ČESKÉ REPUBLIKY:** *Národní číselník laboratorních položek (NČLP).* MZ ČR. [online]. 5. duben 2019 [cit. 2019-07-14]. Dostupné z <http://dastacr.cz/info-6.html>, 2019.
- [200] **ZÁMEČNÍK, Miroslav:** *Nové směry rozvoje standardů DASTA.* Záměry a projekty MZ ČR v oblasti standardů elektronického zdravotnictví. [online]. 13. listopad 2018 [cit. 2019-07-14]. Dostupné z <http://nsez.cz/Soubor.ashx?souborID=35506&typ=application/vnd.ms-powerpoint&nazev=Nov%C3%A9%20sm%C4%9Bry%20rozvoje%20standard%C5%AF%20DASTA.ppt>, 2018.
- [201] **SEINER, Miroslav a ZÁMEČNÍK, Miroslav:** *Současný vývoj a záměry rozvoje datového standardu MZ ČR.* [online]. 22. červen 2012 [cit. 2019-07-14]. Dostupné z <http://dastacr.cz/doc/Průzkum%20k%20DASTA%20011%20pro%20MEDSOFT.pdf>, 2012.
- [202] **DOLIN, R. H., ALSCHULER, L., BOYER, S., BEEBE, C., BEHLEN, F. M., BIRON, P. V. a SHVO, A. Shabo:** *HL7 clinical document architecture, release 2.* Journal of the American Medical Informatics Association, 13(1):30, 2006. DOI 10.1197/jamia.M1888.

- [203] **DOLIN, R. H., ROGERS, B. a JAFFE, C.:** *Health level seven interoperability strategy: big data, incrementally structured*. *Methods of information in medicine*, 54(01):75–82, 2015. ISSN 0026-1270. DOI 10.3414/me14-01-0030.
- [204] **BENSON, Tim a GRIEVE, Grahame:** *Principles of health interoperability: SNOMED CT, HL7 and FHIR*. Springer, London, 2016, ISBN 978-3-319-30368-0. DOI 10.1007/978-3-319-30370-3.
- [205] **BOONE, Keith W.:** *The CDA TM Book*. Springer London, London, 2014, ISBN 978-1-44-715887-5.
- [206] **SEIDL, Libor:** *Procesy vývoje standardů HL7 a adopce v České republice*. [online]. 19. listopad 2018 [cit. 2019-07-14]. Dostupné z <http://nsez.cz/Soubor.ashx?souborID=35507&typ=application/vnd.openxmlformats-officedocument.presentation.presentation.pptx>, 2018.
- [207] **ZÁMEČNÍK, Miroslav, STRAKA, Luděk a JABOR, Antonín:** *DASTA a projekty e-Health, další rozvoj*. [online]. 1. únor 2013 [cit. 2019-07-14]. Dostupné z <https://www.dastacr.cz/dasta/hypertext/MZAXB.htm>, 2013.
- [208] **ZÁMEČNÍK, Miroslav:** *Standardy a DASTA – z průzkumu ČSZIVI (Pracovní den ČSZIVI, Lékařský dům, 30. 5. 2016. ZIVI ČLS*. [online]. 30. květen 2016 [cit. 2019-07-14]. Dostupné z https://www.dastacr.cz/doc/Pruzkum%20mezi%20tvurci%20IS%20po%20linii%20konference%20DASTA_VYVOJ.pdf, 2016.
- [209] **ZÁMEČNÍK, Miroslav:** *Aktuální otázky k Datovému standardu MZ ČR*. České národní fórum pro eHealth. [online]. 26. dubna 2011 [cit. 2019-07-14]. Dostupné z http://www.ehealthforum.cz/files/dasta_mzcr_2011.pdf, 2011.
- [210] **MINISTERSTVO ZDRAVOTNICTVÍ ČESKÉ REPUBLIKY:** *Seznam novinek a změn v blocích a číselnících v DS 4.16.00*. Datový standard Ministerstva zdravotnictví ČR – verze 4. [online]. 11. červen 2019 [cit. 2019-07-14]. Dostupné z http://ciselniky.dasta.mzcr.cz/CD_DS4/hypertext/DSBFY.htm, 2019.
- [211] **EUROPEAN COMMISSION:** *Commission Recommendation of 6. 2. 2019 on a European Electronic Health Record exchange format*, 2019.
- [212] **NATIONAL ELECTRICAL MANUFACTURERS ASSOCIATION:** *DICOM Standard – Current Edition*. [online]. Virginia, 10. června 2019 [cit. 2019-06-10]. Dostupné z <https://www.dicomstandard.org/current>, 2019.
- [213] **KEMP, Bob, VÄRRI, A., ROSA, A. C., NIELSEN, K. D. a GADE, J.:** *A simple format for exchange of digitized polygraphic recordings*. *Electroencephalography and Clinical Neurophysiology*, 82:391–393, May 1992. ISSN 0013-4694.

- [214] **KEMP, Bob** a **OLIVAN, Jesus**: *European data format 'plus' (EDF+), an EDF alike standard format for the exchange of physiological data*. *Clinical Neurophysiology*, 114:1755–1761, September 2003. ISSN 1388-2457.
- [215] **VELDE, Maarten van de**: *EEProbe Reference Manual*. [online]. 18. březen 2003 [cit. 2019-06-10]. Dostupné z <http://eeprobe.ant-neuro.com/Pdf/EEProbe-manual.pdf>, 2003.
- [216] **MATTMÜLLER, Rudi**: *Inomed takes over EEG systems of Walter Graphtek*. [online]. 25. květen 2012 [cit. 2019-07-15]. Dostupné z https://www.en.inomed.com/fileadmin/media/unternehmen/Presse/Pressemitteilung/inomed_takes_over_EEG_systems_of_Walter_Graphtek.pdf, 2012.
- [217] **ZVÁNOVEC, Vít**: *Převodní tabulka: zákon o ochraně osobních údajů – obecné nařízení o ochraně osobních údajů*, 2017. Úřad pro ochranu osobních údajů. [online] 20. září 2017 [cit. 2019-07-08]. Dostupné z https://www.uouu.cz/assets/File.ashx?id_org=200144&id_dokumenty=38589.
- [218] **ANNAS, George J.** a kol.: *HIPAA regulations – a new era of medical-record privacy?* *New England Journal of Medicine*, 348(15):1486–1490, 4/2003. DOI 10.1056/NEJMLim035027.
- [219] **HIPAA COMPLIANCE ASSISTANCE: Summary of the HIPAA privacy rule**. Office for Civil Rights, 2003. [online]. 13. květen 2003 [cit. 2019-06-10]. Dostupné z <https://www.hhs.gov/sites/default/files/privacysummary.pdf>.
- [220] **LUXTON, David D.**, **KAYL, Robert A.** a **MISHKIND, Matthew C.**: *mHealth data security: The need for HIPAA-compliant standardization*. *Telemedicine and e-Health*, 18(4):284–288, 2012. DOI 10.1089/tmj.2011.0180.
- [221] **FERNÁNDEZ-ALEMÁN, José Luis**, **SEÑOR, Inmaculada Carrión**, **OLIVER LOZOYA, Pedro Ángel** a **TOVAL, Ambrosio**: *Security and privacy in electronic health records: A systematic literature review*. *Journal of Biomedical Informatics*, 46(3):541–562, 2013. ISSN 1532-0464. DOI 10.1016/j.jbi.2012.12.003. Dostupné z <http://www.sciencedirect.com/science/article/pii/S1532046412001864>.
- [222] **DANCIU, Ioana**, **COWAN, James D.**, **BASFORD, Melissa**, **WANG, Xiaoming**, **SAIP, Alexander**, **OSGOOD, Susan**, **SHIREY-RICE, Jana**, **KIRBY, Jacqueline** a **HARRIS, Paul A.**: *Secondary use of clinical data: The Vanderbilt approach*. *Journal of Biomedical Informatics*, 52:28–35, 2014. Special Section: Methods in Clinical Research Informatics. ISSN 1532-0464. DOI 10.1016/j.jbi.2014.02.003. Dostupné z <http://www.sciencedirect.com/science/article/pii/S1532046414000392>.
- [223] **CHASSANG, Gauthier**: *The impact of the EU general data protection regulation on scientific research*. *E-cancer Medical Science*, 11(709), 2017. DOI 10.3332/ecancer.2017.709.

- [224] **EUROPEAN UNION AGENCY FOR FUNDAMENTAL RIGHTS AND COUNCIL OF EUROPE: Handbook on European data protection law: 2018 edition.** Publication Office of the European Union, Luxembourg, 2018, ISBN 978-9-29-491901-4. DOI 10.2811/343461. [online]. 16. květen 2018 [cit. 2019-07-13]. Dostupné z https://fra.europa.eu/sites/default/files/fra_uploads/fra-coe-edps-2018-handbook-data-protection_en.pdf.
- [225] **SWEENEY, Latanya: *k-anonymity: A model for protecting privacy.*** International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, 10(05):557–570, 2002. DOI 10.1142/S0218488502001648.
- [226] **LEFEVRE, Kristen, DEWITT, David J a RAMAKRISHNAN, Raghu: Mondrian multidimensional *k-anonymity.*** V *Proceedings of the 22nd International Conference on Data Engineering (ICDE)*, svazek 6, strany 1–11, 2006, ISBN 0-7695-2570-9. DOI 10.1109/ICDE.2006.101.
- [227] **ROHAN, Vladimir, BAXA, Jan, TUPY, Radek, CERNA, Lenka, SEVCIK, Petr, FRIESL, Michal, POLIVKA JR, Jiri, POLIVKA, Jiri a FERDA, Jiri: Length of occlusion predicts recanalization and outcome after intravenous thrombolysis in middle cerebral artery stroke.** Stroke, 45(7):2010–2017, 2014. DOI 10.1161/STROKEAHA.114.005731.
- [228] **KÖCHER, Martin, ŠAŇÁK, Daniel, ZAPLETALOVÁ, Jana, CIHLÁŘ, Filip, CZERNÝ, Daniel, ČERNÍK, David, DURAS, Petr, ENDRYCH, Ladislav, HERZIG, Roman, LACMAN, Jiří, LOJÍK, Miroslav, OSTRÝ, Svatopluk, PÁDR, Radek, ROHAN, Vladimír, ŠKORŇA, Miroslav, ŠRÁMEK, Martin, ŠTĚRBA, Luděk, VÁCLAVÍK, Daniel, VANÍČEK, Jiří, VOLNÝ, Ondřej a TOMEK, Aleš: Mechanical Thrombectomy for Acute Ischemic Stroke in Czech Republic: Technical Results from the Year 2016.** Cardiovascular and interventional radiology, strany 1901–1908, 2018, ISSN 0174-1551. DOI 10.1007/s00270-018-2068-z.
- [229] **TRUELSEN, T, PIECHOWSKI-JÓŹWIAK, B, BONITA, R, MATHERS, C, BOGOUSLAVSKY, J a BOYSEN, G: Stroke incidence and prevalence in Europe: a review of available data.** European Journal of Neurology, 13(6):581–598, 2006. DOI 10.1111/j.1468-1331.2006.01138.x.
- [230] **KATAN, Mira a LUFT, Andreas: Global burden of stroke.** V *Seminars in neurology*, svazek 38, strany 208–211. Thieme Medical Publishers, 2018. DOI 10.1055/s-0038-1649503.
- [231] **HODGE, Victoria, JESSOP, Mark, FLETCHER, Martyn, WEEKS, Michael, TURNER, Aaron, JACKSON, Tom, INGRAM, Colin, SMITH, Leslie a AUSTIN, Jim: A Digital Repository and Execution Platform for Interactive Scholarly Publications in Neuroscience.** Neuroinformatics, 14(1):23–40, 2016. DOI 10.1007/s12021-015-9276-3.
- [232] **EGLLEN, Stephen John, WEEKS, Michael, JESSOP, Mark, SIMONOTTO, Jennifer, JACKSON, Tom a SERNAGOR, Evelyne: A data repository and analysis framework for spontaneous neural activity recordings in developing retina.** Gigascience, 3(1):3, 2014. DOI 10.1186/2047-217X-3-3.

- [233] **WEEKS, Michael, JESSOP, Mark, FLETCHER, Martyn, HODGE, Victoria, JACKSON, Tom a AUSTIN, Jim:** *The CARMEN software as a service infrastructure*. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 371(1983), 2013, ISSN 1364-503X. DOI 10.1098/rsta.2012.0080.
- [234] **MIKULIK, Robert, CASO, Valeria a WAHLGREN, Nils:** *Past and future of stroke care in Europe*. Oruen–CNS Journal, 2:19–26, 2017, ISSN 2059-2442.
- [235] **MIKULIK, Robert, BAR, M., GRECU, A., HERZIG, R., NEUMANN, J., SANAK, D., SKODA, O., SKOLOUDIK, D., SVOBODOVA, V., TOMEK, A. a kol.:** *The registry of stroke care quality (RES-Q): The first nation-wide data on stroke care quality*. Journal of the Neurological Sciences, 381:91, 2017. DOI 10.1016/j.jns.2017.08.302.
- [236] **POLIVKA, Jiri, POLIVKA, Jiri (jr.) a ROHAN, Vladimir:** *Predictive and individualized management of stroke—success story in Czech Republic*. EPMA Journal, 9(4):393–401, 2018. DOI 10.1007/s13167-018-0150-x.
- [237] **NORRVING, Bo, BRAY, Benjamin D., ASPLUND, Kjell, HEUSCHMANN, Peter, LANGHORNE, Peter, RUDD, Anthony G., WAGNER, Markus, WIEDMANN, Silke a WOLFE, Charles D. A.:** *Cross-national key performance measures of the quality of acute stroke care in Western Europe*. Stroke, 46(10):2891–2895, 2015. DOI 10.1161/STROKEAHA.115.008811.
- [238] **WAHLGREN, Nils, AHMED, Niaz, DÁVALOS, Antoni, FORD, Gary A, GROND, Martin, HACKE, Werner, HENNERICI, Michael G, KASTE, Markku, KUELKENS, Sonja, LARRUE, Vincent, LEES, Kennedy R, ROINE, Risto O, SOINNE, Lauri, TONI, Danilo a VANHOOREN, Geert:** *Thrombolysis with alteplase for acute ischaemic stroke in the Safe Implementation of Thrombolysis in Stroke-Monitoring Study (SITS-MOST): an observational study*. The Lancet, 369(9558):275–282, 2007. ISSN 0140-6736. DOI 10.1016/S0140-6736(07)60149-4.
- [239] **MAZYA, Michael, EGIDO, José A., FORD, Gary A., LEES, Kennedy R., MIKULIK, Robert, TONI, Danilo, WAHLGREN, Nils a AHMED, Niaz:** *Predicting the risk of symptomatic intracerebral hemorrhage in ischemic stroke treated with intravenous alteplase: safe Implementation of Treatments in Stroke (SITS) symptomatic intracerebral hemorrhage risk score*. Stroke, 43(6):1524–1531, 2012. DOI 10.1161/STROKEAHA.111.644815.
- [240] **SAFE IMPLEMENTATION OF THROMBOLYSIS IN STROKE: SITS International – reducing the global burden of stroke.** [online]. 8. únor 2019 [cit. 2019-07-18]. Dostupné z https://sitsinternational.org/media/1407/sits-presentation_20190207.pdf, 2019.

- [241] **NOONAN, Vanessa Karine, KWON, Brian K., SORIL, Lesley, FEHLINGS, Michael G., HURLBERT, Jeanne, TOWNSON, Andrea, JOHNSON, M. a DVORAK, Marcel:** *The Rick Hansen spinal cord injury registry (RHSCIR): A national patient-registry*. *Spinal cord*, 50(1):22, 2012. DOI 10.1038/sc.2011.109.
- [242] **THIBAUT-HALMAN, Ginette, RIVERS, Carly S., BAILEY, Christopher S., TSAI, Eve C., DREW, Brian, NOONAN, Vanessa K., FEHLINGS, Michael G., DVORAK, Marcel F., DILINUERKUEBAN, KWON, Brian K. a kol.:** *Predicting recruitment feasibility for acute spinal cord injury clinical trials in Canada using national registry data*. *Journal of neurotrauma*, 34(3):599–606, 2017. DOI 10.1089/neu.2016.4568.
- [243] **PHAN, Philippe, BUDHRAM, Brandon, ZHANG, Qiong, RIVERS, Carly S., NOONAN, Vanessa K., PLASHKES, Tova, WAI, Eugene K., PAQUET, Jérôme, ROFFEY, Darren M., TSAI, Eve a kol.:** *Highlighting discrepancies in walking prediction accuracy for patients with traumatic spinal cord injury: an evaluation of validated prediction models using a Canadian Multicenter Spinal Cord Injury Registry*. *The Spine Journal*, 19(4):703–710, 2019. DOI 10.1016/j.spinee.2018.08.016.
- [244] **TANSLEY, Robert, BASS, Mick a SMITH, MacKenzie:** *DSpace as an Open Archival Information System: Current Status and Future Directions*. V **KOCH, Traugott a SØLVBERG, Ingeborg Torvik** (editoři): *Research and Advanced Technology for Digital Libraries*, svazek 2769 z *Lecture Notes in Computer Science*, strany 446–460. Springer Berlin Heidelberg, Trondheim, Norway, 2003, ISBN 978-3-540-40726-3. DOI 10.1007/978-3-540-45175-4_41.
- [245] **SMITH, MacKenzie, RODGERS, Richard, WALKER, Julie a TANSLEY, Robert:** *DSpace: A Year in the Life of an Open Source Digital Repository System*. V **HEERY, Rachel a LYON, Liz** (editoři): *Research and Advanced Technology for Digital Libraries*, svazek 3232 z *Lecture Notes in Computer Science*, strany 38–44. Springer Berlin Heidelberg, Bath, UK, September 2004, ISBN 978-3-540-23013-7. DOI 10.1007/978-3-540-30230-8_4.
- [246] **PATEL, Bhavesh:** *DSpace Diagram and Process Flow*, 2012. [online]. 8. leden 2012 [cit. 2019-07-06]. Dostupné z <https://issuu.com/bhaveshpatel.info/docs/dspace-diagram>.
- [247] **CORTESE, Claudio, BOLLINI, Andrea, GROppo, Emilia a MORNATI, Susanna:** *Extending DSpace to fulfil the requirements of Digital Libraries for Cultural Heritage Management*, 1/2017. [online]. 24. leden 2017 [cit. 2019-07-08].
- [248] **BECKER, Pascal Nicolas:** *Illuminating DSpace's Linked Data Support*, 06/2016. [online]. 13. červen 2016 [cit. 2019-07-08]. Dostupné z https://www.the-library-code.de/events/or16/Illuminating_DSpaces_LD_support.pdf.

- [249] **KAMPFMEYER, Ulrich:** *ECM: Enterprise Content Management*. Project Consult, Köln, 2006, ISBN 978-3-93-653409-8.
- [250] **ORGANIZATION FOR THE ADVANCEMENT OF STRUCTURED INFORMATION STANDARDS:** *Content Management Interoperability Services (CMIS) Version 1.1 Plus Errata 01: OASIS Standard Incorporating Approved Errata 01*. [online]. 19. září 2015 [cit. 2019-07-18]. Dostupné z <http://docs.oasis-open.org/cmisis/CMIS/v1.1/errata01/os/CMIS-v1.1-errata01-os-complete.pdf>, 2015.
- [251] **MARIN, Mike A. a BROWN, Jay A.:** *Implementing a Case Management Modeling and Notation (CMMN) System using a Content Management Interoperability Services (CMIS) compliant repository*. ArXiv, 2015. DOI 10.13140/RG.2.1.3214.9927.
- [252] **BERGLJUNG, Martin:** *Alfresco CMIS*. Packt Publishing Ltd, 2014, ISBN 978-1-78-216352-7.
- [253] **MULLER, Florian, BROWN, Jay a POTTS, Jeff:** *CMIS and Apache Chemistry in action*. Manning Publ., 7/2013, ISBN 978-16-1729-1159.
- [254] **WESTERLUND, Parvaneh, ANDERSSON, Ingemar, PÄIVÄRINTA, Tero a NILSSON, Jörgen:** *Towards automated pre-ingest workflow for bridging information systems and digital preservation services*. Records Management Journal, 2019. DOI 10.1108/RMJ-05-2018-0011.
- [255] **RICK HANSEN INSTITUTE:** *Secure medical research online web application for data collection*. [online]. Canada, 19. červenec 2019 [cit. 2019-07-19]. Dostupné z <https://www.rhigrp.net/>, 2019.
- [256] **JONES, Talon, MILLER, Rebecca, STREET, John T. a SAWATZKY, Bonita:** *Validation of the Oswestry Disability Index for pain and disability in arthrogryposis multiplex congenita*. Annals of physical and rehabilitation medicine, 62(2):92–97, 2019, ISSN 1877-0657. DOI 10.1016/j.rehab.2018.05.1319.
- [257] **NOONAN, Vanessa K., CHAN, Elaine, BASSETT-SPIERS, Kent, BERLOWITZ, David J., BIERING-SØRENSEN, Fin, CHARLIFUE, Susan, GRACO, Marnie, HAYES, Keith C., HORSEWELL, Jane a JOSHI, Phalgun:** *Facilitators and barriers to international collaboration in spinal cord injury: results from a survey of clinicians and researchers*. Journal of neurotrauma, 35(3):478–485, 2018. DOI 10.1089/neu.2017.5036.
- [258] **PŘEMYSL, Janíček, JIŘÍ, Marek a kol.:** *Expertní inženýrství v systémovém pojetí*. Grada Publishing, 2013. ISBN: 978-80-247-4127-7.
- [259] **JAROŠ, Patrik:** *Převodník Excel – OWL*, 2018. Bakalářská práce, Západočeská univerzita v Plzni, Fakulta aplikovaných věd, Plzeň, 2018.
- [260] **KRYL, Martin:** *Získání a uchování metadat z heterogenních dat*. Diplomová práce, Západočeská univerzita v Plzni, Fakulta aplikovaných věd, Plzeň, 2015.

- [261] **OPENEHR FOUNDATION:** *openEHR Architecture Overview*. [online]. openEHR Foundation. 14. prosinec 2018 [cit. 2019-11-30]. Dostupné z https://specifications.openehr.org/releases/BASE/latest/architecture_overview.html, 2018.
- [262] **KAHN, Jr, Charles, CHANNIN, David a RUBIN, Daniel:** *An ontology for PACS integration*. Journal of digital imaging: the official journal of the Society for Computer Applications in Radiology, 19:316–27, 01/2007. DOI 10.1007/s10278-006-0627-3.
- [263] **KAHN, Charles Jr., LANGLOTZ, Curtis, CHANNIN, David a RUBIN, Daniel:** *Informatics in Radiology: An Information Model of the DICOM Standard*. Radiographics : a review publication of the Radiological Society of North America, Inc, 31:295–304, 10/2010. DOI 10.1148/rg.311105085.
- [264] **RODRÍGUEZ GONZÁLEZ, David, CARPENTER, Trevor, HEMERT, JanoĽ. a WARDLAW, Joanna:** *An open source toolkit for medical imaging de-identification*. European Radiology, 20:1896–1904, 2010. ISSN 0938-7994. DOI 10.1007/s00330-010-1745-3.
- [265] **ŠMUCR, Jan:** *Grafická tvorba SPARQL*. Diplomová práce, Západočeská univerzita v Plzni, Fakulta aplikovaných věd, Plzeň, 2014.
- [266] **SOBĚHART, Michel:** *Sparkle – rozšíření textového editoru SPARQL dotazů*, 2014. Oborový projekt. Západočeská univerzita. Fakulta aplikovaných věd, Plzeň, 2014.
- [267] **KAZÁK, Josef:** *Sparkle – rozšíření nástroje pro tvorbu SPARQL dotazů*. Diplomová práce, Západočeská univerzita v Plzni, Fakulta aplikovaných věd, Plzeň, 2017.
- [268] **HLAVÁČOVÁ, Klára:** *Rozšíření Sparkle o podporu SPARQL Endpointu a využití titulků při tvorbě dotazu*. Diplomová práce, Západočeská univerzita v Plzni, Fakulta aplikovaných věd, Plzeň, 2018.
- [269] **KRUŽEJ, Martin:** *Rozšíření Sparkle o grafickou tvorbu a vizualizaci dotazů*. Diplomová práce, Západočeská univerzita v Plzni, Fakulta aplikovaných věd, Plzeň, 2019.
- [270] **ANDARGOLI, Amirhossein Eslami, SCHEEPERS, Helana, RAJENDRAN, Diana a SOHAL, Amrik:** *Health information systems evaluation frameworks: A systematic review*. International Journal of Medical Informatics, 97:195–209, 2017. ISSN 1386-5056. DOI 10.1016/j.ijmedinf.2016.10.008. Dostupné z <http://www.sciencedirect.com/science/article/pii/S1386505616302234>.
- [271] **GESULGA, Jaillah Mae, BERJAME, Almarie, MOQUIALA, Kristelle Sheen a GALIDO, Adrian:** *Barriers to Electronic Health Record System Implementation and Information Systems Resources: A Structured Review*. Procedia Computer Science, 124:544–551, 2017. 4th Information Systems International Conference 2017, ISICO 2017, 6–8 November 2017, Bali,

Indonesia. ISSN 1877-0509. DOI 10.1016/j.procs.2017.12.188. Dostupné z <http://www.sciencedirect.com/science/article/pii/S1877050917329563>.

- [272] **ABRIL-GONZALEZ, Mauricio, PORTILLA, Fernando A a JARAMILLO-MEJIA, Marta C:** *Standard Health Level Seven for Odontological Digital Imaging*. *Telemedicine and e-Health*, 23(1):63–70, 2017. DOI 10.1089/tmj.2015.0251.
- [273] **THURSTON, Janet:** *Meaningful Use of Electronic Health Records*. *The Journal for Nurse Practitioners*, 10(7):510–513, 2014. ISSN 1555-4155. DOI 10.1016/j.nurpra.2014.05.012.
- [274] **BEN-ASSULI, Ofir:** *Electronic health records, adoption, quality of care, legal and privacy issues and their implementation in emergency departments*. *Health Policy*, 119(3):287–297, 2015. ISSN 0168-8510. DOI 10.1016/j.healthpol.2014.11.014.
- [275] **VUOKKO, Riikka, MÄKELÄ-BENGS, Päivi, HYPÖNEN, Hannele, LINDQVIST, Minna a DOUPI, Persephone:** *Impacts of structuring the electronic health record: Results of a systematic literature review from the perspective of secondary use of patient data*. *International Journal of Medical Informatics*, 97:293–303, 2017. ISSN 1386-5056. DOI 10.1016/j.ijmedinf.2016.10.004.
- [276] **MARTIN-RODILLA, Patricia a GONZALEZ-PEREZ, Cesar:** *Metainformation scenarios in Digital Humanities: Characterization and conceptual modelling strategies*. *Information Systems*, 84:29 – 48, 2019, ISSN 0306-4379. DOI 10.1016/j.is.2019.04.009.
- [277] **HALPIN, Harry:** *Semantic Insecurity: Security and the Semantic Web*. V *Privacy and the Semantic Web – Policy and Technology*, strany 1–10, 2017. [online] PrivOn 2017 – Workshop Society, Privacy and the Semantic Web – Policy and Technology, 29. prosinec 2017, Vienna, Austria, [cit. 2019-06-10]. Dostupné z <https://hal.inria.fr/hal-01673291>.
- [278] **KIRRANE, Sabrina, VILLATA, Serena a D'AQUIN, Mathieu:** *Privacy, security and policies: A review of problems and solutions with semantic web technologies*. *Semantic Web*, 9(2):153–161, 2018. DOI 10.3233/SW-180289.
- [279] **O'RIAIN, Seán, HARTH, Andreas a CURRY, Edward:** *Linked data driven information systems as an enabler for integrating financial data*. V *Information Systems for Global Financial Markets: Emerging Developments and Effects*, strany 239–270. IGI Global, 2012. ISBN 978-1613501627. DOI 10.4018/978-1-61350-162-7.ch010.
- [280] **NARDON, FB a MOURA, LA:** *Knowledge sharing and information integration in healthcare using ontologies and deductive databases*. V **FIESCHI, M, COIERA, E a LI, YCJ** (editoři): *MEDINFO 2004: Proceedings of the 11th world congress on medical informatics, pt 1 and 2*, svazek 107 z *Studies in health technology and informatics*, strany 62–66.

Int Med Informat Assoc, 2004. 11th World Congress on Medical Informatics, Amer Med Informat Assoc, San Francisco, CA, Sep 07-11, 2004. ISBN 1-58603-444-8. ISSN 0926-9630.

- [281] **LINDEMANN, Gabriela, SCHMIDT, Danilo, SCHRADER, Thomas a KEUNE, Dietmar**: *The Resource Description Framework (RDF) as a Modern Structure for Medical Data*. V **ARDIL, C** (editor): *Proceedings of World Academy of Science, Engineering and Technology*, svazek 25 z *Proceedings of World Academy of Science Engineering and Technology*, strany 240–243, 2007. ISSN 1307-6884. Venice, Italy, Nov 23–25, 2007.
- [282] **MUECKE, Roland, LOEBE, Matthias, KNUTH, Magnus a LOEBE, Frank**: *A Semantic Model for Representing Items in Clinical Trials*. V *2009 22nd IEEE International Symposium on Computer-based Medical Systems*, IEEE International Symposium on Computer-Based Medical Systems, strany 382–389. IEEE, 2009, ISBN 978-1-4244-4879-1. ISSN 1063-7125. Albuquerque, NM, AUG 03-04, 2009.
- [283] **BHATT, Mehul, RAHAYU, Wenny, SONI, Sury Prakash a WOUTERS, Carlo**: *Ontology driven semantic profiling and retrieval in medical information systems*. *Journal of web semantics*, 7(4):317–331, ECDEC 2009, ISSN 1570-8268.
- [284] **CIUCIU, Ioana, KANG, Han, MEERSMAN, Robert, SCHMID, Jerome, MAGNENAT-THALMANN, Nadia, GUITIAN, Jose Antonio Iglesias a GOBETTI, Enrico**: *Collaborative Semantic Content Management: An Ongoing Case Study for Imaging Applications*. V **TOME, E** (editor): *Proceedings of the 11th European Conference on Knowledge Management*, svazek 2, strany 257–267, 2010, ISBN 978-1-906638-70-2. Univ Lusiada Vila Nova Famalicao, Famalicao, Portugal, Sep 02-03, 2010.
- [285] **PIERCE, Christopher D., BOOTH, David, OGBUJI, Chimezie, DEATON, Chris, BLACKSTONE, Eugene a LENAT, Doug**: *SemanticDB: A Semantic Web Infrastructure for Clinical Research and Quality Reporting*. *Current Bioinformatics*, 7(3):267–277, 9/2012, ISSN 1574-8936.
- [286] **WORLD WIDE WEB CONSORTIUM**: *Health Care and Life Science (HCLS) Linked Data Guide*, 2012. [online]. 19. listopad 2012 [cit. 2019-06-15, Dostupné z <https://www.w3.org/2001/sw/hcls/notes/hcls-rdf-guide/>].
- [287] **LANDWEHRMEYER, Georg B., FITZER-ATTAS, Cheryl J., GIULIANO, Joseph D., GONÇALVES, Nilza, ANDERSON, Karen E., CARDOSO, Francisco, FERREIRA, Joaquim J., MESTRE, Tiago A., STOUT, Julie C. a SAMPAIO, Cristina**: *Data analytics from Enroll-HD, a global clinical research platform for Huntington's disease*. *Movement disorders clinical practice*, 4(2):212–224, 2017. DOI 10.1002/mdc3.12388.

- [288] **MANDEL, Joshua C, KREDA, David A, MANDL, Kenneth D, KOHANE, Isaac S a RAMONI, Rachel B:** *SMART on FHIR: a standards-based, interoperable apps platform for electronic health records.* Journal of the American Medical Informatics Association, 23(5):899–908, 2016. DOI 10.1093/jamia/ocv189.
- [289] **BENSON, Tim:** *Principles of Health Interoperability HL7 and SNOMED.* Health Information Technology Standards Series. Springer London, London, April 2012, ISBN 978-1-44-712800-7.
- [290] **BENSON, Tim a GRIEVE, Grahame:** *Principles of FHIR. V Principles of Health Interoperability,* strany 329–348. Springer, 2016. ISBN 978-3-319-30370-3. ISSN 2199-2517. DOI 10.1007/978-3-319-30370-3.
- [291] **KUO, Jessica Wan Yi a KUO, Alex Mu Hsing:** *Integration of Health Information Systems Using HL7: A Case Study. V ITCH,* strany 188–194. IOS Press, 2017. DOI 10.3233/978-1-61499-742-9-188.

Seznam obrázků

3.1	Funkční model OAIS	25
4.1	RDF trojice	28
5.1	Data a vztahy pro léčbu a sledování pacienta po CMP	45
5.2	Data a jejich vztahy pro diagnostiku IBD	46
5.3	Diagram funkcí DSpace	49
7.1	Diagram ontologií platformy MRE	56
7.2	MRE Ontology	57
7.3	DASTA Ontology	61
7.4	DICOM Ontology	63
7.5	Population Ontology	65
7.6	Stroke Ontology	67
7.7	IBDS Ontology	68
7.8	IBDT Ontology	69
7.9	TDB Ontology	70
8.1	Schéma výzkumné platformy	71
8.2	Proces zpracování dat z vnějšího vstupního bodu	76
8.3	Proces ochrany osobních údajů	77
8.4	Interakce platformy s okolím	80
B.1	Tabulka pro návrh tříd a atributů	142
B.2	Tabulka pro návrh slovníků	143
D.1	Tvorba dotazu prostřednictvím formulářových prvků	146
D.2	Textový editor se zvýrazněním syntaxe	146
D.3	Vizualizace dotazu	147
D.4	Výsledky dotazu s texty z ontologie místo URI/IRI	147
D.5	Vývojový diagram generátoru webových formulářů	148
D.6	Z ontologie dynamicky generovaný formulář	149
D.7	Modální okno pro editaci instance v generovaném formuláři	149
D.8	Tvorba SPARQL pro temporální data	150
D.9	DataTable s ontologiemi a RDF daty	150
E.1	De-identifikace DICOM snímku s mapou CT perfuze 1	151
E.2	De-identifikace DICOM snímku s mapou CT perfuze 2	152

Seznam tabulek

9.1	Rozsah dat a metadat archivu dle projektu a typu dat	85
9.2	Rychlosti zpracování metadat.	87
9.3	Rychlost importu metadat v RDF	88

Výpisy

3.1	Příklady URI, URL a URN	8
4.1	Syntax RDF/XML	30
4.2	Syntax Notation 3	30
4.3	SPARQL	31
7.1	Schéma jmenných prostorů ontologií v MRE	55
7.2	Schéma URI/IRI pro RDF zdroje v MRE	55
8.1	Schéma a příklady adresářové struktury archivu	73
A.1	Používané existující ontologie a RDF slovníky	140
A.2	Jmenné prostory ontologií platformy MRE a řešených projektů . .	140
B.1	Příklad hodnoty číselníku NIHSS pro postižení zraku	141
C.1	Číselník ÚZIS pro pohlaví	144
C.2	RDF data pro položku pohlaví muž z číselníku ÚZIS	144
D.1	Získání atributů z DICOM Ontology dotazem	145
D.2	Získání referenčního informačního modelu z DICOM Ontology . .	145

Seznam zkratek

ACL	Access Control List
ACR	American College of Radiology
AIC	Archival Information Collection (OAIS)
AIP	Archival Information Package (OAIS)
AIU	Archival Information Unit (OAIS)
ANSI	American National Standards Institute
API	Application Programming Interface
ASCII	American Standard Code for Information Interchange
AVI	Audio Video Interleave
BIDS	Brain Imaging Data Structure
BLOB	Binary Large Object
CBF	Cerebral blood flow
CBV	Cerebral blood volume
CCSDS	Consultative Committee for Space Data Systems
CDA	Clinical Document Architecture
CERIF	Common European Research Information Format
CIS	Customs Information System
CISOB	Číselník obcí
CLI	Command Line Interface
CLOB	Character Large Object
CMIS	Content Management Interoperability Services
CMMN	Case Management Modeling and Notation
CMP	Cévní mozková příhoda
CMS	Content Management System
CPU	Centrální procesorová jednotka (Central Processing Unit)
CRIS	Current Research Information Systems
CRON	Systémový nástroj v Linuxu/Unixu pro spouštění úloh ve stanoveném čase a intervalu

CRP	Laboratorní vyšetření C-reaktivní protein
CRUD	Create, Read, Update, Delete
CSCNS	Cerebrovaskulární sekce České neurologické společnosti
CSS	Cascading Style Sheets
CSV	Comma Separated Values
CT	Počítačová tomografie
CURL	Consortium of University Research Libraries
CWA	Closed World Assumption
DAML	DARPA Agent Markup Language
DAO	Data Access Object
DARPA	Defense Advanced Research Projects Agency
DASTA	Datový standard
DC	Dublin Core
DCAT	Data Catalog Vocabulary
DCC	Digital Curation Centre
DCMI	Dublin Core Metadata Initiative
DFD	Data Flow Diagram
DICOM	Digital Imaging and Communications in Medicine
DIP	Dissemination Information Package (OAIS)
DL	Description Logic
DLP	Description Logic Programs
DO	DICOM Ontology
DOGMA	Developing Ontology-Grounded Methods and Applications
DOI	Digital Object Identifier
DRBD	Distributed Replicated Block Device
DS	Data Standard
DS	Datový standard
DSA	Data Seal of Approval
DTD	Document Type Definition

DWC	Darwin Core
EAC	Encoded Archival Context
EAD	Encoded Archival Description
ECHO	Echokardiografické vyšetření
ECM	Enterprise Content Management
EDF	European Data Format
EDM	Europeana Data Model
EDPS	European Data Protection Supervisor
EEG	Elektroencefalogram
EHR	Electronic Health Record
EIF	European Interoperability Framework
EKG	Elektrokardiogram
EML	Ecological Metadata Language
EMR	Electronic medical record
EOF	End of File
ESO-EAST	European Stroke Organisation – Enhancing and Accelerating Stroke Treatment
ETL	Export, Transform, Load
EU	Evropská unie
EURODAC	European Dactyloscopy or fingerprint identification
EUROPOL	Evropský policejní úřad
EXE	Formát spustitelného souboru pod operačními systémy Microsoft Windows
EXIF	EXchangeable Image File format
EZD	elektronická zdravotní dokumentace
FGDC	Federal Geographic Data Committee
FHIR	Fast Health Interoperability Resources
FLAC	Free Lossless Audio Codec
FOAF	Friend of a Friend

FRA	European Union Agency for Fundamental Rights and Council of Europe
FRBR	Functional Requirements for Bibliographic Records
FW	Laboratorní vyšetření sedimentace erytrocytů (Fahræus Westergren)
GCI	General Concept Inclusion axioms in the ontology
GDPR	obecné nařzení o ochraně osobních údajů (General Data Protection Regulation)
GPG	GNU Privacy Guard
GRP	Global Research Platform
GUI	Graphical User Interface
HCLS	Health Care and Life Science
HDD	Hard Disc Drive
HDF	HL7 Version 3 Development Framework
HIE	health information exchange
HIPAA	Health Insurance Portability and Accountability Act
HIS	Zdravotnický informační systém (health information systems)
HL7	Health Level Seven
HSTS	HTTP Strict Transport Security
HTML	HyperText Markup Language
HTML5	HyperText Markup Language version 5
HTTP	Hypertext Transfer Protocol
HTTPS	Hypertext Transfer Protocol Secure
HW	Hardware
IBD	Inflammatory Bowel Disease
ICD	International Classification of Diseases, viz MKN
IEC	International Electrotechnical Commission
IFCC	International Federation of Clinical Chemistry and Laboratory Medicine
IMIS	Internal Market Information System
IRI	Internationalized Resource Identifier

IS	Informační systém
ISA	Interoperability solutions for public administrations, businesses and citizens
ISAD	International Standard Archival Description
ISBN	International Standard Book Number
ISC	International Science Council
ISO	International Organization for Standardization
ISSN	International Standard Serial Number
JDBC	Java Database Connectivity
JPEG	Joint Photographic Experts Group
JS	JavaScript
JSON	JavaScript Object Notation
KIF	Knowledge Interchange Format
KU	klinická událost
KVM	Kernel-based Virtual Machine
LAU	Místní samosprávné jednotky
LD	Linked Data
LDCI	Linked Data Competency Index
LOD	Linked Open Data
LODD	Linked Open Drug Data
LOINC	Logical Observation Identifiers Names and Codes
LOM	Learning Object Metadata
LZW	Lempel-Ziv-Welch bezztrátový kompresní algoritmus
MAD	Metadata Authority Description
MADS	Metadata Authority Description Schema
MARC	MAchine-Readable Cataloging
MDR	Registr metadat (Metadata Registry)
MeSH	Medical Subject Headings
METS	Metadata Encoding and Transmission Standard

MIME	Internet Media Types
MIP	Maximum Intensity Projection
MKN	Mezinárodní kód nemoci
MODS	Metadata Object Description Schema
MPEG	Moving Picture Experts Group
MR	Magnetická rezonance
MRE	Medical Research & Education
MRI	Magnetic Resonance Imaging
MS	Microsoft
MSH	Message Header (HL7)
MTT	Mean Transit Time
MZ	Ministerstvo zdravotnictví
N-TRIPLE	řádkový čistě textový formát pro reprezentaci RDF/XML dat
N3	zkratka člověku přívětivé non-XML serializace RDF modelu
NCAL	Nepomuk Calendar Ontology
NCO	Nepomuk Contact Ontology
NEMA	National Electrical Manufacturers Association
NEXIF	Nepomuk EXIF Ontology
NFO	Nepomuk File Ontology
NID3	Nepomuk ID3 Ontology
NIE	Nepomuk Information Element Ontology
NIH	National Institutes of Health
NIHSS	National Institutes of Health – Stroke Scale
NIS	Nemocniční informační systém
NISO	National Information Standards Organization
NIX-ZD	Zavedení přeshraničních služeb eHealth v České republice
NKOD	Národní katalog otevřených dat
NMO	Nepomuk Message Ontology
NSDL	National Science Digital Library

NUTS	Nomenklatura územních statistických jednotek
NZIS	Národní zdravotnický informační systém
NČLP	Národní číselník laboratorních položek
OAI	Open Archives Initiative
OAI-PMH	Open Archives Initiative Protocol for Metadata Harvesting
OAIS	Open Archival Information System
OASIS	Organization for the Advancement of Structured Information Standards
OBO	Open Biological and Biomedical Ontologies
ODBC	Open Database Connectivity
OIL	Ontology Inference Layer
ONIX	ONline Information eXchange
OSI	Open Systems Interconnection
OWA	Open World Asumption
OWL	Web Ontology Language
PACS	Picture Archiving and Communication System
PAIMAS	Producer-Archive Interface Methodology Abstract Standard
PDF	Portable Document Format
PET	Positron Emission Tomography
PNG	Portable Network Graphics
POJO	Plain old Java object
PREMIS	PREservation Metadata: Implementation Strategies
PROV	Provenance ontology
PSG	Polysomnographie
PŠČ	poštovní směrovací číslo
PURL	Persistent URL
RAID	Redundant Array of Independent Disks
RAM	Random-Access-Memory
RDA	Resource Description and Access

RDF	Resource Description Framework
RDFS	RDF Schema
RDQL	Dotazovací jazyk pro RDF
RES-Q	REgistry of Stroke Care Quality
REST	Representational State Transfer
RFC	Request for Comments
RFID	Radio Frequency Identification
RHSCIR	Rick Hansen Spinal Cord Injury Registry
RIF	Rule Interchange Format
RIM	Reference Information Model
RLE	Run-Length Encoding
RLG	Research Libraries Group
ROM	Read-Only Memory
RQL	RDF Query Language
RTF	Rich Text Format
RTG	rentgen, rentgenový
SaaS	Software as a Service
SADL	Semantic Application Design Language
SANSA	Semantic Analytics Stack
SAP	Systémy, Aplikace a Produkty v oblasti zpracování dat
SCI	Spinal Cord Injury
SEDI	SEmantic DIcom
SGML	Standard Generalized Markup Language
SGS	Studentská grantová soutěž
SHA	Secure Hash Algorithm
SI	Mezinárodní systém jednotek, z francouzského „Le Système International d’Unités“
SIP	Submission Information Package (OAIS)
SIS	Schengen Information System

SITS	Safe Implementation of Treatments in Stroke
SKOS	Simple Knowledge Organization System
SLIVER07	Segmentation of the Liver Competition 2007
SMART	Substitutable Medical Applications and Reusable Technologies
SMB	Server Message Block
SNOMED CT	Systematized Nomenclature of Medicine – Clinical Terms
SOAP	Simple Object Access Protocol
SONO	Sonography
SPARQL	rekurzivní akronym pro SPARQL Protocol and RDF Query Language
SPSS	Statistical Package for the Social Sciences
SQL	Structured Query Language
SSD	Solid State Disc
SSL	Secure Sockets Layer
STL	Standard Triangle Language, stereolitografie
SŘBD	system řízení báze dat
TCIA	The Cancer Imaging Archive
TDB	Tissue Database
TDKIV	Česká terminologická databáze knihovnictví a informační vědy
TGN	Thesaurus of Geographic Names
TIFF	Tagged Image File Format
TLS	Transport Layer Security
TSV	tabulátory oddělené hodnoty (Tab Separated Values)
TTP	Time To Peak
TURTLE	textová syntaxe RDF kompatibilní s N-TRIPLES a N3
UC	Ulcerózní kolitida
UI	User Interface
ULS	ultra-large scale
UML	Unified Modeling Language

UOOU	Úřad pro ochranu osobních údajů
URI	Uniform Resource Identifier
URL	Uniform Resource Locator
URN	Uniform Resource Name
USA	United States of America
UTF	UCS/Unicode Transformation Format
UUID	Universally Unique IDentifier
VIS	Visa Information System
VRA	Visual Resources Association
VRAF	Visual Resources Association Foundation
VTK	The Visualization Toolkit
W3C	World Wide Web Consortium
WDC	World Data Centre
WG1	Formát Walter Graphtek pro signálová data
WHO	Světová zdravotnická organizace (World Health Organisation)
WSDL	Web Services Description Language
WWW	World Wide Web
XMDR	eXtended Metadata Registry
XML	eXtensible Markup Language
XSD	XML Schema
XSLT	eXtensible Stylesheet Language
XUL	XML User Interface Language
Z39.50	Protokol klient-server pro vyhledávání a získávání informací ze vzdálených počítačových databází
ZD	zdravotnická dokumentace
ZIP	poštovní směrovací číslo (ZIP code)
ZUJ	Základní územní celky
ZZ	zdravotnické zařízení
ZČU	Západočeská univerzita

ÚNMZ	Úřad pro technickou normalizaci, metrologii a státní zkušebnictví
ÚZIS	Ústav zdravotnických informací a statistiky ČR
ČAS	Česká agentura pro standardizaci
ČR	Česká republika
ČSN	Česká technická norma
ČSSZ	Česká správa sociálního zabezpečení
ČSU	Český statistický úřad

A Ontologie a jmenné prostory

Výpis A.1: Používané existující ontologie a RDF slovníky

```
PREFIX acl: <http://www.w3.org/ns/auth/acl#>
PREFIX dc: <http://purl.org/dc/elements/1.1/>
PREFIX dcterms: <http://purl.org/dc/terms/>
PREFIX dctype: <http://purl.org/dc/dcmitype/>
PREFIX fn: <http://www.w3.org/2005/xpath-functions#>
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
PREFIX fsl: <http://www.w3.org/2004/09/fresnel#>
PREFIX geo: <http://www.w3.org/2003/01/geo/wgs84_pos#>
PREFIX nfo: <http://www.semanticdesktop.org/ontologies/2007/03/22/nfo#>
PREFIX org: <http://www.w3.org/ns/org#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX owlTime: <http://www.w3.org/2006/time#>
PREFIX prov: <http://www.w3.org/ns/prov#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX skos: <http://www.w3.org/2004/02/skos/core#>
PREFIX time: <http://www.w3.org/2006/time#>
PREFIX tzont: <http://www.w3.org/2006/timezone#>
PREFIX vann: <http://purl.org/vocab/vann/>
PREFIX vcard: <http://www.w3.org/2006/vcard/ns#>
PREFIX void: <http://rdfs.org/ns/void#>
PREFIX vs: <http://www.w3.org/2003/06/sw-vocab-status/ns#>
PREFIX xml: <http://www.w3.org/XML/1998/namespace>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
```

Výpis A.2: Jmenné prostory ontologií platformy MRE a řešených projektů

```
PREFIX dcm: <https://mre.zcu.cz/ontology/dcm.owl#>
PREFIX dscl: <https://mre.zcu.cz/ontology/dscl.owl#>
PREFIX ds: <https://mre.zcu.cz/ontology/dasta.owl#>
PREFIX form: <https://mre.zcu.cz/ontology/form.owl#>
PREFIX id: <https://mre.zcu.cz/id/>
PREFIX ibd: <https://mre.zcu.cz/ontology/ibd.owl#>
PREFIX ibdt: <https://mre.zcu.cz/ontology/ibdt.owl#>
PREFIX image: <https://mre.zcu.cz/ontology/image.owl#>
PREFIX imm: <https://mre.zcu.cz/ontology/image-mapping.owl#>
PREFIX mre: <https://mre.zcu.cz/ontology/mre.owl#>
PREFIX nihss: <https://mre.zcu.cz/ontology/nihss.owl#>
PREFIX pop: <https://mre.zcu.cz/ontology/pop.owl#>
PREFIX sits: <https://mre.zcu.cz/ontology/sits.owl#>
PREFIX stroke: <https://mre.zcu.cz/ontology/stroke.owl#>
PREFIX tdb: <https://mre.zcu.cz/ontology/tdb.owl#>
```

B Návrh ontologie v tabulkovém procesoru

Příloha navazuje na kapitolu 6. Tabulkový procesor se ukázal jako vhodný praktický nástroj při návrhu ontologií nebo RDF slovníků při multidisciplinární spolupráci.

První tabulka na obrázku B.1 ukazuje návrh tříd a atributů se sloupci: A identifikaci řádky, B stav atributu, C logická skupina tříd, D anglický název třídy, E český název třídy, F URI třídy, G český název vlastnosti, H anglický název pro vlastnost, I URI vlastnosti, Jtyp atributu (*Object*, *Datatype* a *Annotation*), K obor hodnot, L kardinalita, M jednotka, N poznámky a O popis atributu. Druhý řádek tabulky ukazuje v jaké vlastnosti bude hodnota sloupce v OWL reprezentována. Obsah sloupců F, I a J bude přímo v OWL. Sloupce kardinality L a jednotky M jsou jako výchozí bod pro vzájemnou shodu a případnou budoucí dokumentaci.

Druhá tabulka na obrázku B.2 je zaměřena na zjednodušení a zrychlení tvorby číselníků. Konkrétní hodnoty slovníku jsou realizovány tabulkou se sloupci: A vygenerované URI ze sloupců C a D, B identifikace řádky, C identifikační číslo slovníku, D pořadové číslo položky pro slovník, E český název slovníku, F třída, G označení výchozí hodnoty slovníku hodnotou *true/false*, H skóre nebo váha položky, I kódové označení hodnoty slovníku, J český název položky, K český popis pro položku, L anglický název položky, M popis pro položku v angličtině a N stav atributu. Sloupce G, H a I jsou nepovinné, mohou sloužit při konverzi dat nebo pro zobrazení v uživatelských aplikacích nebo reportech místo celého (mnohdy dlouhého) názvu položky slovníku.

Výpis B.1: Příklad hodnoty číselníku NIHSS pro postižení zraku

```
<owl:NamedIndividual rdf:about="https://mre.zcu.cz/ontology/stroke.owl#v00530001">
  <rdf:type rdf:resource="https://mre.zcu.cz/ontology/stroke.owl#NIHSSBestGazeVocab"/>
  <form:code>1</form:code>
  <form:score rdf:datatype="http://www.w3.org/2001/XMLSchema#integer">1</form:score>
  <form:vocabularyId rdf:datatype="http://www.w3.org/2001/XMLSchema#integer">53</form:vocabularyId>
  <form:vocabularyItemId rdf:datatype="http://www.w3.org/2001/XMLSchema#integer">1</form:vocabularyItemId>
  <form:vocabularyLineNumber rdf:datatype="http://www.w3.org/2001/XMLSchema#integer">1765</form:vocabularyLineNumber>
  <rdfs:label>1 = Partial paralysis of the view</rdfs:label>
  <rdfs:label xml:lang="en">1 = Partial paralysis of the view</rdfs:label>
  <rdfs:label xml:lang="cs">1 = ...</rdfs:label>
  <vs:term_status>stable</vs:term_status>
</owl:NamedIndividual>
```

A	B	C	D	E	F	G	H	J	K	L	M	N	O
line	stav	Skupina	Název třídy (en)	Název třídy (cs)	Class URI	Název atributu (cs)	Název atributu (en)	Property URI	Range	Kard	Jednotka	Poznámka	Popis atributu (description)
	vs:term status		rdfs:label	rdfs:label		rdfs:label	rdfs:label		rdfs:range			rdfs:comment	dc:description
1													
2													
3	1	instabile	Pacient	Adresa	ds:PermanentAddress	by adresa	address type	ds:addressType	ds:ContactTypeVocab	1	NE		
4	2	instabile	Pacient	Adresa	ds:PermanentAddress	obec	city	ds:city	pop:PopulationCityVocab	0-1	NE		
5	3	instabile	Pacient	Adresa	ds:PermanentAddress	země	country	ds:country	pop:PopulationCountryVocab	0-1	NE		
6	4	instabile	Pacient	Adresa	ds:PermanentAddress	kraj	district	ds:district	pop:PopulationDistrictVocab	0-1	NE		
7	5	instabile	Pacient	Kontakt	ds>Contact	email	email	ds:contactEmail	xsd:string	0-1	NE		
8	6	instabile	Pacient	Kontakt	ds>Contact	telefon	telephone/cell	ds:contactPhoneNumber	xsd:string	0-1	NE		
9	7	instabile	Pacient	Kontakt	ds>Contact	by kontaktu	contact type	ds:contactType	ds:ContactTypeVocab	1	NE		
10	8	instabile	Pacient	Pacient	ds:Patient	adresa	address	ds:address	ds:Address	0-N	NE		
11	9	instabile	Pacient	Pacient	ds:Patient	kontakt	contact	ds:contact	ds:Contact	0-N	NE		
12	10	instabile	Pacient	Pacient	ds:Patient	země narození	country of birth	ds:countryBirth	pop:PopulationCountryVocab	1	NE		
13	11	instabile	Pacient	Pacient	ds:Patient	datum narození	birth date	ds:dateimeBirth	xsd:date	1	NE		
14	12	instabile	Pacient	Pacient	ds:Patient	jméno	first name	ds:firstName	xsd:string	1	NE		
15	13	instabile	Pacient	Pacient	ds:Patient	příjmení	last name	ds:lastName	xsd:string	1	NE		
16	14	instabile	Pacient	Pacient	ds:Patient	rodné číslo	birth number	ds:patientID	xsd:string	1	NE		
17	15	instabile	Pacient	Pacient	ds:Patient	pohlaví	gender	ds:sex	xsd:string	1	NE		
18													
19													

Obrázek B.1: Tabulka pro návrh tříd a atributů

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	URI	ID	Vocab ID	Item ID	Vocabulary - Title (cs)	Vocabulary - Type	Default value?	Score	Value Code	Label (cs)	Description (cs)	Label (en)	Description (en)	Status
289	ibdt:007800001	287	78	1	Klasifikace – chování choroby	ibdt:ClassificationBehaviorVocab			B1	restrikující, nepenetrující		Non-structuring, non-penetrating		unstable
290	ibdt:007800002	288	78	2	Klasifikace – chování choroby	ibdt:ClassificationBehaviorVocab			B2	strikující		Striking		unstable
291	ibdt:007800003	289	78	3	Klasifikace – chování choroby	ibdt:ClassificationBehaviorVocab			B3	penetrující		Penetrating		unstable
292	ibdt:007800004	290	78	4	Klasifikace – chování choroby	ibdt:ClassificationBehaviorVocab			B2B3	penetrující a strikující	současně nebo v návaznosti	Penetrating at	Simultaneous	unstable
293	ibdt:007800005	291	78	5	Klasifikace – chování choroby	ibdt:ClassificationBehaviorVocab			P	perianální postižení		Perianal disability		unstable
294	ibdt:007900001	292	79	1	Klasifikace – lokalizace	ibdt:ClassificationLocalisationVocab			L1	distální 1/3 ilea +/- limitované postižení céka		Distal 1/3 ilea +/- limited aff		unstable
295	ibdt:007900002	293	79	2	Klasifikace – lokalizace	ibdt:ClassificationLocalisationVocab			L2	kolon		Columns		unstable
296	ibdt:007900003	294	79	3	Klasifikace – lokalizace	ibdt:ClassificationLocalisationVocab			L3	ileokolon		Ileocolon		unstable
297	ibdt:007900004	295	79	4	Klasifikace – lokalizace	ibdt:ClassificationLocalisationVocab			L4a	horní GIT a postižení proximálně od ligamentum Treitz		Upper GIT and affection prox		unstable
298	ibdt:007900005	296	79	5	Klasifikace – lokalizace	ibdt:ClassificationLocalisationVocab			L4b	horní GIT a postižení distálně od ligamentum Treitz a proximálně		Upper GIT and distal aff		unstable
299	ibdt:008000001	297	80	1	Klasifikace – Parksova perianální píštěle	ibdt:ClassificationParksVocab				superficiální		Superficial		unstable
300	ibdt:008000002	298	80	2	Klasifikace – Parksova perianální píštěle	ibdt:ClassificationParksVocab				interfunkterická		Interfunicteric		unstable
301	ibdt:008000003	299	80	3	Klasifikace – Parksova perianální píštěle	ibdt:ClassificationParksVocab				transsfunkterická		Transsfunicteric		unstable
302	ibdt:008000004	300	80	4	Klasifikace – Parksova perianální píštěle	ibdt:ClassificationParksVocab				suprasfinkterická		Suprasfunicteric		unstable
303	ibdt:008000005	301	80	5	Klasifikace – Parksova perianální píštěle	ibdt:ClassificationParksVocab				extrasfinkterická		Extrasfunicteric		unstable
304	ibdt:008000006	302	80	6	Klasifikace – Parksova perianální píštěle	ibdt:ClassificationParksVocab				supralevatorová		Supralelevatorová		unstable
305	ibdt:008000007	303	80	7	Klasifikace – Parksova perianální píštěle	ibdt:ClassificationParksVocab				infralevatorová		Infralelevator		unstable

Obrázek B.2: Tabulka pro návrh slovníků

C Tvorba RDF slovníků z existujících dat

Výpis C.1: Číselník ÚZIS pro pohlaví (zkráceno)

```
<ciselnik xmlns:xsi="..." xsi:noNamespaceSchemaLocation=".../uzis-4.13.01.xsd"
  nazev="Pohlavi" zkratka="POHLAV" spravce="uzis" dat_gen="2017-11-24"
  cas_gen="10:00:00" ver_nclp="02.67.01" ver_ds="04.19.02"
  uzis_verze="201940" u_plati_od="2019-10-01" popis="Číselník použitý v...">
<VETA kod="1" popis="muž" plati_od="2015-01-01" skupina_platnosti_01="1"
  skupina_platnosti_02="1" skupina_platnosti_03="1"/>
<VETA kod="2" popis="žena" plati_od="2015-01-01" skupina_platnosti_01="1"
  skupina_platnosti_02="1" skupina_platnosti_03="1"/>
<VETA kod="3" popis="nelze specifikovat" plati_od="2015-01-01"
  skupina_platnosti_01="1" skupina_platnosti_02="1"
  skupina_platnosti_03="0"/>
<VETA kod="0" popis="neudáno" plati_od="2015-01-01" skupina_platnosti_01="1"
  skupina_platnosti_02="0" skupina_platnosti_03="0"/>
</ciselnik>
```

Výpis C.2: RDF data pro položku pohlaví muž z číselníku ÚZIS

```
<dscl:POHLAV rdf:about="https://mre.zcu.cz/ontology/dscl.owl#POHLAV_1">
<rdfs:label xml:lang="cs">muž</rdfs:label>
<skos:prefLabel xml:lang="cs">muž</skos:prefLabel>
<dc:identifier>1</dc:identifier>
<dc:title xml:lang="cs">muž</dc:title>
<dscl:popis>muž</dscl:popis>
<dscl:plati_od rdf:datatype="http://www.w3.org/2001/XMLSchema#date"
  >2015-01-01</dscl:plati_od>
<dscl:kod rdf:datatype="http://www.w3.org/2001/XMLSchema#integer"
  >1</dscl:kod>
<dscl:skupina_platnosti_01 rdf:datatype="http://www.w3.org/2001/XMLSchema#
  integer">1</dscl:skupina_platnosti_01>
<dscl:skupina_platnosti_02 rdf:datatype="http://www.w3.org/2001/XMLSchema#
  integer">1</dscl:skupina_platnosti_02>
<dscl:skupina_platnosti_03 rdf:datatype="http://www.w3.org/2001/XMLSchema#
  integer">1</dscl:skupina_platnosti_03>
</dscl:POHLAV>
```

D Praktické využití ontologií

D.1 Extrakce metadat – konfigurace z ontologie

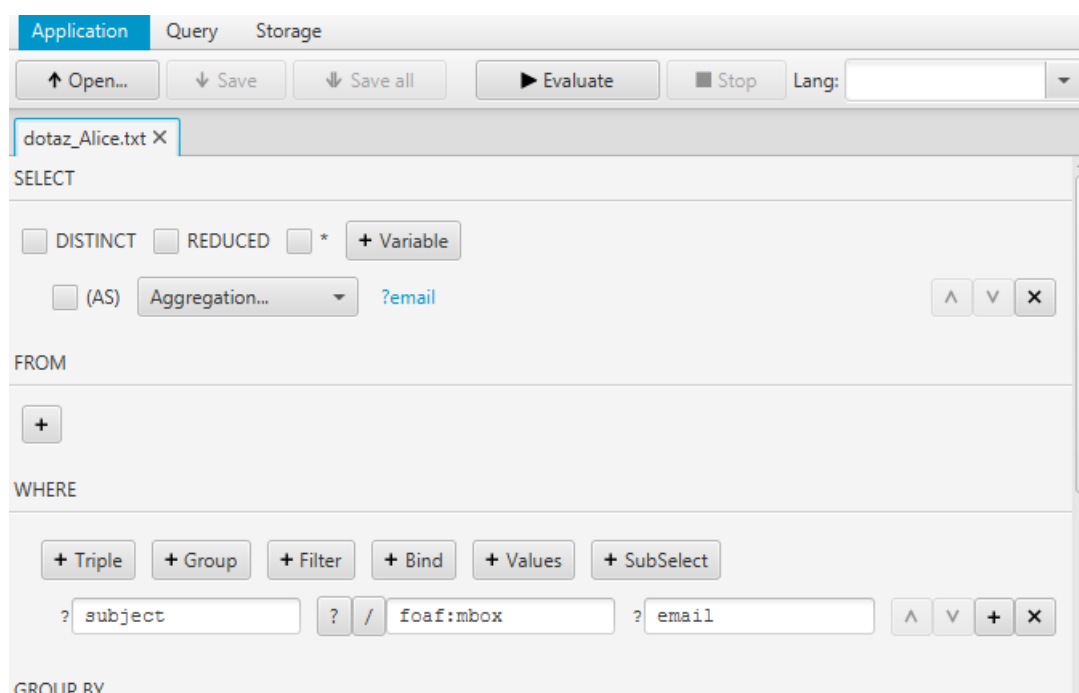
Výpis D.1: Získání atributů z DICOM Ontology dotazem

```
SELECT ?attribute ?attributeLabel ?attributeGroup ?attributeElement
WHERE {
  ?attribute rdf:type dcm:DICOM_Element_CLASS .
  ?attribute rdfs:label ?attributeLabel .
  ?attribute dcm:Group ?attributeGroup .
  ?attribute dcm:Element ?attributeElement .
}
```

Výpis D.2: Získání referenčního informačního modelu z DICOM Ontology

```
SELECT ?realWorld ?realWorldLabel ?dicomIE ?dicomIELabel ?moduleAttributeLine
      ?attribute ?attributeLabel ?attributeGroup ?attributeElement ?attributeVR
WHERE {
  ?realWorld rdf:type dcm:Real_World_Object_CLASS .
  ?realWorld rdfs:label ?realWorldLabel .
  FILTER(LANG(?realWorldLabel) = "\")
  ?realWorld dcm:has_DICOM_IE ?dicomIE .
  ?dicomIE dcm:Module_Attributes_or_Macros ?moduleAttributeLine .
  ?dicomIE rdfs:label ?dicomIELabel .
  ?moduleAttributeLine dcm:Attribute_Data_Element_or_Macro ?attribute .
  ?attribute rdfs:label ?attributeLabel .
  ?attribute dcm:Group ?attributeGroup .
  ?attribute dcm:Element ?attributeElement .
  ?attribute dcm:Data_Element_VR ?attributeVR .
}
```

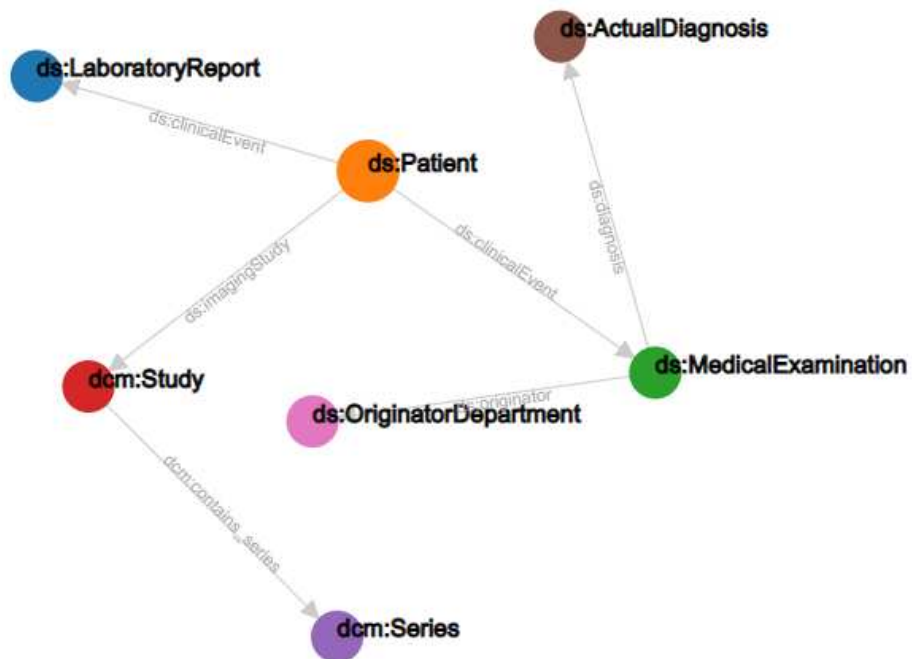
D.2 Příklady aplikace Sparkle



Obrázek D.1: Tvorba dotazu prostřednictvím formulářových prvků [265, 268]

```
1 PREFIX dbo: <http://dbpedia.org/ontology/>
2 PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
3 PREFIX dbpedia2: <http://dbpedia.org/property/>
4
5 SELECT ?battle (SUM(?casualty) AS ?totalCasualty) ?lat ?long (MAX(?date) AS ?mdate)
6 (GROUP_CONCAT(?label ; separator='; ' ) AS ?mlabel)
7 WHERE
8 { ?battle <http://purl.org/dc/terms/subject>
9 <http://dbpedia.org/resource/Category:Battles_of_the_Thirty_Years'_War>
10 OPTIONAL
11 { ?battle rdfs:label ?label
12 FILTER ( ( lang(?label) = "" ) || ( lang(?label) = "en" ) )
13 }
14 OPTIONAL
15 { ?battle dbpedia2:casualties ?casualty }
16 OPTIONAL
17 { ?battle <http://www.w3.org/2003/01/geo/wgs84_pos#lat> ?lat }
18 OPTIONAL
19 { ?battle <http://www.w3.org/2003/01/geo/wgs84_pos#long> ?long }
20 ?battle dbo:date ?date
21 }
22 GROUP BY ?battle ?lat ?long
23 ORDER BY ?battle
24 LIMIT 20
```

Obrázek D.2: Textový editor se zvýrazněním syntaxe. Červeně zvýrazněna chyba představující aplikaci neznámý zdroj dle [268]



Obrázek D.3: Ilustrace vizualizace dotazu [269]

Results

Output attributes:

Output languages:

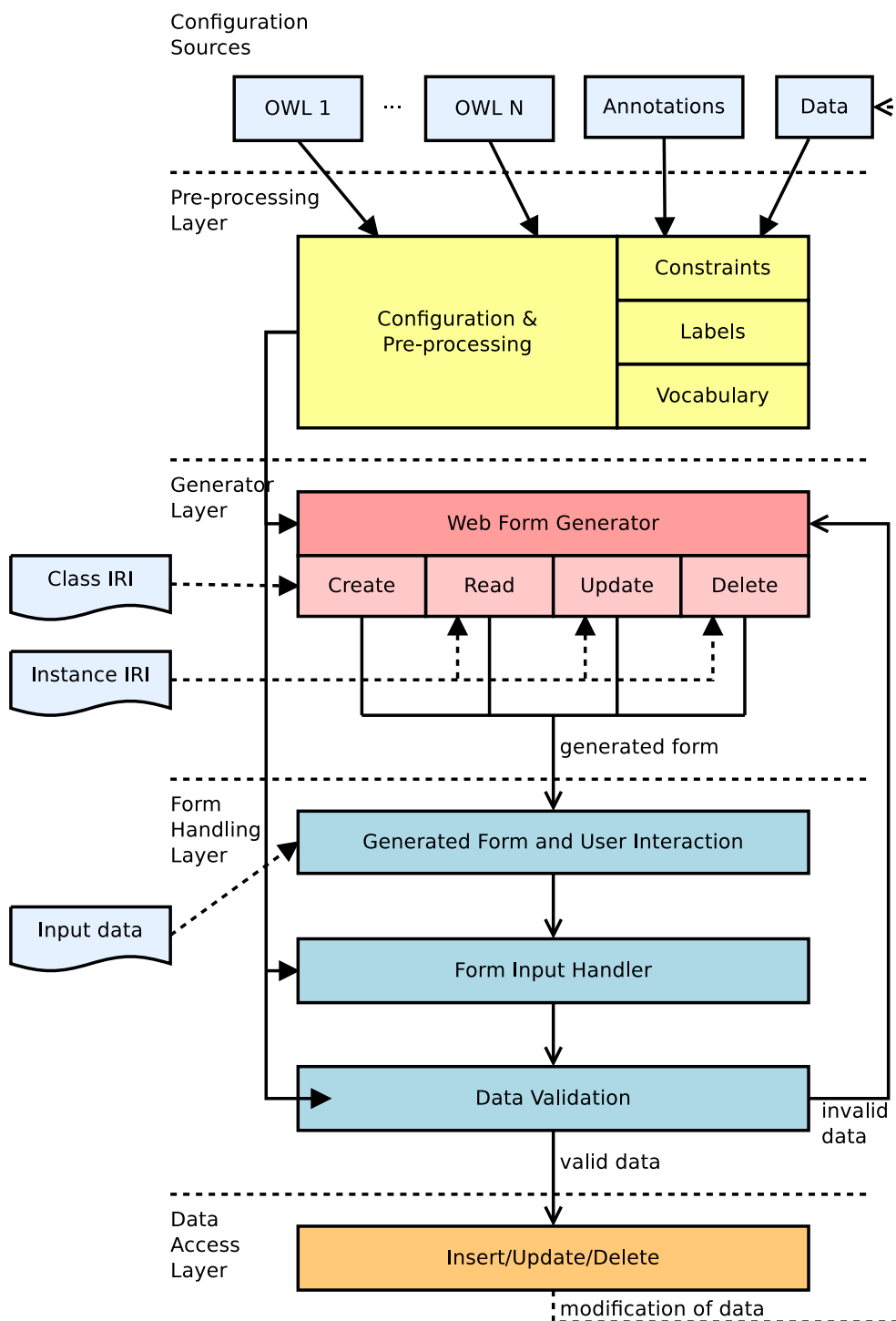
Table JSON

ActualDiagnosis_diagOrder	ActualDiagnosis_datetimeEvent	ActualDiagnosis_diagDetail	Patient_patientID	CT_inTimepoint
1	2013-09-20	Mozkový infarkt způsobený embolií mozkových tepen	123	po 24 hodinách
1	2013-09-20	Mozkový infarkt způsobený embolií mozkových tepen	123	příjem
1	2013-09-19	Diabetes mellitus 2. typu bez komplikací na PAD, těžká obezita	123	po 24 hodinách
1	2013-09-19	Diabetes mellitus 2. typu bez komplikací na PAD, těžká obezita	123	příjem
1	2013-08-28	Směšená hyperlipidemie	892	příjem

Obrázek D.4: Příklad výsledků včetně srozumitelných textů místo URI hodnot číselníku [269]

D.3 Generování webových formulářů z ontologie

Diagram na obrázku D.5 ukazuje princip. Příklady formulářů jsou na obrázcích D.6 a D.7, které jsou vygenerovány na základě *IBDT Ontology*. [VKKK17]



Obrázek D.5: Aplikační vrstvy a vývojový diagram pro získávání a modifikace RDF dat ve formulářích založených na ontologiích.

Klasifikace wPCDAI

bolest břicha	žádné
omezení denní aktivity	bez limitace, cítí se dobře/zdravě
datum	2017-08-28
sedimentace erytrocytů	<20 mm/hod.
extraintestinální manifestace	žádné
fyzikální vyšetření perianální oblasti	normální nález nebo asymptomatické hrbolky
wPCDAI skóre	0
albumin v séru	≥3,5 g/dl
stolice za den	žádná nebo max. 1 průjmovitá stolice, bez krve
hmotnost	přírůstek na váze, event. stabilní váha

[Odeslat](#)

Obrázek D.6: Z ontologie dynamicky generovaný formulář pro klasifikaci dětských pacientů dle wPCDAI.

Endoskopie

datum 2017-09-05

anus normální

endoskopie CD [+ Přidat](#)

Změnit

Endoskopický nález CD

oblast colon ascendens

rozsah postižení <50 %

rozsah ulcerace <10 %

zúžení mnohačetné, prostupné pro endoskop

ulcerace 0,5–2 cm

datum 2017-08-28

[Změnit](#) [Zrušit](#)

jejunum normální

Obrázek D.7: Formulář pro přímou editaci instance v objektové vlastnosti v modálním okně.

D.4 Tvorba dotazů pro temporální data

Podmínka nová/úpravy

Období, časový bod:

Třída:

Vlastnost:

Seznam podmínek

Sloupec	Zobrazit?	Název v ontologii	Název sloupce	Strojový název	Operace
1.	<input checked="" type="checkbox"/>	MRE Stroke ID	stroke id	stroke_id	<input type="checkbox"/> <input type="checkbox"/>
2.	<input checked="" type="checkbox"/>	SITS ID	stroke idSITS	stroke_idSITS	<input type="checkbox"/> <input type="checkbox"/>
3.	<input checked="" type="checkbox"/>	typ iktu	stroke strokeType_t24h	stroke_strokeType_stroke_t24h	<input type="checkbox"/> <input type="checkbox"/>
4.	<input checked="" type="checkbox"/>	CT provedeno?	stroke ctDone_t24h	stroke_ctDone_stroke_t24h	<input type="checkbox"/> <input type="checkbox"/>
5.	<input checked="" type="checkbox"/>	CTA – uzávěr	stroke ctOcclusion_t24h	stroke_ctOcclusion_stroke_t24h	<input type="checkbox"/> <input type="checkbox"/>

Obrázek D.8: Uživatelské rozhraní *Query Builder* pro přidávání nových podmínek (sloupců) při tvorbě SELECT dotazu ve SPARQL nad temporálními daty.

D.5 DataTable, ontologie a RDF data

Vzorky

Zobraz záznamů: 25

Copy Excel CSV

Headat:

Operace	ID	Kód	Pacient	Datum operace	Třída	T	N	M	G	U	K-RAS	B-RAF	CIN	MSI	Číslo biopsie	Diagnóza	R	Výsledek	Onkologická léčba	Prevenční	Stupeň léčby	Typ léčby	Průběh kontroly - 2019	Průběh kontroly - 2019	Datum ošetření	Uvěř z důvodu onemocnění	Vzorek zobrazen	
+	1299		lg23Jt9+ZRHUJtF6R9...	07-23	ZT	4	1	1	2	1320	X	X	X		KRCA	P	ano	ano	ano	N/A				05-14	07-28	07-28	ne	ano
+	734		KF6X7N9AJTJZ2Y93A6...	05-14	ZT	2	0	1	1	014	X	X	X	X	KRCA	P	ano	ano	ano	N/A				05-14	07-28	07-28	ne	ano
+	865		BF6mF7WkTRGj7CE09...	03-27	TT	3	1	1	N/A	613	X	X	X	X	KRCA	P	ano	ano	ano	N/A				03-27	07-08	07-08	ne	ano
+	409		TQw4mTECP8A6e6r9e...	03-31	TT	3	X	1	2	014	W	W	X	X	KRCA	P	ano	ano	ano	N/A				03-31	07-08	07-08	ne	ano
+	936		u5vmy4wvMC76825A...	04-08	TT	4	0	1	3	013	X	X	X	X	KRCA	P	ano	ano	ano	N/A				04-08	07-08	07-08	ne	ano
+	1183		qP76R93J334P6R660...	07-17	ZT	2	1	1	1	014	X	X	X	X	KRCA	P	ano	ano	ano	N/A				07-17	07-08	07-08	ne	ano
+	302		M6L2u7Aqjg7f7m6u6...	02-06	ZT	3	1	1	2	2104	X	X	X	X	KRCA	P	ano	ano	ano	N/A				02-06	07-08	07-08	ne	ano
+	732		KF6X7N9AJTJZ2Y93A6...	05-14	TT	2	0	1	1	014	X	X	X	X	KRCA	P	ano	ano	ano	N/A				05-14	07-08	07-08	ne	ano
+	534		qP76R93J334P6R660...	11-12	ZT	3	1	1	3	712	X	X	X	X	KRCA	P	ano	ano	ano	N/A				11-12	07-08	07-08	ne	ano
+	396		Qm9ym9F259P9H6D2...	07-20	TT	3	0	0	2	010	X	X	X	X	KRCA	R	ano	ano	ano	N/A				07-20	07-08	07-08	ne	ano
+	391		Qm9ym9F259P9H6D2...	07-20	TT	3	0	0	2	010	X	X	X	X	KRCA	R	ano	ano	ano	N/A				07-20	07-08	07-08	ne	ano
+	2032		KR2666M4P4516CC5...	08-13	TT	3	1	1	1	114	W	X	X	X	KRCA	P	ano	ano	ano	N/A				08-13	07-08	07-08	ne	ano
+	868		BF6mF7WkTRGj7CE09...	03-27	ZT	3	1	1	N/A	613	X	X	X	X	KRCA	P	ano	ano	ano	N/A				03-27	07-08	07-08	ne	ano
+	414		TQw4mTECP8A6e6r9e...	03-27	ZT	3	X	1	2	014	W	W	X	X	KRCA	P	ano	ano	ano	N/A				03-27	07-08	07-08	ne	ano
+	1172		TQw4mTECP8A6e6r9e...	04-26	TT	2	X	1	2	013	W	W	X	X	KRCA	P	ano	ano	ano	N/A				04-26	07-08	07-08	ne	ano
+	414		TQw4mTECP8A6e6r9e...	03-31	ZT	3	X	1	2	014	W	W	X	X	KRCA	P	ano	ano	ano	N/A				03-31	07-08	07-08	ne	ano
+	139		R5A4L5JED3JGD8116...	04-01	TT	3	1	X	2	117	X	X	X	X	KRCA	R	ano	ano	ano	N/A				04-01	07-08	07-08	ne	ano
+	238		0e6RISUR966666666...	02-06	TT	3	1	1	2	2104	X	X	X	X	KRCA	R	ano	ano	ano	N/A				02-06	07-08	07-08	ne	ano
+	349		Zy9mP9uL3K2E68...	05-07	TT	3	1	0	2	322	X	X	X	X	KRCA	R	ano	ano	ano	N/A				05-07	07-08	07-08	ne	ano
+	1176		TQw4mTECP8A6e6r9e...	06-28	ZT	2	X	1	2	013	X	W	X	X	KRCA	P	ano	ano	ano	N/A				06-28	07-08	07-08	ne	ano
+	539		qP76R93J334P6R660...	11-12	ZT	3	1	1	3	712	X	X	X	X	KRCA	P	ano	ano	ano	N/A				11-12	07-08	07-08	ne	ano
+	143		R5A4L5JED3JGD8116...	04-01	ZT	3	1	X	2	117	X	X	X	X	KRCA	R	ano	ano	ano	N/A				04-01	07-08	07-08	ne	ano
+	242		0e6RISUR966666666...	02-06	ZT	3	1	0	2	110	X	X	X	X	KRCA	R	ano	ano	ano	N/A				02-06	07-08	07-08	ne	ano
+	296		M6L2u7Aqjg7f7m6u6...	02-06	TT	3	1	1	2	2104	X	X	X	X	KRCA	P	ano	ano	ano	N/A				02-06	07-08	07-08	ne	ano
+	353		Zy9mP9uL3K2E68...	05-07	TT	3	1	0	2	322	X	X	X	X	KRCA	R	ano	ano	ano	N/A				05-07	07-08	07-08	ne	ano
+	298		M6L2u7Aqjg7f7m6u6...	02-06	TT	3	1	1	2	2104	X	X	X	X	KRCA	P	ano	ano	ano	N/A				02-06	07-08	07-08	ne	ano

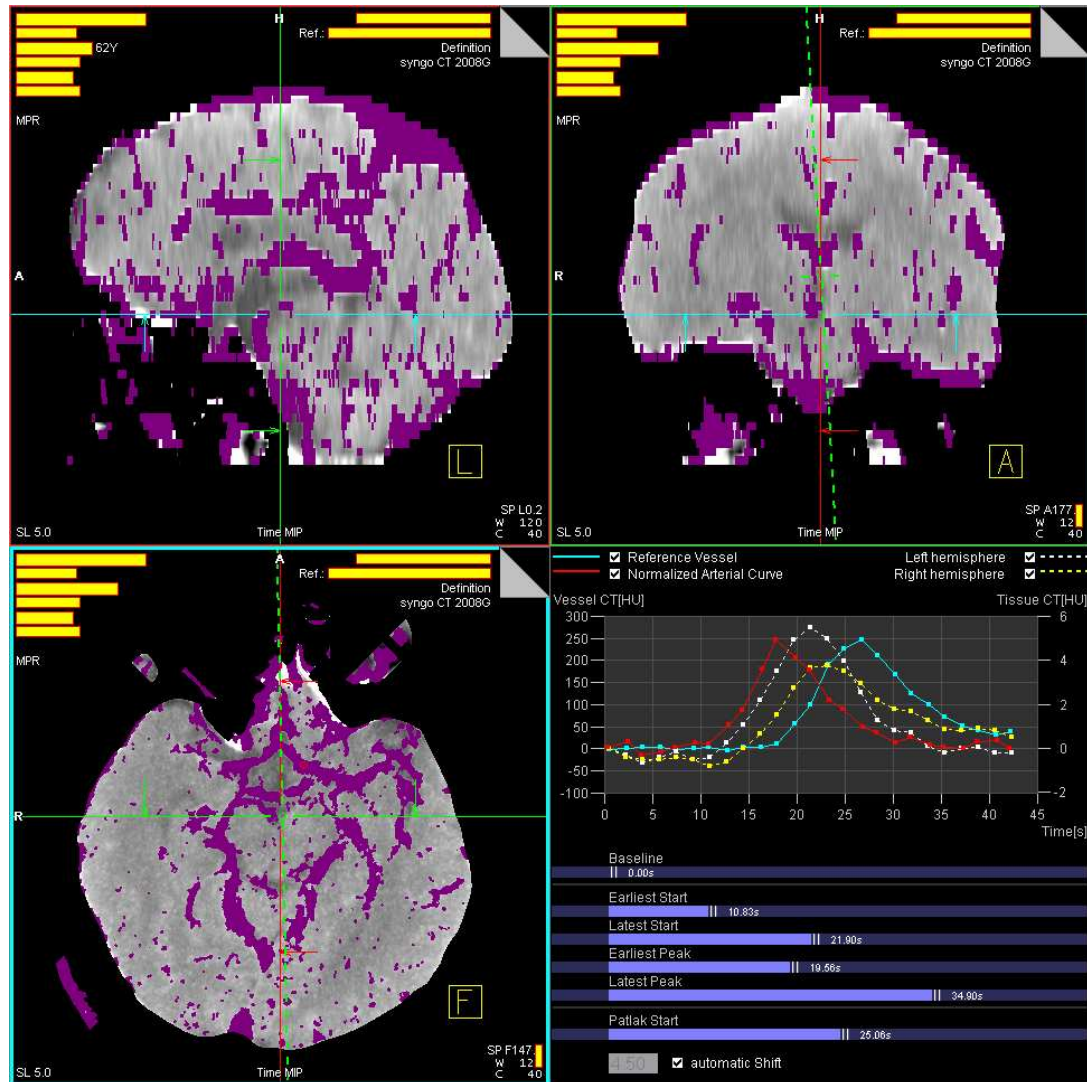
Zobrazují 1 až 25 z celkem 2,905 záznamů

Předchozí 1 2 3 4 5 ... 117 Další

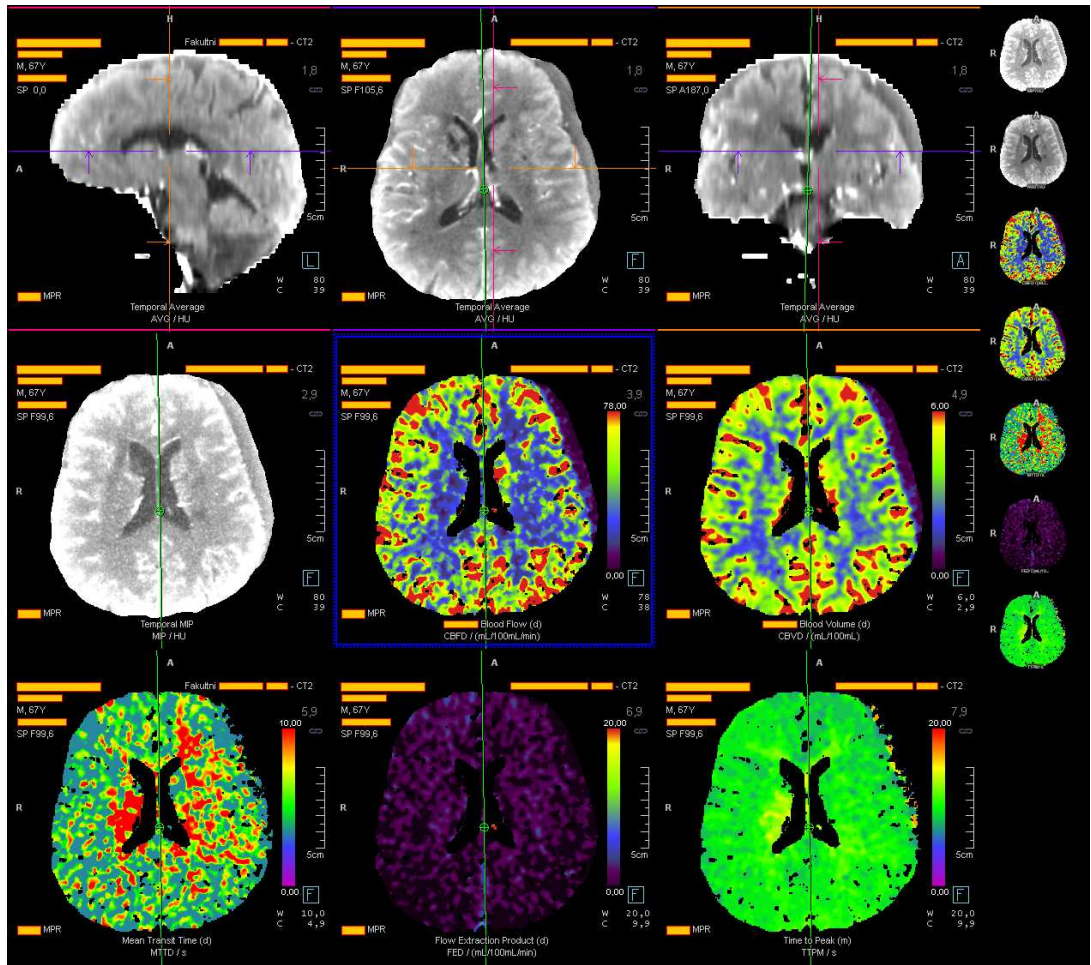
Obrázek D.9: Databáze tkáňových vzorků používající DataTable s ontologiemi a RDF daty

E Ochrana osobních údajů v obrazových vyšetřeních

Obrázky E.1 a E.2 ukazují DICOM snímky po provedení de-identifikace vypálených informací v obrazové části.



Obrázek E.1: Mapy CT perfuze zobrazující vývoj průtoku krve v mozku v čase (Time MIP)



Obrázek E.2: Mapy CT perfuze zobrazující průběh průtoku krve mozkiem v čase

F Metriky ontologií

Expresivita u ontologií v tabulce je popsána následujícím číselným označením:

1. ALCROF(D),
2. AL,
3. ALO,
4. ALCROIF(D),
5. ALCHI(D),
6. ALROF(D),
7. AL(D),
8. ALI(D),
9. ALF(D).

Jednotlivé znaky mají následující vysvětlení expresivity použitého jazyka v ontologii:

AL Základní jazyk, který umožňuje atomickou negaci, průnik konceptů, univerzální omezení, a omezenou existenciální kvantifikaci.

F Funkční vlastnosti (unikátní výskyt).

E Plná existenciální kvalifikace.

U Spojování konceptů.

C Komplexní negace konceptů.

H Hierarchie (`rdfs:subPropertyOf`).

R Omezené axiomy pro začlenění; reflexivita, ireflexivita a disjunktnost.

O Vyčíslené třídy pro omezení hodnot objektů (`owl:oneOf`, `owl:hasValue`).

I Inverzní vlastnosti.

N Kardinalita (`owl:cardinality`, `owl:maxCardinality`).

Q Kvalifikovaná omezení kardinality.

(D) Použití vlastností datového typu, datových hodnot nebo typů.

Ontology	dasta	dscl	dscl-data	form	ibds	ibdt	mre	nihss	pop	sits	stroke	tdb
Metrics												
Axiom	5 828	7 887	6 446 864	223	8 361	20 039	1 085	870	145	4 539	12 462	1 221
Logical axiom count	1 618	470	406 999	0	2 070	3 219	373	172	33	1 203	1 489	147
Declaration axioms count	937	496	0	33	1 180	2 294	183	150	29	777	1 400	164
Class count	298	472	457	0	337	513	67	18	7	221	94	17
Object property count	127	0	0	0	197	387	41	16	3	80	115	17
Data property count	232	0	0	0	274	442	83	6	8	172	109	19
Individual count	296	2	390 719	0	379	956	0	106	0	292	1 041	74
Annotation Property count	30	33	227	45	39	50	16	17	12	26	51	47
DL expressivity	(1)	(2)	(3)	(2)	(4)	(1)	(5)	(6)	(9)	(6)	(7)	(8)
Class axioms												
SubClassOf	246	469	0	0	266	246	43	15	0	223	0	0
EquivalentClasses	4	0	0	0	4	4	2	0	0	1	0	0
DisjointClasses	0	0	0	0	0	0	1	0	0	0	0	0
Hidden GCI Count	2	0	0	0	2	2	2	0	0	0	0	0
Object property axioms												
InverseObjectProperties	0	0	0	0	2	0	5	0	0	0	0	1
DisjointObjectProperties	2	0	0	0	2	2	0	1	0	1	0	0
FunctionalObjectProperty	67	0	0	0	113	67	0	16	2	29	0	0
ObjectPropertyDomain	176	0	0	0	256	436	93	16	4	86	115	17
ObjectPropertyRange	97	0	0	0	167	356	39	16	3	51	115	17
Data Property Axioms												
SubDataPropertyOf	0	0	0	0	0	0	1	0	0	0	0	0
EquivalentDataProperties	0	0	0	0	0	0	1	0	0	0	0	0
DisjointDataProperties	3	0	0	0	3	3	0	1	0	2	0	0

Ontology	dasta	dscl	dscl-data	form	ibds	ibdt	mre	nihss	pop	sits	stroke	tdb
FunctionalDataProperty	78	0	0	0	118	78	0	6	8	21	0	0
DataPropertyDomain	252	0	0	0	303	463	111	6	8	173	109	19
DataPropertyRange	107	0	0	0	152	318	53	6	8	46	109	19
Individual Axioms												
ClassAssertion	502	0	368 185	0	600	1 162	0	69	0	502	1 041	74
DataPropertyAssertion	14	0	0	0	14	14	0	0	0	14	0	0
SameIndividual	33	1	38 814	0	33	33	0	19	0	33	0	0
Annotation axioms												
AnnotationAssertion	3 222	6 921	6 039 865	160	5 007	14 418	521	516	83	2 508	9 535	880
AnnotationPropertyDomain	30	0	0	0	30	41	3	30	0	30	3	0
AnnotationPropertyRangeOf	3	0	0	30	13	14	4	2	0	3	35	30

G Klíčové požadavky a metriky

Následující požadavky jsou klíčové pro zajištění standardizace metadat v rozsáhlých informačních systémech. Pořadí jednotlivých požadavků v seznamu není přikládán význam. Každý požadavek obsahuje jednu nebo více možných metrik.

Akceptace dat

K_1 Heterogenní medicínská data pocházejí z jednoho nebo více zdrojů, u nichž bude jednoznačně definována shoda (obou) stran o tom, jaká data a v jakých formátech se jedná.

$M_{K_1}^1$ Schopnost přijímat data z více zdrojů současně.

$M_{K_1}^2$ Datové zdroje musí sdílet/předávat pouze data, která jsou očekávaná. Ostatní nebo neznámá data musí být odmítnuta.

Ochrana osobních údajů

K_2 Zajištění ochrany osobních údajů pro sekundární využití medicínských dat pro výzkumné účely.

$M_{K_2}^1$ Odstranění identifikátorů (přímých, nepřímých) osob (pseudonymizace, šifrování, anonymizace) z dat.

$M_{K_2}^2$ Tvorba slepých kopií dat (opakované nebo dodatečné provedení metod ochrany osobních údajů).

Archivace

K_3 Archivace heterogenních dat a metadat.

$M_{K_3}^1$ Uživatel má mít možnost uložit soubor libovolného datového formátu.

Metadata

K_4 Data a případná metadata popisovat/anotovat prostřednictvím sémanticky interoperabilní reprezentace metadat.

$M_{K_4}^1$ Unifikovaný přístup metadatům.

$M_{K_4}^2$ Interoperabilita metadat.

$M_{K_4}^3$ Možnost popisu sémantiky.

$M_{K_4}^4$ Uživatel může doplnit vlastní metadata s popisem.

K_5 Sdílení metadat jednotným způsobem.

$M_{K_5}^1$ Povinná existence aplikačního rozhraní pro přístup k metadatům.

$M_{K_5}^2$ Volitelně jeden nebo více dalších způsobů zpřístupnění dat a metadat uživatelům (např. prostřednictvím vyhledávání).

$M_{K_5}^3$ Zpřístupnění vazeb a vztahů v datech.

$M_{K_5}^4$ Poslední metrikou je rychlost získání metadat, aby byla v časech řádově desítek nebo stovek milisekund.

K_6 Sdílení dat i metadat projektu s dalšími projekty nebo systémy.

$M_{K_6}^1$ Povinnou součástí je sdílení všech potřebných slovníků, číselníků, datových schémat, ontologií nebo dalších způsobů popisu (dokumentace), jež jsou za účelem interoperability dat a metadat používány.

$M_{K_6}^2$ Existence možnosti sdílení (výměna, předávání) dat a metadat.

Nezávislost

K_7 Vedení více projektů nezávisle na sobě a současně.

$M_{K_7}^1$ Minimálně dva projekty mohou být vedeny souběžně. Není rozlišováno, zda striktně odděleně nebo se stejnými zdrojovými medicínskými daty.

Softwarové prostředky

K_8 Rozšiřitelnost popisu metadaty pro softwarové prostředky, postupy zpracování dat i vztahů mezi daty.

$M_{K_8}^1$ Pro budoucí rozšiřitelnost platformy je vhodné, aby bylo možné popsat vedle dat také softwarové prostředky používané ke zpracování dat.

$M_{K_8}^2$ Důležité jsou specifikace spuštění softwarových nástrojů včetně parametrů.

$M_{K_8}^3$ Pro dlouhodobý udržitelný výzkum (budoucí opakované zpracování) je významná možnost popisu postupů/procesů dosažení výsledných souborů. Za metriku lze uvažovat existenci i jen možnosti budoucího rozšíření o takovou možnost.

Zabezpečení

K_9 Zabezpečení dat včetně jejich přenosu.

$M_{K_9}^1$ Vyžadována je zabezpečená komunikace při přenosu dat a metadat.

H Hodnocení metrik klíčových požadavků

Požadavek		Hodnocení
$M_{K_1}^1$	Přijímat data z více zdrojů.	ano
$M_{K_1}^2$	Akceptace dat.	ano (AnonMed)
$M_{K_2}^1$	Odstranění osobních údajů.	ano (AnonMed, PureImage)
$M_{K_2}^2$	Tvorba slepých kopií dat.	ano (AnonMed)
$M_{K_3}^1$	Archivace libovolného souboru.	ano (podpora v MetaMed)
$M_{K_3}^2$	Vlastní metadata uživatele.	ano
$M_{K_4}^1$	Unifikovaný přístup metadatům.	ano (OWL)
$M_{K_4}^2$	Interoperabilita metadat.	ano
$M_{K_4}^3$	Možnost popisu sémantiky.	ano (OWL)
$M_{K_5}^1$	Aplikačního rozhraní metadata.	ano (SPARQL Endpoint)
$M_{K_5}^2$	Další zpřístupnění dat a metadat.	ne (plánováno)
$M_{K_5}^3$	Zpřístupnění vazeb a vztahů.	ano (SPARQL Endpoint)
$M_{K_5}^4$	Rychlost získání metadat.	ano (viz 9)
$M_{K_6}^1$	Vedení více projektů souběžně.	ano
$M_{K_7}^1$	Popis software.	ano
$M_{K_7}^2$	Popis spuštění úloh.	částečně
$M_{K_7}^3$	Popis postupu dosažení výsledku.	částečně
$M_{K_8}^1$	Sdílení schéma metadat.	ano (OWL, RDF, SPARQL Endpoint)
$M_{K_8}^2$	Jednotné sdílení dat a metadat.	ano (SPARQL Endpoint)
$M_{K_9}^1$	Zabezpečená komunikace.	ano (HTTPS, SSH, TLS)

I Obsah DVD

data/ Příklady anonymních medicínských dat. K dispozici je vždy adresář se zdrojovými daty a tomu odpovídající metadata v RDF souborech ve formátech N-TRIPLE (*.nt), TURTLE (*.ttl) a RDF/XML (*.xml).

data-imaging/ Klinická data obrazového vyšetření (CT) při diagnostice pacienta s podezřením na CMP dle 5.3. K dispozici je soubor ve formátu DASTA (uvádějící odkazy na DICOM snímky) a metadata.

data-laboratory/ Výsledky laboratorních vyšetření (4x) ve shodě s 5.4.

data-stroke/ Záznam diagnostiky a sledování průběhu léčby pacienta po cévní mozkové příhodě. Příklad ilustruje data ve formátu DASTA doplněná o strukturovaná metadata ve *StrokeXML* (5.3).

population.ttl Ukázka dat používajících *Population Ontology* dle 7.6.

images/ Obrázky použité v disertační práci

mre-doc/ Celková dokumentace (<https://mre.zcu.cz>) projektů platformy MRE (Apache Maven, JavaDoc) a ontologií (owl2doc; v `./ontology/`).

mre-ontology/ Ontologie (OWL) a RDF slovníky.

dasta.owl DASTA Ontology (7.3)

dsc1.owl DASTA Code-lists (7.3), generováno `dsc12rdf` z číselníků formátu DASTA (MZČR, ÚZIS, ...). V souboru `dsc1.rdf.gz` je obsah číselníků.

dcm.owl DICOM Ontology (viz 7.4)

form.owl Form Ontology (7.5).

ibd.owl IBDS Ontology (7.8.1).

ibdt.owl IBDT Ontology (7.8.2).

mre.owl MRE Ontology (7.2).

nihss.owl NIHSS Ontology (7.7)

pop.owl Population Ontology (7.6)

sits.owl SITS Ontology (7.7)

stroke.owl Stroke Ontology (7.7)

tdb.owl Tissue Database Ontology (7.9)

mre-ontology-metrics/ Metriky ontologií (získány prostřednictvím Protége).

mre-ontology-visual/ Vizualizace ontologií OWL a RDF slovníků ve formátech PDF a SVG.

mre-src/ Zdrojové soubory projektů tvořících platformu MRE.

mre-sw/ Softwarové nástroje platformy.

ontology-design/ Návrh ontologie prostřednictvím tabulkového procesoru (viz kapitola 6 a příloha B).

- Příklad listu (`ontology-design-example-attributes.ods`) s třídami a atributy.
- Příklad listu (`ontology-design-example-vocabularies.ods`) pro číselníky hodnot.
- Návrh *Stroke Ontology* (`ontology-design-stroke.xlsx`) v tabulkovém procesoru.
- *Stroke Ontology* ve formátu OWL (`ontology-design-stroke.owl`).

vysledky/ Příklady výsledků a ilustrace použití ontologií v praxi.

owl-description/ Příklad anotace ontologie *Stroke Ontology*.

pureimage/ Příklad jednotlivých kroků algoritmu v *PureImage* (8.3.2) a DICOM snímky po de-identifikaci (zabarvení) míst s osobními údaji prostřednictvím *PureImage*.

query-builder/ Schéma modelu, ukázková data a výsledky ilustrují tvorbu SPARQL prostřednictvím webové aplikace *Query Builder* (8.6.5). Používá se *Stroke Ontology* (7.7, `stroke.owl`) pro tvorbu výběrových dotazů ve SPARQL. Znalost RDF dat, ontologií ani dotazovacího jazyka není po uživateli vyžadována. Přidání sloupce vyžaduje pouze (1) výběr období (`stroke:Timepoint`), (2) výběr třídy, (3) výběr požadované vlastnosti (sloupce). Generátor vše potřebné (titulky/názvy, vztahy) převezme přímo z ontologie. Uživatel nemusí skutečný dotaz vidět a data v CSV nebo MS Excel může přímo stáhnout.

rdf-dasta-code-list/ Číselník ÚZIS pro pohlaví (`Pohlavi.xml`) a odpovídající definice OWL třídy `dsc1:POHLAV` a hodnot číselníku v RDF. Číselník je používán ve formátu DASTA.

rdf-research/ Příklad popisu výzkumného projektu (`research.ttl`), sady dat (`dataset.ttl`) a software (`application.ttl`).

sparkle/ Snímky z aplikace *Sparkle* (8.6.4) pro tvorbu dotazů ve SPARQL 1.0 a 1.1. Na formulářích založený přístup a klasický textový editor se zvýrazňováním syntaxe a množstvím našeptávačů.

tdb/ Databáze tkáňových vzorků (8.6.7) poskytuje přehled v prostředí DataTables. Některá data mohou být šifrována, aby bylo možné zobrazení jinými lékaři bez nutnosti znát identitu.

web-form-generator/ Ukázková data z *IBDT Ontology* pro dynamicky generované webové formuláře prostřednictvím *Web Form Generator* a vytvořená nová RDF data.

J Aktivita autora

Časopisy

- [VKKK19] Petr VCELAK, Martin KRYL, Michal KRATOCHVIL a Jana KLECKOVA: *Identification and Classification of DICOM Files with Burned-in Text Content*. International Journal of Medical Informatics, 126:128–137, June 2019. ISSN 1386-5056. [Scopus, WOS; IF_{2018} 2,731; IF_{2018}^{5y} 3,146; SJR_{2018} 0,96; Q1 v kategorii *Health Informatics*]. Citováno v [RPV+20, KCC+20].
- [SSC+17] Jan SCHWARZ, Josef SYKORA, Dominika CVALINOVA, Renata POMAHACOVA, Jana KLECKOVA, Martin KRYL a Petr VCELAK: *Inflammatory bowel disease incidence in Czech children: A regional prospective study, 2000–2015*. World Journal of Gastroenterology, 23(22):4090–4101, June 2017. ISSN 1007-9327. [Scopus, WOS; IF_{2017} 3,300; SJR_{2018} 1,19; Q1 v kategoriích *Gastroenterology a Medicine (miscellaneous)*]. Citováno v [JPK+20, RTT+20, RRH+19, MHSS18, CHM+18, Hin17].
- [KKK+16a] Tomas KOUTNY, Michal KRCMA, Josef KOHOUT, Petr JEZEK, Jana VARNUSKOVA, Petr VCELAK a Jan STRNADEK: *On-Line Blood Glucose Level Calculation*. Procedia Computer Science, 2016(98):228–235, October 2016. ISSN: 1877-0509. DOI 10.1016/j.procs.2016.09.037. [Scopus, WOS; SJR_{2018} 0,28]. Citováno v [SM19, ASI+18].
- [SJV+15] Miroslava SVOBODOVA, Miroslav JIRIK, Petr VCELAK, Vladimir LUKES, Tomas RYBA, Petr HOSEK, Kristyna BAJCUROVA, Jaroslav LUDVIK, Hynek MIRKA, Zbynek TONAR, Vladislav TRESKA a Vaclav LISKA: *Software LISA – virtuální resekcce jater pro urychlení a usnadnění předoperačního plánování (Software LISA – Virtual Liver Resection to Accelerate and Facilitate Preoperative Planning)*. Rozhledy v chirurgii, 94(11):485–490, December 2015. ISSN 0035-9351. [Časopis indexován v *Seznamu recenzovaných neimpaktovaných periodik vydávaných v ČR (2015)*].
- [VKR11] Petr VCELAK, Jana KLECKOVA a Vladimir ROHAN: *Privacy and Security Issues in Cerebrovascular Diseases Data Research*. International Journal for Information Security Research (IJISR), 1:27–34, March 2011. ISSN 2042-4639.
- [RSC+11] Vladimir ROHAN, Petr SEVCIK, Lenka CERNA, Petr VCELAK a Jiri (jr.) POLIVKA: *Intravenózní trombolytická léčba ischemické cévní mozkové příhody při uzávěru vnitřní krkavice. V Česká a Slovenská neurologie a neurochirurgie, svazek 74, strany 27–27, November 2011. ISSN 1210-7859. ISSN online: 1802-4041.*

[Časopis indexován v *Seznamu recenzovaných neimpaktovaných periodik vydávaných v ČR*].

- [RSS⁺10] **Vladimir ROHAN, Petr SEVCIK, J. SUCHÁ, H. VACOVSKÁ, J. MRAČEK, I. HOLEČKOVÁ, P. LAVIČKA, Petr VČELÁK a Jiří POLÍVKA:** *Maligní edém po trombolýze infarktu v povodí střední mozkové tepny.* V *Česká a Slovenská neurologie a neurochirurgie*, svazek 73, strana 12, November 2010. ISSN 1210-7859. ISSN online: 1802-4041. [Časopis indexován v *Seznamu recenzovaných neimpaktovaných periodik vydávaných v ČR*].

Konference WoS/Scopus

- [VKK18] **Petr VCELAK, Martin KRYL a Jana KLECKOVA:** *SPARQL Query-Builder for Medical Temporal Data.* V *Proceedings 2018 11th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI 2018)*, strany 1–5, October 2018, ISBN 978-1-53-867603-5. Beijing University of Chemical Technology, Beijing Institute of Technology, Eastern China Normal University, 2018. [IEEE Explore, Scopus, WOS].
- [VKRK17] **Petr VCELAK, Martin KRYL, Ladislav RACAK a Jana KLECKOVA:** *Acquisition of Confidential Patient Data Over Shared Mobile Device.* V *10th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC2017)*, strany 334–339, February 2017, ISBN 978-9-89-758213-4. Setúbal: ScitePress. [Scopus, WOS].
- [VKKK17] **Petr VCELAK, Martin KRYL, Michal KRATOCHVIL a Jana KLECKOVA:** *Ontology-based Web Forms – Acquisition and Modification of Structured Data.* V *Proceedings 2017 10th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI 2017)*, strany 1–5, October 2017, ISBN 978-1-53-861936-0. Shanghai: School of Information Science and Technology, East China Normal University, 2017. [IEEE Explore, Scopus, WOS]. Citováno v [AA20, ZL18].
- [VKK14] **Petr VCELAK, Michal KRATOCHVIL a Jana KLECKOVA:** *Ontology-driven Information System and Management of Heterogeneous Research Data.* V *2014 7th International Conference on Biomedical Engineering and Informatics (BMEI)*, strany 766–770, November 2014, ISBN 978-1-47-995838-2. [IEEE Explore, Scopus, WOS].
- [VKK13] **Petr VCELAK, Michal KRATOCHVIL a Jana KLECKOVA:** *Collaboration and Research Information System Used for Sustainable Long-Term Research.* V *2013 Fourth World Congress on Software Engineering*, strany 307–310, December 2013, ISBN 978-1-47-992883-5. DOI 10.1109/WCSE.2013.56. [Scopus, WOS]. Citováno v [WGH⁺18].

- [VKKR12] **Petr VCELAK, Michal KRATOCHVIL, Jana KLECKOVA a Vladimir ROHAN:** *MetaMed – Medical Meta Data Extraction and Manipulation Tool Used in the Semantically Interoperable Research Information System*. V *2012 5rd International Conference on Biomedical Engineering and Informatics (BMEI)*, strany 1281–1285, November 2012, ISBN 978-1-46-731182-3. [IEEE Explore, Scopus]. Citováno v [GEPCG20, Kly19, Nia15, CYH⁺15].
- [KVKR12] **Michal KRATOCHVIL, Petr VCELAK, Jana KLECKOVA a Vladimir ROHAN:** *MedIDEA – Medical Image Data Extraction and Analysis*. V *2012 5th International Conference on Biomedical Engineering and Informatics (BMEI)*, strany 225–229, October 2012, ISBN 978-1-46-731182-3. [IEEE Explore, Scopus].
- [KVK12] **Michal KRATOCHVIL, Petr VCELAK a Jana KLECKOVA:** *Unified parallel Experiment Interface for Medical Research System*. V *Proceeding of the International Conference on Health Informatics*, svazek 1, strany 449–452, October 2012, ISBN 978-9-89-842588-1. DOI 10.5252/0003873404490452. [IEEE Explore, Scopus].
- [VK11b] **Petr VCELAK a Jana KLECKOVA:** *Semantically Interoperable Research Medical Data and Meta Data Extraction Strategy*. V *2011 4th International Conference on Biomedical Engineering and Informatics (BMEI)*, svazek 4, strany 1955–1959, October 2011, ISBN 978-1-4577-1447-4. DOI 10.1109/BMEI.2011.6098665. [IEEE Explore, Scopus]. Citováno v [SVB⁺19] (U.S. Patent).
- [VKR10] **Petr VCELAK, Jana KLECKOVA a Vladimir ROHAN:** *Experimental Database Implementation for the Cerebrovascular Diseases Research Integrates Together Different Kinds of Medical Data*. V *1st International Multi-Conference on Innovative Developments in ICT: International Conference on e-Health Services and Technologies*, strany 111–114, July–August 2010. [Scopus, WOS].
- [VKR10] **Petr VCELAK, Jana KLECKOVA a Vladimir ROHAN:** *Cerebrovascular Diseases Research Database*. V **YU, W AND ZHANG, M AND WANG, L AND SONG, Y** (editor): *2010 3RD International Conference on Biomedical Engineering and Informatics (BMEI 2010)*, VOLS 1-7, strany 2687–2691, 345 E 47th St, New York, NY 10017 USA, October 2010. IEEE Engn Med & Biol Soc, IEEE, ISBN 978-1-42-446496-8. 2010 3rd International Conference on Biomedical Engineering and Informatics (BMEI 2010), Yantai Univ, Yantai, Peoples Republic China, Oct 16–18, 2010. [IEEE Explore, Scopus, WOS]. Citováno v [Tuo14].
- [VKR10] **Petr VCELAK, Jana KLECKOVA a Vladimir ROHAN:** *Cerebrovascular Diseases Research Based on Heterogeneous Medical Data Mining and Knowledge Base*. V *2010 International Conference for Internet Technology and Secured Transactions (ICITST)*, strany

1–6, November 2010, ISBN 978-1-42-448862-9. [IEEE Explore, Scopus].

Ostatní

- [KKV19] **Jana KLECKOVA, Martin KRYL a Petr VCELAK**: *IBDS – Extension of the prospective database and operations automation (IBDS – Rozšíření prospektivní databáze a automatizace činnosti)*. Technická zpráva, Západočeská univerzita, Plzeň, 12/2019. Souhrnná výzkumná zpráva.
- [KVK18] **Martin KRYL, Petr VCELAK a Jana KLECKOVA**: *Interactive Exploration and Querying of RDF Data*. V *The Third International Conference on Informatics and Assistive Technologies for Health-Care, Medical Support and Wellbeing (HEALTHINFO 2018)*, strany 65–68, October 2018, ISBN 978-1-61-208675-0. ISSN 2519-8491.
- [KV17] **Jana KLECKOVA a Petr VCELAK**: *Secondary use of medical data (Problematika sekundárního využití medicínských dat)*. V **Arnost MOTYCKA** (editor): *INFORMATIKA XXX: Sborník abstraktů z mezinárodní odborné pedagogicky zaměřené konference*, strany 35–36. Vydavatelství Mendelovy univerzity, Brno, January 2017, ISBN 978-8-07-509512-1. EUNIS-CZ, Provozně ekonomická fakulta. Luhačovice, Česká republika.
- [KKV17] **Jana KLECKOVA, Martin KRYL a Petr VCELAK**: *IBDS – Analysis and design of a prospective database (IBDS – Analýza a návrh prospektivní databáze)*. Technická zpráva, Západočeská univerzita, Plzeň, 12/2017. Souhrnná výzkumná zpráva.
- [KKK⁺16b] **Tomas KOUTNY, Michal KRCMA, Josef KOHOUT, Petr JEZEK, Jana VARNUSKOVA, Petr VCELAK, Jan STRNADEK a Marek RASOCHA**: *Diabetes.zcu.cz – Internetový portál pro výpočet glykémie z hodnot CGMS (Blood glucose level reconstruction using continuous glucose monitoring in the subcutaneous tissue)*, October 2016. Autorizovaný software.
- [VK16] **Petr VCELAK a Jana KLECKOVA**: *Automatická identifikace a odstranění citlivých osobních údajů z obrazových dat ve formátu DICOM pro účely sekundárního využití (Automatic personal data identification and removal from imaging examination in the DICOM file format for secondary purposes)*. V *56. studentská vědecká konference*, strana 67, April 2016.
- [KKV16] **Jana KLECKOVA, Martin KRYL a Petr VCELAK**: *Analysis and design of a prospective database (Analýza a návrh prospektivní databáze)*. Technická zpráva, Západočeská univerzita, Plzeň, 12/2016. Souhrnná výzkumná zpráva.

- [VSK15] **Petr VCELAK, Jan SMUCR a Jana KLECKOVA:** *Sparkle – Enriched Graphical SPARQL Query Builder*, June 2015. Autorizovaný software.
- [VK12] **Petr VCELAK a Jana KLECKOVA:** *Extrakce metadat z medicínských dat (Medical Meta Data Extraction and Manipulation Project)*, October 2012. Autorizovaný software.
- [VK11c] **Petr VCELAK a Jana KLECKOVA:** *Automatic Real-Patient Medical Data De-Identification for Reserarch Purposes*. An International Journal of Science, Engineering and Technology, 52:506–510, October 2011. ISSN 2010-376X. Citováno v [DFKR⁺16].
- [VK11a] **Petr VCELAK a Jana KLECKOVA:** *Medical Data Integration Ontology Used in the Cerebrovascular Diseases Research*. V *Proceedings of the 8th Extended Semantic Web Conference on The Semantic Web*. Springer-Verlag, Berlin, Heidelberg, May 2011, ISBN 978-3-642-21033-4. 8th Extended Semantic Web Conference.
- [Vce11] **Petr VCELAK:** *Rozsáhlé informační systémy – standardizace metadat (Meta Data Standardisation in a Large Information Systems)*. Odborná práce ke státní doktorské zkoušce DCSE/TR-2011-05, Západočeská univerzita v Plzni, Katedra informatiky a výpočetní techniky, Západočeská univerzita v Plzni, Univerzitní 8, 30614 Plzeň, Česká republika, September 2011.
- [VK10] **Petr VCELAK a Jana KLECKOVA:** *Medical Data De-Identification Tool (Anonymizace medicínských dat)*, December 2010. Autorizovaný software.
- [KVK10] **Pavel KODL, Petr VCELAK a Jana KLECKOVA:** *Medical Data Retrieval from Hospitals Over Encrypted Connection (Šifrovaný přenos medicínských dat z nemocnic)*, December 2010. Autorizovaný software.
- [VPM⁺09] **Petr VCELAK, Jiri (jr.) POLIVKA, Petr MAULE, Petr KRATOCHVIL a Jana KLECKOVA:** *Experimental Database System for the Vascular Brain Diseases Research*. V *Front. Neur. Conference Abstract: Neuroinformatics 2009*, October 2009. DOI 10.3389/conf.neuro.11.2009.08.002. Citováno v [GEPCG20].

Citace v publikacích

- [RPV⁺20] **Leonardo RUNDO, Roberto PIRRONE, Salvatore VITABILE, Evis SALA a Orazio GAMBINO:** *Recent advances of HCI in decision-making tasks for optimized clinical workflows and precision medicine*. Journal of Biomedical Informatics, 8/2020, ISSN 1532-0464. DOI 10.1016/j.jbi.2020.103479. [IF₂₀₁₉ 3,526]. Citováno [VKKK19].

- [RTT⁺20] **Stephen E ROBERTS, Kymberley THORNE, N THAPAR, Ilse Julia BROEKAERT, M A BENNINGA, Jernej DOLINSEK, E MAS, Erasmo MIELE, Rok OREL, C PIENAR, Carmen RIBES-KONINCKX, M THOMSON, Christos TZIVINIKOS, Sian MORRISON-REES, Ann JOHN a John G WILLIAMS:** *A Systematic Review and Meta-analysis of Paediatric Inflammatory Bowel Disease Incidence and Prevalence Across Europe.* Journal of Crohn's and Colitis, 2/2020, ISSN 1873-9946. DOI 10.1093/ecco-jcc/jjaa037. [*IF*₂₀₁₉ 8,658; *IF*₂₀₁₈ 7,827]. Citováno [SSC⁺17].
- [KCC⁺20] **Surajit KUNDU, Santam CHAKRABORTY, Sanjoy CHATTERJEE, Syamantak DAS, Rimpa Basu ACHARI, Jayanta MUKHOPADHYAY, Partha Pratim DAS, Indranil MALLICK, Moses ARUNSINGH, Tapesh BHATTACHARYYAA a Soumendranath RAY:** *De-Identification of Radiomics Data Retaining Longitudinal Temporal Information.* Journal of Medical Systems, 44(99):1–15, 2020. DOI 10.1007/s10916-020-01563-0. [*IF*₂₀₁₉ 3,058; *IF*₂₀₁₉^{5yr} 3,072]. Citováno [VKKK19].
- [JPK⁺20] **Petr JABANDZIEV, Tereza PINKASOVA, Lumir KUNOVSKY, Jan PAPEZ, Martin JOUZA, Bara KARLINOVA, Martina NOVACKOVA, Milan URIK, Stefania AULICKA, Ondrej SLABY, Julia BOHOSOVA, Katerina BAJEROVA, Milan BAJER a Ajay GOEL:** *Regional Incidence of Inflammatory Bowel Disease in a Czech Pediatric Population: 16 Years of Experience (2002–2017).* Journal of Pediatric Gastroenterology and Nutrition, 70(5):586–592, 5/2020, ISSN 0277-2116. DOI 10.1097/mpg.0000000000002660. [*IF* 2,937]. Citováno [SSC⁺17].
- [GEPCG20] **Kristina GAGALOVA, M. ELIZALDE, Elodie PORTALES-CASAMAR a Matthias GÖRGES:** *What You Need to Know Before Implementing a Clinical Research Data Warehouse: Comparative Review of Integrated Data Repositories in Health Care Institutions.* Journal of Medical Internet Research, 4:p1–p27, 8/2020, ISSN 2561-326X. JMIR Form Res: e17687. DOI 10.2196/17687. [*IF*₂₀₁₉ 5,03]. Citováno [VPM⁺09, VKKR12].
- [AA20] **Sahin AYDIN a Mehmet N AYDIN:** *Ontology-based data acquisition model development for agricultural open data platforms and implementation of OWL2MVC tool.* Computers and Electronics in Agriculture, 175:105589, 7/2020, ISSN 0168-1699. DOI 10.1016/j.compag.2020.105589. [*IF*₂₀₁₉ 3,858]. Citováno [VKKK17].
- [SM19] **Christoph STACH a Bernhard MITSCHANG:** *ECHOES: A Fail-Safe, Conflict Handling, and Scalable Data Management Mechanism for the Internet of Things.* V *European Conference on Advances in Databases and Information Systems*, svazek 11695,

strany 373–389. Springer, 2019, ISBN 978-3-030-28729-0. DOI 10.1007/978-3-030-28730-6_23. Citováno [KKK⁺16a].

- [SVB⁺19] **Adam SEERING, Rajat VENKATESH, Charles Edward BEAR, Shilpa LAWANDE a Andrew Allinson LAMB**: *User-defined Loading of Data onto a Database*, 11/2019. U.S. Patent 10474658, 12. listopad 2019. European Patent (EP) 2856348 B1, 23. říjen 2019. China Patent (CN) 104508660 A, 8. duben 2015. [Patent; podáno 4. červen 2012]. Citováno [VK11b].
- [RRH⁺19] **Hollian RICHARDSON, Glenn RHODES, Peter HENRYS, Luigi SEDDA, Andrew J. WEIGHTMAN a Roger W. PICKUP**: *Presence of Mycobacterium avium subspecies paratuberculosis monitored over varying temporal and spatial scales in river catchments: persistent routes for human exposure*. *Microorganisms*, 7(5):136, 2019. DOI 10.3390/microorganisms7050136. [IF₂₀₁₈ 4,167]. Citováno [SSC⁺17].
- [Kly19] **Vladimir KLYUCHIKOV**: *Language Integrated Query as a Canonical Data Model for Virtual Data Integration*. V **Alexander ELIZAROV, Boris NOVIKOV a Sergey STUPNIKOV** (editoři): *Data Analytics and Management in Data Intensive Domains*, strana 306, 2019. XXI International Conference DAMDID/RCDL 2019 (October 15–18, 2019, Kazan, Russia). Citováno [VKKR12].
- [ZL18] **Tao ZHANG a Jixiang LU**: *The Application of Text Analytics in Electric Power Dispatching*. V *2018 International Conference on Power System Technology (POWERCON)*, strany 4186–4189. IEEE, 11/2018. DOI 10.1109/POWERCON.2018.8601990. Citováno [VKKK17].
- [WGH⁺18] **Jinsong WU, Song GUO, Huawei HUANG, William LIU a Yong XIANG**: *Information and Communications Technologies for Sustainable Development Goals: State-of-the-Art, Needs and Perspectives*. *IEEE Communications Surveys and Tutorials*, 20(3):2389–2406, 2018. ISSN 1553-877X. DOI 10.1109/COMST.2018.2812301. [IF₂₀₁₉ 23,7; IF₂₀₁₇ 20,230]. Citováno [VKK13].
- [MHŠ18] **Klára MARATOVÁ, Ondřej HRADSKÝ, Ondřej SOUČEK a Zdeněk ŠUMNÍK**: *Vitamin D a jeho suplementace u dětských pacientů se zánětlivým střevním onemocněním*. *Pediatric pro praxi*, 19(4):190–194, 2018. ISSN 1213-0494 (tištěná verze), ISSN 1803-5264 (on-line verze). Dostupné z <https://www.pediatricpropraxi.cz/pdfs/ped/2018/04/02.pdf>. [Scopus]. Citováno [SSC⁺17].
- [CHM⁺18] **Raguraj CHANDRADEVAN, Tatyana HOFMEKLER, Kajari MONDAL, Nusrat HARUN, Suresh VENKATESWARAN, Hari K SOMINENI, Cortney R BALLENGEE, Mi Ok KIM, Anne GRIFFITHS, Joshua D NOE a kol.**: *Evolution of Pediatric Inflammatory Bowel Disease Unclassified (IBD-U): Incorporated*

With Serological and Gene Expression Profiles. Inflammatory bowel diseases, 24(10):2285–2290, 2018. DOI 10.1093/ibd/izy136. [IF 4,261]. Citováno [SSC⁺17].

[ASI⁺18] **Ganjar ALFIAN, Muhammad SYAFRUDIN, Muhammad IJAZ, M. SYAEKHONI, Norma FITRIYANI a Jongtae RHEE:** *A personalized healthcare monitoring system for diabetic patients by utilizing BLE-based sensors and real-time data processing.* *Sensors*, 18(7):2183, 2018, ISSN 1424-8220. DOI 10.3390/s18072183. [IF₂₀₁₈ 3,031; IF₂₀₁₉ 3,275]. Citováno [KKK⁺16a].

[Hin17] **Martin HINGANT:** *Maladies inflammatoires chroniques de l'intestin pédiatriques sur l'île de la Réunion: épidémiologie descriptive et suivi à 1 an.* disertace, Interne des hôpitaux de l'île de la Réunion, Médecine humaine et pathologie, 2017. <dumas-01635805>. Citováno [SSC⁺17].

[DFKR⁺16] **Dina DEMNER-FUSHMAN, Marc D KOHLI, Marc B ROSENMAN, Sonya E SHOOSHAN, Laritza RODRIGUEZ, Sameer ANTANI, George R THOMA a Clement J MCDONALD:** *Preparing a collection of radiology examinations for distribution and retrieval.* *Journal of the American Medical Informatics Association*, 23(2):304–310, 3/2016. DOI 10.1093/jamia/ocv080. [IF₂₀₁₉ 4,112]. Citováno [VK11c].

[Nia15] **Olegas NIAKSU:** *CRISP data mining methodology extension for medical domain.* *Baltic Journal of Modern Computing*, 3(2):92, 2015, ISSN 2255-8942. [Scopus]. Citováno [VKKR12].

[CYH⁺15] **Jui Jen CHEN, Hung Chi YEN, Shu Hua HUANG, Yen Hsiang CHANG, Huang Chung CHEN a Shin Chun HUANG:** *Applying ontology and information extraction techniques to develop a nuclear medicine examination report auditing system in department of nuclear medicine.* V *MDA*, strany 1–16, 2015. 10th International Conference on Mass Data Analysis of Signals and Images with Applications in Medicine, r/g/b Biotechnology, Food Industries and Dietetics, Biometry and Security, and Agriculture, MDA 2015; Hamburg; Germany; 11 July 2015 through 14 July 2015. ISBN 978-394295234-7. Citováno [VKKR12].

[Tuo14] **Lauri TUOVINEN:** *From Machine Learning to Learning with Machines.* disertace, University of Oulu, Oulu, 2014. ISBN 978-9-52-620523-6. ISSN 0355-3213. Dostupné z <http://jultika.oulu.fi/files/isbn9789526205243.pdf>. Citováno [VKR10].

Výukové aktivity

- 2019/2020
 - Informační technologie (KIV/ITZK) – Př, Cv, ZS,
 - Informační technologie (KIV/ITZS) – Cv, ZS,
 - Informatika a výpočetní technika (KIV/IVT) – Cv, ZS i LS,
 - * nahrazuje předměty KIV/IFYER, KIV/IVTP a KIV/ITZS,
 - Informatika a výpočetní technika (KIV/IFYER) – Cv, LS,
 - garant předmětů KIV: DB1-E, IFYER, IT, IT1, IT1K, IT2, ITR, ITRK, ITVP, ITVPK, ITZK, ITZS, IVT, IVTU, VTZL.
- 2018/2019
 - Informační technologie 1 (KIV/IT1) – Př, Cv, ZS,
 - Informační technologie 2 (KIV/IT2) – Př, Cv, LS,
 - Informační technologie (KIV/ITZK) – Př, Cv, ZS,
 - Informační technologie (KIV/ITZS) – Cv, ZS,
 - Informatika a výpočetní technika (KIV/IFYER) – Cv, LS.
- 2017/2018
 - Informační technologie (KIV/IT) – Př, Cv, ZS,
 - Informační technologie 1 (KIV/IT1) – Př, Cv, ZS,
 - Informační technologie 2 (KIV/IT2) – Př, Cv, LS,
 - Informační technologie (KIV/ITZK) – Př, Cv, ZS,
 - Informační technologie (KIV/ITZS) – Cv, ZS,
 - Informatika a výpočetní technika (KIV/IFYER) – Cv, LS.
- 2016/2017
 - Databázové systémy 2 (KIV/DB2) – Cv, LS,
 - Informační technologie (KIV/IT) – Př, Cv, ZS,
 - Informační technologie (KIV/ITZK) – Př, Cv, ZS,
 - Informační technologie (KIV/ITZS) – Cv, ZS,
 - Informatika a výpočetní technika (KIV/IFYER) – Cv, LS,
 - Základy informačních systémů (KIV/ZIS) – Cv, ZS.
- 2015/2016
 - Informační technologie 1 (KIV/IT1) – Př, Cv, ZS,
 - Informační technologie 2 (KIV/IT2) – Př, Cv, LS,
 - Informační technologie (KIV/ITZK) – Př, Cv, ZS,
 - Informační technologie (KIV/ITZS) – Cv, ZS,

- Informatika a výpočetní technika (KIV/IFYER) – Cv, LS.
- 2014/2015
 - Informační technologie (KIV/IT) – Př, Cv, ZS,
 - Informační technologie 1 (KIV/IT1) – Př, Cv, ZS,
 - Informační technologie 2 (KIV/IT2) – Př, Cv, LS,
 - Informační technologie (KIV/ITZK) – Př, Cv, ZS,
 - Informační technologie (KIV/ITZS) – Cv, ZS,
 - Informatika a výpočetní technika (KIV/IFYER) – Cv, LS.
- 2013/2014
 - Informační technologie (KIV/IT) – Př, Cv, ZS,
 - Informační technologie 1 (KIV/IT1) – Př, Cv, ZS,
 - Informační technologie 2 (KIV/IT2) – Př, Cv, LS,
 - Informační technologie (KIV/ITVP) – Cv, ZS,
 - Informační technologie (KIV/ITVPK) – Př, Cv, ZS,
 - Informatika a výpočetní technika (KIV/IFYER) – Cv, LS.
- 2012/2013
 - Informační technologie (KIV/IT) – Př, Cv, ZS,
 - Informační technologie 1 (KIV/IT1) – Př, Cv, ZS,
 - Informační technologie 2 (KIV/IT2) – Př, Cv, LS,
 - Informační technologie (KIV/ITVP) – Cv, ZS,
 - Informační technologie (KIV/ITVPK) – Př, Cv, ZS,
 - Informatika a výpočetní technika (KIV/IFYER) – Cv, LS.
- 2011/2012
 - Informační technologie (KIV/ITVP) – Cv, ZS,
 - Informační technologie (KIV/ITVPK) – Př, Cv, ZS,
 - Informatika a výpočetní technika (KIV/IFYER) – Cv, LS.
- 2010/2011
 - Informační technologie (KIV/ITVP) – Cv, ZS.
- 2009/2010
 - Informační technologie (KIV/ITVP) – Cv, ZS,
 - Informatika a výpočetní technika (KIV/IFYER) – Cv, LS.
- 2008/2009
 - Informační technologie (KIV/ITVP) – Cv, ZS,
 - Úvod do počítačových sítí (KIV/UPS) – Cv, ZS.

Spolupráce na projektech

1. *Aplikace moderních technologií v medicíně a průmyslu (AMTMI)*, období 05/2018–10/2022, CZ.02.1.01/0.0/0.0/17_048/0007280.
2. *PUNTIS – Podpora udržitelnosti centra NTIS – Nové technologie pro informační společnost (LO1506)*, Ministerstvo školství, mládeže a tělovýchovy, 1. 7. 2015–30. 6. 2020.
3. *LiverCloud – H2020-SC1-DTH-2018-1: LiverCloud. Research and Innovation Framework Programme*. Splněny podmínky, nevybráno k financování.
4. *IBDS – analýza a návrh prospektivní databáze*, Ministerstvo školství, mládeže a tělovýchovy, 12. 6. 2017–31. 10. 2017.
5. *Věda, která léčí – výzkum Crohnovy nemoci*. Spolupráce Fakulty aplikovaných věd ZČU, Fakultní nemocnice Plzeň a hnutím Na vlastních nohou – Stonožka. Období 2016–2017.
6. *Obtaining information from heterogeneous data (IA7)*, Technologická ověření výsledků výzkumu a vývoje I. (CZ.1.05/3.1.00/14.0298), Podpora pre-seed aktivit, Komercializace a popularizace VaV, Komercializace výsledků výzkumných organizací a ochrana jejich duševního vlastnictví, období řešení 1. 1. 2014–30. 9. 2015.
7. *Zvyšování resekability maligních ložiskových procesů pomocí metod zpřesňujících měření perfúzních parametrů zbytkového jaterního parenchymu – počítačem asistované diagnostiky a softwarového modelování (IGA MZ ČR NT 13326)* období řešení 2012–2015.
8. *Metody a aplikace biomedicínské informatiky (SGS-2013-039)*, Studentská grantová soutěž 2013, období řešení 1. 2. 2013–31. 12. 2015.
9. *Mikrostrukturálně orientované hierarchické modelování prokrvení mozku pro vyhodnocení CT perfúzního vyšetření (GA106/09/0740)*, Grantová agentura České republiky (GA ČR), období řešení 1. 1. 2009–31. 12. 2012.
10. *Systematická podpora popularizace výzkumu a vývoje ZČU (CZ.1.07/2.3.00/35.0014)*, OP Vzdělávání pro konkurenceschopnost, období řešení 1. 7. 2012–30. 6. 2014.
11. *Metody a aplikace biomedicínské informatiky (SGS-2010-038)*, Studentská grantová soutěž 2010, období řešení 1. 4. 2010–31. 12. 2012.
12. *Podpora biomechaniky na FAV (SGS-2010-077)*, Studentská grantová soutěž 2010, období řešení 1. 4. 2010–31. 12. 2012.
13. *Rozvoj kompetencí pro komunikaci vědy (SCICOM) (CZ.1.07/2.3.00/09.0151)*, OP Vzdělávání pro konkurenceschopnost, období řešení 13. 7. 2009–30. 3. 2011.